

THESE

En vue de l'obtention du : **DOCTORAT**

Structure de Recherche : Laboratoire Conception et Systèmes (Electronique,
Signaux et Informatique)

Discipline : Informatique

Spécialité : Informatique et Intelligence Artificielle

Présentée et soutenue le 20/12/2023 par :

Imane LASRI

**Nouvelles Approches Émotionnelles basées sur les Algorithmes de
l'Intelligence Artificielle et du Big Data pour l'Amélioration et l'Optimisation
des Processus de Prise de Décision : Cas d'Utilisation dans l'Enseignement
Inclusif**

Devant le jury

Noureddine ZAHID	PES, Université Mohammed V, Faculté des Sciences de Rabat	Président
Jaafar ABOUCHABAKA	PES, Université Ibn Tofail, Faculté des Sciences de Kénitra	Rapporteur/Examineur
Bouabid El OUAHIDI	PES, Université Mohammed V, Faculté des Sciences de Rabat	Rapporteur/Examineur
Soumia ZITI	PES, Université Mohammed V, Faculté des Sciences de Rabat	Rapporteuse/Examinatrice
Anouar RIAD SOLH	PH, Université Mohammed V, Faculté des Sciences de Rabat	Co-directeur de Thèse
Mourad EL BELKACEMI	PES, Université Mohammed V, Faculté des Sciences de Rabat	Directeur de Thèse

Année Universitaire : 2023/2024

Je dédie cette thèse

À la mémoire de ma mère El Hassania

Que je sens maintenant sa fierté - là où elle est

Que dieu la garde dans son vaste paradis.

À ma deuxième mère Aicha qui m'a entouré

d'amour, d'affection et qui fait tout pour ma réussite.

À ma grand-mère Saadia pour son attention particulière,

ses prières et son amour inconditionnel que dieu te donne

bonne santé et longue vie parmi nous.

À mon père Mustapha, pour son amour et ses conseils.

À mes oncles Khalid et Houssine qui m'ont soutenue

moralement et émotionnellement jusqu'ici.

À mon mari Naoufal, pour son soutien

quotidien indéfectible.

Remerciement

Les travaux présentés dans cette thèse ont été réalisés au sein du laboratoire Conception et Système (Électronique, Signaux et Informatique) de la Faculté des Sciences de Rabat, sous la direction éclairée du chef du laboratoire, Monsieur **Aziz Ettouhami** Professeur d'Enseignement Supérieur à la Faculté des Sciences de Rabat.

Je tiens à remercier tout d'abord mon directeur de thèse, Monsieur **Mourad EL BELKACEMI** Professeur d'Enseignement Supérieur à la Faculté des Sciences de Rabat, pour son encadrement, sa patience, sa grande disponibilité, ses qualités humaines et scientifiques et toute l'aide qu'il m'a procuré durant l'élaboration de ce travail.

Je remercie bien vivement aussi mon co-directeur de thèse Monsieur **Anouar RIAD SOLH** Professeur Habilité à la Faculté des Sciences des Rabat, pour son encadrement, ses encouragements, et sa passion pour la recherche. Il a énormément contribué à ce que ma thèse se déroule dans d'excellentes conditions scientifiques et dans une bonne ambiance générale.

Je souhaite exprimer ma profonde gratitude à Monsieur **Noureddine ZAHID** Professeur d'Enseignement Supérieur à la Faculté des Sciences de Rabat, pour sa disponibilité, et de m'avoir fait l'honneur d'être président de ma soutenance.

Je remercie vivement Monsieur **Jaafar ABOUCHABAKA** Professeur d'Enseignement Supérieur à la Faculté des Sciences de Kénitra, d'avoir accepté d'être le rapporteur et examinateur de cette thèse.

Je tiens à exprimer ma profonde gratitude envers Monsieur **Bouabid El OUAHIDI**, Professeur d'Enseignement Supérieur à la Faculté des Sciences de Rabat, pour avoir accepté d'évaluer mon travail en tant que rapporteur et examinateur de cette thèse. Sa contribution active, ses conseils judicieux et son dévouement envers l'évaluation de ce travail ont été d'une valeur inestimable.

Je tiens également à exprimer ma sincère gratitude envers Madame **Soumia ZITI**, Professeure d'Enseignement Supérieur à la Faculté des Sciences de Rabat, d'avoir accepté d'être rapportrice et examinatrice de cette thèse, contribuant ainsi de manière significative à son évaluation.

Enfin je ne saurais oublier tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce travail.

Résumé

À l'ère actuelle, l'adaptation de l'apprentissage aux besoins des apprenants est essentielle pour évaluer la qualité des systèmes éducatifs. C'est dans cet esprit, conscient de la diversité des apprenants, que notre thèse se focalise sur l'importance de l'adaptation de l'apprentissage, étroitement liée à l'évaluation de la qualité des systèmes éducatifs, tout en accordant une attention particulière à l'éducation inclusive. Dans ce contexte, deux volets ont été explorés. Le premier volet se focalise sur l'amélioration de l'évaluation formative en utilisant deux systèmes de reconnaissance des émotions faciales (FER). Le premier utilise un réseau de neurones convolutifs (CNN) pour reconnaître les émotions faciales des étudiants en temps réel. Le deuxième système FER repose sur un réseau de neurones convolutifs profonds VGG-16, basé sur l'apprentissage par transfert, pour détecter l'engagement des étudiants sourds et malentendants en se basant sur leurs expressions faciales. Les systèmes ont été évalués auprès d'étudiants en master et en licence fondamentale d'excellence à la Faculté des Sciences de Rabat (FSR). Le deuxième volet de notre recherche se concentre sur le développement de systèmes d'analyse des sentiments en utilisant l'apprentissage automatique et les technologies des Big Data pour évaluer la qualité des systèmes éducatifs. Nous avons exploré différentes approches, y compris l'analyse des tweets pour mesurer la réputation des universités marocaines, ainsi que l'opinion des utilisateurs de Twitter sur l'apprentissage à distance dans l'enseignement supérieur. Les résultats de ces systèmes fournissent des indications précieuses pour évaluer en profondeur la qualité des systèmes éducatifs.

Mots-clefs : Reconnaissance des émotions faciales, Éducation inclusive, Engagement des étudiants, Surdit , Analyse des sentiments, Éducation de qualit , Syst me de gestion intelligente de la classe, Apprentissage automatique, Apprentissage profond, Big data.

Abstract

In today's era, the adaptation of learning to the needs of learners is essential for evaluating the quality of educational systems. It is in this spirit, aware of the diversity of learners, that our thesis focuses on the importance of adapting learning, which is closely linked to the evaluation of the quality of educational systems, while paying particular attention to inclusive education. In this context, two strands have been explored. The first strand focuses on improving formative assessment using two facial emotion recognition (FER) systems. The first one uses a convolutional neural network (CNN) to recognize students' facial emotions in real time. The second FER system relies on a VGG-16 deep convolutional neural network, based on transfer learning, to detect the engagement of deaf and hard-of-hearing students based on their facial expressions. The systems were evaluated with students in the Master's and Bachelor's degree programs of excellence at the Faculty of Sciences of Rabat (FSR). The second part of our research focuses on the development of sentiment analysis systems using machine learning and Big Data technologies to evaluate the quality of educational systems. We explored different approaches, including analyzing tweets to measure the reputation of Moroccan universities, as well as Twitter users' opinions on distance learning in higher education. The results of these systems provide valuable insights for in-depth evaluation of the quality of educational systems.

Keywords : Facial emotion recognition, Inclusive education, Student engagement, Deafness, Sentiment analysis, Quality education, Intelligent classroom management system, Machine learning, Deep learning, Big data.

Liste des figures

Figure 1.1 – Exemple d’une régression linéaire.	10
Figure 1.2 – Fonction Sigmoidale.	11
Figure 1.3 – Visualisation d’hyperplan de la machine à vecteurs de support (SVM).	13
Figure 1.4 – Structure d’une cellule neuronale, montrant les moyens de communication	19
Figure 1.5 – Réseau de neurone simple proposé par Warren McCulloch et Walter Pitts [46]	19
Figure 1.6 – Modélisation d’un neurone artificiel appelé le Perceptron.	20
Figure 1.7 – Types de perceptrons.	20
Figure 1.8 – Architecture d’un bloc Vanilla LSTM.	21
Figure 1.9 – Architecture de Bi-LSTM.	23
Figure 1.10 – Architecture générique de CNN.	24
Figure 1.11 – Fonction d’unité linéaire rectifiée (ReLU).	24
Figure 1.12 – Exemple de max pooling (a) et average pooling (b) de taille 2×2.	25
Figure 1.13 – Architecture du modèle VGG-16.	26
Figure 1.14 – Système limbique.	29
Figure 1.15 – Architecture d’un système de reconnaissance automatique des expressions faciales.	30
Figure 2.1 – Triangle didactique de Houssaye [69].	36
Figure 2.2 – Triangle DidacTIC.	37
Figure 2.3 – Quatre types de modèle éducatif.	40
Figure 2.4 – Structure de notre système de reconnaissance des expressions faciales.	43
Figure 2.5 – Détection de visages avec Haar Cascades.	44
Figure 2.6 – Exemple d’image de la base de données FER2013 avant et après augmentation.	44
Figure 2.7 – Architecture de notre modèle CNN.	44
Figure 2.8 – Exemples d’images de base de données FER2013.	46
Figure 2.9 – Matrice de confusion normalisée du modèle proposée sur la base de données FER 2013.	48
Figure 2.10 – Rapport de classification de notre approche proposée sur la base de données FER2013. ...	48
Figure 2.11 – Résultats de la reconnaissance des émotions faciales des étudiants.	51
Figure 2.12 – Architecture globale de notre système proposé pour la reconnaissance des émotions et la détection de l’engagement des étudiants sourds et malentendants.	56
Figure 2.13 – Résumé de l’optimisation du modèle VGG-16 proposé.	58
Figure 2.14 – Validation croisée 10 fois (10-fold cross-validation (CV)).	60
Figure 2.15 – Exemples d’images de base de données JAFFE.	61
Figure 2.16 – Illustrations d’images de la base de données KDEF.	62
Figure 2.17 – Comparaison de tous les optimiseurs en utilisant le modèle VGG-16 sur la base de données JAFFE.	63
Figure 2.18 – Comparaison de tous les optimiseurs en utilisant le modèle VGG-16 sur la base de données KDEF.	63
Figure 2.19 – Précision (a) et perte (b) d’apprentissage et de test du modèle VGG-16 sur la base de données JAFFE.	65
Figure 2.20 – Précision (a) et perte (b) d’apprentissage et de test du modèle VGG-16 sur la base de données KDEF.	66
Figure 2.21 – Matrice de confusion de modèle VGG-16 sur la base de données JAFFE.	67
Figure 2.22 – Matrice de confusion de modèle VGG-16 sur la base de données KDEF.	67
Figure 2.23 – Reconnaissance des émotions faciales et détection de l’engagement de 7 étudiants sourds et	

malentendants en licence fondamentale d'excellence "Électronique, Informatique et Robotique".	71
Figure 2.24 – Niveau d'engagement de 7 étudiants sourds et malentendants en licence fondamentale d'excellence "Électronique, Informatique et Robotique".	71
Figure 3.1 – Cadre UNESCO pour la compréhension de la qualité de l'éducation.	77
Figure 3.2 – Différentes tâches de l'analyse du sentiment.	80
Figure 3.3 – Différents niveaux d'analyse du sentiment.	84
Figure 4.1 – Architecture globale de l'analyse des sentiments.	98
Figure 4.2 – Chronologie des tweets sur l'apprentissage à distance.	100
Figure 4.3 – Résumé du modèle Bi-LSTM basé sur l'auto-attention.	103
Figure 4.4 – Matrice de confusion normalisée du modèle Bi-LSTM basé sur l'auto-attention.	104
Figure 4.5 – Mesures de performance du modèle proposé avec différents modèles d'apprentissage profond.	105
Figure 4.6 – Résumé du modèle LSTM.	106
Figure 4.7 – Résumé du modèle Bi-LSTM.	106
Figure 4.8 – Résumé du modèle CNN-Bi-LSTM.	107
Figure 4.9 – Nuage de mots fréquents dans l'ensemble de données.	107
Figure 4.10 – Distribution des sentiments prédits pour les tweets relatifs à l'apprentissage à distance.	108
Figure 4.11 – Sentiments prédits à l'égard de l'apprentissage à distance à différents niveaux de l'éducation.	109
Figure 4.12 – Distribution des sentiments prédits dans les tweets sur l'enseignement à distance par six disciplines académiques.	110
Figure 4.13 – Les 20 bigrammes les plus fréquents des tweets positifs sur l'apprentissage à distance dans l'enseignement supérieur.	111
Figure 4.14 – Les 20 bigrammes les plus fréquents des tweets négatifs sur l'apprentissage à distance dans l'enseignement supérieur.	112
Figure 4.15 – Architecture d'analyse des sentiments en temps réel à partir de Twitter.	116
Figure 4.16 – Architecture d'Apache Kafka.	121
Figure 4.17 – Architecture d'Apache Spark.	122
Figure 4.19 – Mesures de performance des six algorithmes d'apprentissage automatique avec TF-IDF.	124
Figure 4.18 – Mesures de performance des six algorithmes d'apprentissage automatique avec Count Vectorizer.	125
Figure 4.20 – Matrice de confusion normalisée du classificateur Forêts aléatoires (RF) avec TF-IDF.	126
Figure 4.21 – Nuage de mots fréquents dans l'ensemble des données.	127
Figure 4.22 – Diagramme à barres montrant le nombre de tweets en temps réel reçus par jour et par sentiment.	128
Figure 4.23 – Diagramme à barres horizontales empilées montrant le sentiment prédit des universités Marocaines.	128
Figure 4.24 – Distribution des sentiments prédits des tweets relatifs aux universités Marocaines.	129
Figure 4.25 – Pourcentage de tweets positifs par université.	130
Figure 4.26 – Répartition géographique des tweets en temps réel.	131
Figure 4.27 – Tableau de bord produit par Kibana pour la visualisation des tweets en temps réel.	132

Liste des tableaux

Tableau 2.1 – Configuration du modèle CNN proposé.	45
Tableau 2.2 – Nombre d’images pour chaque émotion dans la base de données FER2013.	46
Tableau 2.3 – Hyperparamètres du modèle VGG-16.	59
Tableau 2.4 – Poids émotionnel.	60
Tableau 2.5 – Détection d’engagement à partir de l’indice de concentration (CI).	60
Tableau 2.6 – Précision de la reconnaissance de notre modèle proposé sur les bases de données JAFFE et KDEF en utilisant différents optimiseurs.	64
Tableau 2.7 – Précision, Recall et F1-score de notre modèle proposé sur les bases de données JAFFE et KDEF.	64
Tableau 2.8 – Comparaison de performance du modèle proposé avec huit modèles DCNN pré-entraînés.	68
Tableau 2.9 – Comparaison entre la méthode proposée et les méthodes existantes sur les bases de données JAFFE et KDEF.	70
Tableau 4.1 – Exemple de mise en œuvre des étapes de prétraitement sur un tweet.	100
Tableau 4.2 – Comparaison des performances du modèle proposé avec différents modèles d’apprentissage profond.	105
Tableau 4.3 – Nombre de tweets dans chaque étape éducative par sentiment.	108
Tableau 4.4 – Exemple de tweets relatifs à l’enseignement supérieur avec leur sentiment prédit.	109
Tableau 4.5 – Nombre de tweets pour chaque université marocaine.	117
Tableau 4.6 – Exemple de mise en œuvre des étapes de prétraitement sur un tweet.	118
Tableau 4.7 – Résultats des performances des six classificateurs avec TF-IDF et Count Vectorizer.	125

Table des matières

Dédicace	i
Remerciement	ii
Résumé	iii
Abstract	iv
Liste des figures	v
Liste des tableaux	vii
Table des matières	viii
Introduction générale	1
Partie I. Adaptation des méthodes d'apprentissage aux étudiants.....	7
Chapitre 1. État de l'art des algorithmes de Machine learning et de Deep learning dans le contexte de l'enseignement	8
1.1. Introduction	8
1.2. Algorithmes de l'apprentissage automatique	8
1.2.1. Régression linéaire	10
1.2.2. Régression logistique	11
1.2.3. Machines à vecteurs de support (SVM)	12
1.2.4. Naïve bayes	15
1.2.5. Arbres de décisions	16
1.2.6. Forêts aléatoires.....	17
1.2.7. Boosting de gradient extrême (XGBoost).....	17
1.3. Algorithmes de l'apprentissage profond.....	18
1.3.1. Principe de base des réseaux de neurones artificiels.....	18
1.3.2. Réseaux de longue mémoire à court terme (LSTM).....	21
1.3.3. Réseaux bidirectionnels de longue mémoire à court terme (Bi-LSTM).....	22
1.3.4. Réseau de neurone convolutif (CNN).....	23
1.3.5. Réseaux de neurones convolutifs profonds (DCNN).....	25
1.4. Reconnaissance des émotions faciales.....	28
1.4.1. Définition de l'émotion	28
1.4.2. Neurophysiologie des émotions : le système limbique.....	28
1.4.3. Caractéristiques typiques des émotions universelles sur le visage	29
1.4.4. Système de reconnaissance automatique des expressions faciales	30
1.5. Applications de la reconnaissance des émotions faciales dans le domaine de l'enseignement	33
1.6. Conclusion	34
Chapitre 2. Techniques de la reconnaissance automatique des émotions faciales pour une approche pédagogique différenciée et inclusive	35
2.1. Introduction	35
2.2. Définition du concept enseignement / apprentissage	35
2.3. Evolution des théories d'apprentissage à l'ère numérique	37

2.4. Principes fondamentaux de la pédagogie différenciée	38
2.4.1. Formes de différenciation pédagogique	38
2.4.2. Evaluation dans la pédagogie différenciée.....	39
2.5. Principes fondamentaux de la pédagogie inclusive	40
2.6. Reconnaissance des émotions faciales des étudiants en temps réel en utilisant CNN	41
2.6.1. Introduction	41
2.6.2. Travaux connexes.....	42
2.6.3. Système d'analyse des émotions faciales des étudiants en utilisant CNN	43
2.6.4. Bases de données d'expressions faciales	45
2.6.5. Détails de l'implémentation du système proposé	46
2.6.6. Résultats expérimentaux	47
2.7. Détection d'engagement des étudiants sourds et malentendants en temps réel à partir de leurs expressions faciales avec VGG-16 et l'apprentissage par transfert	52
2.7.1. Introduction	52
2.7.2. Travaux connexes.....	54
2.7.3. Système de détection d'engagement des étudiants sourds et malentendants à partir de leurs expressions faciales en utilisant le VGG-16 et l'apprentissage par transfert	56
2.7.4. Bases de données d'expressions faciales	60
2.7.5. Détails de l'implémentation du système proposé	62
2.7.6. Résultats expérimentaux	62
2.8. Conclusion et discussion.....	71
Partie II. Evaluation de la qualité des systèmes éducatifs.....	73
Chapitre 3. Concepts fondamentaux de l'Analyse de Sentiment pour l'Évaluation des Systèmes Éducatifs	74
3.1. Introduction.....	74
3.3. Fondements théoriques de l'analyse de sentiment	78
3.3.1. Définition de l'opinion	78
3.3.2. Types d'opinion.....	79
3.3.3. Tâches d'analyse du sentiment.....	79
3.3.4. Niveaux d'analyse du sentiment	83
3.3.5. Prétraitement du texte.....	86
3.3.6. Extraction de caractéristiques textuelles	87
3.3.7. Classification du sentiment	87
3.4. Etat de l'art de l'application de l'analyse du sentiment dans le domaine de l'enseignement supérieur	89
3.5. Conclusion	91
Chapitre 4. Approches d'analyse de sentiment pour mesurer la pertinence des systèmes éducatifs	92
4.1. Introduction.....	92
4.2. Analyse de la satisfaction relative à l'enseignement à distance en temps réel	92
4.2.1. Introduction	92
4.2.2. Travaux connexes.....	95
4.2.3. Système d'analyse du sentiment en temps réel sur les tweets relatifs à l'apprentissage à distance dans l'enseignement supérieur	97
4.2.4. Détails de l'implémentation du système proposé	103
4.2.5. Résultats expérimentaux	104
4.3. Analyse de la réputation des universités marocaines en temps réel	112
4.3.1. Introduction	112
4.3.2. Travaux connexes.....	114
4.3.3. Système d'analyse du sentiment en temps réel sur les tweets relatifs aux universités marocaines.....	115
4.3.3.1. Phase de création d'un modèle d'analyse de sentiment.....	116

4.3.3.2.Phase de création d'un pipeline d'analyse de sentiments en temps réel en utilisant les technologies de Big data.....	120
4.3.4. Détails de l'implémentation du système proposé	123
4.3.5. Résultats expérimentaux	124
4.4. Conclusion et discussion.....	133
Conclusion générale.....	134
Bibliographie	139
Liste des publications.....	159

Introduction générale

Au cours des dernières décennies, l'accès à une éducation de qualité pour tous les apprenants, indépendamment de leur diversité et de leur hétérogénéité, est devenu un objectif majeur à l'échelle mondiale. Pourtant, bien que de nombreux pays aient établi des systèmes éducatifs pour répondre à cette exigence universelle, la qualité de ces systèmes varie considérablement. Dans ce contexte, l'évaluation de la qualité des systèmes éducatifs est essentielle pour garantir une éducation efficace et équitable pour tous les apprenants, en prenant en compte leur diversité en termes de capacités, de personnalités, de stratégies d'apprentissage, d'attitudes et de motivations différentes lorsqu'ils entrent en classe [1]. Ainsi, il est primordial de tenir compte de cette diversité en termes de culture, d'intelligence, de langue, de milieu économique, de genre, d'intérêts et de handicap pour améliorer l'apprentissage.

L'éducation inclusive est un aspect important de cette démarche, qui implique la mise en place d'une pédagogie différenciée pour adapter des stratégies d'enseignement aux intérêts et aux besoins d'apprentissage différents des étudiants. Les classes hétérogènes avec une pédagogie différenciée préparent mieux les étudiants aux situations réelles et conduisent à un apprentissage significatif, durable et transférable [2]. Par conséquent, la décision de l'enseignant d'offrir à quel étudiant quel type d'enseignement différencié doit dépendre d'une bonne connaissance par l'enseignant de la manière dont les étudiants diffèrent sur ces stratégies d'apprentissage. Comme ce dernier est influencé par divers facteurs tels que les émotions, le stress, l'attention, la concentration, et l'engagement les enseignants peuvent être biaisés lorsqu'ils utilisent des méthodes d'évaluation formative traditionnelles telles que les questionnaires, les projets pratiques, les projets individuels ou de groupe, les examens oraux, et les présentations pour évaluer les capacités des étudiants, en particulier pour ceux qui proviennent de milieux défavorisés ou de minorités ethniques [3]. De ce fait, ces méthodes d'évaluation formative traditionnelles peuvent être utilisées de manière subjective pour évaluer les stratégies d'apprentissage. Toutefois, il est important de compléter ces méthodes avec des outils d'évaluation supplémentaires basées sur l'intelligence artificielle (IA). L'utilisation de l'IA dans l'éducation, en particulier dans l'enseignement inclusif, a suscité un grand intérêt avec les avancées des technologies éducatives. L'IA est capable d'apprendre, de s'adapter, de s'auto-corriger et d'analyser les

données, ce qui permet aux enseignants de créer des environnements d'apprentissage personnalisés, adaptatifs et participatifs, ainsi que des systèmes de tutorat intelligents pour les apprenants. En outre, les enseignants peuvent recevoir en temps réel des données pertinentes sur l'état d'apprentissage des étudiants, y compris leur niveau de concentration et leur état émotionnel. Cette capacité leur permet d'adapter leur enseignement en fonction des besoins individuels de chaque étudiant pendant toute l'année scolaire.

Les états émotionnels d'un individu, qui sont des réactions et des interactions à certains stimuli, sont influencés par divers facteurs tels que les intentions, les normes, les antécédents, les capacités cognitives, les états physiques ou psychologiques, les tendances à l'action et les sentiments subjectifs. Ekman [4] a proposé des expressions faciales universelles (joie, colère, peur, surprise, tristesse, dégoût, et neutre) et a affirmé qu'elles sont présentées chez tous les êtres humains quelle que soit leur culture. Les émotions ont un fort impact sur l'apprentissage et jouent donc un rôle essentiel dans l'éducation. Elles n'ont pas seulement un impact sur l'apprentissage traditionnel en classe, mais aussi sur différents types d'apprentissage tels que l'apprentissage en ligne. Elles doivent donc être prises en compte pour améliorer le processus d'apprentissage et pour détecter le niveau de concentration et d'engagement. Caraway et al. [5] ont étudié la relation entre la peur de l'échec et l'engagement scolaire. Ils ont constaté que la peur de l'échec permettait de prédire de manière significative une baisse de la moyenne des étudiants. En outre, l'anxiété liée aux examens était négativement liée aux notes, mais elle n'était pas significativement liée à l'engagement des étudiants. D'autres études ont décrit la relation entre les émotions d'un étudiant exprimées explicitement par son visage et ses résultats académiques, et elles ont démontré l'importance de prendre en compte la relation quadratique entre les émotions positives d'un étudiant pendant l'apprentissage et ses résultats [6,7].

Un bon apprentissage peut avoir un impact positif sur les résultats d'un étudiant, tandis que des difficultés d'apprentissage peuvent faire baisser les résultats d'un étudiant et l'amener à abandonner l'école. Ainsi, l'élaboration d'un système de reconnaissance des expressions faciales (FER) en temps réel peut donc avoir une influence significative dans le domaine de l'éducation. Un système de reconnaissance automatique des expressions faciales (FER) est un système informatique qui tente de classer automatiquement les changements ou les déformations des composants du visage dans des classes abstraites. En général, un système FER comprend trois étapes : la détection des visages, l'extraction et la représentation des caractéristiques faciales, et la reconnaissance des expressions faciales. L'étape de détection des visages cherche à localiser automatiquement la région du visage dans une image ou une séquence d'image d'entrée. En tant que première étape du système FER, sa fiabilité a une influence majeure sur les performances et une facilité d'utilisation

de l'ensemble du système. Après la détection de visage, l'étape suivante consiste à extraire et à représenter l'information sur l'expression faciale à reconnaître en utilisant des caractéristiques géométriques et d'apparence. Le processus d'extraction forme une description de haut niveau de l'expression en fonction des données des pixels de l'image. Cette description, communément appelée "vecteur de caractéristiques", est utilisée pour la classification ultérieure des expressions faciales.

L'état émotionnel, la concentration et l'engagement des étudiants ont récemment fait l'objet d'une grande attention, car ils ont été associés à d'importants indicateurs de la qualité de l'enseignement supérieur, tels que la rétention, les résultats académiques, la satisfaction [8], l'apprentissage et la réputation des universités. Cette réputation joue un rôle clé dans l'évaluation de la qualité des systèmes éducatifs et est fréquemment utilisée par les parents et les étudiants potentiels pour prendre des décisions concernant les établissements d'enseignement supérieur de leur choix. Les universités ayant une forte réputation sont également plus susceptibles d'attirer des financements publics, tels que les subventions de recherche, ce qui peut contribuer à leur excellence académique. Par conséquent, la réputation des universités est un indicateur important de la qualité des services offerts par les systèmes éducatifs et doit être considérée lors des évaluations de la qualité de l'enseignement supérieur. De plus, la capacité des universités à offrir une éducation de qualité est étroitement liée à leur réputation, y compris en période de crise comme la pandémie de COVID-19, où les universités ont dû s'adapter rapidement en offrant des cours à distance ou en adoptant un modèle d'enseignement hybride. Cependant, la réputation des universités peut être influencée par divers facteurs, notamment les performances des étudiants, la qualité des programmes, les ressources et les infrastructures disponibles, ainsi que l'engagement des enseignants et des chercheurs.

Afin d'obtenir une vue d'ensemble plus complète de la qualité de l'enseignement et de l'apprentissage dans un pays ou une région donnée, il est possible d'exploiter des outils d'intelligence artificielle tels que l'analyse de sentiment pour évaluer les perceptions et les opinions des différents acteurs du système éducatif, y compris les étudiants, les enseignants, les parents et les employeurs. L'analyse des sentiments, également appelé "opinion mining", est défini comme le domaine d'étude qui analyse les opinions, les évaluations, les attitudes et les émotions des personnes à partir du texte. Il s'agit de l'un des domaines de recherche les plus actifs dans le traitement du langage naturel (NLP) et il est également utilisé dans l'exploration des données, l'exploration du web, et l'exploration de textes. L'analyse des sentiments peut être effectuée au niveau du mot, de phrase, d'aspect, du concept, ou du document. L'une des tâches les plus populaires de l'analyse des sentiments est la

classification de la polarité (positif, négatif et neutre). Les méthodes de classification proposées dans la littérature peuvent être classées en trois groupes : approches basées sur des lexiques, apprentissage automatique, et approches hybrides.

Jusqu'à présent, aucune étude n'a été menée dans le domaine de l'éducation pour détecter l'engagement des étudiants sourds ou malentendants à partir de leurs expressions faciales, ni pour analyser les sentiments liés à l'enseignement à distance après la pandémie de COVID-19 en général. De plus, aucune étude n'a spécifiquement examiné l'analyse des émotions faciales des étudiants ou la mesure de la réputation des universités dans l'enseignement supérieur marocain en utilisant l'analyse de sentiment à partir de Twitter. Cette thèse a pour objectif de combler ces lacunes en améliorant la prise de décision dans l'enseignement supérieur marocain, avec pour but principal d'adapter l'apprentissage aux apprenants en utilisant une pédagogie différenciée inclusive et une évaluation formative basée sur la reconnaissance des émotions faciales (FER). En outre, cette thèse vise à évaluer la qualité des systèmes éducatifs et à fournir des indicateurs sur la réputation des universités marocaines et l'enseignement à distance après la pandémie de COVID-19 en se basant sur l'analyse des sentiments.

Pour atteindre ces objectifs, cette thèse se compose de deux parties distinctes. La première partie présente deux systèmes de reconnaissance des émotions faciales (FER) visant à améliorer l'évaluation formative et à favoriser la pédagogie différenciée dans des environnements inclusifs, appliqués sur deux groupes d'étudiants distincts : les étudiants normaux et les étudiants sourds ou malentendants. Ces systèmes permettent aux enseignants d'adapter leur approche pédagogique en accord avec les exigences individuelles de chaque élève, ce qui réduit l'écart de performance entre les étudiants handicapés et non handicapés. Le premier système utilise un réseau de neurones convolutif (CNN) pour reconnaître en temps réel les émotions des étudiants normaux à partir de leurs expressions faciales, tandis que le deuxième système FER utilise le Visual Geometry Group (VGG-16) et l'apprentissage par transfert pour détecter le niveau d'engagement des étudiants sourds et malentendants en temps réel à partir de leurs émotions faciales.

La deuxième partie de cette thèse propose deux systèmes d'analyse des sentiments pour évaluer la qualité des systèmes éducatifs. Le premier système permet d'analyser en temps réel les tweets relatifs à douze universités publiques marocaines en utilisant l'apprentissage automatique et le Big Data, en utilisant six algorithmes d'apprentissage automatique tels que la régression logistique, la machine à vecteurs de support (SVM), le Naïve Bayes multinomial, les arbres de décision, les forêts aléatoires et le Boosting de Gradient Extrême (XGBoost). Le deuxième système d'analyse des sentiments est basé sur l'apprentissage profond et permet d'analyser les sentiments des utilisateurs de Twitter à l'égard de

l'apprentissage à distance après la pandémie de COVID-19 dans six disciplines universitaires (psychologie, informatique, mathématiques, physique, chimie et biologie). Ce système utilise un modèle Bi-LSTM basé sur self-attention dont la performance est comparée avec d'autres modèles d'apprentissage profond, notamment les modèles LSTM, Bi-LSTM et CNN-Bi-LSTM.

L'adoption de ces systèmes par les universités marocaines peut avoir un impact considérable sur l'amélioration de l'éducation en renforçant la qualité de l'enseignement supérieur et en offrant une expérience d'apprentissage positive à tous les étudiants, ce qui contribue de manière significative à la réalisation de l'objectif de développement durable (ODD) de l'ONU, en particulier l'ODD 4 qui vise à garantir une éducation de qualité inclusive et équitable pour tous [9]. De plus, ces avancées technologiques peuvent contribuer de manière significative à la croissance durable du pays en stimulant l'innovation dans le secteur de l'éducation et en permettant aux universités marocaines de se positionner comme des acteurs majeurs sur la scène nationale et internationale, renforçant ainsi leur réputation. En outre, ces initiatives technologiques peuvent favoriser l'égalité des chances en permettant aux étudiants sourds et malentendants de participer pleinement à l'apprentissage en classe grâce à des expériences d'apprentissage adaptées à leurs besoins, telles que les systèmes de reconnaissance des émotions faciales qui peuvent améliorer leur concentration, leur compréhension et leurs performances académiques.

La première partie de ce mémoire comprend les chapitres suivants :

- Le chapitre 1 dresse dans un premier temps une description des différents algorithmes d'apprentissage automatique, et d'apprentissage profond utilisés dans la reconnaissance des émotions faciales et l'analyse des sentiments tout en mettant en exergue leurs avantages, leurs inconvénients, et leurs applications dans différents domaines. Le chapitre présente également les notions de base des émotions, les méthodes de reconnaissance des expressions faciales, suivies d'une analyse de l'état de l'art de la reconnaissance des émotions faciales dans le contexte de l'enseignement.
- Le chapitre 2 définit les concepts d'enseignement et d'apprentissage. Il explore ensuite en détail diverses stratégies pour mettre en place une pédagogie différenciée et inclusive, en mettant particulièrement l'accent sur la reconnaissance des émotions faciales. Deux nouvelles approches qui utilisent l'apprentissage profond ont été proposées pour reconnaître les émotions faciales et détecter l'engagement des étudiants.

La deuxième partie de ce mémoire comprend les chapitres suivants :

- Le chapitre 3 présente les bases théoriques de l'analyse de sentiment comme une méthode potentielle d'évaluation de la qualité des systèmes éducatifs. Les différentes tâches, niveaux et étapes du processus d'analyse y sont détaillés, ainsi que l'état de l'art de son application dans le domaine de l'enseignement supérieur.
- Le chapitre 4 détaille une méthodologie fondée sur l'apprentissage profond pour examiner les opinions des utilisateurs de Twitter concernant l'apprentissage à distance après la pandémie de COVID-19. Ensuite il décrit en détail la mise en place d'une analyse de sentiment en temps réel des tweets portant sur les universités publiques marocaines.

Partie I

**Adaptation des méthodes
d'apprentissage aux étudiants**

Chapitre 1

État de l'art des algorithmes de Machine learning et de Deep learning dans le contexte de l'enseignement

1.1. Introduction

L'intelligence artificielle (IA) a révolutionné les technologies de l'information et a profondément impacté notre mode de vie. Grâce à des avancées majeures dans les domaines de l'apprentissage automatique (Machine Learning) et de l'apprentissage profond (Deep Learning), l'IA permet aux ordinateurs d'apprendre à partir de données et de résoudre des problèmes complexes de manière approximative. Cette flexibilité et cette robustesse ont conduit à une application généralisée de ces algorithmes dans de nombreux domaines, allant de la robotique au contrôle des vols d'avion. Dans ce chapitre, nous allons explorer les algorithmes d'apprentissage automatique dans la section 1.2, ainsi que les algorithmes d'apprentissage profond dans la section 1.3. Nous examinerons également les avantages et les inconvénients de chaque algorithme, afin de mieux comprendre leurs capacités et leurs limites respectives. De plus, nous nous pencherons sur les diverses applications de ces algorithmes dans différents domaines, afin d'illustrer leur impact concret sur notre société. Une attention particulière sera accordée à la reconnaissance des émotions faciales, un domaine d'application de l'IA qui suscite un intérêt croissant (section 1.4).

1.2. Algorithmes de l'apprentissage automatique

L'apprentissage automatique ou Machine Learning est un sous-domaine de l'intelligence artificielle qui fournit aux machines la capacité d'apprendre automatiquement à partir de données historiques afin de reconnaître des caractéristiques, trouver des résolutions pour des nouveaux problèmes et faire des prédictions sans être spécifiquement programmée. L'objectif est de trouver une fonction de prédiction f représentée par un

algorithme d'apprentissage afin de déterminer la relation entre le jeu de données X et la variable à prédire Y représentée comme suit :

$$Y = f(X) \tag{1.1}$$

Les algorithmes d'apprentissage sont regroupés en quatre catégories principales, à savoir supervisé, non supervisé, semi-supervisé, et par renforcement. L'apprentissage supervisé consiste à utiliser des données d'apprentissage annotées pour déduire une fonction prédictive qui met en relation un groupe de variables d'entrée et de variables de sorties respectives. Ce processus d'entraînement permet au modèle d'apprendre à partir des données afin de prédire la valeur de sortie d'une manière précise. Les tâches supervisées les plus fréquentes sont la "classification" qui prédit une catégorie, par exemple les prédictions météorologiques, et la "régression" qui prédit une valeur de sortie continue en fonction de la variable indépendante d'entrée. Par exemple, la prédiction du sentiment à partir d'un texte ou d'une image.

L'apprentissage non supervisé a pour objectif d'examiner plusieurs occurrences d'un vecteur X pour apprendre la distribution de probabilité $p(X)$ définie comme suit :

$$p(X) = \prod_{i=1}^n p(x_i | x_1, \dots, x_{i-1}) \tag{1.2}$$

Cette formule signifie que nous pouvons résoudre le problème non supervisé de la modélisation de $p(X)$ en le divisant en n problèmes d'apprentissage supervisé. Ce type d'apprentissage est largement utilisé pour l'extraction de caractéristiques génératives et le regroupement de résultats. Les tâches d'apprentissage non supervisé les plus courantes incluent le regroupement, l'assimilation de caractéristiques, l'évaluation de la densité, la diminution de la dimensionnalité, la détection d'anomalies, et l'identification de règles d'association.

L'apprentissage semi-supervisé combine des méthodes supervisées et non supervisées en utilisant à la fois des données étiquetées et non étiquetées. Les données étiquetées sont utilisées pour mettre à jour les poids du modèle en utilisant l'apprentissage supervisé, tandis que les données non étiquetées sont regroupées en se basant sur des similitudes dans un espace de caractéristiques. Les domaines d'application de l'apprentissage semi-supervisé incluent la traduction automatique, la détection des fraudes, l'étiquetage des données et la classification des textes. L'apprentissage par renforcement est une forme d'apprentissage automatique dans laquelle les logiciels et les machines évaluent automatiquement le comportement optimal dans un environnement donné pour améliorer leur efficacité. Ce type d'apprentissage se base sur des récompenses et des pénalités, et vise à utiliser les

informations de l'environnement pour prendre des mesures qui augmentent les récompenses ou minimisent les risques. Il est utilisé pour former des modèles d'intelligence artificielle (IA) qui optimisent l'efficacité opérationnelle de systèmes sophistiqués tels que la robotique et la conduite autonome.

Les principaux algorithmes d'apprentissage supervisé comprennent la régression linéaire, la régression logistique, les machines à vecteurs de support (SVM), l'algorithme de Naïve Bayes, les arbres de décision, les forêts aléatoires et le Boosting de gradient extrême (XGBoost).

1.2.1. Régression linéaire

La régression linéaire simple est un algorithme d'apprentissage supervisé permettant de prédire la relation linéaire qui peut exister entre deux variables continues, l'une agissant en tant que variable indépendante et la seconde en tant que variable dépendante. Le modèle de régression linéaire simple est représenté graphiquement par une droite de régression qui correspond le mieux aux données (Figure 1.1). Cette droite est définie par l'équation suivante :

$$y = w^T x + b \quad (1.3)$$

Où $w \in \mathbb{R}^n$ est un vecteur de paramètres w_i , $x \in \mathbb{R}^n$ est un vecteur de caractéristiques x_i , b est le paramètre de biais, et y est la valeur prédite par le modèle. Le modèle de régression linéaire simple est largement utilisé pour la prévision des changements climatiques [10], et l'analyse des marchés boursiers [11].

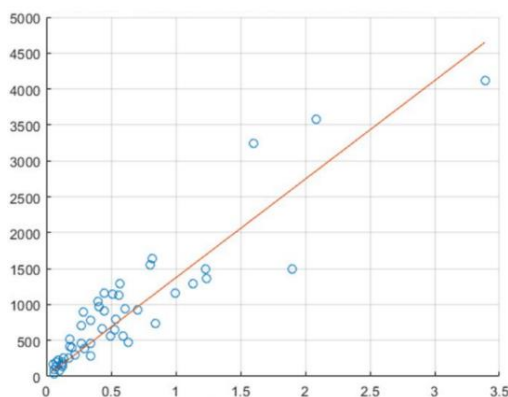


Figure 1.1 – Exemple d'une régression linéaire.

La régression linéaire multiple est une généralisation de la régression linéaire simple. Elle consiste à trouver la relation entre deux variables indépendantes explicatives ou plus et une variable dépendante. Le modèle de régression linéaire multiple est représenté par l'équation suivante :

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p + \varepsilon \quad (1.4)$$

Où y est la variable dépendante prédite par le modèle, x_1, \dots, x_n sont les variables indépendantes du modèle, b_0 est l'interception, b_1, \dots, b_p sont les coefficients de régression pour les variables indépendantes explicatives et ε est le terme d'erreur qui représente la portion de la variable dépendante qui ne peut pas être expliquée par les variables indépendantes. La régression linéaire multiple est utilisée pour évaluer la qualité de l'air [12] et dans des domaines médicaux [13, 14]. Cependant, elle peut avoir du mal à s'adapter aux ensembles de données complexes, car la relation entre les variables n'est pas toujours linéaire dans la plupart des scénarios réels. De plus, les valeurs aberrantes peuvent avoir des effets considérables sur la régression, ce qui limite sa précision.

1.2.2. Régression logistique

La régression logistique est un algorithme de classification supervisée utilisé lorsque la variable à prédire est binaire, prenant seulement deux valeurs comme 0 ou 1, oui ou non. Il est similaire à la régression linéaire, mais utilise le logarithme des odds ou des cotes de la variable à prédire pour représenter la courbe. Son équation est la suivante :

$$p = \frac{1}{1+e^{-(b_0+b_1x)}} \quad (1.5)$$

Où b_0 représente une constante qui décale la courbe vers la droite et la gauche, tandis que la pente b_1 détermine l'inclinaison de la courbe. La fonction Sigmoïde, définie sur \mathbb{R} à valeur dans $[0,1]$, est un cas particulier de la régression logistique lorsque $b_0=0$ et $b_1=1$. Elle correspond graphiquement à une courbe en forme S qui a pour limite 0 et 1 quand x tend vers $-\infty$ et $+\infty$ et $y=0.5$ quand x égale à 0 (Figure 1.2), et elle s'écrit de la manière suivante :

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (1.6)$$

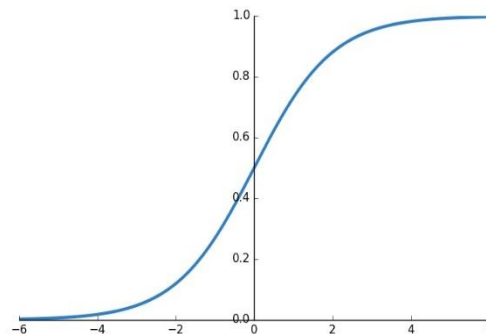


Figure 1.2 – Fonction Sigmoïde.

La régression logistique multiple est utilisée pour prédire une seule variable binaire en utilisant une ou plusieurs variables explicatives définies comme suit :

$$p = \frac{1}{1+e^{-(b_0+b_1x+b_2x+\dots+b_px_p)}} \quad (1.7)$$

La régression logistique multinomiale est une technique de classification qui étend l'algorithme de régression logistique pour résoudre les problèmes des résultats multi-classes en utilisant une ou plusieurs variables indépendantes. Ce modèle utilise la fonction Softmax pour prédire les probabilités de la variable dépendante catégorique, qui a deux ou plusieurs classes $i = 1, \dots, k$. La fonction Softmax représentée dans l'équation ci-dessous, donne toujours un résultat entre 0 et 1.

$$Softmax(y_i) = \frac{e^{y_i}}{\sum_{j=1}^k e^{y_j}} \quad (1.8)$$

La régression logistique binomiale a été utilisée dans divers domaines, tels que la justice [15] et le transport [16]. La régression logistique multiple a été appliquée pour identifier la source des éruptions d'eau dans les mines [17] et les facteurs associés à la mortalité infantile [18]. De plus, la régression logistique multinomiale a été utilisée pour évaluer le risque de crédit bancaire [19, 20] et l'insécurité alimentaire dans les ménages ruraux au Népal [21]. Bien que la régression logistique soit facile à mettre en œuvre, sa principale limitation réside dans sa capacité à résoudre uniquement des problèmes où les données sont linéairement séparables.

1.2.3. Machines à vecteurs de support (SVM)

Les machines à vecteurs de support (SVM) sont un algorithme d'apprentissage supervisé inventé par Vapnik et Lerner [22], sous le nom de la méthode du portrait généralisé en 1963, pour la classification et a été étendu à la régression. Le SVM binaire est un algorithme qui sépare un ensemble de données D en deux classes avec une droite appelée hyperplan séparateur.

$$D = \{(x_1|y_1), (x_2|y_2), \dots, (x_m|y_m)\} \quad (1.9)$$

Où x_i est un vecteur réel à n dimensions et y_i indique la classe 1 ou -1 à laquelle le point x_i appartient. L'équation de l'hyperplan séparateur est représentée comme suit :

$$h(x) = w^T x + b \quad (1.10)$$

Où w est le vecteur de poids, b est le biais, et x est l'entrée à classer. Pour chaque point x_i dans D :

$$\begin{cases} h(x_i) > 0 & \text{si } y_i = 1 \\ h(x_i) < 0 & \text{si } y_i = -1 \end{cases} \quad (1.11)$$

Ces conditions sont équivalentes à :

$$y_i(h(x_i)) > 0 \Leftrightarrow y(w^T x_i + b) > 0 \quad (1.12)$$

Le machine à vecteurs des support (SVM) binaire cherche l'hyperplan séparateur $h(x)$ qui passe au milieu entre les points de la classe 1 les plus proches de ce séparateur qui se trouvent à la même distance de l'hyperplan séparateur que les points les plus proches de la classe -1. Cette distance est appelée marge. Un exemple d'hyperplan optimal maximisant la marge dans un espace bidimensionnel est représenté dans la figure 1.3.

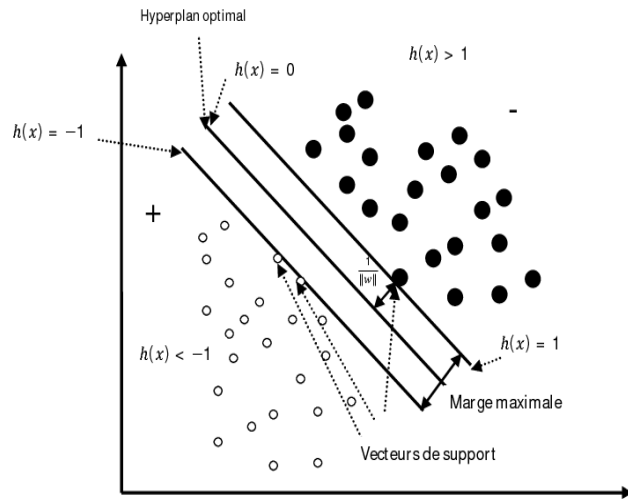


Figure 1.3 – Visualisation d'hyperplan de la machine à vecteurs de support (SVM).

Les éléments les plus proches de l'hyperplan séparateur sont appelé "vecteurs de support". Dans le cas où les données sont linéairement séparables, il n'y a pas des données d'apprentissage dans la marge, c'est-à-dire entre les deux hyperplans représentés comme suit :

$$w^T x_i + b = 1 \quad \text{et} \quad w^T x_i + b = -1 \quad (1.13)$$

Par conséquent, les équations (1.11) et (1.12) sont révisées en équations (1.14) et (1.15) suivantes :

$$\begin{cases} h(x_i) \geq 1 & \text{si } y_i = 1 \\ h(x_i) \leq -1 & \text{si } y_i = -1 \end{cases} \quad (1.14)$$

$$y_i(h(x_i)) \geq 1 \Leftrightarrow y(w^T x_i + b) \geq 1, \quad \forall (x_i, y_i) \in D \quad (1.15)$$

Si x_s est un vecteur de support et $H = \{x | w^T x + b = 0\}$ est un hyperplan séparateur alors la marge est définie par :

$$\text{Marge} = 2d(x, H) = 2 \frac{|w^T x_s + b|}{\|w\|} \quad (1.16)$$

En imposant pour les vecteurs de support x_s la condition de normalisation suivante $|w^T x_s + b| = 1$, alors l'équation (1.16) devient :

$$\text{Marge} = \frac{2}{\|w\|} \quad (1.17)$$

L'optimisation de la marge se transforme en une minimisation de $\|w\|$. On arrive donc au problème d'optimisation quadratique suivant :

$$\begin{cases} \min_{w, b} \frac{1}{2} \|w\|^2 \\ \text{tel que } y(w^T x_i + b) \geq 1, \forall (x_i, y_i) \in D \end{cases} \quad (1.18)$$

La solution de ce problème (1.18) donne les multiplicateurs de Lagrange optimaux α_i^* en appliquant la technique des multiplicateurs de Lagrange [23]. En substituant le vecteur de poids w par sa valeur optimisée $\sum_{i=1}^n \alpha_i x_i$, la fonction de décision permet de classer une nouvelle observation x , est représentée par :

$$h(x) = \sum_{i=1}^n \alpha_i^* y_i x_i^T x + b \quad (1.19)$$

Lorsque les données ne peuvent pas être séparées linéairement, c'est-à-dire qu'aucun hyperplan ne peut diviser les données en deux catégories, on utilise la technique du noyau (ou Kernel trick) pour transformer un problème non linéaire en un problème linéaire en utilisant une fonction non linéaire $\varphi: E \rightarrow E'$. L'hyperplan de séparation devient une fonction linéaire dans l'espace E' transformé et une fonction non linéaire dans l'espace E . L'obtention de l'hyperplan dans l'espace E' requiert la connaissance de la fonction noyau $k: E \times E' \rightarrow \mathbb{R}$ représentée comme suit :

$$k(x, x') = \varphi(x)^T \cdot \varphi(x') \quad (1.20)$$

Selon le théorème de Mercer [24], la fonction noyau k est valide, si et seulement si elle est symétrique continue et semi-définie positive. La fonction de classification dans le cas non linéaire en utilisant la fonction k , est définie comme suit :

$$h(x) = \sum_{i=1}^n \alpha_i^* y_i k(x_i, x) + b \quad (1.21)$$

Parmi les fonctions noyau les plus fréquemment employées, il y a :

- Le noyau à fonction de base radiale gaussienne (RBF) :

$$k(x, x') = e^{-\gamma \|x - x'\|^2}, \gamma > 0 \quad (1.22)$$

- Le noyau polynomial de degré d :

$$k(x, x') = (x \cdot x' + 1)^d \quad (1.23)$$

- Le noyau Sigmoidé :

$$k(x, x') = \tanh(ax^T x' + b) \quad (1.24)$$

L'algorithme machine à vecteurs de support (SVM) a été utilisé pour la classification des textes [25, 26], et des images [27], ainsi que la reconnaissance vocale [28], et suivi d'objets [29]. Le SVM fonctionne relativement bien lorsqu'il existe une marge claire entre les classes. Et parmi les faiblesses du SVM c'est qu'il n'est pas très performant lorsque les ensembles de données sont grands ou comporte plus de bruit. Cependant, le choix du bon noyau est une autre faiblesse de cet algorithme qui peut être surmontée par faire varier la fonction noyau dans la recherche d'hyperparamètres.

1.2.4. Naïve bayes

L'algorithme Naïve Bayes est un algorithme d'apprentissage supervisé basé sur le théorème de Bayes, exprimé par l'équation (1.25), avec une hypothèse d'indépendance entre les variables explicatives.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1.25)$$

Où A et B sont deux évènements, P(A) est la probabilité a priori de A, P(B) est la probabilité a priori de B avec P(B) ≠ 0, P(A|B) est la probabilité a posteriori de A sachant B, et P(B|A) est la probabilité a posteriori de B sachant A. Étant donné un vecteur de caractéristiques $X = (x_1, x_2, \dots, x_n)$ et une variable cible y, l'équation (1.25) sera :

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (1.26)$$

Cette dernière peut être exprimée comme suit :

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)p(y)}{P(x_1)P(x_2)\dots p(x_n)} \quad (1.27)$$

Ce qui est équivalent à :

$$P(y|x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i|y)}{P(x_1)P(x_2)\dots p(x_n)} \quad (1.28)$$

Et comme le dénominateur reste constant pour toutes les valeurs, la probabilité postérieure $P(y|x_1, \dots, x_n)$ est représentée comme suit :

$$P(y|x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i|y) \quad (1.29)$$

Le classificateur Naïve Bayes utilise la règle de décision nommée la règle de maximum de probabilité a posteriori, définie dans l'équation (1.30), pour trouver la sortie y avec la probabilité maximale.

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y) \quad (1.30)$$

L'algorithme Naïve Bayes multinomial (MNB) est une instance de l'algorithme Naïve Bayes, généralement appliquée au texte dans lequel les données w_i sont représentées sous forme d'un vecteur de caractéristiques W indiquant les occurrences de chaque terme i apparu dans le document D déjà affecté à une classe parmi les classes $\mathcal{C} = (c_1, \dots, c_k)$. Dans ce cas, la probabilité a posteriori est définie comme suit :

$$P(W|c_k) = \frac{(\sum_{i=1}^n w_i)!}{\prod_{i=1}^n w_i!} \prod_{i=1}^n P_{k_i}^{w_i} \quad (1.31)$$

La règle de maximum de probabilité a posteriori du classificateur Naïve Bayes multinomial est représentée dans l'équation suivante :

$$c_k = \operatorname{argmax}_{c_k} P(c_k) \prod_{i=1}^n P(w_i|c_k) \quad (1.32)$$

L'algorithme Naïve Bayes est utilisé dans divers domaines, tels que la médecine [30-32] et l'analyse de sentiment [33]. Il est apprécié pour sa simplicité, sa faible demande en données d'apprentissage et sa rapidité pour des prédictions en temps réel. Cependant, il présente comme principale limite son hypothèse d'indépendance mutuelle entre les attributs, qui peut être irréaliste dans des situations réelles où les prédicteurs sont rarement totalement indépendants.

1.2.5. Arbres de décisions

L'arbre de décision est un algorithme d'apprentissage supervisé composé de nœuds et de feuilles. Chaque nœud représente un attribut spécifique, tandis que chaque feuille désigne une classe. L'algorithme utilise l'entropie, définie par l'équation (1.33), à chaque niveau de l'arbre pour évaluer la diversité de la variable cible dans chaque catégorie en utilisant une ou plusieurs variables aux niveaux supérieurs de la hiérarchie. Le gain d'information, défini par l'équation (1.34), est une mesure de la réduction de l'entropie. La variable qui présente la plus grande réduction d'entropie, c'est-à-dire le gain d'information le plus important, est sélectionnée comme nœud de l'arbre à chaque niveau de décision.

$$\operatorname{Entropy}(AC) = -\sum_{i=1}^n P_i \log_2 P_i \quad (1.33)$$

$$\operatorname{Gain}(AC, V) = \operatorname{Entropy}(AC) - \sum_{v \in \text{valeurs}(V)} \frac{|AC_v|}{|AC|} \operatorname{Entropy}(AC_v) \quad (1.34)$$

Où, AC est l'attribut cible, n est le nombre de catégories de l'attribut cible, P_i est proportion du nombre de catégories i , $Valeurs(V)$ est le domaine énuméré de variable V , et AC_v est le sous-ensemble de AC où la valeur correspondante de AC est v .

Les arbres de décision ont été employés dans divers domaines, tels que la médecine [34, 35] et l'économie [36]. Ils sont intuitifs et faciles à interpréter. Cependant, ils sont sensibles aux variations des données, ce qui peut entraîner des prédictions considérablement différentes. De plus, les arbres de décision sont moins efficaces pour résoudre les problèmes de régression.

1.2.6. Forêts aléatoires

Une forêt aléatoire [37] est une méthode d'apprentissage supervisé utilisée pour la classification et la régression. Elle combine l'approche de Bagging avec la sélection aléatoire de caractéristiques pour construire un ensemble d'arbres de décision avec une variabilité contrôlée. Chaque arbre de l'ensemble agit comme un classificateur de base pour déterminer la classe d'une instance non annotée. Les prédictions sont effectuées en utilisant un vote majoritaire parmi les classificateurs. Les forêts aléatoires ont été appliquées à divers domaines tels que la neuroimagerie dans la maladie d'Alzheimer [38], la prédiction du roulement du personnel [39], et la prédiction du rayonnement solaire [40]. Les forêts aléatoires sont moins susceptibles de surapprentissage que les arbres de décision, mais elles peuvent être sensibles à de petits changements dans les données et l'algorithme peut être plus complexe que d'autres algorithmes.

1.2.7. Boosting de gradient extrême (XGBoost)

Le Boosting de gradient extrême (XGBoost) est un algorithme d'apprentissage supervisé d'ensemble similaire aux forêts aléatoires. Il construit un modèle final en combinant plusieurs modèles individuels, généralement des arbres de décision boostés par le gradient. Le XGBoost utilise le gradient pour minimiser la fonction de perte, similaire aux réseaux de neurones, en optimisant les poids avec une descente de gradient extrême. Il utilise également des approximations plus détaillées en calculant les gradients de second ordre pour minimiser la perte et applique une régularisation avancée (L1 et L2) pour réduire le surapprentissage et améliorer la généralisation. Le XGBoost a été appliqué dans les domaines de la médecine [41-43] et de la finance [44, 45]. Il est rapide à interpréter, évolutif et capable de gérer de grandes quantités de données. Cependant, un ajustement incorrect des paramètres peut entraîner du surapprentissage.

1.3. Algorithmes de l'apprentissage profond

L'apprentissage profond, également connu sous le nom de Deep Learning (DL), représente une subdivision de l'apprentissage automatique qui exploite des couches de neurones pour modéliser des données. Contrairement aux méthodes traditionnelles qui nécessitent une extraction manuelle des caractéristiques, le deep learning utilise des transformations non linéaires pour extraire et transformer les caractéristiques de manière automatique. Il permet aux ordinateurs d'apprendre à partir d'une représentation hiérarchique des données, où les caractéristiques de haut niveau sont dérivées des caractéristiques de bas niveau. Cela élimine le besoin d'une extraction manuelle des caractéristiques et permet d'obtenir de bons résultats en passant directement les données au réseau de neurones.

1.3.1. Principe de base des réseaux de neurones artificiels

En biologie, l'apprentissage implique l'établissement de nouvelles connexions entre les neurones ou à modifier les connexions existantes. Les cellules neuronales émettent des signaux électriques par l'intermédiaire de l'axone lorsque le potentiel dans le corps de la cellule atteint un certain niveau appelé potentiel d'action. Le cerveau humain contient 8×10^{10} neurones, avec en moyenne chaque neurone connecté à 10000 autres neurones. Les synapses s'adaptent en ajustant la quantité de transmetteurs et de récepteurs, mettant en œuvre la plasticité du cerveau. Les synapses réalisent la transmission chimique des signaux, bien qu'elles soient lentes, elles fonctionnent en parallèle. Ces principes biologiques ont inspiré la construction de réseaux de neurones artificiels, appelés perceptrons, qui modélisent les neurones qui additionnent les signaux d'entrée, l'adaptabilité des synapses et une fonction d'activation similaire au corps d'une cellule neuronale. Les réseaux de neurones artificiels capturent également la plasticité et la redondance du cerveau, notamment dans les réseaux plus grands et plus profonds (Figure 1.4).

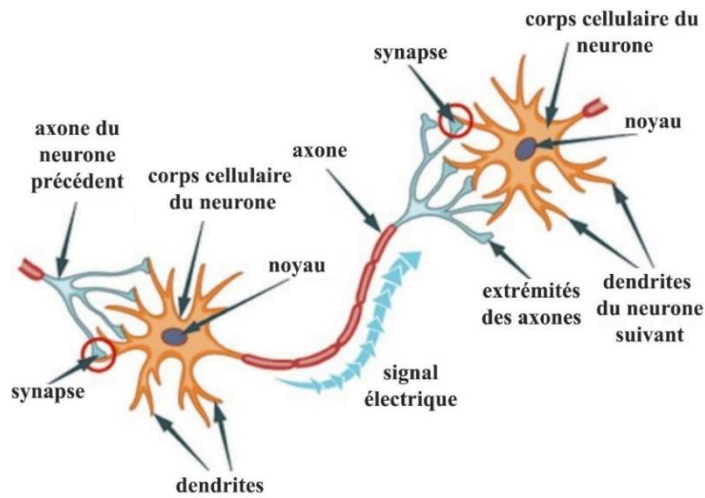


Figure 1.4 – Structure d’une cellule neuronale, montrant les moyens de communication entre deux cellules connectées.

Le premier modèle informatique d’un neurone a été proposé par les mathématiciens Warren McCulloch et Walter Pitts [46] en 1943, dans le but de reproduire approximativement le fonctionnement du cerveau humain (Figure 1.5). Leur approche ne peut pas traiter les entrées non booléennes et donne un poids égal à chaque entrée.

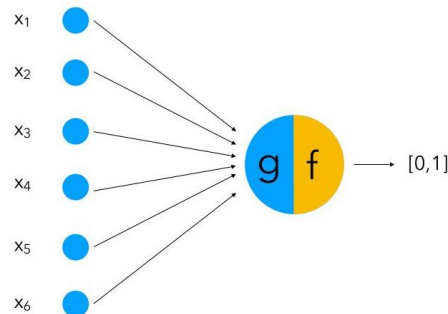


Figure 1.5 – Réseau de neurone simple proposé par Warren McCulloch et Walter Pitts [46].

En 1957, Frank Rosenblatt a proposé les perceptrons pour résoudre les problèmes rencontrés par le neurone de McCulloch-Pitts. Un perceptron (Figure 1.6) est un réseau de neurones composé d’entrées pondérées et d’une fonction d’activation qui génère une sortie à partir de la somme pondérée des entrées. L’activation de sortie du neurone o_j , est calculée dans l’équation (1.35), où φ est la fonction d’activation.

$$o_j = \varphi\left(\sum_{i=1}^n w_{ij}x_i + \theta_j\right) \quad (1.35)$$

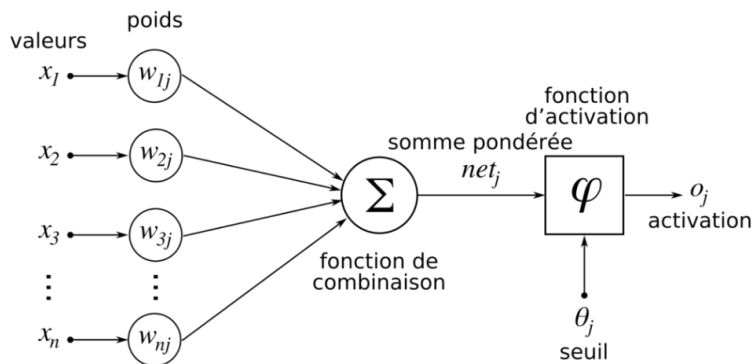


Figure 1.6 – Modélisation d'un neurone artificiel appelé le Perceptron.

Un réseau de neurones à propagation avant (feedforward neural network) est un type de réseau neuronal artificiel où les informations circulent dans une seule direction. Il est composé de trois types de couches : la couche d'entrée qui reçoit les données brutes, les couches cachées qui transmettent les informations entre les couches d'entrée et de sortie, et la couche de sortie qui envoie les informations du réseau vers l'extérieur. Autrement dit, un neurone de la couche L reçoit des entrées de tous les neurones du niveau précédent $L - 1$. La couche de sortie est spéciale et est définie en fonction de la nature du problème à résoudre. Par exemple, pour la régression, la couche de sortie peut ne comprendre qu'un seul neurone, dont l'activation modélise la variable o_j . Pour la classification, la couche de sortie peut comporter autant de neurones que de valeurs de classe, où chaque activation rend compte d'une probabilité de classe. Il existe deux types de réseaux à propagation avant : le perceptron mono-couche, doté d'une seule couche cachée (Figure 1.7a), et le perceptron multi-couche qui comporte plusieurs couches cachées (Figure 1.7b).

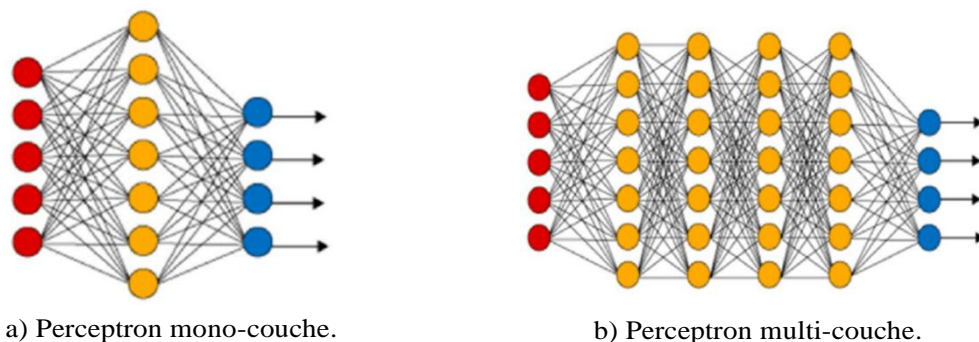


Figure 1.7 – Types de perceptrons.

La phase d'apprentissage d'un réseau de neurones consiste à minimiser une fonction de coût qui évalue l'erreur entre les données de formation et les données prédites par le réseau. Cette procédure est appelée rétropropagation (ou backpropagation [48]) et utilise une descente de gradient. Le taux d'apprentissage (ou learning rate) est un hyperparamètre crucial dans la descente de gradient pour optimiser les réseaux de neurones profonds. Un

choix inapproprié des hyperparamètres peut entraîner un surapprentissage, un problème courant lors de l'apprentissage de modèles de deep learning. Pour le surmonter, la technique du Dropout peut être utilisée pour désactiver aléatoirement certains neurones du réseau à chaque itération, avec une probabilité donnée p . Différentes architectures profondes sont utilisées pour représenter diverses sources de données. Celles-ci incluent les réseaux de longue mémoire à court terme (LSTM), les réseaux bidirectionnels de longue mémoire à court terme (Bi-LSTM), les réseaux de neurones convolutifs (CNN) et les réseaux de neurones convolutifs profonds (DCNN).

1.3.2. Réseaux de longue mémoire à court terme (LSTM)

Le LSTM (Long Short-Term Memory) est un type de réseau de neurones récurrents (RNN) qui possède une capacité de mémoire améliorée, permettant de se souvenir des sorties de chaque nœud sur une période plus longue afin de produire efficacement les résultats des nœuds suivants. Ce modèle a été spécialement conçu pour résoudre les problèmes de disparition du gradient (Vanishing Gradient Problem) qui se produisent généralement lors de l'apprentissage de dépendances à long terme, même avec de longs décalages temporels. Le Vanilla LSTM [49] est l'architecture LSTM la plus populaire, où chaque unité est composée d'une cellule, d'une porte d'entrée, d'une porte de sortie et d'une porte d'oubli [50] (Figure 1.8).

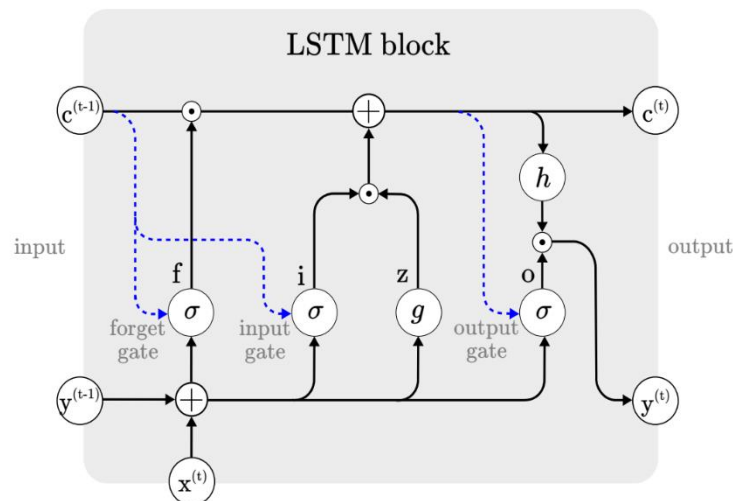


Figure 1.8 – Architecture d'un bloc Vanilla LSTM.

- *Entrée du bloc (Block input)* : combine l'entrée actuelle $x^{(t)}$ et la sortie de l'unité LSTM $y^{(t-1)}$ lors de la dernière itération pour mettre à jour la composante d'entrée du bloc.
- *Porte d'entrée du bloc (Input gate)* : met à jour la porte d'entrée en combinant

l'entrée actuelle $x^{(t)}$, la sortie $y^{(t-1)}$ et la valeur de la cellule $c^{(t-1)}$ de la dernière itération.

- *Porte d'oubli (Forget gate)* : détermine quelles informations doivent être supprimées des états cellulaires précédents $c^{(t-1)}$.
- *Cellule (Cell)* : calcule la valeur de la cellule en combinant l'entrée du bloc, la porte d'entrée et la porte d'oubli avec la valeur de la cellule précédente.
- *Porte de sortie (Output gate)* : calcule la porte de sortie en combinant l'entrée actuelle, la sortie précédente et la valeur de la cellule précédente.
- *Sortie du bloc (Block output)* : calcule la sortie du bloc en combinant la valeur actuelle de la cellule avec la valeur actuelle de la porte de sortie.

LSTM est utilisée dans divers domaines. Par exemple, elle a été utilisée dans [51] pour modéliser les séries temporelles d'observations et prédire la demande de taxi à New York. Dans [52], les auteurs ont démontré que LSTM peut également détecter l'activité à partir de données provenant de capteurs de santé portables, tels que ceux obtenus via un ordinateur portable. Dans [53, 54], deux couches LSTM ont été utilisées pour construire un classificateur binaire et apprendre le contexte des phrases. Cependant, un inconvénient majeur de cet algorithme réside dans le fait qu'il requiert une quantité de mémoire plus importante pour l'entraînement.

1.3.3. Réseaux bidirectionnels de longue mémoire à court terme (Bi-LSTM)

Bidirectional LSTM (Bi-LSTM) est un modèle de traitement de séquence utilisant deux LSTM : un fonctionne dans la direction avant (Forward direction) et l'autre dans la direction arrière (Backward direction), comme illustré dans la figure 1.9. Les LSTM forward et backward sont décrits par les équations (1.42) et (1.43), respectivement, où m représente la longueur maximale des caractéristiques.

$$\vec{h}_{f_{lstm}} = \overset{\rightarrow}{LSTM}(p_i), \quad i \in [1, m] \quad (1.42)$$

$$\overset{\leftarrow}{h}_{b_{lstm}} = \overset{\leftarrow}{LSTM}(p_i), \quad i \in [m, 1] \quad (1.43)$$

Le résultat de la concaténation de l'équation (1.42) et l'équation (1.43) est représenté comme suit :

$$h_t = LSTM[\vec{h}_{f_{lstm}}, \overleftarrow{h}_{b_{lstm}}] \quad (1.44)$$

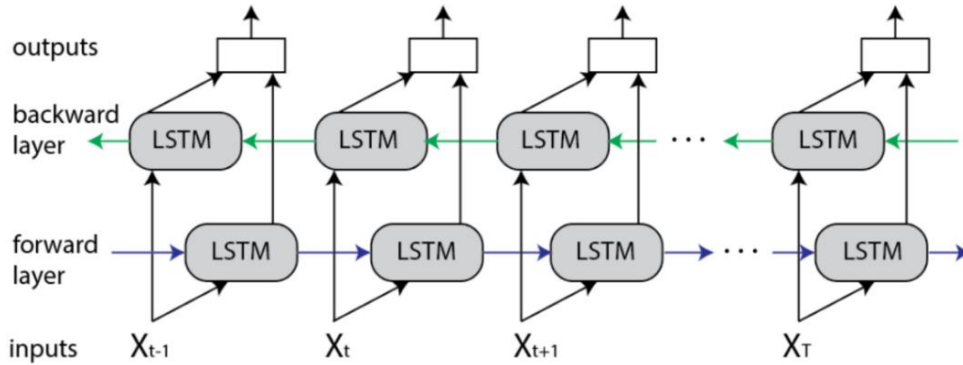


Figure 1.9 – Architecture de Bi-LSTM.

Bi-LSTM a été utilisé dans [55] avec l'architecture CBOW de Word2Vec afin d'apprendre de manière efficace la fonction générale de l'embedding pour des contextes de phrases de longueur variable autour du mot cible. Bi-LSTM a été aussi utilisé dans [56] pour effectuer une classification des sentiments et découvrir diverses questions liées aux opinions publiques de COVID-19. Tout comme le LSTM, Bi-LSTM est coûteux en temps et en mémoire pour les données volumineuses.

1.3.4. Réseau de neurone convolutif (CNN)

Le réseau de neurones convolutif (CNN) est un type de réseau de neurones artificiels inspiré des réseaux de neurones biologiques. Il est particulièrement conçu pour reconnaître des données bidimensionnelles telles que des images et des vidéos. Les images peuvent être directement utilisées comme entrée du réseau, ce qui évite le processus complexe d'extraction de caractéristiques et de reconstruction des données dans les algorithmes traditionnels de reconnaissance d'images. La figure 1.10 présente un exemple de réseau de neurones convolutifs à deux couches cachées. Chaque couche de ce réseau prend les sorties de la couche précédente comme des entrées. L'architecture générique de CNN, comprend différentes couches comme des couches de convolution, des couches de pooling, et des couches entièrement connectées.

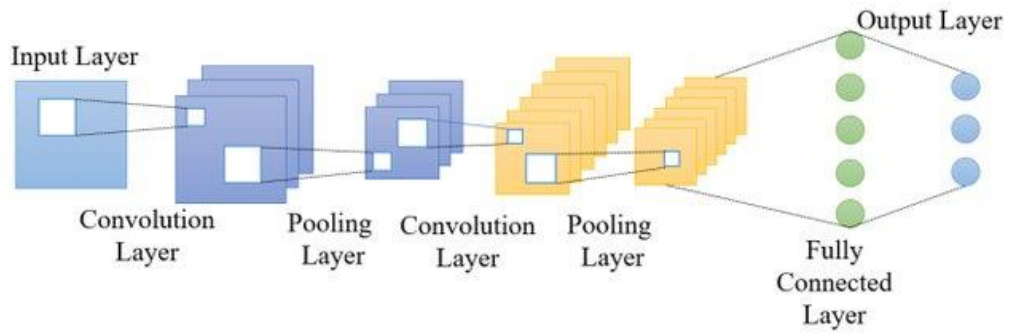


Figure 1.10 – Architecture générique de CNN.

- *Couche de convolution* : est la première couche qui recueille les caractéristiques d'une image d'entrée. Le but principal de la convolution est de préserver la relation spatiale entre les pixels en apprenant les caractéristiques de l'image à travers l'utilisation de petits carrés de données d'entrée. Elle réalise un produit scalaire entre deux matrices où l'une est l'image et l'autre est un noyau. La formule de convolution est représentée dans l'équation suivante :

$$\begin{aligned} y[m,n] &= h[m,n] * x[m,n] \\ &= \sum^j \sum^i h[i,j]x[m-i,n-j] \end{aligned} \quad (1.45)$$

Où $y[m,n]$ est la sortie de la couche suivante, h est l'image de l'entrée, x est la matrice de filtre, m et n sont les indices des lignes et des colonnes de la matrice de résultat, et $*$ est l'opération de convolution. Chaque couche de convolution est suivie d'une fonction d'activation non linéaire telle que la fonction d'unité linéaire rectifiée (ou rectified linear activation function (ReLU)) illustrée dans la figure 1.11 et définie mathématiquement comme suit :

$$R(x) = \begin{cases} \max(0, x), & \text{si } x \geq 0 \\ 0, & \text{si } x < 0 \end{cases} \quad (1.46)$$

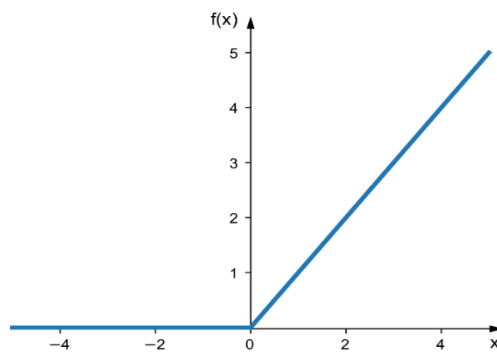


Figure 1.11 – Fonction d'unité linéaire rectifiée (ReLU).

- *Couche de pooling* : réduit la dimensionnalité de chaque carte de caractéristique (ou feature map) en conservant les informations les plus importantes afin de rendre le réseau invariant aux petites transformations, distorsions et translations de l'image d'entrée. Il existe différents types de pooling tels que max pooling et average pooling. Le max pooling renvoie la valeur maximale de toutes les valeurs à une position du noyau dans une image. Et average pooling calcule la moyenne des éléments présents dans la région de la carte des caractéristiques couverte par le noyau. Les figures 1.12 présentent un exemple de max pooling et d'average pooling.

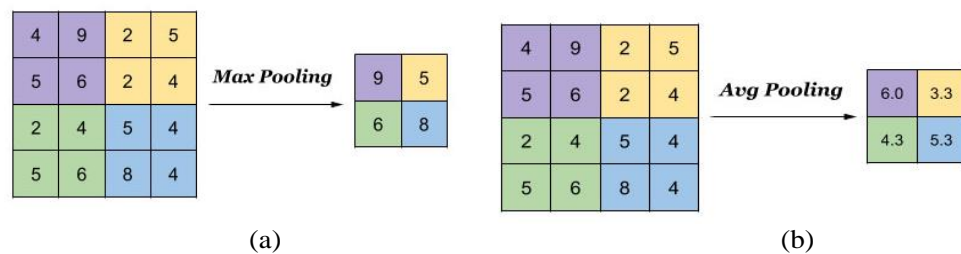


Figure 1.12 – Exemple de max pooling (a) et average pooling (b) de taille 2×2.

- *Couche entièrement connectée* : est un perceptron multicouche traditionnel qui emploie une fonction d'activation dans la couche de sortie. Le terme "Fully Connected" signifie que chaque neurone dans la couche précédente est relié à tous les neurones de la couche suivante. L'objectif de la couche entièrement connectée est d'exploiter les résultats des couches de convolution et de pooling afin de catégoriser l'image d'entrée dans diverses classes en se basant sur l'ensemble des données d'apprentissage. Ainsi, les couches de convolution et de pooling agissent comme des extracteurs de caractéristiques de l'image d'entrée, tandis que la couche entièrement connectée agit comme un classificateur.

Parmi les limitations de CNN c'est qu'il ne code pas la position et l'orientation de l'objet et qu'il manque de capacité à être spatialement invariant par rapport aux données d'entrée.

1.3.5. Réseaux de neurones convolutifs profonds (DCNN)

Les réseaux de neurones convolutifs profonds (DCNN) se distinguent des réseaux de neurones convolutifs normaux (CNN) par leur architecture, qui comporte généralement plus de 5 couches cachées. Cette structure permet aux DCNN d'extraire des caractéristiques plus complexes et d'améliorer la précision des prédictions.

La technique du transfert d'apprentissage (ou bien transfer learning) a amélioré

l'efficacité de l'exploitation de réseaux profonds pré-entraînés dans diverses applications telles que la reconnaissance des émotions à partir des images, du texte, ou de la parole. Cette technique consiste à réutiliser un modèle pré-entraîné sur un nouveau problème de classification. Le modèle original entraîné a besoin d'une généralisation élevée pour s'adapter à des données non vues. L'apprentissage par transfert signifie que l'apprentissage du modèle n'a pas besoin de recommencer à zéro pour chaque nouvelle tâche, il applique simplement un modèle pré-entraîné sur un ensemble de données énorme, et les poids obtenus à partir de ce modèle peuvent être utilisés pour toutes tâches, ce qui permet d'économiser le temps d'apprentissage. Les réseaux de neurones convolutifs profonds (DCNNs) sont très utilisés dans le domaine du traitement d'image [57-59]. Il existe plusieurs types de DCNN à savoir :

- *Visual Geometry Group 16 (VGG-16)* : est une architecture populaire de réseau de neurones convolutifs profonds (DCNN) proposée par Simonyan et al [60] de l'université d'Oxford. À ce jour, il est considéré comme l'un des meilleurs modèles de vision et a été le modèle gagnant du concours de la compétition ILSVR (ImageNet) de 2014. L'architecture VGG-16 est composée de 13 couches de convolution d'un filtre (3×3) avec stride 1, 5 couches de max pooling avec stride 2, deux couches entièrement connectées avec 4096 canaux, suivies d'une autre couche entièrement connectée avec 1000 canaux, et la dernière couche est la couche Softmax (Figure 1.13).

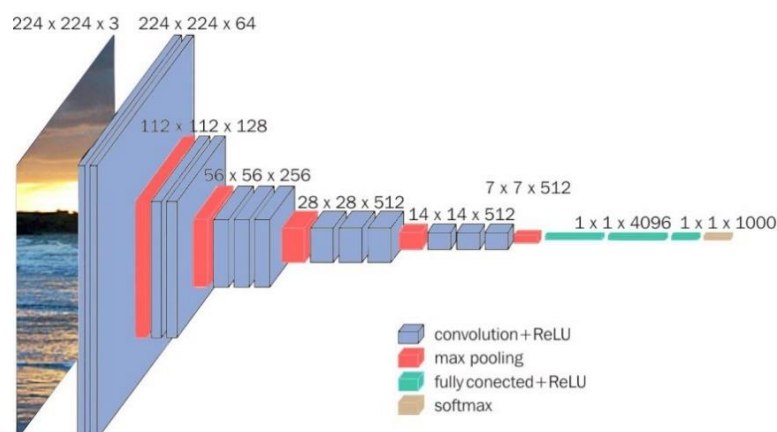


Figure 1.13 – Architecture du modèle VGG-16.

- *Visual Geometry Group 19 (VGG-19)* : est un modèle similaire à l'architecture de VGG-16 avec 3 couches de convolution supplémentaires. Il se compose de 16 couches de convolution, 5 couches de max pooling, deux couches entièrement

connectées et une couche Softmax.

- *Inception V3* : est un réseau de neurones convolutif profond proposé par Szegedy et al. [61], et possède 48 couches profondes. Il contient de nombreux blocs asymétriques et symétriques, où chaque bloc a des convolutions, max pooling, average pooling, dropout, couches concaténées, et des couches entièrement connectées. L'architecture d'Inception V3 contient des couches de convolutions factorisées afin de minimiser le nombre de paramètres sans diminuer l'efficacité du réseau et l'ajout de la couche batch normalization à la couche entièrement connectée.
- *Xception [62]* : est une expansion de la structure Inception, qui utilise des convolutions séparables en profondeur (depthwise separable convolutions) pour remplacer les modules Inception. Les convolutions séparables en profondeur consistent en une convolution spatiale indépendante par canal, suivie d'une convolution (1x1) pour capturer les corrélations inter-canaux. L'Xception est composée de 36 couches de convolution qui servent à extraire les fonctions du réseau. Ces couches sont organisées en 14 modules, et chaque module est doté de connexions résiduelles linéaires autour de lui, à l'exception des premiers et derniers modules.
- *ResNet ou Residual networks [63]* : est un réseau de neurones convolutifs profond qui fait référence au concept d'ajout de l'apprentissage résiduel au réseau de neurones convolutif traditionnel afin de résoudre le problème du Vanishing Gradient dans les réseaux profonds. Cela permet aux gradients de circuler à travers les couches de convolution et également de se propager en connectant l'entrée de la $n^{ième}$ couche à la sortie de la $(n + a)^{ième}$ couche. Cette connexion résiduelle est représentée par l'équation suivante :

$$x_i = f(x) + x \quad (1.47)$$

Où x_i la sortie du $i^{ième}$ bloc résiduel, tandis que f est fonction qui définit l'ensemble des couches du bloc résiduel. Il existe différentes architectures de ResNet : ResNet-50, ResNet-18, ResNet-34, ResNet-101, et ResNet152.

- *MobileNet [64]* : est un réseau de neurones convolutifs profond conçu pour les applications mobiles et de vision embarquée. Il utilise des convolutions séparables en profondeur pour former un réseau de neurones convolutifs profond léger et à faible latence. Le modèle MobileNet est constitué de 28

couches, si l'on considère les convolutions en profondeur et les convolutions ponctuelles comme des couches séparées.

- *DenseNet ou densely connected convolutional network [65]* : est un réseau de neurones convolutifs profond qui utilise des connexions denses entre les couches par le biais des blocs denses, où chaque couche est connectée à chaque autre couche. L'architecture des DenseNets est divisée en 4 blocs denses avec un nombre variable de couches, où les tailles des cartes de caractéristiques demeurent invariables au sein d'un bloc. Par exemple, le DenseNet-121 possède (6, 12, 24, 16) couches denses, le DenseNet-169 a (6, 12, 32, 32) couches denses, le DenseNet-201 a (6, 12, 48, 32) couches denses, tandis que le DenseNet-264 a (6, 12, 64, 48).

1.4. Reconnaissance des émotions faciales

1.4.1. Définition de l'émotion

L'émotion, selon le psychologue Scherer [66], est un processus dynamique et relativement bref au cours duquel des changements massifs et interdépendants se produisent dans plusieurs sous-systèmes de l'organisme en réponse à un événement déclencheur d'importance majeure pour l'individu. Les composantes psychologiques et physiologiques interagissent de manière spécifique pendant cet épisode émotionnel, permettant à l'organisme de faire face à la situation d'urgence créée par l'événement. L'importance de l'événement déclencheur, qui suscite une émotion, est évaluée par la personne en fonction de critères personnels tels que les besoins, les objectifs et les valeurs importantes.

1.4.2. Neurophysiologie des émotions : le système limbique

Le système limbique est une partie du cerveau impliquée dans nos réponses émotionnelles et comportementales liées à la survie. Deux structures principales de ce système sont l'hippocampe et l'amygdale (Figure 1.14). L'hippocampe est essentiellement responsable de la mémoire et de l'orientation spatiale. Il joue un rôle clé dans la création de souvenirs épisodiques et l'association de souvenirs avec nos sens. L'hippocampe est également important pour l'apprentissage et la neurogenèse, la création de nouveaux neurones. L'amygdale, située à côté de l'hippocampe, est en forme d'amande et joue un rôle central dans nos réactions émotionnelles, associant un contenu émotionnel à nos souvenirs. Elle est impliquée dans la détermination de la solidité du stockage des souvenirs et dans la

formation de nouveaux souvenirs.

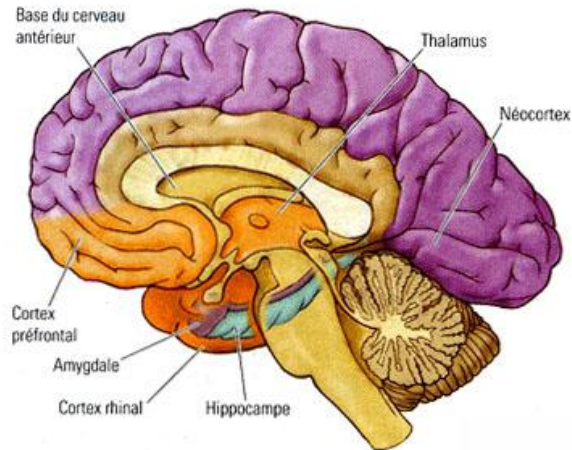


Figure 1.14 – Système limbique.

1.4.3. Caractéristiques typiques des émotions universelles sur le visage

Les expressions faciales sont l'un des aspects les plus importants de la communication humaine. Le visage est chargé de communiquer non seulement des pensées ou des idées, mais aussi des émotions. Les expressions faciales sont des configurations de différents mouvements micromoteurs (petits muscles) du visage qui sont utilisés pour déduire l'état émotionnel discret d'une personne. Les expériences de Ekman [4] sur les expressions faciales associées aux émotions ont permis de conclure que les six émotions de base (joie, peur, surprise, tristesse, colère, et dégoût) sont universellement reconnues, indépendamment de l'âge, du genre et du contexte culturel. La joie se manifeste par un sourire de Duchenne [67] avec les yeux rétrécis, les joues relevées et les lèvres tirées en arrière. Ensuite, la peur se distingue de la surprise par des sourcils plus droits et horizontaux, une paupière supérieure plus haute et des lèvres tendues. Par ailleurs, la surprise se caractérise par des sourcils levés, des paupières supérieures relevées, des paupières inférieures neutres et des mâchoires baissées. En revanche, la tristesse se traduit par le redressement des coins internes des sourcils, des paupières supérieures droites et un regard vers le bas, ainsi que des coins des lèvres tirés vers le bas. En outre, la colère se manifeste par des sourcils abaissés et rapprochés, des yeux éblouissants et des commissures des lèvres rétrécies. Finalement, le dégoût se caractérise par un froncement du nez, des sourcils baissés, des rides sur le nez, une lèvre supérieure relevée en forme de 'U' inversé et une lèvre inférieure relevée.

1.4.4. Système de reconnaissance automatique des expressions faciales

Un système de reconnaissance automatique des expressions faciales (FER) est généralement composé de trois phases, comme illustré dans la figure 1.15. Dans la première phase, le système détecte le visage dans une image d'entrée ou une séquence vidéo. Dans la deuxième étape, après que le visage a été localisé dans l'image ou la séquence vidéo, un ensemble des caractéristiques est extrait pour expliquer l'expression faciale. À la fin, en se basant sur l'ensemble de caractéristiques qui ont été extraites de la région du visage dans la phase précédente, l'image sera affectée à une catégorie d'expressions à l'aide d'un classificateur. Il existe deux catégories de systèmes de reconnaissance automatique des expressions faciales (FER) : un système basé sur les méthodes traditionnelles et un système basé sur les méthodes de deep learning.

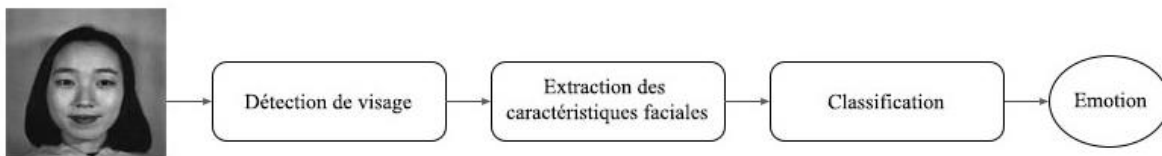


Figure 1.15 – Architecture d'un système de reconnaissance automatique des expressions faciales.

- *La détection de visages* : est la première étape d'un système de reconnaissance automatique des expressions faciales. Elle consiste à identifier le cadre entourant le visage dans une image ou une vidéo, en tenant compte des différences de pose, d'illumination et d'échelle, tout en éliminant autant que possible l'arrière-plan. Bien que les humains puissent détecter naturellement les visages avec peu d'effort, cette tâche est complexe pour les machines et nécessite plusieurs étapes informatiques. Les avancées récentes dans les technologies informatiques ont permis d'améliorer la recherche dans ce domaine, aboutissant à de nombreux algorithmes de détection de visages. Ces algorithmes peuvent être classés en quatre catégories : méthodes basées sur la connaissance, méthodes basées sur les caractéristiques invariantes, méthodes basées sur la mise en correspondance et méthode basée sur l'apparence.
 - *Méthodes basées sur la connaissance* : reposent sur la compréhension humaine de la géométrie et de la disposition des traits du visage. Elles utilisent la symétrie naturelle du visage et l'ordre dans lequel les traits apparaissent pour décrire leur forme, leur taille, leur texture et leurs relations les uns avec les autres. Cependant, ces méthodes rencontrent des difficultés lorsque la pose ou l'orientation de la tête varie, et il est complexe de traduire ces connaissances humaines en règles précises et

significatives.

- *Méthodes basées sur les caractéristiques invariantes* : visent à trouver des caractéristiques structurelles du visage qui restent constantes malgré les variations d'éclairage. Elles emploient des caractéristiques locales du visage, la texture, la forme et la couleur de la peau pour détecter les visages. Ces approches peuvent être limitées en cas de corruption ou de déformation des caractéristiques de l'image. Parmi les principaux algorithmes de cette catégorie, on trouve l'algorithme de détection de visage Viola-Jones (VJ) [68], basé sur des caractéristiques de type Haar, et les détecteurs de visage basés sur l'apprentissage profond, tels que le R-CNN plus rapide et le détecteur à coup unique (SSD) [69-72].
- *Méthodes basées sur la mise en correspondance* : consistent à détecter le contour de la tête, généralement de forme elliptique, en utilisant des filtres, des détecteurs de bord ou des silhouettes. Ensuite, les contours des caractéristiques locales du visage sont extraits en utilisant la connaissance de la géométrie du visage. Enfin, la corrélation entre les caractéristiques extraites de l'image et les modèles prédéfinis de visage et de caractéristiques faciales est calculée pour déterminer la présence d'un visage. Les méthodes basées sur des modèles prédéfinis sont sensibles aux variations d'échelle, de forme et de pose, mais des méthodes de modèles déformables ont été proposées pour y faire face [73].
- *Méthode basée sur l'apparence* : elle utilise un grand nombre d'exemples représentant différentes variations de visage. Elles sont fondées sur l'analyse statistique et les techniques d'apprentissage automatique pour détecter les visages dans une image. Ces méthodes fonctionnent en classifiant les régions de l'image comme "visage" ou "non visage" en se basant sur les propriétés statistiques des modèles de luminosité des pixels. Parmi les méthodes couramment utilisées, on trouve les eigenfaces, LDA, les réseaux de neurones, les machines à vecteurs de support (SVM) et les modèles de Markov cachés [74-76].
- *L'extraction des caractéristiques faciales* : est la deuxième phase du système de reconnaissance automatique des expressions faciales, elle consiste à trouver et à représenter les caractéristiques d'une image pour un traitement ultérieur. Les

performances du système dépendent des caractéristiques extraites, car même les meilleures techniques de classification peuvent échouer si des caractéristiques insignifiantes sont utilisées. Les méthodes traditionnelles d'extraction des caractéristiques faciales peuvent être séparées en deux classes principales : les caractéristiques géométriques et les caractéristiques basées sur l'apparence.

- *Caractéristiques géométriques* : sont capables de décrire la forme et l'emplacement des composants du visage tels que la bouche, le nez, les yeux et les sourcils. Trois méthodes couramment utilisées sont les modèles de forme actifs (ASM) [77], les modèles actifs d'apparence (AAM) [78] et le Scale-Invariant Feature Transform (SIFT) [79]. Cependant, l'extraction de ces caractéristiques nécessite une détection précise des points caractéristiques, ce qui peut être difficile à réaliser dans des environnements réels. De plus, ces méthodes ont tendance à ignorer les changements de texture de la peau, tels que les rides et les sillons, qui sont souvent accentués avec l'âge.
- *Caractéristiques d'apparence* : utilisent l'ensemble du visage ou des régions spécifiques telles que les yeux, le nez et la bouche pour extraire des informations. Différentes techniques telles que la transformée en ondelettes de Gabor [80], l'histogramme des gradients orientés (HOGs) [81], le motif binaire local (LBP) [82] ou le Scale-Invariant Feature Transform (SIFT) sont utilisées pour extraire ces caractéristiques. Certaines approches combinent des classificateurs de Haar avec des histogrammes de motifs binaires locaux (LBP) pour la détection des visages [83], tandis que d'autres utilisent la couleur et la profondeur avec des capteurs Kinect [84]. Cependant, les caractéristiques basées sur l'apparence peuvent être sensibles à l'illumination, ce qui peut introduire des erreurs dans le processus de reconnaissance.

En général, les approches traditionnelles d'extraction de caractéristiques faciales sont sensibles aux variations de la pose, de l'âge et de l'apparence du visage. Dans ces systèmes basés sur l'apprentissage profond, l'extraction de caractéristiques s'effectue de manière automatique grâce à des approches spécifiques. Par exemple, le deep belief network (DBN) a été utilisé pour l'extraction de caractéristiques faciales [85][86]. D'autres chercheurs ont utilisé des réseaux bidirectionnels de longue mémoire à court terme (Bi-LSTM) en combinaison avec des CNN pour extraire des caractéristiques faciales [87].

- *La classification* : est la phase finale d'un système de reconnaissance automatique des expressions faciales (FER) dans lequel le classificateur catégorise les expressions de joie, de tristesse, de surprise, de colère, de peur, de dégoût et neutres. Dans la tâche de classification des émotions faciales, plusieurs algorithmes de machine learning et de deep learning ont été appliqués pour obtenir des résultats plus précis. Hegde et al. [88] ont calculé les caractéristiques de texture des images avec le filtre de Gabor. La classification est effectuée à l'aide de l'algorithme des machines à vecteurs de support (SVM) avec un noyau de fonction de base radiale gaussienne (RBF). Le SVM a été aussi utilisé par [89, 90] pour la classification des émotions. L'ACP et l'analyse en composantes indépendantes (ACI) ont été utilisées pour extraire des caractéristiques dans [91], puis un système de classification en deux étapes pour l'expression faciale est conçu à l'aide de modèles de Markov cachés et de l'analyse discriminante linéaire. L'algorithme d'arbres de décisions basé sur le gain d'information a été utilisé dans [92] pour la classification des émotions. Les perceptrons multicouches (MLP) [93, 94], et d'autres classificateurs ont également été utilisés pour la reconnaissance des émotions. Des travaux basés sur Deep Belief Networks (DBN) ont été utilisés pour la classification des expressions faciales [95]. Dans [96], une approche est proposée pour la reconnaissance d'expression faciale à partir d'une vidéo basée sur un CNN spatio-temporel. Ainsi, un réseau de neurones convolutifs profond DCNN contenant trois modules et Inception-ResNet ont été utilisés dans [97] pour capturer les relations spatiales de l'expression faciale dans les images.

1.5. Applications de la reconnaissance des émotions faciales dans le domaine de l'enseignement

L'état émotionnel des étudiants, les performances de l'enseignant et les résultats d'apprentissage sont liés. La reconnaissance des émotions doit être intégrée dans les systèmes d'apprentissage pour une éducation de qualité, en raison des liens inextricables entre la cognition et les émotions. D'ailleurs, plusieurs chercheurs ont étudié le rôle de la cognition dans la génération des émotions. Dans [98] les auteurs ont examiné la relation entre les facteurs émotionnels et la performance des étudiants, concluant que les étudiants en bonne santé émotionnelle ont des chances de réussir à l'université. L'un des mécanismes fondamentaux de l'apprentissage est l'engagement actif de l'étudiant. Plus précisément, un

étudiant est généralement plus actif lorsqu'il s'engage dans le processus d'apprentissage de manière pratique, plutôt que d'être simplement passif ou totalement réceptif. Par exemple, lorsque l'espacement entre les yeux du visage est plus grand, cela indique que l'étudiant est très intéressé par le contenu du cours et qu'il est très concentré. Par ailleurs, lorsque l'espacement entre les yeux du visage devient plus petit, cela indique que l'étudiant n'est plus intéressé ni concentré. D'autres chercheurs pensent que la motivation est sensible au contexte et peut être encouragée en classe. Ainsi, de nombreux travaux ont été développés dans le sens de la motivation des étudiants à apprendre dans les systèmes informatiques éducatifs [99]. Dans le contexte de l'enseignement supérieur, les chercheurs ont utilisé des algorithmes d'apprentissage profond pour identifier les émotions faciales des étudiants dans une salle de classe. Li et al. [100] ont proposé une étude à méthode mixte qui combine des questionnaires psychologiques avec l'analyse des émotions faciales pour évaluer de manière exhaustive la santé mentale des étudiants dans l'université. Les auteurs ont employé un réseau de neurones convolutif (CNN) avec 9 couches, dont 4 couches de convolution, 3 couches de max pooling et deux couches entièrement connectées. Le modèle proposé a obtenu une précision de 62.98% sur FER2013, et a été testé sur 350 étudiants de l'université de médecine de Xuzhou des classes de 2020-2022, dont 187 étudiants masculins et 163 étudiants féminins d'un âge moyen de 22 ans. Bhatti et al. [101] ont proposé une approche de reconnaissance des émotions faciales d'enseignant dans une salle de classe en utilisant un modèle d'apprentissage profond par anticipation (feedforward), dans lequel des caractéristiques profondes sont extraites à l'aide de plusieurs CNN et le RELM (regularized extreme learning machine) a été utilisé comme classificateur. RELM avec DenseNet201 a obtenu les meilleures précisions de 96.8%, 86.59% et 62.74% sur les bases de données JAFFE, CK+ et FER2013, respectivement.

1.6. Conclusion

Dans ce chapitre, nous avons examiné en détail les algorithmes les plus couramment utilisés en apprentissage automatique et en apprentissage profond. Nous avons étudié leurs avantages et inconvénients, ainsi que leurs applications dans divers domaines. En outre, nous avons exploré les principes fondamentaux de la reconnaissance faciale des émotions. Le prochain chapitre se concentrera sur le renforcement des connaissances acquises en examinant plus en détail les principes fondamentaux des systèmes de reconnaissance des émotions faciales utilisés dans le contexte de cette thèse. Cela nous permettra d'évaluer comment ces algorithmes peuvent être exploités pour évaluer automatiquement les émotions des apprenants dans un environnement différencié et inclusif.

Chapitre 2

Techniques de la reconnaissance automatique des émotions faciales pour une approche pédagogique différenciée et inclusive

2.1. Introduction

L'utilisation des technologies de l'information et de la communication dans l'éducation a conduit à une évolution du concept d'enseignement/apprentissage vers une pédagogie différenciée et inclusive. Cette approche vise à répondre aux besoins individuels des apprenants, en adaptant les stratégies pédagogiques de manière continue. Cependant, mettre en place une telle pédagogie nécessite des outils et des méthodes d'évaluation adaptés aux contextes différenciés et inclusifs. C'est là qu'intervient la reconnaissance faciale des émotions (FER) en tant que méthode d'évaluation formative automatique. Dans ce chapitre, nous allons explorer les fondements de la pédagogie différenciée et inclusive, ainsi que l'évolution des théories d'apprentissage à l'ère numérique. Nous introduirons ensuite deux nouvelles méthodes centrées sur l'apprentissage profond pour la détection des émotions faciales des étudiants. La première approche [102], présentée dans la section 2.6, utilise l'algorithme de réseaux de neurones convolutifs (CNN). La seconde approche [103], décrite dans la section 2.7, vise à détecter l'engagement des étudiants sourds et malentendants à partir de leurs expressions faciales en temps réel avec l'algorithme VGG-16 et l'apprentissage par transfert.

2.2. Définition du concept enseignement / apprentissage

Au fil de l'histoire, les besoins et les exigences éducatives ont connu une croissance et une mondialisation, avec des structures institutionnelles comparables. En même temps, des

espaces ont été créés pour délimiter une frontière entre le monde extérieur, où se fait l'acquisition, et l'espace didactique, qui est celui de l'enseignement-apprentissage.

L'enseignement/apprentissage implique un ensemble d'actions visant à faciliter la transmission de connaissances et de compétences aux apprenants. Ces actions peuvent inclure la planification des activités d'enseignement, la sélection de ressources pédagogiques appropriées, la mise en place d'un environnement approprié à l'apprentissage, la proposition de situations d'apprentissage engageantes et l'évaluation de la compréhension des apprenants. L'enseignement est un processus interpersonnel et intentionnel qui utilise essentiellement la communication comme moyen de transmettre un savoir ou un savoir-faire [104]. Son but est de dispenser des connaissances, avec l'enseignant comme acteur principal. En revanche, l'apprentissage est un processus dynamique et systématique, englobant une interaction entre l'étudiant et son environnement, ainsi qu'un ajustement successif des connaissances acquises. Prost [105] souligne que la notion d'apprentissage a l'avantage de transférer le rôle de l'enseignant à l'action de l'élève. L'apprentissage est donc un processus dynamique et interactif qui engage activement l'apprenant dans sa propre formation et la construction de ses connaissances. Pour dispenser une leçon efficace, l'enseignant doit avoir une connaissance solide en pédagogie et en didactique et établir un contrat didactique avec ses élèves, qui définit de manière précise les fonctions et les obligations de chaque partie impliquée dans le processus d'apprentissage. La pédagogie, implique la gestion et la signification du travail en classe ainsi que la réflexion sur les ressources et les objectifs. D'autre part, la didactique concerne plutôt les méthodes d'enseignement spécifiques à chaque discipline, y compris les difficultés spécifiques à un domaine particulier. Selon le modèle didactique classique de Chevallard [106], la situation pédagogique est un triangle composé de l'enseignant, du savoir et de l'apprenant, tel que présenté par Houssaye [107] dans la figure 2.1.

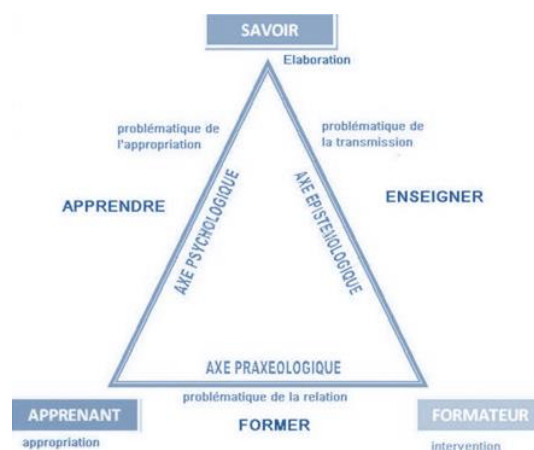


Figure 2.1 – Triangle didactique de Houssaye [69].

En 2002, Paquelin [108] a proposé une vision renouvelée du triangle didactique traditionnel en intégrant les technologies de l'information et de la communication (TIC) dans le processus d'enseignement-apprentissage (Figure 2.2).

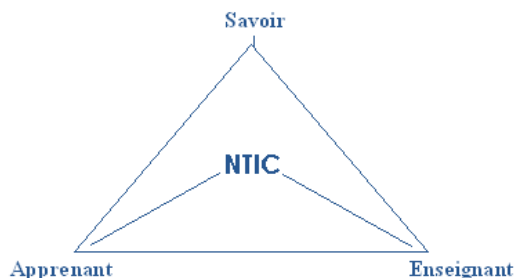


Figure 2.2 – Triangle DidacTIC.

2.3. Evolution des théories d'apprentissage à l'ère numérique

Au cours des cinquante dernières années, les théories d'apprentissage ont connu une évolution marquée, passant de l'approche behavioriste à l'approche cognitive, puis aux approches constructiviste, socio-constructiviste et connectiviste [109]. Cette transition a eu un impact significatif sur les objectifs de l'apprentissage, le rôle de l'enseignant et les processus cognitifs internes [110]. Le behaviorisme, centré sur les comportements observables, a été remplacé par l'approche cognitive, centrée sur les processus mentaux internes engagés dans l'apprentissage [111]. Le cognitivisme, apparu dans les années 1960, reconnaît l'apprenant en tant qu'un système actif de traitement de l'information et met en avant l'importance des processus mentaux tels que la mémoire, l'attention et la résolution de problèmes [112]. Cependant, le modèle cognitiviste présente une limitation importante, car la motivation des apprenants joue également un rôle clé dans un apprentissage efficace. Le constructivisme, émergé dans les années 1970, propose un apprentissage social et actif, où les étudiants ont un rôle essentiel dans la construction de la connaissance à partir de leur expérience [113]. Les enseignants agissent comme facilitateurs et créent des environnements d'apprentissage stimulants, reconnaissant l'importance du contexte social et culturel [114]. Le socio-constructivisme, une extension du constructivisme apparu dans les années 1980, se concentre sur l'interaction sociale et la participation active des apprenants, influencé par les travaux de Vygotsky et Bruner [115][116]. Ces approches favorisent l'apprentissage interactif et la participation active des apprenants, et ont un fort potentiel pour les technologies éducatives.

Le connectivisme, une nouvelle théorie pédagogique apparue dans les années 2000, tient compte de l'influence des technologies de l'information et de la communication sur

l'apprentissage [117]. Les chercheurs Siemens et Downes remettent en question les théories d'apprentissage existantes et proposent une approche adaptée à l'apprentissage en ligne et à un monde de plus en plus connecté [118]. Le connectivisme reconnaît que l'apprentissage se produit dans des environnements ambigus et en constante évolution, où les liens facilitant l'apprentissage sont plus importants que notre niveau de connaissances actuel [119]. Cette évolution des théories d'apprentissage a entraîné une redéfinition de la pédagogie traditionnelle, intégrant des méthodes plus différenciées et inclusives, inspirées des principes du constructivisme et du socio-constructivisme.

2.4. Principes fondamentaux de la pédagogie différenciée

La pédagogie différenciée est une pratique éducative qui se base sur l'hétérogénéité naturelle des individus et qui vise à personnaliser l'enseignement pour répondre aux besoins particuliers de chaque apprenant. Cette approche s'applique à tous les étudiants, qu'ils aient des difficultés ou un rythme de travail rapide. Cela n'implique pas la distinction des objectifs, mais plutôt donner à tous les étudiants la possibilité d'atteindre les mêmes objectifs par des approches variées. [120]. La pédagogie différenciée a connu une évolution notable depuis sa première apparition au début du XXe siècle. Parkhurst [121] a mené l'une des premières mises en œuvre de la pédagogie différenciée à l'école de Dalton dans le Massachusetts en 1905 en individualisant le travail scolaire en prenant en compte le niveau et la personnalité individuelle de chaque apprenant. Dans les années 1920, Dottrens [122] et Freinet [123] ont élaboré des programmes de différenciation pédagogique. La pédagogie différenciée a ensuite connu une augmentation de sa popularité dans les années 1960-1970 grâce à Bloom [124], et Tomlinson [125]. Depuis les années 2000, de nouveaux dispositifs ont été créés pour satisfaire les besoins des apprenants, et la pédagogie différenciée est actuellement développée dans de nombreux pays, notamment le Maroc, où des mesures de différenciation pédagogique sont prévues pour l'enseignement préscolaire [126].

2.4.1. Formes de différenciation pédagogique

La pédagogie différenciée, telle que présentée par Przesmycki [127], se base sur trois dispositifs liés aux aspects fondamentaux du triangle didactique : la personne, le savoir et l'institution.

- *Différenciation des processus d'apprentissage* : propose une méthode d'apprentissage personnalisée pour chaque étudiant en utilisant des pratiques de travail autonome, en répartissant les étudiants en groupes travaillant

simultanément sur les mêmes objectifs avec des pratiques adaptées à leurs besoins individuels après une analyse approfondie de l'hétérogénéité des étudiants.

- *Différenciation des contenus d'apprentissage* : consiste à personnaliser les contenus en fonction des besoins et des difficultés individuelles de chaque étudiant. Les étudiants sont regroupés en fonction d'objectifs cognitifs, méthodologiques et comportementaux, sélectionnés par l'équipe pédagogique ou l'enseignant à partir d'un ensemble commun d'objectifs.
- *Différenciation des structures* : implique la création de groupes d'étudiants dans des structures différentes de la classe traditionnelle, offrant ainsi aux étudiants la possibilité d'explorer de nouveaux types de regroupements, de lieux et d'animateurs, favorisant ainsi de nouvelles interactions sociales. Les étudiants peuvent être répartis selon leurs besoins spécifiques, de leur niveau de réussite ou de difficulté, ou encore inclus dans une classe hétérogène globale ou dans des modules d'assistance en petits groupes.

2.4.2. Evaluation dans la pédagogie différenciée

La différenciation pédagogique implique la création d'un plan d'action personnalisé pour chaque apprenant, en utilisant des méthodes et des adaptations spécifiques. Il est important que les étudiants ne voient pas la différenciation comme une forme de punition, mais plutôt comme un outil pour apprendre de manière optimale. Dans ce contexte, l'évaluation joue un rôle important pour mesurer les progrès des étudiants, identifier leurs besoins spécifiques et ajuster les stratégies pédagogiques en conséquence. Meirieu [128] identifie quatre types d'évaluation, chacun ayant un rôle spécifique dans la situation d'enseignement-apprentissage :

- *L'évaluation diagnostique* : consiste en une évaluation initiale effectuée par l'enseignant au début de l'apprentissage afin d'identifier les besoins, les capacités, les méthodes de travail, la maîtrise de situations ou d'outils de chaque apprenant.
- *L'évaluation formative* : permet de faire d'accomplir un bilan intermédiaire de l'apprentissage de l'étudiant pendant la période d'apprentissage, afin d'ajuster les méthodes d'enseignement. Elle permet à l'enseignant de suivre la progression de l'apprenant, d'identifier les éléments qui fonctionnent bien et de détecter les obstacles ou les situations de handicap qui peuvent affecter l'apprentissage.

- *L'évaluation sommative* : est une évaluation conclusive qui permet de vérifier si les objectifs d'apprentissage ont été atteints. Elle permet d'établir un bilan des compétences et des connaissances acquises par l'étudiant par rapport aux objectifs définis en amont. Cette évaluation est exprimée sous la forme d'une note qui est attribuée en fonction de critères d'évaluation prédéfinis.
- *L'auto-évaluation* : est une méthode permettant à l'apprenant d'évaluer lui-même ses compétences et ses connaissances en fonction de critères proposés par l'enseignant. La forme la plus courante d'auto-évaluation est une grille d'acquisition établie en amont par le professeur. Cette évaluation permet à l'étudiant de développer une capacité à se situer dans sa formation et à estimer son niveau d'acquisition. Si elle n'aboutit pas à une notation chiffrée, l'auto-évaluation peut être utilisée comme une évaluation formative.

2.5. Principes fondamentaux de la pédagogie inclusive

Au fil des décennies, les modèles d'éducation proposés aux étudiants, notamment ceux ayant des besoins éducatifs spécifiques tels que les étudiants en situation de handicap, ont connu une évolution. Les chercheurs ont généralement réparti les systèmes éducatifs en quatre catégories : exclusion, ségrégation, intégration et inclusion, comme illustré dans la figure 2.3.

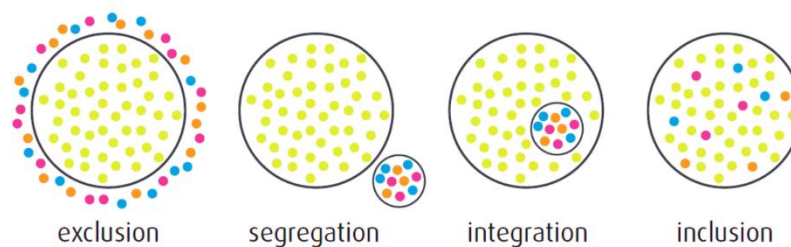


Figure 2.3 – Quatre types de modèle éducatif.

L'éducation peut être caractérisée par différentes catégories : l'exclusion, la ségrégation, l'intégration et l'inclusion. L'exclusion se produit lorsque les étudiants sont empêchés d'accéder à l'éducation, tandis que la ségrégation implique l'éducation de différents groupes dans des environnements séparés. L'intégration consiste à placer des étudiants ayant des besoins spécifiques dans des environnements éducatifs ordinaires avec quelques adaptations. En revanche, l'inclusion vise à éliminer les obstacles qui limitent la présence, la participation et la réussite de tous les apprenants. La pédagogie inclusive reconnaît et valorise les différences de tous les apprenants, qu'il s'agisse de leur origine linguistique,

ethnique, culturelle, religieuse, de leur genre, de leur déficience ou de leurs aptitudes particulières. Cependant, la mise en place de la pédagogie inclusive engendre trois défis majeurs : les ressources, les attitudes et le défi pédagogique. Il est essentiel de prendre en compte le rôle des émotions faciales dans l'élaboration des cours pour maximiser l'attention des apprenants [129], améliorer leur apprentissage et leur rétention à long terme du matériel [130-132]. Les émotions jouent un rôle important dans l'apprentissage. Les émotions positives favorisent l'apprentissage et la réussite scolaire [133], tandis que les émotions négatives, comme le stress excessif, nuisent à l'apprentissage et à la performance cognitive [134]. La pédagogie intelligente, qui utilise des technologies d'apprentissage adaptatif et d'intelligence artificielle, permet de personnaliser l'enseignement en fonction des caractéristiques d'apprentissage propres à chaque étudiant. Ces technologies permettent également de collecter et d'exploiter les données sur le comportement des étudiants, y compris leurs expressions faciales, pour ajuster la pédagogie en temps réel. L'incorporation de ces outils dans le processus d'évaluation de la pédagogie différenciée favorise une éducation inclusive et équitable qui répond aux besoins éducatifs individuels de chaque étudiant.

2.6. Reconnaissance des émotions faciales des étudiants en temps réel en utilisant CNN

2.6.1. Introduction

Ces dernières années, la reconnaissance des expressions faciales a suscité une grande attention en raison de son impact dans plusieurs domaines, notamment la médecine, la robotique et l'éducation. Selon diverses recherches, l'émotion joue un rôle important dans l'éducation. Un enseignant utilise des examens, des questionnaires et des observations comme sources de feedback, mais ces méthodes classiques sont souvent peu efficaces. En utilisant l'expression faciale des élèves, l'enseignant peut ajuster sa stratégie et ses supports pédagogiques pour aider à favoriser l'apprentissage des étudiants. L'objectif de ce travail est de mettre en œuvre la reconnaissance d'émotions dans l'enseignement en réalisant un système automatique qui analyse les expressions faciales des élèves en se basant sur le réseau de neurones convolutif (CNN), qui est très employé dans le domaine du traitement d'images. Il consiste en un traitement d'image en plusieurs étapes pour extraire des caractéristiques. Notre système comprend trois phases : la détection des visages, la normalisation et la reconnaissance de l'émotion qui devrait être l'une de ces sept émotions (neutre, colère, peur, tristesse, joie, surprise et dégoût).

Dans ce qui suit, nous allons commencer tout d'abord par présenter les travaux connexes (sous-section 2.6.2). Ensuite, nous décrivons le système de reconnaissance des émotions faciales proposé (sous-section 2.6.3). Puis nous allons présenter une description de la base de données d'expressions faciales utilisée (sous-section 2.6.4), suivie par les détails de l'implémentation du système dans la sous-section 2.6.5. Enfin, nos résultats expérimentaux, les comparaisons et la discussion sont présentés dans la sous-section 2.6.6.

2.6.2. Travaux connexes

De nombreux chercheurs s'intéressent à l'amélioration de l'environnement d'apprentissage avec la reconnaissance des émotions du visage (FER). Tang et al. [135] ont proposé un système capable d'analyser les expressions faciales des étudiants afin d'évaluer l'effet de l'enseignement en classe. Le système est composé de cinq phases : acquisition de données, détection des visages, reconnaissance des visages, reconnaissance des expressions, et post-traitement. L'approche utilise K plus proches voisins (KNN) pour la classification et la séquence uniforme d'histogrammes binaires de Gabor locaux (ULGBPHS) pour l'analyse des motifs. Savva et al. [136] ont proposé une application Web qui analyse l'émotion des étudiants dans une salle de classe. L'application utilise des webcams installées dans les salles de classe pour collecter des enregistrements en direct, puis ils ont appliqué des algorithmes d'apprentissage automatique à ces enregistrements. Whitehill et al. [137] ont proposé une approche qui reconnaît l'engagement à partir des expressions faciales des étudiants. L'approche utilise des caractéristiques de Gabor et l'algorithme SVM pour identifier l'engagement des étudiants qui interagissent avec un logiciel d'entraînement aux compétences cognitives. Les auteurs ont obtenu des étiquettes à partir de vidéos annotées par des juges humains. Ensuite, les auteurs de [138] ont utilisé l'apprentissage automatique pour analyser les expressions faciales des élèves dans un laboratoire informatique d'une école, où les élèves interagissaient avec un jeu éducatif visant à expliquer des concepts fondamentaux de la mécanique classique.

Dans [139], les auteurs ont proposé un système qui identifie et surveille l'émotion de l'élève et lui donne un feedback en temps réel afin d'améliorer l'environnement d'apprentissage en ligne. Le système utilise le modèle de mouvement de la tête et des yeux pour déduire des informations pertinentes et de comprendre l'humeur des étudiants dans un environnement d'apprentissage en ligne. Ayvaz et al. [140] ont développé un système de reconnaissance des émotions faciales (FERS), qui reconnaît les états émotionnels et la motivation des étudiants dans l'apprentissage en ligne par vidéoconférence. Le système utilise 4 algorithmes d'apprentissage supervisé (SVM, KNN, forêts aléatoires et arbres de

décision) et les meilleures précisions ont été obtenues en utilisant KNN et SVM. Kim et al. [141] ont proposé un système capable de produire en temps réel des recommandations à l'enseignant afin de renforcer la mémorisation et la qualité de leurs cours en lui permettant de modifier en temps réel son comportement, non comme le langage corporel et les expressions faciales. Les auteurs dans [142] ont proposé un modèle qui reconnaît les émotions faciales dans l'environnement d'apprentissage virtuel basé sur la méthode des cascades Haar pour identifier la bouche et les yeux en utilisant la base de données JAFFE.

2.6.3. Système d'analyse des émotions faciales des étudiants en utilisant CNN

Dans cette sous-section, nous décrivons notre système d'analyser des émotions à partir des expressions faciales des étudiants en temps réel en utilisant le réseau de neurones convolutif (CNN) (sous-section 1.3.4). Tout d'abord, le système détecte le visage à partir de l'image d'entrée et ces visages détectés sont recadrés et normalisés à une taille de 48×48. Ensuite, ces images de visage sont comme entrée du CNN. Enfin, la sortie est le résultat de la reconnaissance de l'expression faciale (neutre, colère, peur, tristesse, joie, surprise ou dégoût). La figure 2.4 présente la structure de notre approche proposée.

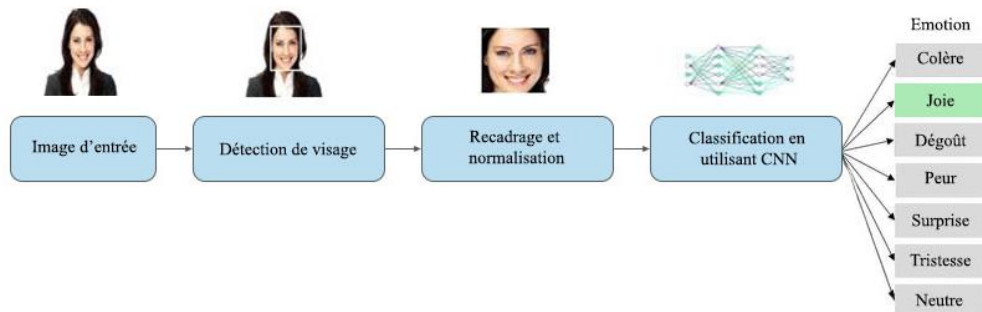


Figure 2.4 – Structure de notre système de reconnaissance des expressions faciales.

Nous avons utilisé la bibliothèque OpenCV [143] pour capturer les images en direct de la caméra web et pour détecter les visages des étudiants en se basant sur la méthode Haar Cascades, comme le montre la figure 2.5. La méthode Haar Cascades utilise l'algorithme d'apprentissage Adaboost inventé par Freund et al. [144], qui ont remporté le prix Gödel 2003 pour leurs travaux. L'algorithme d'apprentissage Adaboost choisit un petit nombre de traits significatifs à partir d'un grand ensemble de caractéristiques afin de fournir un résultat efficace des classificateurs.

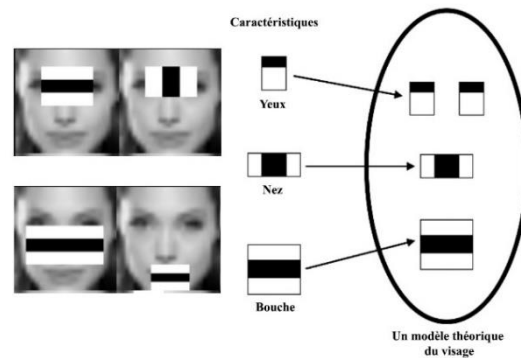


Figure 2.5 – Détection de visages avec Haar Cascades.

Nous avons ensuite utilisé des techniques d’augmentation des données pour transformer les images d’apprentissage par rotation, décalage, cisaillement, zoom et retournement. La figure 2.6 montre un exemple d’image de la base de données FER2013 avant et après transformations.

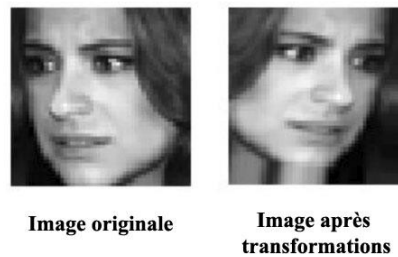


Figure 2.6 – Exemple d’image de la base de données FER2013 avant et après augmentation.

Ces images sont passés au modèle CNN afin de les analyser et identifier l’émotion qui correspond à l’expression faciale de visage. La figure 2.7 représente l’architecture de notre modèle CNN. Il contient 4 couches de convolution, 4 couches de max pooling pour extraire les caractéristiques, et 2 couches entièrement connectées puis la couche Softmax avec 7 classes d’émotions (neutre, colère, peur, tristesse, joie, surprise ou dégoût).

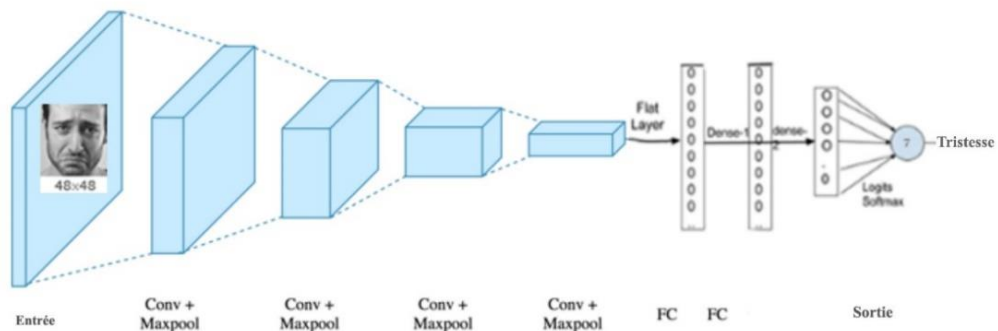


Figure 2.7 – Architecture de notre modèle CNN.

L'image d'entrée est en niveaux de gris avec une taille de 48×48. Pour chaque couche de convolution, nous avons utilisé des filtres 3×3 avec un pas (stride) de 2. Pour les couches de pooling, nous avons utilisé le max pooling et des noyaux 2×2 avec stride 2. Ainsi, pour introduire la non-linéarité dans notre modèle, nous avons utilisé la fonction d'activation (ReLU), définie dans l'équation 1.46, qui est la fonction d'activation la plus utilisée récemment. Ensuite, nous avons utilisé la normalisation des lots (batch normalization) pour normaliser l'activation de la couche précédente à chaque batch et une régularisation L2 pour appliquer des pénalités sur les différents paramètres du modèle. Le Tableau 2.1 présente la configuration de notre modèle.

Tableau 2.1 – Configuration du modèle CNN proposé.

Type de la couche	Taille	Pas
Image d'entrée	48x48	-
Convolution 1	3x3	2
Max pooling 1	2x2	2
Convolution 2	3x3	2
Max pooling 2	2x2	2
Convolution 3	3x3	2
Max pooling 3	2x2	2
Convolution 4	3x3	2
Max pooling 4	2x2	2
Entièrement connectées	-	-
Entièrement connectées	-	-

2.6.4. Bases de données d'expressions faciales

Pour entraîner notre architecture CNN, nous avons utilisé la base de données FER2013 [145] qui se compose de 35887 images en niveaux de gris, dont 28709 images pour l'apprentissage, 3589 images pour le test et 3589 images pour la validation. Sept émotions sont présentes dans cette base de données : colère, dégoût, peur, joie, tristesse, surprise et neutre. La base de données FER2013 a été introduite dans le défi de reconnaissance d'expression de l'ICML2013 et a été générée en collectant des images provenant des recherches Google qui représentaient chaque émotion. Les images sont avec une résolution de 48×48 pixels, et il y a de grandes différences dans l'âge et la posture des visages, y compris non seulement des images de visages réels, mais aussi des images de visages de

dessins animés, qui sont très difficiles. La figure 2.8 illustre une sélection d'images de la base de données FER2013. Les images d'une même colonne représentent des expressions identiques, à savoir la colère, le dégoût, la peur, la joie, la tristesse, la surprise, ainsi que le neutre. Le nombre d'images pour chaque émotion est représenté dans le Tableau 2.2.



Figure 2.8 – Exemples d'images de base de données FER2013.

Tableau 2.2 – Nombre d'images pour chaque émotion dans la base de données FER2013.

Numéro d'émotion	Émotion	Nombre d'images
0	Colère	4593
1	Dégoût	547
2	Peur	5121
3	Joie	8989
4	Tristesse	6077
5	Surprise	4002
6	Neutre	6198

2.6.5. Détails de l'implémentation du système proposé

Dans Keras [146], nous avons utilisé la classe ImageDataGenerator pour effectuer l'augmentation de l'image, comme illustré par la figure 3.3. Cette classe nous a permis de transformer les images d'apprentissage par rotation, déplacement, cisaillement, zoom et retournement. La configuration utilisée est la suivante : `rotation_range = 10`, `width_shift_range = 0.1`, `zoom_range = 0.1`, `height_shift_range = 0.1` et `horizontal_flip = True`.

Ensuite, pour entraîner notre modèle CNN, nous avons divisé la base de données FER2013 en 80% de données d'apprentissage et 20% de données de test. Puis nous avons compilé le modèle en utilisant l'optimiseur Stochastic Gradient Descent (SGD) [147]. À chaque époque, Keras vérifie si notre modèle est plus performant que les modèles des époques précédentes. Si c'est le cas, les meilleurs poids du modèle sont enregistrés dans un fichier. Cela nous permettra de charger directement les poids sans avoir à réentraîner le modèle si nous voulons l'utiliser dans une autre situation.

2.6.6. Résultats expérimentaux

Pour évaluer l'efficacité et la qualité de notre approche proposée, nous avons calculé un ensemble spécifique de mesures de performance, à savoir : l'accuracy, la précision, le rappel et le score F1, comme le montre la figure 2.10. Les formules correspondantes pour chacune de ces mesures sont définies dans les équations (2.1), (2.2), (2.3) et (2.4), où TP (True Positive) représente les vrais positifs, TN (True Negative) représente les vrais négatifs, FP (False Positive) représente les faux positifs, et FN (False Negative) représente les faux négatifs.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

$$F1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (2.4)$$

Nous avons obtenu un taux de précision (accuracy) de 70% aux 106 époques sur la base de données FER2013. La figure 2.9 montre la matrice de confusion de test avec un schéma de validation de 80-20%.

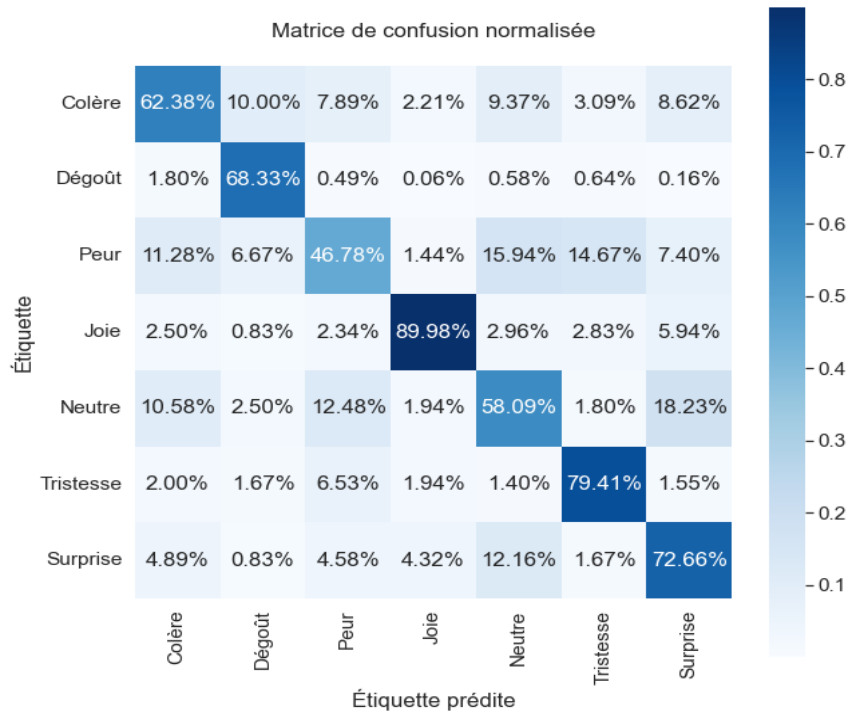


Figure 2.9 – Matrice de confusion normalisée du modèle proposée sur la base de données FER 2013.

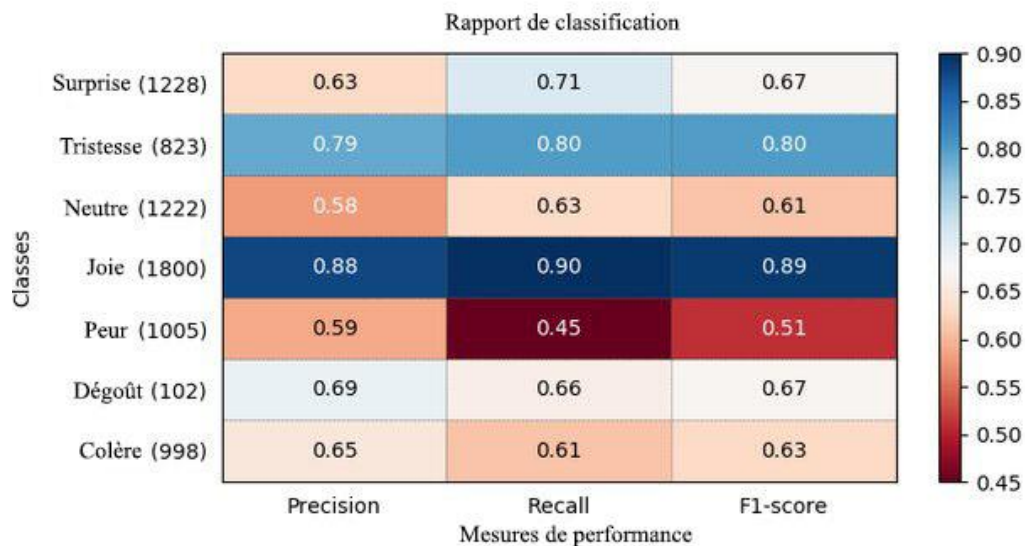


Figure 2.10 – Rapport de classification de notre approche proposée sur la base de données FER2013.

La figure 2.9 montre que le modèle a atteint des performances exceptionnelles dans la reconnaissance de l'émotion de joie, avec une précision de 89.98%. Cela démontre sa capacité élevée à identifier les expressions faciales associées à la joie. Cependant, des confusions ont été observées, notamment avec les émotions de surprise, atteignant 5.94%. La classe de tristesse présente également une précision relativement bonne de 79.41%, mais

des erreurs de prédiction ont été constatées, notamment des confusions avec l'émotion de peur, qui s'élèvent à 6.53%. Des confusions importantes ont été observées pour l'émotion de surprise, avec des pourcentages élevés de confusions avec la neutralité (12.16%), la colère (4.89%) et la peur (4.58%). Des erreurs de prédiction ont également été observées pour l'émotion de dégoût, avec des confusions importantes avec l'émotion de colère, représentant 1.80%. De plus, le modèle a montré des difficultés à distinguer l'émotion de colère des autres émotions, notamment du dégoût (10.0%), de la peur (7.89%), de la neutralité (9.37%) et de la surprise (8.62%). L'émotion neutre a été prédite avec une précision de 58.09%, mais le modèle a rencontré des difficultés à la distinguer des autres émotions, notamment de la surprise (18.23%), de la peur (12.48%) et de la colère (10.58%). Pour l'émotion de peur, le modèle a prédit correctement 46.78% des émotions de peur, mais des erreurs de prédiction importantes ont été observées, avec des confusions fréquentes avec les émotions de neutralité (15.94%), de tristesse (14.67%) et de colère (11.28%).

En complément de l'analyse de la matrice de confusion, le rapport de classification dans la figure 2.10 fournit des informations supplémentaires sur les performances du modèle. Pour l'émotion joie, nous avons une précision élevée de 88%, un rappel de 90% et un F1-score de 89%, ce qui suggère une performance solide avec des prédictions positives correctes à 88% et une bonne identification des instances réellement positives. Pour l'émotion tristesse, nous avons une précision de 79%, un rappel de 80% et un F1-score de 80%, ce qui indique une performance solide avec des prédictions positives correctes à 79% et une bonne identification des instances réellement positives. L'émotion dégoût présente une précision de 69%, un rappel de 66% et un F1-score de 67%, montrant une performance moyenne avec des prédictions positives correctes à 69% et une identification correcte de 66% des instances réellement positives. L'émotion colère présente une précision de 65%, un rappel de 61% et un F1-score de 63%, indiquant une performance moyenne avec des prédictions positives correctes à 65% et une identification correcte de 61% des instances réellement positives. L'émotion surprise présente une précision de 63%, un rappel de 71% et un F1-score de 67%. Cela suggère que 63% des prédictions positives pour cette émotion étaient correctes, 71% des instances réellement positives ont été identifiées, et le modèle a une performance globale équilibrée pour cette émotion. L'émotion peur présente une précision de 59%, un rappel de 45% et un F1-score de 51%, indiquant une performance inférieure avec un pourcentage relativement bas de prédictions positives correctes et une identification limitée des instances réellement positives. Enfin, l'émotion neutre présente une précision de 58%, un rappel de 63% et un F1-score de 61%, ce qui indique des prédictions positives correctes à 58% et une identification correcte de 63% des instances

réellement positives.

En reliant les résultats de la matrice de confusion et du rapport de classification, nous pouvons observer des similitudes dans les performances du modèle pour certaines émotions. Par exemple, les confusions observées entre l'émotion de joie et l'émotion de surprise sont corroborées à la fois par la matrice de confusion (5.94%) et par le rappel de l'émotion surprise dans le rapport de classification (71%). Ces confusions peuvent être dues à des similitudes dans les expressions faciales ou à des caractéristiques subtiles qui rendent difficile la distinction entre la joie et la surprise. De même, les confusions entre l'émotion de tristesse et l'émotion de peur sont reflétées à la fois dans la matrice de confusion (6.53%) et dans le rappel de l'émotion tristesse du rapport de classification (80%). Ces confusions peuvent être attribuées à des expressions faciales similaires entre la tristesse et la peur en raison des petites différences inter-classes, telles que des sourcils froncés ou des lèvres légèrement tendues. Cela indique que les expressions faciales associées à la peur peuvent être similaires à celles des autres émotions, ce qui rend la distinction plus difficile pour le modèle. De plus, le nombre limité d'échantillons d'entraînement pour l'émotion de peur peut également affecter sa précision de prédiction. Ainsi, plusieurs facteurs tels que l'occlusion du visage, la luminosité aléatoire, les visages non-frontaux, etc., peuvent avoir un impact négatif sur les performances du modèle quand il est question de distinguer entre les émotions de peur et de tristesse. Enfin, les confusions entre la neutralité et les autres émotions peuvent être dues à des variations subtiles dans les expressions faciales.

Nous avons validé notre modèle sur un groupe de neuf étudiants (six garçons et trois filles) en master "Ingénierie des données et développement logiciel" au cours de l'année universitaire 2018/2019 à la Faculté des Sciences de Rabat (FSR), Université Mohammed V de Rabat. Il convient de souligner que ces étudiants ont exprimé volontairement leur désir de participer à cette expérience. Parmi ces étudiants, deux d'entre eux portaient des lunettes. L'expérience s'est déroulée dans l'amphithéâtre Ibn al-Haytham, situé au département de physique de la Faculté des Sciences de Rabat. La figure 2.11 montre les résultats de reconnaissance des émotions de ces étudiants. Malgré les conditions telles que l'occlusion du visage, la luminosité variable, et les lunettes, notre modèle a donné une meilleure précision et a réussi à prédire correctement les émotions. Les étiquettes des émotions prédites sont représentées par un texte rouge, et la barre rouge indique la probabilité de chaque émotion. Nous constatons que 5 étudiants étaient heureux, 3 étaient neutres et un étudiant était en colère.

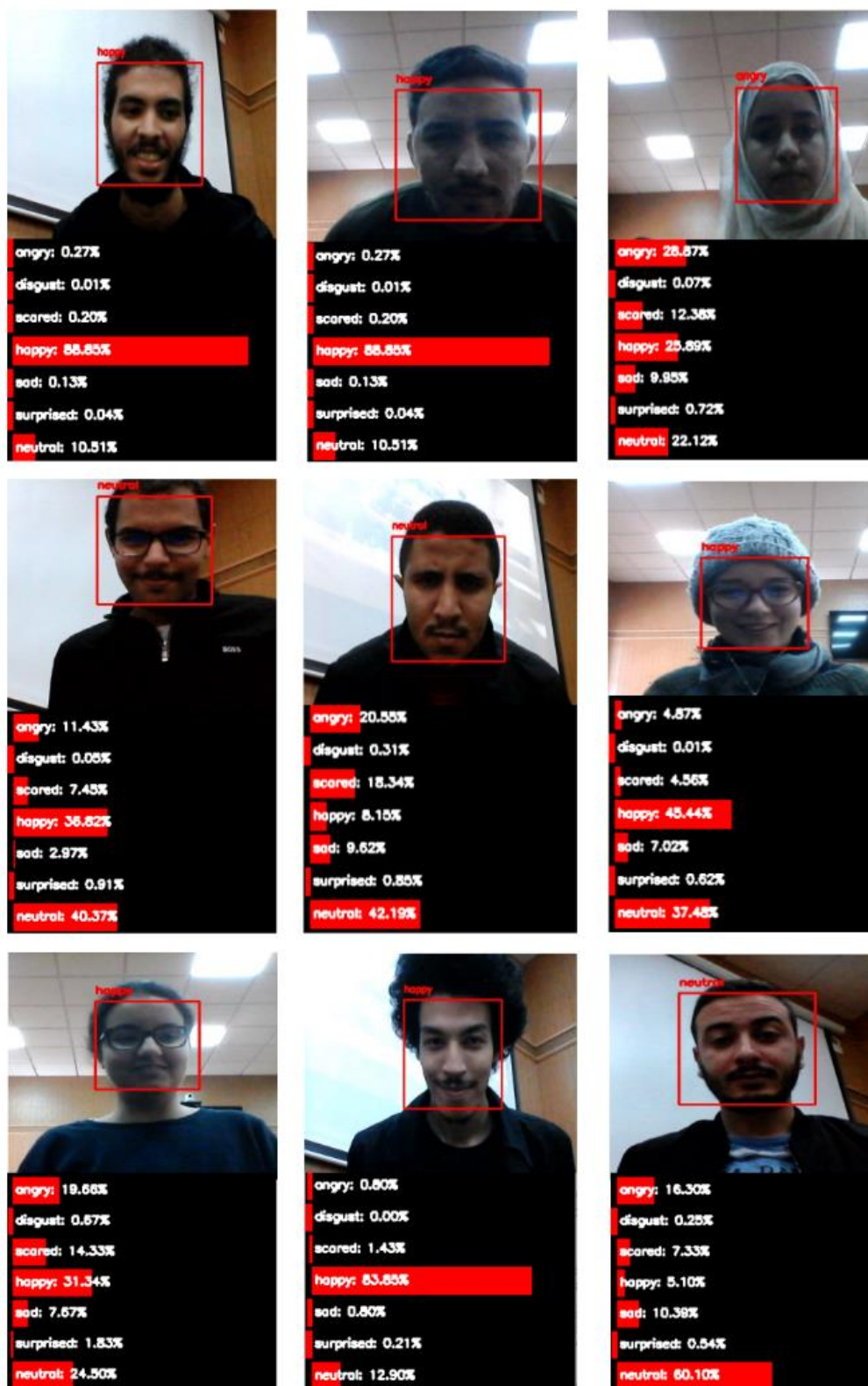


Figure 2.11 – Résultats de la reconnaissance des émotions faciales des étudiants.

2.7. Détection d'engagement des étudiants sourds et malentendants en temps réel à partir de leurs expressions faciales avec VGG-16 et l'apprentissage par transfert

2.7.1. Introduction

Les expressions faciales sont l'un des moyens les plus importants pour les êtres humains d'exprimer des émotions et des intentions sans dire un mot. Elles constituent une forme de communication non verbale, surtout pour les personnes sourdes, car elles sont utilisées dans la langue des signes pour exprimer des fonctions grammaticales et des émotions. Par conséquent, la reconnaissance des émotions à partir des expressions faciales est devenue un domaine bien étudié. En psychologie, Ekman et Friesen [4] ont identifié six émotions universelles (joie, tristesse, dégoût, peur, surprise et colère), chacune ayant une expression faciale unique qui peut être reconnue automatiquement par des algorithmes de vision par ordinateur.

Récemment, dans le secteur de l'éducation, la reconnaissance des émotions faciales peut être utilisée pour surveiller l'engagement des élèves en classe. ELlaban et al. [148] ont également proposé un système en temps réel pour la reconnaissance des expressions faciales des apprenants dans l'environnement d'apprentissage en ligne. Thomas et Jayagopi [149] ont classifié le niveau d'engagement des étudiants en analysant les indices comportementaux à partir de leurs expressions faciales, de leurs mouvements de tête et de leur comportement de regard. Aslan et al. [150] ont présenté un système d'engagement des étudiants en temps réel qui améliore les pratiques de classe des instructeurs.

L'une des principales questions de recherche axées sur l'exploration de données éducatives (EDM) est de savoir comment les étudiants sourds et malentendants sont engagés dans une classe. La question de l'engagement est importante et peut influencer l'enseignement et les résultats des élèves dans différents environnements d'apprentissage, tels que les salles de classe et les cours en ligne massivement ouverts (MOOC). Le manque d'engagement des élèves sourds et malentendants peut être dû à diverses raisons, notamment le fait que les élèves qui se fient à la lecture labiale n'ont pas toujours le temps de traiter les informations précédentes lorsque les enseignants ne font pas de pause avant de passer à un nouveau sujet. Et les enseignants qui ne connaissent pas la langue des signes peuvent rencontrer des difficultés lorsqu'ils communiquent avec des étudiants sourds ou malentendants. En outre, certaines universités ne sont pas en mesure de fournir à leurs étudiants sourds ou malentendants des technologies d'assistance. Par conséquent, pour améliorer l'efficacité du processus d'apprentissage, les enseignants peuvent suivre le niveau d'engagement de chaque étudiant. L'engagement des étudiants sourds et malentendants peut

être évalué à l'aide de questionnaires ou de systèmes automatisés basés sur l'observation de la parole, les mouvements oculaires et la reconnaissance des émotions faciales (FER) afin d'améliorer la pédagogie et l'expérience d'apprentissage.

L'extraction de caractéristiques d'expressions faciales à partir des images faciales et la reconnaissance de différentes expressions faciales à l'aide d'un classificateur entraîné est un processus complexe. Les techniques classiques de FER comportent trois étapes principales : le prétraitement des images, l'extraction des caractéristiques et la reconnaissance des émotions. Pendant la phase de prétraitement, la région du visage est détectée, puis découpée dans l'image d'entrée. Ensuite, l'image du visage est réduite pour éliminer le bruit, redimensionnée et normalisée. L'étape d'extraction des caractéristiques de l'image traitée est une étape importante, qui consiste à trouver diverses caractéristiques spatiales et temporelles à partir des composants du visage. Enfin, les méthodes traditionnelles d'apprentissage automatique (ML) et les méthodes d'apprentissage profond (DL) classent l'image d'entrée en utilisant les caractéristiques extraites pour comprendre les émotions. Les méthodes traditionnelles d'apprentissage automatique (ML) visent à détecter la région du visage dans l'image et à en extraire les caractéristiques, puis à classer l'image d'entrée en utilisant les caractéristiques extraites. Les méthodes d'apprentissage profond, en particulier les réseaux neuronaux convolutifs (CNN) et les réseaux neuronaux convolutifs profonds (DCNN), effectuent la tâche de FER en combinant les étapes d'extraction de caractéristiques et de classification dans un seul processus opérationnel composite. D'autres approches d'apprentissage profond comprennent des réseaux DCNN pré-entraînés, tels que VGG-16 [60], VGG-19 [60], Inception v3 [61], Xception [62], Resnet-50 [63], DenseNet-121 [65], DenseNet-169 [65], et MobileNet [64] qui réduisent le long processus d'apprentissage en utilisant des poids pré-entraînés.

À notre connaissance, aucune recherche n'a été effectuée jusqu'à présent sur la détection d'engagement des étudiants sourds et malentendants en utilisant des algorithmes d'apprentissage automatique ou d'apprentissage profond. Dans la présente étude, nous tentons de relever ce défi, pour la première fois, en proposant un système automatique qui évalue l'engagement des étudiants sourds et malentendants à partir de leurs expressions faciales en se basant sur les réseaux de neurones convolutifs profonds (DCNN) et l'apprentissage par transfert (TL). Les images faciales des élèves sont obtenues à l'aide d'une caméra placée dans la salle de classe. Ce système peut aider les enseignants à observer la réaction des élèves sourds ou malentendants sur un sujet particulier pendant un cours, à ajuster la méthodologie d'enseignement en fonction de la compréhension des étudiants, à identifier les étudiants sourds ou malentendants qui ne sont pas engagés et qui ont besoin

de séances de soutien scolaire afin d'améliorer la gestion de la classe et d'économiser le temps et les ressources. Un modèle ImageNet pré-entraîné VGG-16 a été utilisé dans notre approche FER proposée en remplaçant ses couches supérieures par d'autres couches denses, des couches dropout et des couches de normalisation par lots (batch normalization). Ensuite, nous avons ajusté le modèle VGG-16 sur deux ensembles de données d'images faciales : Japanese female facial expression (JAFFE) et Karolinska directed emotional faces (KDEF). Ensuite, les performances du modèle proposé sont évaluées et comparées à celles de sept modèles DCNN pré-entraînés (VGG-19, Inception V3, DenseNet-121, DenseNet-161, MobileNet, ResNet-50, Xception) sur les bases de données JAFFE et KDEF.

Dans ce qui suit, nous allons exposer les travaux connexes (sous-section 2.7.2). Une description de notre approche proposée sera présentée dans la sous-section 2.7.3. Ensuite, nous allons présenter une description des bases de données d'expressions faciales utilisées (sous-section 2.7.4), suivie par les détails de l'implémentation du système proposé dans la sous-section 2.7.5, Enfin, une présentation et une discussion de nos résultats expérimentaux, ainsi qu'une comparaison avec les méthodes de l'état de l'art seront présentés dans la sous-section 2.7.6.

2.7.2. Travaux connexes

Plusieurs techniques ont été réalisées sur l'application de l'apprentissage automatique et des méthodes d'apprentissage profond pour analyser les expressions faciales humaines au cours des dernières décennies. Les travaux antérieurs sur la reconnaissance des émotions faciales s'appuyaient sur des méthodes d'apprentissage automatique traditionnelles telles que les machines à vecteur de support (SVM), le K plus proches voisins (KNN) et les réseaux de neurones (NN) avec différentes techniques d'extraction de caractéristiques.

Lee et al. [151] ont utilisé la transformée en contours (CT) pour l'extraction de caractéristiques et l'algorithme de boosting basé sur l'analyse discriminante régularisée (RDAB) pour la classification. L'approche qu'ils ont proposée a été évaluée sur la base de données JAFFE. Liew et Yairi [152] ont examiné cinq descripteurs de caractéristiques, notamment, Gabor, Haar, le motif binaire local (LBP), l'histogramme des gradients orientés (HOG) et les caractéristiques élémentaires indépendantes binaires (BRIEF), en utilisant SVM, KNN, l'analyse discriminante linéaire (LDA) et le boosting adaptatif (AdaBoost), sur les bases de données CK+, JAFFE, et KDEF. Les auteurs ont identifié HOG comme le meilleur descripteur de caractéristiques et SVM comme le meilleur classificateur. HOG et SVM ont également été employés par Eng et al. [153]. Ils ont utilisé JAFFE et KDEF pour évaluer leur méthode. Holder et Tapamo [154] ont utilisé l'opérateur de gradient de Scharr,

la réduction de la dimensionnalité et l'extraction de composantes faciales pour améliorer le modèle local de gradient ternaire (GLTP), qui a été utilisé pour l'extraction de caractéristiques. Puis ils ont utilisé le SVM pour la classification des caractéristiques sur les ensembles de données CK+ et JAFFE.

Avec le grand succès de l'apprentissage profond pour la classification des images, de nombreux chercheurs ont utilisé des méthodes fondées sur l'apprentissage profond pour la reconnaissance des émotions faciales (FER). Zhao et al. [155] ont utilisé un réseau DBN pour l'apprentissage et l'extraction de caractéristiques à partir d'images d'expressions faciales. Ensuite, un modèle de perceptron multicouche (MLP) est utilisé pour la classification des émotions sur les bases de données JAFFE et CK+. Un réseau BDBN a été proposé par Liu et al. [156] qui combine l'apprentissage des caractéristiques, la sélection d'émotions. Le BDBN a été évalué en utilisant CK+ et JAFFE. Sari et al. [157] ont proposé une architecture CNN standard avec deux couches de convolutions pour la classification des émotions faciales sur CK+, JAFFE, et KDEF. Hamester et al. [158] ont considéré une architecture de réseau de neurones convolutifs multicanaux (MCCNN) évaluée sur JAFFE. Le premier canal est composé d'un CNN standard. Ensuite, le deuxième canal utilise des paramètres pré-entraînés obtenus par un autoencodeur convolutionnel (CAE), qui apprend des filtres de type Gabor. Les deux canaux sont connectés avec une couche entièrement connectée, suivie d'un classificateur de régression logistique. Une approche hybride CNN-RNN est utilisée par Jain et al. [159] pour la reconnaissance des émotions faciales. Un modèle CNN avec six couches convolutives et deux couches entièrement connectées a été utilisé principalement pour l'extraction de caractéristiques. Ensuite, le RNN a été introduit pour classer les émotions faciales en utilisant la base de données JAFFE.

La mesure de l'engagement des étudiants en classe est une préoccupation majeure des enseignants, car elle a une incidence positive sur la qualité de l'enseignement et de l'apprentissage. De nombreuses méthodes ont été appliquées pour détecter l'engagement des élèves à partir de données pédagogiques, de la reconnaissance des émotions faciales et des mouvements de la tête et des yeux. Ayouni et al. [160] ont proposé un système permettant de prédire le niveau d'engagement des étudiants (actif, passif et non engagé) à l'aide de SVM, d'un réseau neuronal artificiel (ANN) et d'un arbre de décision sur les activités enregistrées des étudiants. Le système peut alerter l'enseignant lorsqu'un étudiant a un faible niveau d'engagement par le biais d'e-mails. Shen et al. [161] ont employé un modèle de réseau convolutif attentionnel pour évaluer le niveau d'engagement des étudiants (excellent, pas mal et pas si bien) à partir de leurs expressions faciales dans l'environnement d'apprentissage en ligne. Cependant, à notre connaissance, aucune étude n'a été menée, à

ce jour, sur l'analyse des expressions faciales des étudiants sourds et malentendants en utilisant des algorithmes d'apprentissage automatique ou d'apprentissage profond.

2.7.3. Système de détection d'engagement des étudiants sourds et malentendants à partir de leurs expressions faciales en utilisant le VGG-16 et l'apprentissage par transfert

Nous avons proposé un nouveau système pour la détection en temps réel de l'engagement à partir des expressions faciales des étudiants sourds et malentendants en utilisant un réseau de neurones convolutifs profonds (DCNN) et l'apprentissage par transfert (TL), comme le montre la figure 2.12. Le système peut être utilisé en classe pour aider les enseignants à comprendre l'attention et l'engagement des étudiants sourds et malentendants vis-à-vis du matériel d'apprentissage. Les images des étudiants sont automatiquement analysées par le système pour évaluer leurs états de concentration à partir des expressions faciales avec une caméra Web.

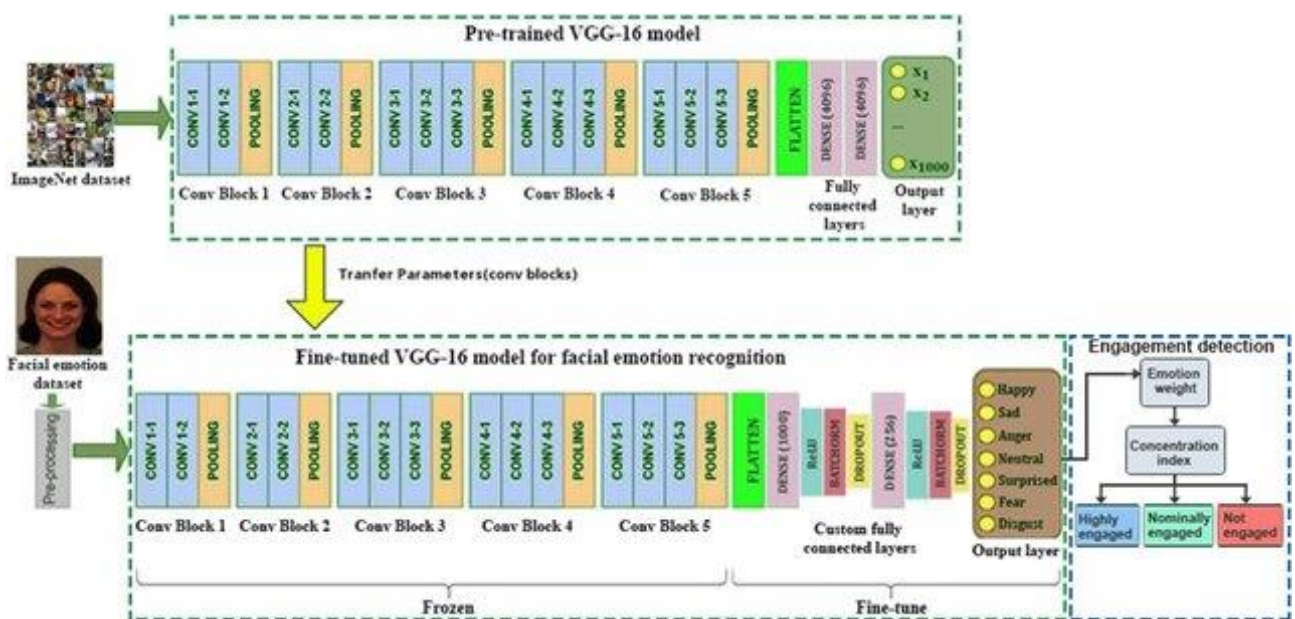


Figure 2.12 – Architecture globale de notre système proposé pour la reconnaissance des émotions et la détection de l'engagement des étudiants sourds et malentendants.

- *Phase de prétraitement et augmentation des images :* Les étapes de prétraitement d'image suivantes sont considérées dans notre travail pour formater les images avant qu'elles ne soient utilisées par les modèles DCNN. Tout d'abord, le classificateur en cascade basé sur les caractéristiques Haar d'OpenCV est utilisé pour détecter et recadrer les visages de chaque image. Il s'agit d'une approche

dans laquelle une fonction en cascade est entraînée à l'aide d'un grand volume d'images positives et négatives pour reconnaître la région du visage dans l'image. Ensuite, les régions de visage ont été redimensionnées en 224×224 pixels, ce qui est la dimension d'entrée par défaut des modèles DCNN pré-entraînés. Enfin, nous avons appliqué certaines techniques d'augmentation de l'image afin d'éviter le surapprentissage et de classer les données non vues par le modèle de manière très précise. Chaque image est retournée horizontalement et pivotée avec un angle de $(-10^\circ \text{ à } 10^\circ)$. Aucune augmentation des données n'a été effectuée sur les images de l'ensemble du test.

- *Phase de reconnaissance des émotions faciales* : Dans la phase de reconnaissance des émotions faciales, le modèle VGG-16 pré-entraîné avec apprentissage par transfert (sous-section 1.3.5.1) est utilisé pour identifier l'émotion dominante exprimée par les visages des étudiants sourds et malentendants à chaque instant. Le Fine-tuning est une technique couramment employée pour l'apprentissage par transfert, où les poids et l'apprentissage du modèle pré-entraîné sont utilisés comme initialiseurs pour une nouvelle tâche. Cette technique est beaucoup plus rapide que la construction d'un modèle complet à partir de zéro et peut contribuer à réduire le risque de surapprentissage (ou bien d'overfitting). Il existe trois méthodes les plus utilisées pour affiner un modèle : entraîner le modèle entier, figer certaines couches et entraîner les autres, ou bien figer la partie de convolution.

Tout d'abord, nous avons figé les blocs de convolution du modèle VGG-16 pré-entraîné afin que leurs poids ne soient pas mis à jour dans le modèle à chaque époque. Ensuite, nous avons remplacé les dernières couches denses par de nouvelles couches denses pour classer une image faciale dans l'une des sept classes d'émotions de base (peur, colère, joie, surprise, tristesse, neutre et dégoût). La taille des canaux des nouvelles couches denses est de 1000 et 256, respectivement. La fonction d'activation RELU, définie dans l'équation (1.46), est incluse après les couches entièrement connectées ajoutées, suivies des couches de normalisation par lots (ou Batch normalization) et des couches d'abandon (ou dropout) avec une probabilité de $p = 0.5$ pour éviter les surapprentissages. En outre, la couche de sortie utilise la fonction d'activation Softmax et l'entropie croisée catégorielle, indiquée dans (2.5) comme fonction de perte pour la classification multi-classes.

$$CE = - \sum_i^C t_i \log(f(s)_i) \quad (2.5)$$

Où C représente le nombre de classes, l'indice i désigne la $i^{ème}$ classe, t_i est l'étiquette, et S_i est la probabilité Softmax pour la $i^{ème}$ classe. La descente de gradient stochastique (ou stochastic gradient descent, SGD) avec le momentum de Nesterov [162], définie dans l'équation (2.6), a été utilisée comme optimiseur de modèle avec un taux d'apprentissage de 0.01 et un momentum de Nesterov de 0.9 pour générer de meilleures performances et des résultats plus rapides. Nous avons fixé la taille du lot à 16 et le nombre d'époques à 150. La figure 2.13 présente un résumé de l'optimisation de notre modèle VGG-16 proposé et le Tableau 2.3 décrit les détails des hyperparamètres du modèle.

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta - \gamma v_{t-1}) \\ \theta &= \theta - v_t \end{aligned} \quad (2.6)$$

Où v est la vélocité et γ est généralement fixé à 0.9.

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 1000)	25089000
batch_normalization (Batch Normalization)	(None, 1000)	4000
dropout (Dropout)	(None, 1000)	0
dense_1 (Dense)	(None, 256)	256256
batch_normalization_1 (Batch Normalization)	(None, 256)	1024
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 7)	1799
=====		
Total params: 40,066,767		
Trainable params: 25,349,567		
Non-trainable params: 14,717,200		

Figure 2.13 – Résumé de l'optimisation du modèle VGG-16 proposé.

Tableau 2.3 – Hyperparamètres du modèle VGG-16.

Paramètre	Valeur
Taille d'entrée	(224 × 224 × 3)
Optimiseur	SGD avec Nesterov momentum
Momentum	0.9
Taux d'apprentissage	0.01
Taille du lot	16
Nombre d'itérations	150
Fonction loss	Categorical cross-entropy
Classificateur	Softmax
Dropout	0.5
Normalisation par lots	Oui
Fonction d'activation	ReLU

- *Phase de détection de l'engagement* : Les émotions faciales reconnues dans la phase de reconnaissance des émotions faciales sont utilisées pour détecter le niveau de concentration des étudiants sourds et malentendants. L'indice de concentration (CI) qui en résulte est défini en multipliant la probabilité des émotions dominantes (DEP) par les poids d'émotion (EW) correspondants, comme indiqué dans l'équation (2.7).

$$CI = DEP \times EW \quad (2.7)$$

Le poids émotionnel est une valeur comprise entre 0 et 1 qui détermine le degré de concentration d'une émotion faciale à un moment donné. Pour obtenir les poids correspondant à chaque émotion, nous avons basé notre analyse sur l'étude réalisée par Sharma et al. [163] qui consiste en l'enregistrement des émotions faciales de 30 étudiants qui sont en train de voir une vidéo informative et de participer à un quiz de 10 questions. Les étudiants ont été regroupés sur la base de leur principale émotion exprimée. Par exemple, si un étudiant exprimait la majorité du temps (plus de 50 % de la durée de la vidéo) une expression neutre, il était inclus dans le groupe des émotions neutres. Les étudiants ont été répartis entre les sept groupes des émotions en fonction des expressions faciales obtenues pendant la vidéo. Le score moyen obtenu dans le quiz pour chaque groupe a été calculé et est présenté dans le Tableau 2.4. Ainsi, le score de chaque groupe reflète, dans une certaine mesure, la relation entre la concentration d'un étudiant

et son état émotionnel.

Selon l'indice de concentration obtenu, le niveau d'engagement (fortement engagé, nominalement engagé ou non engagé) de l'étudiant sourd ou malentendant est évalué par le Tableau 2.5. Il peut être utilisé par les enseignants pour ajuster la leçon en conséquence. Les enseignants peuvent aussi avoir un rapport d'engagement en temps réel de chaque étudiant sourd ou malentendant, qui peut les aider à mieux savoir le niveau des connaissances acquises en classe.

Tableau 2.4 – Poids émotionnel.

Emotion	Colère	Joie	Peur	Surprise	Tristesse	Dégoût	Neutre
Poids	0.25	0.6	0.3	0.6	0.3	0.2	0.9

Tableau 2.5 – Détection d'engagement à partir de l'indice de concentration (CI).

Niveau d'engagement	Indice de concentration (CI)
Fortement engagé	>50%
Nominalement engagé	20-50%
Non engagé	< 20%

2.7.4. Bases de données d'expressions faciales

Dans cette étude, nous avons utilisé deux modes différents pour diviser les bases de données JAFFE et KDEF.

- 80 % des images sont utilisées pour l'apprentissage et 20% pour le test.
- Une validation croisée 10 fois (CV), dans laquelle les ensembles de données sont divisés de manière aléatoire en dix parties, neuf parties sont utilisées pour l'apprentissage du modèle, et le reste est utilisé pour le test, comme l'illustre la figure 2.14.

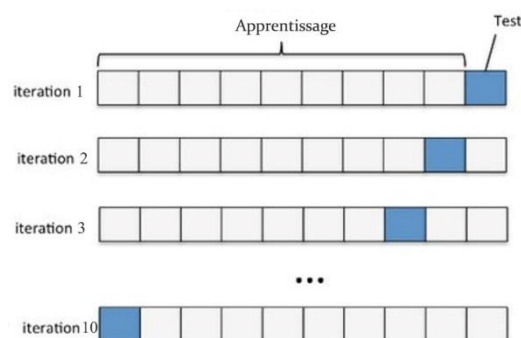


Figure 2.14 – Validation croisée 10 fois (10-fold cross-validation (CV)).

La base de données JAFFE [164] comprend 213 images d'expression faciale en niveaux de gris de 10 modèles féminins japonais différents, prises dans un environnement contrôlé. Chaque modèle a exprimé 7 émotions de base (30 colère, 29 dégoût, 33 peur, 30 joie, 31 tristesse, 30 surprise, et 30 neutre) dans lesquelles chaque expression contient 3 à 4 images par modèle. La figure 2.15 illustre des exemples d'images de la base de données JAFFE, de gauche à droite, représentant les émotions neutre, joie, tristesse, surprise, colère, peur, et dégoût. Les images sont au format .tif avec une résolution de 256×256 pixels, et elles ont été créées dans des conditions d'éclairage similaires et très contrôlées sans aucun facteur environnemental comme les cheveux ou les lunettes.

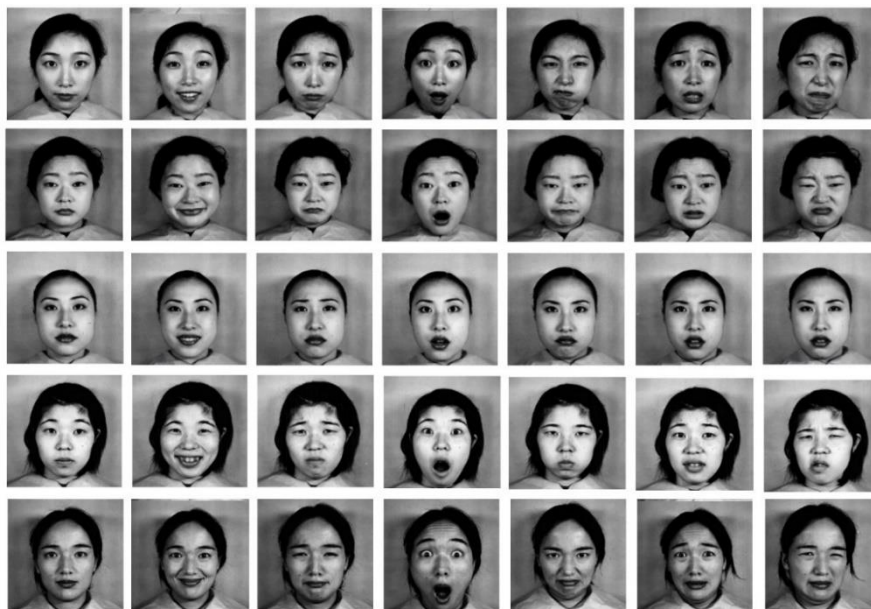


Figure 2.15 – Exemples d'images de base de données JAFFE.

La base de données KDEF [165] a été créée par l'institut Karolinska, département de neurosciences cliniques, section de psychologie à Stockholm, Suède. La base KDEF est une collection de 4900 images de 70 individus (35 femmes et 35 hommes) présentant 7 états émotionnels (joie, tristesse, surprise, colère, peur, dégoût et neutre), photographiés deux fois sous cinq angles différents (-90° , -45° , 0° , $+45^\circ$, et $+90^\circ$). La figure 2.16 montre une sélection d'images de la base de données KDEF. Les images de gauche à droite représentent les émotions neutre, joie, tristesse, surprise, colère, peur, et dégoût, respectivement. Et les images de haut en bas représentent les angles de capture des images 0° , 45° , -45° , 90° , et -90° . Les images sont au format RGB avec une résolution de 562×762 pixels. Différents critères ont été appliqués pour la sélection des acteurs et la procédure de prise de vue, tels que l'âge entre 20 et 30 ans, l'absence de boucles d'oreilles ou de lunettes, et de maquillage

visible pendant la séance photo.



Figure 2.16 – Illustrations d’images de la base de données KDEF.

2.7.5. Détails de l’implémentation du système proposé

Nous avons programmé notre modèle de reconnaissance des émotions faciales des étudiants sourds et malentendants à l’aide du langage de programmation Python et Keras avec un backend TensorFlow [166] pour la classification des images et OpenCV pour le prétraitement des images. Les expériences ont été effectuées au sein de notre laboratoire Conception et Systèmes (Electronique, Signaux et Informatique) sur un ordinateur portable équipé d’une carte graphique Nvidia GTX 1070, d’un processeur Intel Core i7 de 3.20 GHz, d’une mémoire vive de 16 Go et d’un environnement Windows 10 de 64 bits.

2.7.6. Résultats expérimentaux

Dans cette sous-section, la performance de notre modèle proposé est évaluée à l’aide de différents types d’algorithmes d’optimisation sur deux bases de données des expressions faciales : JAFFE et KDEF. Ensuite, la performance du modèle optimisé obtenu est comparée à différents modèles DCNN pré-entraînés. Les optimiseurs comme la descente de gradient stochastique (SGD) [147], SGD avec momentum [152], SGD avec le momentum de Nesterov [167], Adam [168], Adagrad [169] et Adadelta [170] sont utilisés dans le présent travail pour obtenir des performances de modèle optimisées. Les figures 2.17 et 2.18 montrent une comparaison de l’évolution de la perte (loss) pendant la phase d’apprentissage

avec différents algorithmes d'optimisation au fil des époques sur les bases de données JAFFE et KDEF avec une taille de lot de 16 et une entropie croisée catégorielle comme fonction de perte (loss).

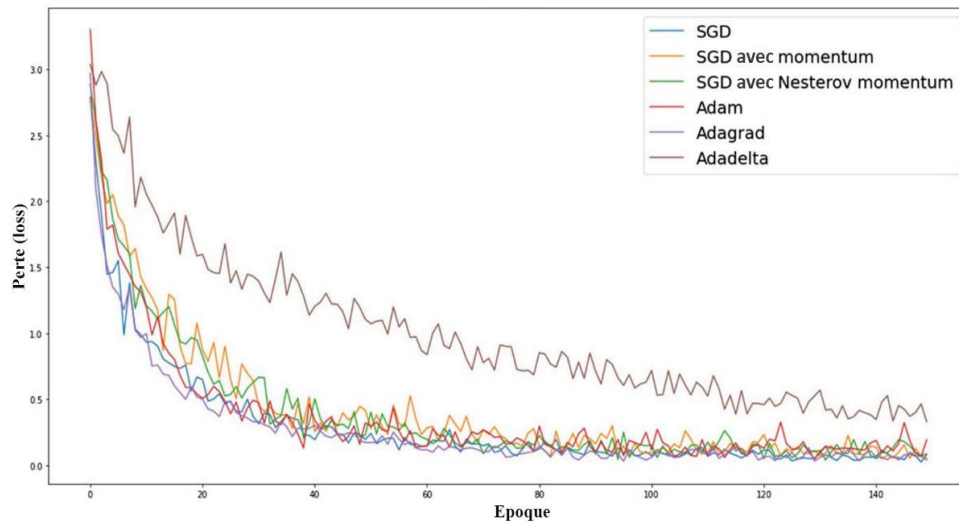


Figure 2.17 – Comparaison de tous les optimiseurs en utilisant le modèle VGG-16 sur la base de données JAFFE.

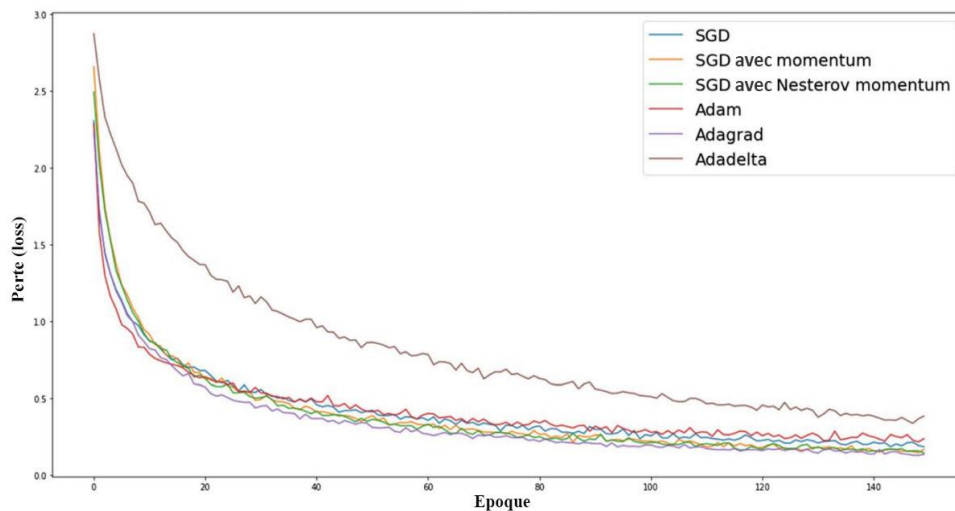


Figure 2.18 – Comparaison de tous les optimiseurs en utilisant le modèle VGG-16 sur la base de données KDEF.

L'algorithme de descente de gradient stochastique (SGD) avec le momentum de Nesterov a montré la meilleure précision de reconnaissance de 97.7% et 86.33% sur les bases de données JAFFE et KDEF, respectivement. De plus, les figures 3.11 et 3.12 montrent que la SGD avec le momentum de Nesterov a les pertes d'apprentissage (training losses) les plus faibles parmi tous les algorithmes d'optimisation, alors qu'Adadelta a les pertes d'apprentissage les plus élevées. Le tableau 2.6 présente la précision de

reconnaissance de notre modèle proposé avec différents optimiseurs en utilisant un schéma de validation de 80-20%.

Tableau 2.6 – Précision de la reconnaissance de notre modèle proposé sur les bases de données JAFFE et KDEF en utilisant différents optimiseurs.

Optimiseur	Précision de la reconnaissance (Accuracy)	
	JAFFE	KDEF
SGD	95.35%	85.50%
SGD avec momentum	95.35%	85.90%
SGD avec le momentum de Nesterov	97.70%	86.33%
Adam	95.35%	83.70%
Adagrad	95.35%	85.80%
Adadelta	90.70%	82.75%

Le tableau 2.7 présente la précision, le recall, et le F1-score du modèle proposé en utilisant l'optimiseur SGD avec le momentum de Nesterov sur les bases de données JAFFE et KDEF.

Tableau 2.7 – Précision, Recall et F1-score de notre modèle proposé sur les bases de données JAFFE et KDEF.

Type d'émotion	JAFFE			KDEF		
	Precision	F1-score	Recall	Precision	F1-score	Recall
Colère	1.00	1.00	1.00	0.84	0.84	0.84
Tristesse	1.00	0.93	0.88	0.80	0.79	0.77
Surprise	1.00	1.00	1.00	0.87	0.89	0.90
Joie	0.88	0.93	1.00	0.98	0.97	0.96
Dégoût	1.00	1.00	1.00	0.95	0.89	0.84
Peur	1.00	1.00	1.00	0.77	0.78	0.79
Neutre	1.00	1.00	1.00	0.86	0.90	0.96

Les figures 2.19 et 2.20 illustrent les graphiques de précision (accuracy) et de perte (loss) lors des phases d'apprentissage et de test en utilisant l'optimiseur SGD avec le momentum de Nesterov sur les bases de données JAFFE et KDEF. L'axe x se réfère au nombre d'époques d'apprentissage du modèle, qui correspond au nombre de cycles d'apprentissage sur l'ensemble complet de données, et l'axe y indique la perte et la précision respectivement. La ligne bleue représente le comportement des données d'apprentissage, avec une précision finale de 99% sur JAFFE et 98% sur KDEF. Tandis que la ligne orange représente le comportement des données de test, avec une précision finale de 97.7% sur JAFFE et 86.33% sur KDEF en utilisant 80% de données pour l'apprentissage et 20% de données pour le test.

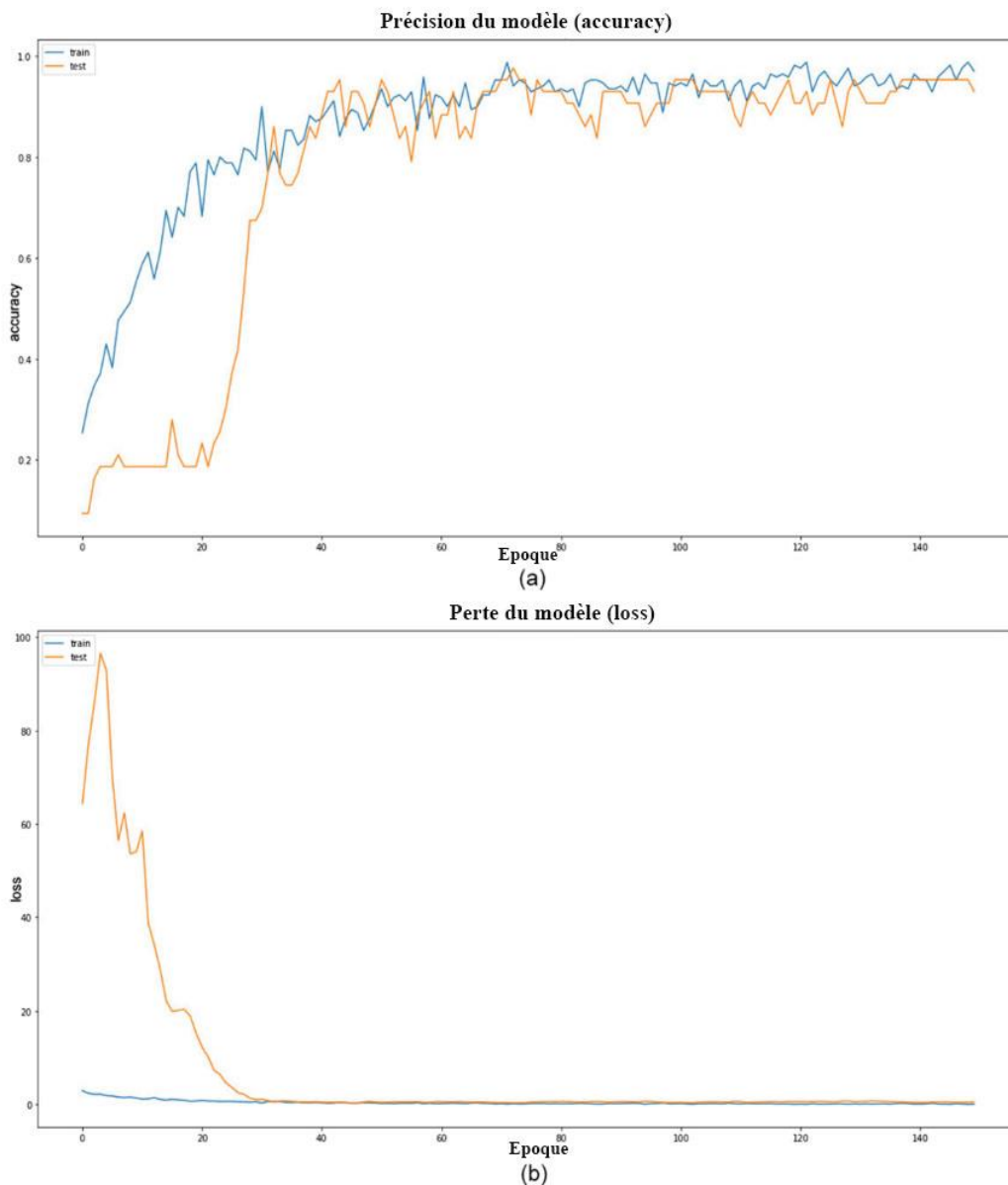


Figure 2.19 – Précision (a) et perte (b) d'apprentissage et de test du modèle VGG-16 sur la base de données JAFFE.

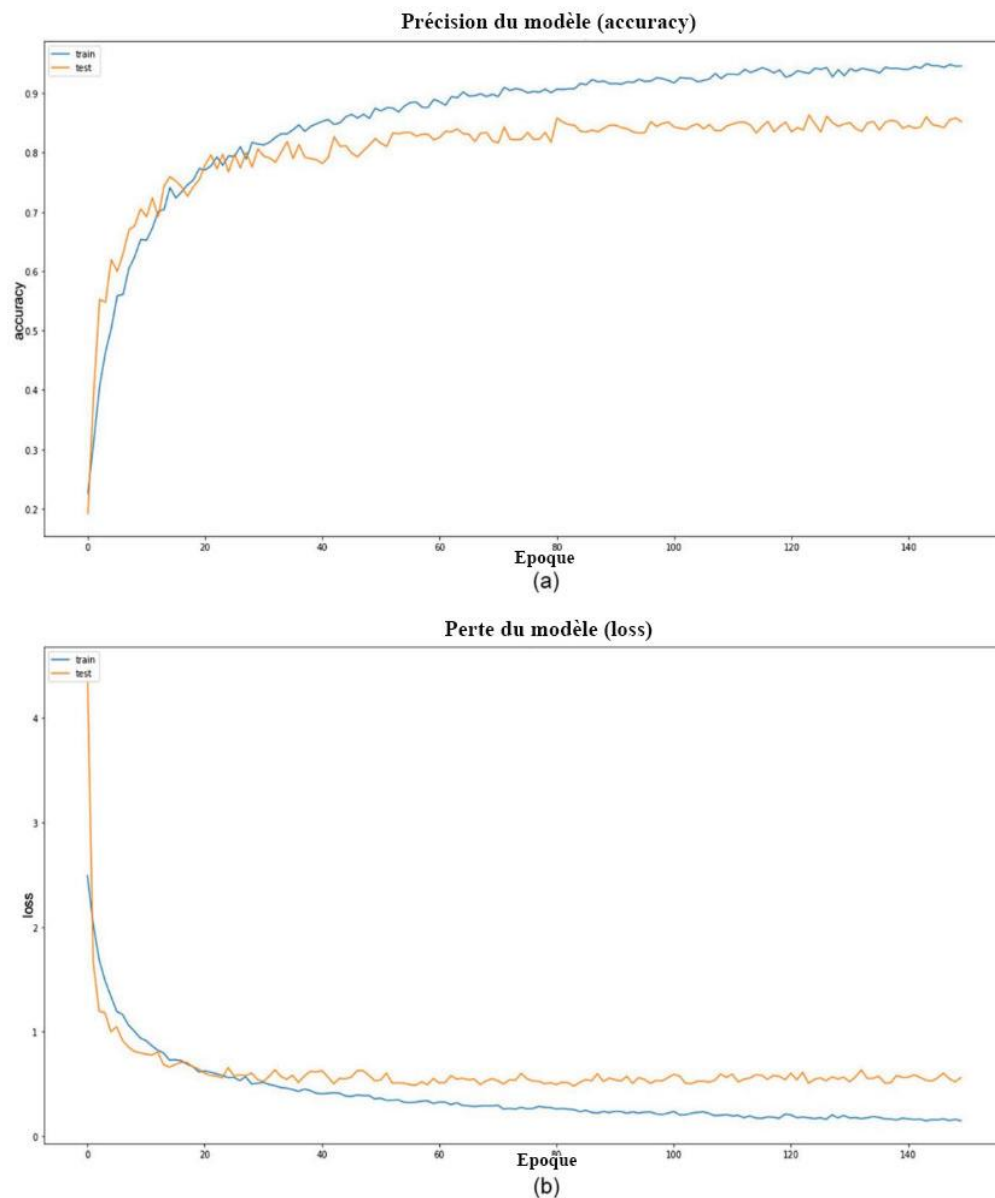


Figure 2.20 – Précision (a) et perte (b) d'apprentissage et de test du modèle VGG-16 sur la base de données KDEF.

Ensuite, afin de présenter et d'observer le fonctionnement de notre approche FER proposée, nous avons utilisé les matrices de confusion pour JAFFE et KDEF validées avec un schéma de validation de 80-20%, comme présenté dans les figures 2.21 et 2.22. Dans ces matrices, les colonnes représentent les classes réelles, et les lignes représentent les classes prédites.

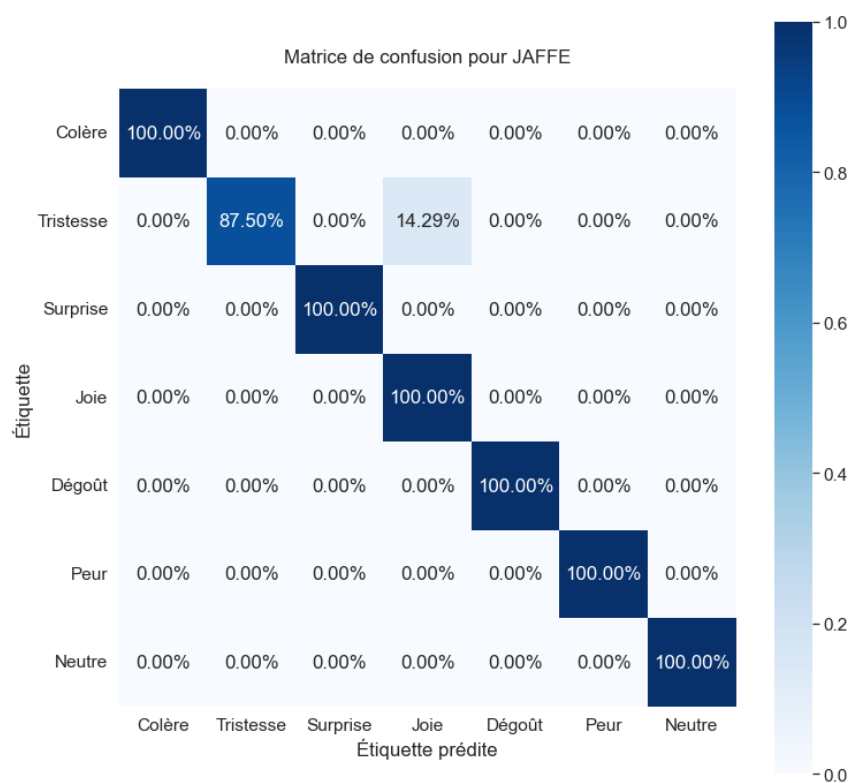


Figure 2.21 – Matrice de confusion de modèle VGG-16 sur la base de données JAFFE.

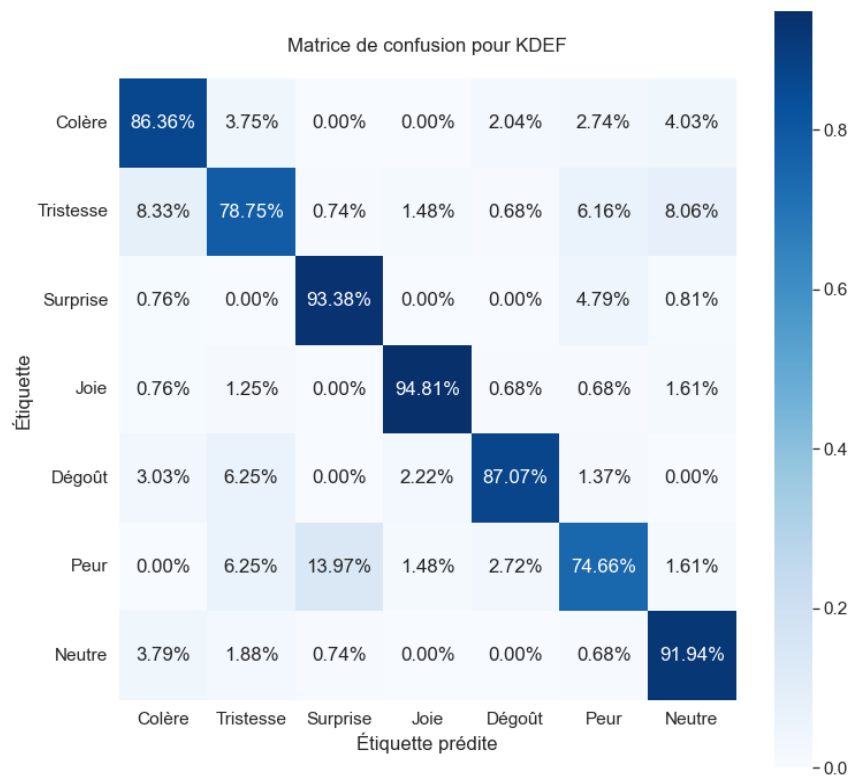


Figure 2.22 – Matrice de confusion de modèle VGG-16 sur la base de données KDEF.

Notre méthode est examinée pour huit modèles DCNN pré-entraînés différents : VGG-16, VGG-19, Inception V3, DenseNet-121, DenseNet-169, MobileNet, ResNet-50 et Xception, comme indiqué dans le Tableau 2.8.

Tableau 2.8 – Comparaison de performance du modèle proposé avec huit modèles DCNN pré-entraînés.

Modèle DCNN pré-entraîné	JAFFE		KDEF	
	20% test	10 Kfold CV	20% test	10 Kfold CV
VGG-16	97.70%	98% ± 0.07	86.33%	99% ± 0.03
VGG-19	97.70%	98% ± 0.04	83.40%	98% ± 0.03
Inception V3	86.04%	97% ± 0.08	72.34%	86% ± 0.02
DenseNet-121	93.02%	98% ± 0.07	80.10%	90% ± 0.02
DenseNet-169	93.02%	98% ± 0.07	79.30%	96% ± 0.03
MobileNet	95.35%	97% ± 0.08	82.70%	92% ± 0.01
ResNet-50	83.72%	98% ± 0.06	50.40%	81% ± 0.05
Xception	86.04%	98% ± 0.07	69.40%	90% ± 0.03

Les résultats montrent que le modèle VGG-16 a obtenu les meilleures précisions de classification de 97.7% et 86.33% pour JAFFE et KDEF sur 20% de données de test, suivi de VGG-19 avec la même précision de 97.7% pour JAFFE et 83.40% pour KDEF. ResNet-50 a présenté les faibles résultats avec une précision de 83.72% pour JAFFE et de 50.4% pour KDEF. De plus, dans le cas de la validation croisée 10 fois (CV), la précision variait de 97% à 99% pour JAFFE et de 81% à 99% pour KDEF. Le modèle VGG-16 a obtenu les meilleures précisions de 98% et 99% pour JAFFE et KDEF, respectivement.

La performance de la méthode FER proposée est comparée à d'autres méthodes de l'état de l'art sur les ensembles de données JAFFE et KDEF. Le Tableau 2.9 présente les méthodes utilisées dans les études antérieures pour reconnaître les émotions faciales, l'année de publication, le total de l'échantillon des deux ensembles de données, la méthode de séparation des données et la précision du test. Dans le cas de validation croisée (leave-one-out (LOOCV)), trois méthodes FER ont été utilisées sur la base de données JAFFE. La méthode de la transformée en contours (CT) avec l'algorithme de boosting basé sur l'analyse discriminante régularisée (RDAB) [151] a obtenu une précision de 96.43%, suivie par la méthode de réseau de neurones convolutifs multicanaux (MCCNN) [158] qui a obtenu

une précision de 95.80%, et la méthode de local de gradient ternaire amélioré (IGLTP) avec SVM [154] qui a présenté une précision de 84.50%. Dans le cas de validation croisée (leave-one-subject-out (LOSO-CV)), le réseau de croyances profond boosted (BDBN) [156] a obtenu une précision de 91.80% sur JAFFE. Pour le mode de séparation de données 90-10%, l'approche de l'histogramme des gradients orientés (HOG) avec SVM [152] a obtenu une précision de 89.50% sur JAFFE et 80.20% sur un échantillon de 980 images de la base de données KDEF. L'approche de l'histogramme des gradients orientés (HOG) avec SVM a été aussi utilisée par [153] avec un mode de séparation de données 70-30% et a obtenu une précision de 76.19% pour JAFFE et 80.95% pour KDEF, suivie par l'approche hybride CNN-RNN [159] qui a obtenu une précision de 94.91% pour JAFFE.

Dans le cas de la validation croisée 10 fois (CV), la méthode de perceptron multicouche (MLP) avec les réseaux de croyance profonds (DBN) [155] a donné une précision de 90.95% pour JAFFE. Cependant, notre approche proposée a obtenu la meilleure précision de reconnaissance, soit 98% pour le JAFFE et de 99% pour le KDEF. Pour le mode de séparation de données 80-20%, notre méthode proposée présente une précision de 97.7% et 86.33% pour JAFFE et KDEF, ce qui surpasse la méthode CNN [157] qui donne une précision de 86.24% pour JAFFE et 82.38% pour KDEF, respectivement. On peut déduire de ce qui précède que notre approche FER proposée avec le modèle d'apprentissage par transfert basé sur VGG-16 a surpassé toute autre méthode de l'état de l'art pour les bases de données JAFFE et KDEF.

Nous avons validé notre approche FER sur des images de sept étudiants sourds et malentendants (4 garçons et 3 filles) en licence fondamentale d'excellence "Électronique, Informatique et Robotique" de l'année universitaire 2020/2021 dans notre Faculté des Sciences (FSR), Université Mohammed V de Rabat. Il convient de souligner que ces étudiants ont volontairement exprimé leur désir de participer à cette expérience. Parmi eux, trois portaient des lunettes et deux étaient dans un angle de prise de vue demi-gauche, comme le montre la figure 2.23. L'expérience a été réalisée dans la salle mécatronique au département physique de la Faculté des Sciences de Rabat. On peut clairement observer à partir de la figure 2.23 que tous les visages ont été reconnus et marqués par des rectangles rouges. Ensuite, chaque émotion est représentée par un texte noir, et la barre rouge représente sa probabilité. L'étiquette de l'émotion dominante avec la valeur maximale est représentée par un texte rouge en haut de chaque rectangle. Par la suite, le type d'engagement calculé à partir de l'émotion dominante de chaque étudiant sourd et malentendant est représenté par un texte blanc en bas de chaque rectangle. Sur les sept visages, cinq ont été étiquetés "Joie" et deux ont été étiquetés "Neutre". En outre, six

étudiants étaient nominalement engagés et un étudiant était fortement engagé.

Tableau 2.9 – Comparaison entre la méthode proposée et les méthodes existantes sur les bases de données JAFFE et KDEF.

Méthodes FER	Méthode de séparation des données	Taille de l'échantillon		Précision du test	
		JAFFE	KDEF	JAFFE	KDEF
Transformée en contours (CT) avec l'algorithme de boosting basé sur l'analyse discriminante régularisée (RDAB) [151]	LOOCV	210	-	96.43%	-
Réseau de croyances profond boosted (BDBN) [156]	LOSO-CV	213	-	91.80%	-
Perceptron multicouche (MLP) avec les réseaux de croyance profonds (DBN) [155]	10-Fold CV	213	-	90.95%	-
Histogramme des gradients orientés (HOG) avec SVM [152]	90-10%	213	980	89.50%	80.20%
Réseau de neurones convolutifs multicanaux (MCCNN) [158]	LOOCV	213	-	95.80%	-
Local de gradient ternaire amélioré (IGLTP) avec SVM [154]	LOOCV	213	-	84.50%	-
Approche hybride CNN-RNN [159]	70-30%	213	-	94.91%	-
Histogramme des gradients orientés (HOG) avec SVM [153]	70-30%	213	4900	76.19%	80.95%
CNN [157]	80-20%	213	490	86.24%	82.38%
Notre approche de VGG-16 basé sur l'apprentissage par transfert	80-20%	213	4900	97.70%	86.33%
	10-Fold CV			98%	99%

La figure 2.24 montre le pourcentage du niveau d'engagement des étudiants sourds et malentendants dans la classe, 85.71% ont été nominalement engagés et 14.29% ont été fortement engagés. D'après les résultats obtenus, l'approche proposée a montré des performances remarquables dans la reconnaissance des expressions faciales et l'évaluation de l'engagement des étudiants sourds et malentendants dans un environnement de classe.



Figure 2.23 – Reconnaissance des émotions faciales et détection de l’engagement de 7 étudiants sourds et malentendants en licence fondamentale d’excellence "Électronique, Informatique et Robotique".

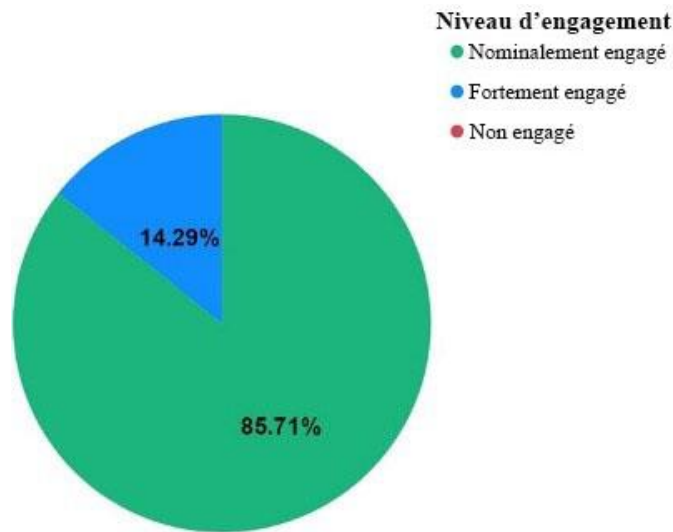


Figure 2.24 – Niveau d’engagement de 7 étudiants sourds et malentendants en licence fondamentale d’excellence "Électronique, Informatique et Robotique".

2.8. Conclusion et discussion

Dans ce chapitre, deux nouvelles approches sont présentées pour aider les enseignants à appliquer efficacement la pédagogie différenciée de manière automatique. La première

approche consiste à classifier en sept émotions de base (peur, colère, joie, surprise, tristesse, neutre et dégoût) les visages des étudiants normaux à l'aide d'un réseau CNN qui comprend 4 couches de convolution, 4 couches de max pooling et 2 couches entièrement connectées, après avoir détecté les visages avec la méthode Haar Cascades. Cette approche a obtenu une précision de 70% sur FER2013. La seconde approche consiste à évaluer en temps réel l'engagement des étudiants à partir de leurs émotions capturées par la caméra, en utilisant un modèle d'apprentissage par transfert basé sur VGG-16 avec une stratégie de fine-tuning. Ce modèle produit trois niveaux d'engagement différents : "nominalement engagé", "fortement engagé" et "non engagé". Il a obtenu la meilleure précision de classification de 97.7% et 86.33% pour JAFFE et KDEF sur 20% des données de test, ainsi que la meilleure précision de classification de 98% et 99% pour JAFFE et KDEF lors de la validation croisée 10 fois (CV). L'optimiseur SGD avec le momentum de Nesterov a montré la perte d'apprentissage la plus faible par rapport aux autres algorithmes d'optimisation.

Partie II

Evaluation de la qualité des systèmes éducatifs

Chapitre 3

Concepts fondamentaux de l'Analyse de Sentiment pour l'Évaluation des Systèmes Éducatifs

3.1. Introduction

Dans un monde en constante évolution, la gestion des systèmes éducatifs représente un défi complexe. Au cours de la dernière décennie, les réformes éducatives ont été courantes à travers le monde, conduisant à une tendance à la restructuration des systèmes éducatifs pour améliorer leur performance. Toutefois, il est désormais essentiel d'évaluer la qualité de ces systèmes éducatifs afin de mesurer leur contribution à l'amélioration générale du niveau de formation, à la lutte contre les inégalités sociales et à l'optimisation du rapport coût/qualité des services.

Face à la variété des pratiques d'évaluation d'un pays à l'autre, de nouvelles méthodes d'évaluation ont émergé. Dans ce chapitre, nous allons définir la qualité de l'éducation (section 3.2). Nous mettrons l'accent sur les fondements théoriques de l'analyse de sentiment en tant que nouvelle méthode d'évaluation de la qualité des systèmes éducatifs, en présentant les différentes tâches, niveaux et étapes du processus d'analyse (section 3.3). Ensuite, nous aborderons l'état de l'art de l'application de l'analyse de sentiment dans le domaine de l'enseignement supérieur (section 3.4). Enfin, nous présenterons une synthèse de ce chapitre dans la section 3.5.

3.2. Définition de la qualité des systèmes éducatifs

La qualité des systèmes éducatifs est essentielle pour assurer un avenir prospère et durable pour les individus et les sociétés, conformément à l'objectif de développement durable des Nations unies pour 2030. Cette notion est complexe et multifacette, nécessitant une approche de pilotage intégrant toutes les composantes pour atteindre les objectifs

pertinents répondant aux besoins d'un environnement social en constante évolution. Afin d'évaluer la qualité d'un système éducatif, il est indispensable de prendre en compte plusieurs relations interdépendantes entre les différentes composantes du système, notamment la faisabilité des objectifs et la pertinence, leur cohérence avec les moyens mobilisés, leur faisabilité, la capacité du système à accomplir ses objectifs de façon efficace, la durabilité et l'équité du système. Pour atteindre ces objectifs, un comité de pilotage peut utiliser des tableaux de bord pour gérer de manière systémique toutes ces composantes.

Les travaux de Bengston [173] soulignent que la qualité de l'éducation est une notion complexe qui est souvent sujette à des débats permanents. Étant donné la variété des systèmes éducatifs et des structures de valeurs, il n'y a pas de définition universelle de la qualité. Scheerens [174] propose une définition simple de la qualité de l'éducation en se référant aux systèmes d'indicateurs internationaux disponibles. Selon lui, la qualité de l'éducation est déterminée par les performances des élèves dans des tests standardisés. Cette définition se concentre sur les résultats plutôt que sur les processus éducatifs. D'un autre côté, Kellaghan et Greaney [175] considèrent la qualité comme l'adéquation ou l'adaptation des objets ou des processus aux fins pour lesquelles ils ont été prévus. Cette définition souligne l'importance de la pertinence et la faisabilité des objectifs éducatifs. Les définitions conventionnelles de la qualité de l'éducation se sont concentrées sur les facteurs tels que la politique, la gestion et l'administration, la planification, les enseignants, les contenus, les méthodologies, le curriculum ainsi que les systèmes d'évaluation. Toutefois, il convient de remarquer que la qualité de l'éducation peut varier selon les contextes, les circonstances, les systèmes et les ressources.

Le Cadre UNESCO pour la compréhension de la qualité de l'éducation [176] est un cadre complet qui identifie les éléments clés des systèmes éducatifs indispensables pour parvenir à une éducation de qualité (Figure 3.1). Le cadre met l'accent sur le fait qu'une éducation de qualité ne se restreint pas seulement à l'excellence académique, mais inclut également le développement de la pensée critique, de l'innovation et d'autres compétences essentielles qui préparent les apprenants aux défis du XXI^e siècle. Le cadre identifie cinq dimensions interdépendantes qui influencent la qualité de l'éducation :

- *Dimension des caractéristiques des apprenants* : fait référence aux besoins, aux capacités et aux antécédents uniques des apprenants. Cette dimension accorde une grande importance à la nécessité de comprendre les styles et les besoins d'apprentissage diversifiés des étudiants et d'adapter les méthodes et les ressources d'enseignement pour répondre à leurs besoins individuels. Les

éducateurs doivent créer un environnement d'apprentissage favorable et inclusif qui favorise le développement de compétences essentielles comme la communication, la pensée critique et la résolution de problèmes.

- *Dimension contextuelle* : souligne l'importance de considérer le contexte social, économique et culturel plus vaste dans lequel l'éducation se déroule. Cela inclut des facteurs tels que le genre, la langue, l'ethnicité, et le statut socio-économique. Il est primordial de comprendre le contexte dans lequel les apprenants vivent pour garantir que l'éducation est pertinente, significative et adaptée à leurs besoins.
- *Dimension des moyens* : fait référence aux ressources et à l'infrastructure nécessaires pour soutenir une éducation de qualité. Cela inclut un financement adéquat, des enseignants bien formés, une direction scolaire de soutien, un accès à la technologie et aux ressources éducatives, ainsi que des environnements d'apprentissage sûrs et sains.
- *Dimension enseignement et apprentissage* : fait référence aux approches pédagogiques, à la conception des programmes et aux stratégies d'évaluation utilisées pour faciliter l'apprentissage. Cette dimension accorde une grande importance à l'usage approprié des méthodes d'enseignement centrées sur l'apprenant et basées sur l'enquête qui encouragent l'apprentissage actif, la pensée critique et la résolution de problèmes. Les stratégies d'enseignement et d'apprentissage efficaces doivent être alignées sur les besoins des apprenants et soutenues par les ressources et l'infrastructure appropriées.
- *Dimension des résultats* : fait référence aux résultats d'apprentissage souhaités et à l'impact de l'éducation sur les apprenants et la société. Les résultats d'apprentissage doivent être pertinents et significatifs pour les apprenants et contribuer à leur développement personnel et à leur réussite future. Ils doivent être mesurés de manière efficace pour évaluer la qualité de l'éducation et identifier les domaines qui nécessitent des améliorations.

La globalisation et les nouvelles règles du jeu dans l'éducation ont entraîné l'impératif de créer des indicateurs pour évaluer la qualité de l'éducation et mesurer les compétences, les connaissances, les comportements et les attitudes appropriées des étudiants. Cependant, Il est complexe d'évaluer la qualité de l'éducation en raison de la complexité à élaborer des indicateurs de contexte sur des questions qualitatives, et à mesurer

des processus dépendant de normes, de la qualité de l'enseignement, du temps consacré, du climat scolaire et du leadership éducatif. Pour surmonter ces défis, l'analyse des sentiments à partir des médias sociaux peut compléter les indicateurs traditionnels en fournissant des données qualitatives sur l'expérience des apprenants et des enseignants qui ne sont pas considérés dans les indicateurs classiques. En effet, des études ont montré que l'éducation pourrait être plus similaire que différente selon les cultures et les pays [177]. L'analyse de sentiment peut également être utilisée pour mesurer la réputation des universités en évaluant les opinions et les perceptions des étudiants, des enseignants et des employeurs sur la qualité de l'innovation, de l'enseignement, de la recherche, des installations et des services proposés par l'université. Cette mesure de la réputation est importante pour les universités, car elle peut influencer le nombre d'étudiants qui postulent à l'université, le choix des employeurs pour le recrutement et les partenariats avec d'autres institutions. En outre, l'analyse de sentiment peut aider à prendre des décisions concernant le passage d'un enseignement entièrement présentiel à un enseignement hybride combinant distance et présentiel adapté aux besoins de la communauté éducative.

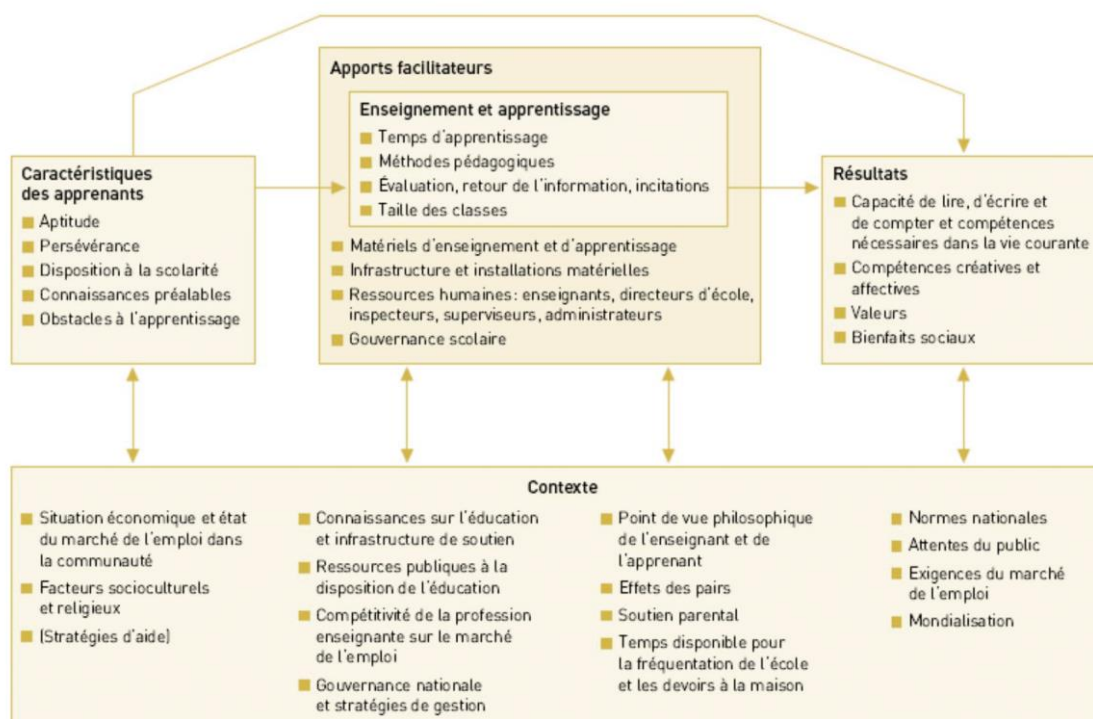


Figure 3.1 – Cadre UNESCO pour la compréhension de la qualité de l'éducation.

3.3. Fondements théoriques de l'analyse de sentiment

Le traitement du langage naturel (ou Natural Language Processing (NLP)) comprend l'étude de modèles mathématiques et informatiques de différents aspects du langage et le développement d'une variété étendue de systèmes. La recherche dans le domaine du traitement du langage naturel est hautement interdisciplinaire, incluant des concepts en informatique, en linguistique, en logique et en psychologie. Le NLP joue un rôle unique dans l'informatique, en particulier dans le sous-domaine de l'intelligence artificielle. De nombreux aspects de ce domaine traitent des caractéristiques linguistiques de l'informatique et la NLP cherche à modéliser le langage de manière informatique. Il existe deux tâches principales dans le domaine de le NLP : la compréhension du langage (ou analyse de texte) et la génération du langage. L'analyse des sentiments, est l'étude des opinions, des attitudes, et des émotions des personnes. Elle est considérée comme l'un des champs de recherche les plus dynamiques dans le domaine du traitement du langage naturel (NLP). Il vise à identifier, extraire, résumer ou catégoriser automatiquement les opinions, évaluations, appréciations, attitudes et émotions exprimées dans un texte à l'égard d'une entité ou d'un événement.

3.3.1. Définition de l'opinion

Les opinions sont généralement des expressions subjectives qui décrivent les sentiments, les appréciations ou les sensations des personnes à l'égard d'entités, d'événements et de leurs propriétés. Il est courant de considérer les phrases neutres d'un document comme objectives et les phrases d'opinion comme subjectives. Les éléments de base d'une opinion exprimée sont :

- *Entité* : elle peut s'agir d'un produit, d'une personne, d'un événement, d'une organisation ou d'un sujet sur lequel une opinion est exprimée (cible d'opinion). Une entité est composée d'une hiérarchie de composants et de sous-composants, chaque composant pouvant avoir un ensemble d'attributs. Par exemple, un téléphone portable est composé d'un écran, d'une batterie, d'autres composants, dont les attributs peuvent être la taille et le poids. Pour simplifier, les composants et les attributs sont tous deux appelés aspects.
- *Porteur de l'opinion* : est la personne ou l'organisation qui détient une opinion spécifique sur une entité particulière. Par exemple, dans les critiques ou les articles de blog, les porteurs sont généralement les auteurs des documents.
- *Sentiment* : est la vue, l'attitude, l'évaluation ou l'appréciation d'un objet par un

porteur d'opinion. Les principales caractéristiques du sentiment sont la polarité et l'intensité. Nous pouvons considérer un sentiment comme un vecteur : la direction du vecteur serait sa polarité et la longueur serait son intensité. La polarité peut être positive ou négative, et l'intensité peut être représentée par une échelle discrète ou continue. Une polarité positive correspond à un sentiment positif tel que l'appréciation, la satisfaction ou la joie. Une polarité négative définit des sentiments opposés : colère, insatisfaction, déception. Certains chercheurs ajoutent également une polarité neutre qui correspondrait à un vecteur nul.

Compte tenu des composantes des opinions présentées ci-dessus, une opinion est définie comme un quintuplet de quatre facteurs $(e_i, a_{ij}, os_{ijkl}, p_k, t_l)$, où e_i est une entité, a_{ij} est un aspect de e_i et os_{ijkl} est l'orientation de l'opinion ou bien le sentiment exprimé par le détenteur p_k sur a_{ij} pendant la période t_l . Les valeurs possibles pour os_{ijkl} sont les catégories positives, négatives et neutres.

3.3.2. Types d'opinion

Il existe deux types d'opinion : les opinions régulières et les opinions comparatives. Une opinion régulière exprime un sentiment à l'égard d'une entité particulière ou d'un aspect de cette entité. Par exemple, la phrase ("Les écouteurs sont confortables à porter.") exprime un sentiment positif à l'égard de la qualité sonore des écouteurs. Les opinions régulières peuvent directement exprimer un sentiment envers une entité, par exemple ("Les écouteurs sont inconfortables à porter."), ou indirectement, par exemple ("Après avoir porté les écouteurs, mes oreilles ont commencé à bourdonner."). Les opinions régulières sont beaucoup plus faciles à détecter et à classer. D'autre part, une opinion comparative compare plusieurs entités sur la base de certains de leurs aspects communs. Par exemple, la phrase ("Les écouteurs de la marque x sont plus confortables à porter que les écouteurs de la marque y.") compare les écouteurs avec une autre marque sur la base de leur confort, et exprime une préférence pour les autres écouteurs.

3.3.3. Tâches d'analyse du sentiment

De nombreuses études définissent l'analyse du sentiment et l'extraction d'opinions comme étant constituées de plusieurs tâches différentes. Pang et al. [178] ont résumé les principales tâches de l'analyse du sentiment en extraction d'opinions, classification des sentiments, détermination de la polarité et résumé d'opinions, tandis que dans [179] l'auteur

a énuméré la classification de la subjectivité et des sentiments, l'analyse des sentiments basée sur les aspects, la construction d'un lexique des sentiments, le résumé d'opinions, l'analyse d'opinions comparatives, la recherche et la récupération d'opinions, la détection de spam d'opinions et la qualité des critiques, comme étant les principales tâches de l'analyse des sentiments. Dans cette section, nous décrivons les tâches communes de l'analyse des sentiments et de l'extraction d'opinions (Figure 3.2).

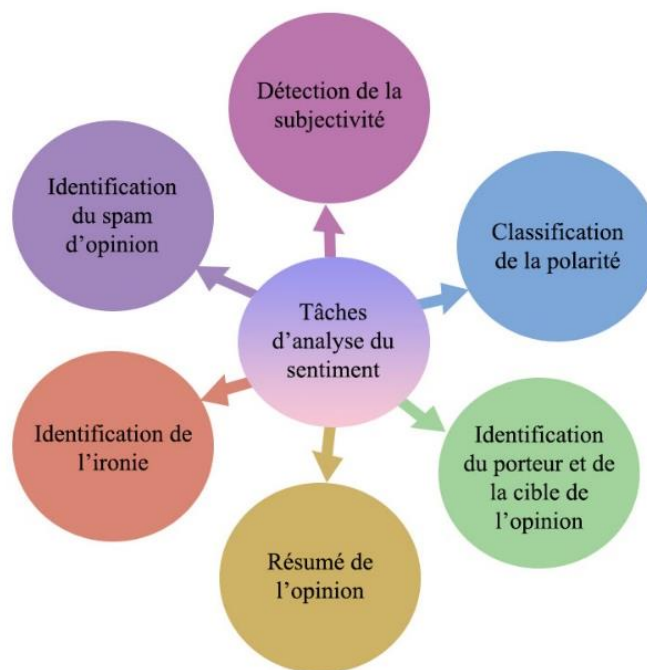


Figure 3.1 – Différentes tâches de l'analyse du sentiment.

Le plan proposé est comme suit : Détection de la subjectivité, classification de la polarité, identification du porteur et de la cible de l'opinion, résumé de l'opinion, identification de l'ironie et identification du spam d'opinion. Nous fournissons une définition de chaque tâche, les problèmes communs et les principales approches.

- *Détection de la subjectivité* : est un problème de détection de la présence d'opinions dans un texte donné. En termes simples, elle détermine ce qui est une opinion et ce qui ne l'est pas. L'analyse de la subjectivité peut traiter les sentiments en séparant les textes polaires des textes neutres. En général, la détection de la subjectivité est considérée comme un problème de classification binaire : pour un texte donné, le système doit renvoyer un message vrai s'il contient des opinions et faux dans le cas contraire. L'approche courante consiste à utiliser un classificateur d'apprentissage automatique formé sur deux ensembles de textes représentant les opinions positives et les opinions négatives.

Il est relativement facile de collecter un ensemble de textes contenant uniquement des faits et un autre contenant des opinions. Pang et Lee [180] ont augmenté le processus de classification de polarité avec une étape supplémentaire de prétraitement où les phrases d'un texte analysé sont classées comme subjectives ou objectives. Les auteurs ont traduit la classification de la subjectivité en un problème de partitionnement de graphe et ont utilisé le théorème min-cut max-flow pour le résoudre. Les phrases étiquetées comme "subjectives" sont extraites et transmises à un classificateur de polarité général (model bag-of-words avec SVM). Ils ont rapporté une amélioration statistiquement significative de la précision de classification de 82.8% à 86.4%.

- *Classification de la polarité* : en général, la tâche principale de la classification de polarité est de déterminer si un texte exprime une attitude positive ou négative de son auteur envers le sujet du texte. Une tâche plus avancée consiste à définir le degré de polarité (faible ou fort) et également à être capable d'identifier une polarité mixte, c'est-à-dire positive et négative à la fois. Les commentaires tels que "pouce levé" ou "pouce baissé", ou "aimer" ou "ne pas aimer" sont des exemples de classification de polarité à deux classes. La classification de polarité est le sujet principal de notre travail, le chapitre 5 (section 5.2) discute plus en détail de nombreux travaux connexes antérieurs.
- *Identification du porteur et de la cible de l'opinion* : l'objectif de cette tâche est de filtrer les opinions qui sont pertinentes pour le sujet donné, puisqu'il peut y avoir plusieurs opinions dans un texte sur différents sujets. Par exemple, s'il s'agit d'une analyse de film et que l'auteur parle de l'expérience d'une sortie au cinéma, nous voulons extraire uniquement les opinions relatives au film et non au cinéma. Connaître le porteur de l'opinion nous aide à estimer les données démographiques ou à recueillir les opinions d'une personne spécifique. Ce dernier point est utile pour la personnalisation de l'utilisateur, c'est-à-dire pour sélectionner les sujets qu'un utilisateur spécifique préfère et éviter les choses qu'il n'aime pas.
- *Résumé de l'opinion* : est un problème bien connu en analyse de texte. La tâche du résumé est de fournir une version plus courte d'un texte en préservant son idée principale. Le résumé d'opinion est une tâche similaire dans le domaine de l'extraction d'opinion. Son objectif principal est d'extraire les opinions du texte et de les présenter sous une forme plus courte. En général, le problème consiste à extraire des phrases d'opinion du texte original en les réordonnant si

nécessaire. Nishikawa et al. [181] ont traduit le problème de résumé d'opinion en une formulation de programmation linéaire en nombres entiers en maximisant la fonction de notation qui mesure les poids d'opinion des phrases sous réserve de la cohérence des phrases extraites.

- *Identification de l'ironie* : est une autre tâche à résoudre pour obtenir un système d'analyse des sentiments précis. La théorie rhétorique distingue plusieurs types d'ironie, parmi lesquels l'ironie verbale est celle qui a fait l'objet de plusieurs recherches en analyse des sentiments. L'ironie verbale est une déclaration dans laquelle le sens employé par le locuteur est fortement différent du sens ostensiblement exprimé. La déclaration ironique implique généralement l'expression explicite d'une attitude ou d'une évaluation, mais avec des indications dans la situation globale du discours que le locuteur a l'intention d'adopter une attitude ou une évaluation très différente, et souvent opposée. Certains dictionnaires définissent le sarcasme comme un type d'ironie verbale, tandis que d'autres affirment qu'il n'y a pas de relation entre ces deux termes. Les chercheurs en linguistique informatique définissent généralement l'ironie comme l'expression d'émotions négatives en utilisant des mots ayant un sens littéral positif (ou vice versa). Carvalho et al. [182] ont signalé que 35% des erreurs de classification de la polarité étaient dues à l'ironie verbale lors de l'application de leurs règles de détection de la polarité à l'analyse de débats politiques. D'autres chercheurs ont également mentionné que l'ironie provoque des erreurs de classification de la polarité, car elle est utilisée pour exprimer des sentiments négatifs avec des phrases positives. Les approches existantes de l'identification de l'ironie dans les textes sont basées sur l'apprentissage automatique. Les auteurs dans [183] ont présenté un ensemble de motifs servant d'indices pour distinguer les phrases ironiques des phrases non ironiques. Les modèles les plus fréquemment observés étaient ceux capturant des interjections, des ponctuations, des guillemets et des indicateurs de rire. Reyes et al. [184] ont appliqué une approche d'apprentissage automatique. La liste des caractéristiques utilisées dans leur méthode se composait des n-grammes, des POS et des caractéristiques de profilage obtenues en recherchant des mots du texte examiné dans différents dictionnaires, tels que des lexiques affectifs. L'identification de l'ironie reste un problème ouvert pour l'analyse des sentiments. Les approches existantes n'appliquent que des méthodes couramment utilisées telles que l'apprentissage supervisé et la comparaison de motifs.

- *Identification du spam d'opinion* : les opinions du public ne servent pas seulement de retour d'information pour les fabricants de produits (fournisseurs de services, partis politiques, etc.), mais elles influencent également la décision des autres personnes lorsqu'elles choisissent un produit. Ainsi, de nombreuses parties sont intéressées par la publication des opinions positives sur elles-mêmes et des opinions négatives sur leurs concurrents. Dans un tel environnement, la question de l'authenticité des opinions se pose, car aujourd'hui le Web offre de nombreux moyens d'exprimer publiquement des opinions sans aucun moyen de vérifier leur fiabilité, ce qui crée une nouvelle direction de recherche appelée l'identification du spam d'opinion. En outre, les chercheurs définissent également le spam d'opinion comme des avis répétitifs de faible qualité créés à des fins publicitaires. Le spam d'opinion peut être nuisible pour les systèmes automatiques qui traitent les opinions, puisqu'il ajoute du bruit aux données d'apprentissage. Un défi typique de ces études est de collecter un corpus annoté suffisamment large. Jindal et Liu [185] ont appliqué une heuristique pour extraire les commentaires mensongers d'Amazon. Pour collecter ces avis, les auteurs ont analysé 5.8 millions d'avis provenant de 2.14 millions d'évaluateurs afin de trouver les auteurs qui ont posté des commentaires répétitifs similaires sur des produits, considérés comme du spam d'opinion. Cependant, dans un cas général, les chercheurs peuvent ne pas disposer d'informations sur l'auteur et tout ce qu'il y a à analyser est le texte de l'avis. Au fur et à mesure que les opinions sur le Web gagnent en importance, elles seront davantage malveillantes, ce qui fait de l'identification du spam d'opinion un problème important pour l'avenir de l'analyse des sentiments.

3.3.4. Niveaux d'analyse du sentiment

Le sentiment peut être exprimé dans des textes de différentes longueurs, allant de textes courts, superficiels et informels, tels que les messages de microblogging, à des textes longs, détaillés et plus formels, comme les articles de blog. Les mots seuls peuvent susciter des sentiments, mais lorsqu'ils sont combinés à d'autres mots pour former une unité lexicale plus large, comme des phrases, différents sentiments peuvent être exprimés. L'analyse des sentiments peut être appliquée à des textes de différents niveaux de granularité : niveau du document, niveau de la phrase, niveau du mot, niveau de l'aspect, et niveau du concept, comme le montre la figure 3.3.

Dans cette section, nous allons décrire en détail chacun de ces niveaux dans les sous-sections suivantes.

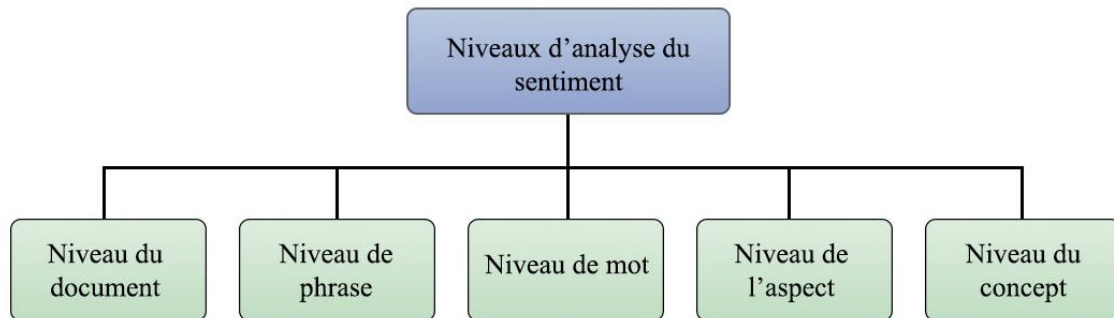


Figure 3.2 – Différents niveaux d'analyse du sentiment.

- *Niveau du document* : la tâche à ce niveau consiste à déterminer si un document texte exprime un sentiment. Le résultat à ce niveau est souvent une information générale sur les documents. Les sentiments de polarité sont finalement résumés sur l'ensemble du document comme positifs ou négatifs. Cependant, l'un des inconvénients de l'analyse des sentiments à ce niveau est que l'on suppose que l'ensemble du document exprime une seule opinion à l'égard d'une seule entité. Dans la pratique, un document d'opinion peut évaluer plusieurs entités et exprimer différents sentiments à leur égard. La plupart des premières études sur l'analyse des sentiments se sont concentrées sur le niveau du document et se sont appuyées sur des ensembles de données tels que les critiques de films et de produits [186-188].
- *Niveau de phrase* : étant donné que le résultat de l'analyse des sentiments au niveau du document est général et ne fournit pas d'informations précises, de nombreuses études ont commencé à utiliser les phrases du document comme une approche pour analyser les opinions [189, 190]. L'analyse des sentiments au niveau des phrases vise à classer les opinions dans chaque phrase. Elle se base sur deux tâches de classification. Premièrement, la classification de la subjectivité vise à distinguer les phrases qui reflètent des informations et des faits (phrases objectives) des phrases qui expriment des points de vue et des opinions (phrases subjectives). La seconde est la classification de la polarité des phrases (positive ou négative).
- *Niveau de mot* : le sentiment peut également résider dans des unités linguistiques encore plus petites, comme des mots (par exemple, génial, affreux). L'analyse à

ce niveau vise à classifier le sentiment général d'un segment de texte en distinguant le sentiment des unigrammes et des bigrammes individuels qu'il contient. Cette tâche implique souvent l'utilisation de lexiques affectifs, qui contiennent des entrées de mots qui véhiculent un biais subjectif, afin d'extraire des mots de sentiment à partir de contextes.

- *Niveau de l'aspect* : la classification des sentiments du texte au niveau du document et de la phrase est essentielle dans de nombreux cas, mais elle ne présente pas tous les détails nécessaires. Par exemple, être positif sur une entité particulière n'implique pas que l'opinion de l'auteur soit positive sur tous les aspects d'une entité. De même, les sentiments négatifs ne représentent pas l'opinion négative de l'auteur sur tous les aspects d'une entité. Pour une comparaison fine de deux ou plusieurs produits de catégories similaires, nous devons identifier les avantages et les inconvénients de divers composants et caractéristiques (aspects). L'extraction d'opinion au niveau des aspects examine l'opinion donnée elle-même au lieu de s'intéresser aux structures linguistiques (document, phrase ou expression). L'objectif de ce niveau est d'identifier et d'extraire les aspects de chaque entité nommée apparaissant dans le texte, puis de leur attribuer une polarité. Un résumé des sentiments sur les différents aspects de l'entité souhaitée est le résultat le plus courant dérivé de ce niveau d'analyse des sentiments. De nombreuses études ont traité de ce niveau [191-193]. Il convient de noter que ce niveau d'exploration d'opinion fournit une analyse plus approfondie de l'entité cible.
- *Niveau du concept* : l'objectif est de fournir de nouvelles approches de l'exploration d'opinion et de l'analyse des sentiments qui permettent un passage plus efficace de l'information textuelle (non structurée) aux données (structurées) traitables par machine, dans n'importe quel domaine. Les méthodes d'analyse des sentiments basées sur les concepts se concentrent sur une analyse sémantique du texte à l'aide d'ontologies web ou de réseaux sémantiques, qui permettent l'agrégation d'informations conceptuelles et affectives associées aux opinions en langage naturel. En s'appuyant sur de vastes bases de connaissances sémantiques, ces approches s'éloignent de l'utilisation modeste des mots-clés et des comptes de cooccurrence des mots. Au lieu de compter les cooccurrences, elles s'appuient sur les caractéristiques implicites associées aux concepts du langage naturel. Contrairement aux techniques purement syntaxiques, les approches basées sur les concepts peuvent également découvrir des sentiments

exprimés de manière subtile, par exemple, tout au long de l'analyse d'idées qui ne véhiculent pas explicitement une émotion, mais qui sont implicitement liées à d'autres concepts qui le font. Cambria et al. [194] ont présenté le niveau conceptuel de l'exploration d'opinion comme une nouvelle avenue dans l'analyse des sentiments. L'analyse des émotions au niveau des concepts est basée sur l'inférence de informations conceptuelles sur l'émotion et le sentiment associés au langage naturel. Poria et al. [195] ont amélioré la précision de la détection de la polarité par une nouvelle approche. Une analyse des commentaires au niveau conceptuel a été proposée et intègre des techniques linguistiques, et des techniques d'apprentissage automatique. Leurs résultats montrent que la méthode proposée a une précision souhaitable et meilleure que les méthodes statistiques classiques. Un dictionnaire de sentiments au niveau conceptuel a été construit par Tsai et al. [196] basé sur la connaissance du sens commun.

3.3.5. Prétraitement du texte

Le prétraitement du texte (ou Text pre-processing) est souvent la première étape d'un système de traitement du langage naturel (NLP). Il s'agit du procédé de nettoyage et de préparation du texte en vue de la classification. Les textes présents sur le web contiennent généralement une quantité importante de perturbations et de parties non informatives, telles que les balises HTML, les scripts et les publicités. De plus, sur le plan lexical, de nombreux termes dans le texte n'influencent pas l'orientation globale de celui-ci. La préservation de ces termes entraîne une dimensionnalité élevée du problème, rendant ainsi la classification plus ardue, car chaque terme du texte est traité comme une dimension. Ainsi, réduire le bruit dans le texte améliore les performances du classificateur et accélère le processus d'analyse des sentiments. L'ensemble du processus de prétraitement du texte comporte plusieurs étapes. Ci-dessous, les étapes de prétraitement sur lesquelles se concentre le travail d'analyse de texte de cette thèse, à savoir :

- *Normalisation de texte* : est une étape importante pour la plupart des problèmes NLP et en particulier pour le problème d'analyse de sentiment. Le texte des critiques (reviews) et des messages sociaux (comme les tweets) est généralement plus bruyant que le texte normal des articles et des blogs. Les gens ont tendance à écrire leurs posts, messages et critiques de manière informelle. Ainsi, dans cette thèse, nous avons utilisé plusieurs étapes de normalisation du texte afin d'utiliser la plupart des informations contenues dans le texte, en supprimant les mots non

significatifs (ou stop words), les chiffres, les URL, les noms d'utilisateur, les caractères spéciaux et non ASCII, puis en mettant le texte en minuscule, et en procédant à la tokenisation, au stemming, à la lemmatisation et à l'annotation des mots.

- *Étiquetage morpho-syntaxique* : aussi appelé POS tagging (part-of-speech tagging) en anglais, est une tâche NLP basée sur la phrase. Une phrase est formée d'une séquence de mots, l'étiquetage morpho-syntaxique (POS) tente d'étiqueter chaque mot avec sa partie correcte du discours. Les étiquettes POS sont utiles, car elles exposent un grand nombre d'informations sur un mot et ses voisins. Par exemple, le fait de savoir si un mot est un substantif ou un verbe nous indique sur les mots voisins probables (les noms sont précédés de déterminants et d'adjectifs, les verbes par les noms) et de la structure syntaxique (les noms font généralement partie de syntagmes nominaux). C'est pourquoi, dans cette thèse, nous avons utilisé les étiquettes POS pour extraire des caractéristiques à partir du texte.

3.3.6. Extraction de caractéristiques textuelles

Il est assez facile pour l'être humain de déterminer le sentiment exprimé dans un texte, mais il est assez difficile d'accomplir la même tâche automatiquement. Afin de détecter le sentiment, les machines s'appuient souvent sur la présence ou la fréquence de caractéristiques textuelles, qui fournissent des informations utiles sur le sentiment exprimé. La deuxième étape de l'analyse du sentiment est l'extraction des caractéristiques du texte, notamment : les mots de sentiment (par exemple, bon, mauvais), la négation (par exemple, pas, jamais) qui peut inverser la polarité des mots de sentiment, les caractéristiques syntaxiques des mots (par exemple, l'étiquetage morpho-syntaxique (POS)). Il existe plusieurs méthodes d'extraction de caractéristiques textuelles telles que le sac de mots (ou Bag of Words en anglais), Count Vectorizer, Term Frequency-Inverse Document Frequency (TF-IDF) [197], Word2Vec, et Global Vectors for Word Representations (Glove) [198]. Nous décrivons les méthodes Count Vectorizer et Term Frequency-Inverse Document Frequency (TF-IDF) utilisées dans cette thèse dans le chapitre 4 (sous-section 4.3.1).

3.3.7. Classification du sentiment

L'étape la plus importante et la plus critique de l'extraction d'opinion est la sélection d'une technique appropriée pour classer les sentiments. La classification des sentiments,

également appelée détermination de la polarité, vise à déterminer la polarité d'un objet (document, phrase, etc.) afin de savoir s'il exprime un sentiment positif, négatif ou neutre envers le sujet. La classification des sentiments a été appliquée aux réseaux sociaux, les blogs, les articles de presse, etc. Les méthodes de classification proposées dans la littérature peuvent être classées dans trois groupes : approches basées sur des lexiques (lexicon-based), apprentissage automatique (machine learning), et approches hybrides.

- *Approches basées sur des lexiques (lexicon-based)* : elles permettent de classer un document, une phrase, ou un aspect comme positif ou négatif sur la base de listes de mots qui représentent les deux classes de polarité de base, telles que bon, merveilleux, beau, étonnant, etc. Ou mauvais, affreux, laid, terrible, etc. L'approche basée sur le lexique utilise un algorithme pour classifier le sentiment général d'un segment de texte, en agrégeant les sentiments des mots individuels tels qu'ils ont été extraits du lexique. Cette méthode est une approche simple, mais naïve, pour déterminer le sentiment général d'un segment de texte. Dans [199], les auteurs ont proposé une approche basée sur le lexique pour annoter des tweets non étiquetés avec des classes de polarité concernant une entité donnée en agrégeant les polarités des mots d'un lexique avec des mots positifs et négatifs en utilisant une fonction de notation. Les tweets étiquetés automatiquement sont ensuite utilisés pour entraîner un classificateur. Une autre approche basée sur le lexique est proposée dans [200]. Les auteurs ont construit un graphe dont les nœuds sont des utilisateurs, des tweets, des mots, des hashtags et des émoticônes. Un sous-ensemble de ces nœuds est étiqueté par une connaissance préalable du sentiment fournie par un lexique de polarité.
- *Approches d'apprentissage automatique* : elles sont classées en deux catégories principales à savoir les techniques supervisées et non supervisées. Le succès de ces deux catégories repose principalement sur la sélection et l'extraction d'un ensemble approprié de caractéristiques utilisées pour détecter les sentiments. L'apprentissage automatique supervisé repose sur un ensemble de données d'apprentissage annotées, où la catégorie à laquelle une instance appartient est connue. L'approche non supervisée (ou clustering), consiste à regrouper les instances d'apprentissage, dont les catégories sont inconnues, en catégories basées sur une certaine mesure de similarité ou de distance inhérente entre les caractéristiques qu'elles possèdent. La classification non supervisée est plus généralisée, mais dans la plupart des cas, la classification supervisée permet d'obtenir une meilleure précision. Pour les deux techniques, les instances non

vues, souvent appelées données de test, sont classées en les comparant aux instances d'apprentissage à l'aide d'algorithmes, appelés classificateurs. De nombreux classificateurs ont été utilisés dans la littérature pour l'analyse de sentiment. Pang, Lee et Vaithyanathan [180], qui ont mené l'une des toutes premières études empiriques pour classifier les opinions exprimées dans les critiques de films, ont évalué trois classificateurs : naïve bayes, maximum entropy, et machines à vecteurs de support (SVM). Les classifieurs SVM sont connus pour leur haute performance et ont été largement utilisés dans les problèmes de classification de sentiments [201, 202, 203]. Des approches d'apprentissage profond ont également été adoptées pour l'analyse des sentiments sur Twitter. Une architecture de réseau de neurones convolutif (CNN) est développée dans [204]. Chaque tweet est représenté comme une matrice dont les colonnes correspondent aux mots du tweet, en préservant l'ordre dans lequel ils apparaissent. Le réseau est formé des couches suivantes : une couche d'entrée avec la matrice de tweet donnée, une couche convolutive unique, une fonction d'activation linéaire rectifiée, une couche de max pooling et une couche de classification Softmax. Les poids du réseau neuronal sont pré-entraînés à l'aide de données annotées par des émoticônes, puis entraînés avec les tweets annotés manuellement lors de la compétition SemEval. Les résultats expérimentaux montrent que la phase de pré-entraînement permet une initialisation correcte des poids du réseau, et a donc un impact positif sur la précision de la classification.

- *Approches hybrides* : elles combinent deux ou plusieurs techniques telles que l'approche basée sur des lexiques, l'apprentissage automatique supervisé, l'apprentissage automatique non supervisé, ou même toutes ces techniques ensemble pour améliorer les performances de la classification des sentiments. Dans certains travaux antérieurs, il a été prouvé que l'application d'une combinaison de plusieurs méthodes de détection des émotions donne de meilleurs résultats que les méthodes individuelles [205, 206].

3.4. Etat de l'art de l'application de l'analyse du sentiment dans le domaine de l'enseignement supérieur

La rétroaction (ou feedback) des étudiants représente une source d'information indispensable, qui peut être utilisée par les enseignants ou les formateurs afin d'améliorer les procédures d'apprentissage et les activités de formation. La popularité et l'importance

de la rétroaction des étudiants ont augmenté, en particulier à l'époque de la pandémie COVID-19, lorsque la plupart des établissements d'enseignement ont remplacé l'apprentissage traditionnel en face à face par un format en ligne. Cependant, en raison de la nature de la langue utilisée par les étudiants et du grand volume d'informations exprimant leurs points de vue et leurs émotions sur différents aspects dans les forums des MOOCs, traiter les opinions des étudiants est une tâche complexe. Une façon de surmonter ces difficultés est de tirer parti des avantages de l'analyse des sentiments et des techniques de fouille d'opinion (ou opinion mining).

L'analyse des sentiments, impliquant l'identification des mots et des phrases exprimant des émotions, a récemment attiré l'attention de nombreux chercheurs, notamment dans le domaine de l'éducation en général et de l'enseignement supérieur en particulier. Les messages sur Twitter (ou tweets) sont considérés comme une ressource riche pour l'analyse des sentiments. La tâche d'analyse de sentiments qui peut être appliquée aux données Twitter est la classification de polarité. En raison de la nature courte des tweets (au maximum 140 caractères), une approche de classification au niveau de la phrase peut être adoptée en supposant que les tweets expriment des opinions sur une seule entité. En outre, l'extraction de messages de Twitter est une tâche simple grâce à l'API publique de Twitter.

De nombreuses études ont été menées pour analyser les tweets publics concernant l'apprentissage en ligne pour l'enseignement supérieur en utilisant des algorithmes d'apprentissage automatique. Remali et al. [207] ont introduit une approche fondée sur l'apprentissage automatique pour analyser les tweets concernant l'apprentissage en ligne (online learning) dans l'enseignement supérieur. Ils ont utilisé quatre algorithmes d'apprentissage automatique pour classer le sentiment de 38602 tweets. Les résultats qu'ils ont obtenus révèlent que SVM avec un rapport de division de 80:20 et Vader ont obtenu la meilleure précision de test. Baragash et al. [208] ont utilisé SVM pour classer 1201 tweets sur l'apprentissage en ligne postés par des étudiants de l'université de Malaisie. Le modèle qu'ils ont proposé a obtenu les meilleures performances. Dans le domaine de l'enseignement supérieur marocain, aucune étude n'a été menée sur l'analyse du sentiment. L'objectif de cette thèse est de pallier cette lacune en élaborant des approches novatrices pour mieux comprendre l'état émotionnel des étudiants et les opinions sur les universités marocaines. Cela favorisera un apprentissage efficace en permettant aux enseignants de mieux interagir avec leurs étudiants. De même, l'analyse de sentiment à partir de Twitter peut fournir des informations utiles aux responsables des universités pour identifier les points forts et les faiblesses de leurs programmes et services, ainsi que pour cibler les domaines nécessitant des améliorations. Cela permettra aux universités de répondre de

manière proactive aux préoccupations des étudiants, améliorant ainsi leur satisfaction et leur expérience globale.

3.5. Conclusion

Dans ce chapitre, nous avons exploré les différentes méthodes d'évaluation de la qualité des systèmes éducatifs, en mettant l'accent sur l'analyse de sentiment comme une nouvelle approche prometteuse pour évaluer la qualité de l'éducation. Nous avons examiné les fondements théoriques de l'analyse de sentiment, présenté les différentes tâches, niveaux et étapes du processus d'analyse, ainsi que l'état de l'art de son application dans le contexte de l'enseignement supérieur. Dans le chapitre suivant, nous allons appliquer l'analyse de sentiment pour étudier la réputation des universités marocaines à partir de Twitter et pour analyser les opinions du public sur l'enseignement à distance dans cinq disciplines académiques après la pandémie de COVID-19. Ces études de cas pratiques fourniront des exemples concrets de l'application de l'analyse de sentiment dans le domaine de l'éducation, tout en illustrant ses avantages et ses limites.

Chapitre 4

Approches d'analyse de sentiment pour mesurer la pertinence des systèmes éducatifs

4.1. Introduction

Dans ce chapitre, nous allons présenter notre système d'analyse de sentiment en temps réel sur les tweets relatifs à l'apprentissage à distance dans l'enseignement supérieur (section 4.2). Nous allons présenter tout d'abord les travaux connexes. Ensuite, nous décrivons le système de reconnaissance des émotions faciales proposé. Puis nous allons explorer les détails de l'implémentation du système. Enfin, nos résultats expérimentaux, sont présentés. Ensuite, nous allons présenter notre système de prédiction du sentiment à partir de Twitter concernant douze universités publiques marocaines en temps réel en utilisant les algorithmes d'apprentissage automatique (section 4.3). Nous allons présenter tout d'abord les travaux connexes. Ensuite, nous décrivons le système de reconnaissance des émotions faciales proposé. Puis nous allons explorer les détails de l'implémentation du système. Enfin, nos résultats expérimentaux, les comparaisons et la discussion sont présentés. La synthèse du chapitre sera l'objet de la section 4.4.

4.2. Analyse de la satisfaction relative à l'enseignement à distance en temps réel

4.2.1. Introduction

Le coronavirus (COVID-19) a été détecté fin décembre 2019 dans la ville chinoise de Wuhan et a provoqué des maladies respiratoires, des maladies graves et même des décès. L'Organisation mondiale de la santé (OMS) a annoncé qu'il s'agissait d'une pandémie mondiale qui présentait une menace sérieuse pour la santé publique et l'économie dans le

monde entier. La plupart des gouvernements du monde ont appliqué plusieurs mesures de prévention pour lutter contre la propagation du COVID-19, telles que la limitation des voyages internationaux, l'adoption de mesures de confinement et la fermeture temporaire d'écoles et d'universités [36]. En conséquence, les établissements d'enseignement ont été contraints de passer à l'apprentissage à distance.

L'apprentissage à distance, également appelé apprentissage en ligne, e-learning et enseignement à distance, est une forme d'éducation dans laquelle les étudiants reçoivent des supports d'apprentissage en ligne via des appareils mobiles compatibles avec Internet [209]. Ses avantages sont l'accessibilité, la flexibilité [210], l'absence de contrainte de temps et de lieu, l'enregistrement des cours et la réduction des coûts financiers [211]. Cependant, divers problèmes ont été signalés, tels que les problèmes d'accès à l'internet, la difficulté à s'adapter aux technologies d'apprentissage à distance et le manque d'interactions entre l'enseignant et l'étudiant [212]. Par conséquent, le niveau de stress, de peur et d'anxiété a augmenté chez les étudiants [213].

Depuis le début de la pandémie, certains établissements d'enseignement supérieur ont continué à dispenser uniquement des cours à distance, d'autres ont décidé d'utiliser l'apprentissage mixte et d'autres encore ont repris l'apprentissage en personne. Cette transition vers l'enseignement à distance a ouvert de nouvelles voies de recherche dans le domaine de l'enseignement supérieur, où plusieurs études basées sur des enquêtes [214, 215, 216] ont cherché à analyser l'impact de l'enseignement à distance sur les étudiants et les enseignants. Cependant, les études basées sur des enquêtes peuvent ne pas fournir une vue d'ensemble en raison de leurs contraintes en termes de taille d'échantillon, de temps et de public. Pour y remédier, les techniques d'analyse des sentiments peuvent traiter de grandes quantités de données, y compris des textes [217], des images [218] et des discours, afin d'extraire les opinions et les émotions des gens à l'égard de l'apprentissage à distance. En reposant sur des techniques d'intelligence artificielle, l'analyse des sentiments permet une compréhension plus nuancée et plus complète de l'impact de l'apprentissage à distance sur les étudiants et les enseignants. Les plateformes de médias sociaux comme Twitter, Facebook et Instagram sont les principaux canaux par lesquels les gens expriment leurs opinions et leurs sentiments sur divers sujets, y compris l'apprentissage en ligne, les services, les instructeurs, les cours et les méthodes d'enseignement. Des études ont utilisé l'analyse des sentiments pour examiner ces opinions, comme l'étude de So et Oh [219] qui a examiné les perceptions des étudiants sur l'apprentissage en ligne pendant la pandémie COVID-19, et l'étude de Kim et al. [220] qui a analysé les messages Twitter liés aux cours en ligne ouverts et massifs (MOOC). Cette approche permet d'obtenir un échantillon plus

important et plus diversifié que les études traditionnelles basées sur des enquêtes [221]. Les établissements d'enseignement supérieur peuvent utiliser des techniques d'analyse des sentiments pour analyser la grande quantité de données générées sur ces plateformes en temps réel. L'analyse des données des médias sociaux peut fournir des indications sur les domaines à améliorer dans l'expérience d'apprentissage en ligne pour les étudiants et les enseignants, permettant ainsi aux établissements d'enseignement supérieur de prendre des décisions fondées sur des données.

De nombreuses recherches existantes sur l'analyse des sentiments à l'égard de l'enseignement à distance dans l'enseignement supérieur s'appuient sur des algorithmes d'apprentissage automatique. Cependant, ces algorithmes ont des limites en termes de précision et de traitement avancé de grands ensembles de données. Il est donc primordial d'explorer de nouvelles approches pour l'analyse des sentiments qui peuvent fournir une analyse plus avancée et plus précise. Les algorithmes basés sur l'apprentissage profond gagnent en popularité dans divers domaines de recherche en raison de leur capacité à traiter de grands ensembles de données et à fournir des prédictions précises. C'est pourquoi nous proposons pour la première fois une étude qui utilise une méthode basée sur l'apprentissage profond pour analyser les sentiments (positifs, négatifs et neutres) des gens sur Twitter concernant l'apprentissage à distance dans l'enseignement supérieur. L'approche proposée comprend sept phases principales : collecte de données sur Twitter, prétraitement des données, annotation des données, extraction des caractéristiques à l'aide de GloVe [198], division stratifiée des données avec un ratio 90:10, classification des sentiments à l'aide d'un modèle Bi-LSTM basé sur l'auto-attention, et évaluation des performances du modèle et comparaison avec les modèles LSTM [222], Bi-LSTM [223], et CNN-Bi-LSTM. L'objectif de cette étude est de comprendre le sentiment du public exprimé sur Twitter à propos de l'apprentissage à distance dans l'enseignement supérieur à l'échelle mondiale après la pandémie. Et d'identifier les facteurs qui expliquent la variation de ce sentiment public concernant six disciplines académiques : les mathématiques, la physique, la biologie, la chimie, l'informatique et la psychologie. Il est important de souligner que ces disciplines ont été choisies en se basant sur les tweets collectés.

Les principales contributions de notre étude sont les suivantes :

- Introduction d'une nouvelle méthode fondée sur l'apprentissage profond pour la prédiction des tweets par rapport à l'enseignement à distance.
- Collecte de données sur Twitter en anglais concernant l'enseignement à distance dans l'enseignement supérieur, puis prétraitement et étiquetage.

- Prédire le sentiment des tweets à l'aide d'un modèle Bi-LSTM basé sur l'auto-attention avec GloVe word embedding.
- Comparaison des performances du modèle proposé avec trois autres modèles d'apprentissage profond.

Dans ce chapitre, nous allons présenter notre système d'analyse du sentiment en temps réel sur les tweets relatifs à l'apprentissage à distance dans l'enseignement supérieur. Nous allons présenter tout d'abord les travaux connexes (section 4.2.2). Ensuite, nous décrivons le système de reconnaissance des émotions faciales proposé (section 4.2.3). Puis nous allons examiner les détails de l'implémentation du système dans la section 4.2.3. Enfin, nos résultats expérimentaux, les sont présentés dans la section 4.2.4.

4.2.2. Travaux connexes

Plusieurs recherches se sont focalisées sur l'analyse de l'opinion à partir de Twitter concernant l'apprentissage à distance pendant la pandémie de COVID-19 à l'aide d'algorithmes d'apprentissage automatique. Kharde et Sonawane [224] ont utilisé des algorithmes d'apprentissage automatique pour classer les tweets positifs et négatifs sur l'apprentissage à distance. Senadhira et al [225] ont réalisé une étude pour analyser les tweets sur l'apprentissage en ligne pendant la pandémie. Les auteurs ont collecté et prétraité 8976 tweets (4486 positifs et 4490 négatifs) pour les soumettre à un réseau neuronal artificiel et à des algorithmes de machine à vecteur de support [226]. D'après les résultats obtenus, l'ANN a obtenu les meilleures performances. Das et al [227] ont proposé une approche hybride pour prédire le sentiment des tweets sur l'apprentissage en ligne en utilisant la technique d'analyse du lexique pour l'analyse du sentiment et quatre modèles d'apprentissage automatique pour la prédiction du sentiment. Le modèle multinomial de Naïve Bayes a été plus performant que les autres modèles. Sahir et al [228] ont réalisé une étude pour analyser les tweets publics sur l'apprentissage en ligne en Indonésie pendant la pandémie d'octobre 2020 à l'aide de l'algorithme de Naïve Bayes. Leurs résultats révèlent que "stress" et "covid" sont les mots qui reviennent le plus souvent dans les tweets.

Trois études différentes ont été réalisées à l'aide d'algorithmes d'apprentissage automatique pour analyser les tweets arabes concernant l'apprentissage à distance en Arabie saoudite. Tout d'abord, Almalki [229] a proposé un modèle qui utilise l'API de Twitter pour collecter 14000 tweets, un algorithme de régression logistique pour la prédiction du sentiment et l'API de Flask pour obtenir le sentiment des tweets. La régression logistique a

surpassé les autres modèles. Il a également été utilisé dans la deuxième étude d'Aljabri et al [230] avec TF-IDF et l'unigramme. Le modèle proposé a donné les meilleurs résultats sur un total de 20827 tweets. Dans la troisième étude, Althagafi et al [231] ont analysé 8176 tweets arabes sur l'apprentissage à distance en Arabie Saoudite en utilisant trois algorithmes d'apprentissage automatique. La forêt aléatoire avec classification multi-classes a donné la meilleure précision de test.

D'autre part, les algorithmes d'apprentissage profond se sont révélés très efficaces dans la classification des sentiments des tweets sur l'apprentissage à distance. Mujahid et al [232] ont analysé 17155 tweets sur l'apprentissage en ligne en utilisant des approches basées sur le lexique pour l'étiquetage des tweets, le sac de mots et TF-IDF pour l'extraction des caractéristiques, la technique de suréchantillonnage minoritaire synthétique (SMOTE) pour l'équilibrage des données, et neuf modèles d'apprentissage automatique pour la classification des sentiments. La combinaison de la fréquence des termes et de la fréquence inverse des documents avec le modèle de machine à vecteur de support donne les meilleurs résultats. En outre, les auteurs ont utilisé des modèles d'apprentissage profond tels que LSTM, CNN, CNN-LSTM et Bi-LSTM pour la classification des sentiments. D'après leurs résultats, les modèles d'apprentissage profond ont obtenu les meilleures performances par rapport aux modèles d'apprentissage automatique. Waheeb et al [37] ont utilisé l'autoencodeur de machine d'apprentissage extrême (ELM-AE) avec LSTM pour analyser les tweets concernant l'apprentissage en ligne et pour détecter les fausses nouvelles COVID-19. Leur premier ensemble de données contenait 60000 tweets étiquetés comme (très positif, positif, neutre, négatif et très négatif), et leur deuxième ensemble de données contenait 6000 articles provenant de plusieurs sites web sur les fausses et vraies nouvelles COVID-19. Leurs résultats montrent que l'ELM-AE avec SMOTE a donné les meilleures performances.

En particulier, plusieurs études ont été menées pour analyser les tweets publics concernant l'apprentissage en ligne dans l'enseignement supérieur à l'aide d'algorithmes d'apprentissage automatique. Remali et al [207] ont introduit une méthode basée sur l'apprentissage automatique pour analyser les tweets sur l'apprentissage en ligne dans l'enseignement supérieur. Ils ont utilisé quatre algorithmes d'apprentissage automatique différents pour classer le sentiment de 38602 tweets postés entre le 23 juillet et le 14 août 2020. Les résultats obtenus révèlent que la machine à vecteur de support avec un ratio de division de 80:20 et Vader ont obtenu la meilleure précision de test. Baragash et al [208] ont utilisé un modèle de machine à vecteur de support pour classer 1201 tweets sur l'apprentissage en ligne postés par des étudiants d'universités malaisiennes. Le modèle

qu'ils ont proposé a obtenu les meilleures performances. Alors que les algorithmes d'apprentissage automatique ont été largement employés dans diverses études pour analyser l'opinion publique sur l'apprentissage à distance pendant la pandémie de COVID-19 à partir de Twitter, certaines études ont démontré de meilleures performances en utilisant des modèles d'apprentissage profond par rapport aux modèles d'apprentissage automatique traditionnels. Ces études ont été menées dans différents pays, tels que l'Indonésie et l'Arabie saoudite. Bien que plusieurs études se soient concentrées sur l'analyse des tweets concernant l'apprentissage en ligne dans l'enseignement supérieur avec l'apprentissage automatique, aucune étude antérieure n'a utilisé d'algorithmes d'apprentissage profond pour analyser le sentiment des tweets concernant l'apprentissage à distance dans l'enseignement supérieur. Cette étude vise donc à combler cette lacune en examinant le sentiment des tweets sur l'apprentissage à distance dans l'enseignement supérieur après la pandémie et en évaluant les performances des algorithmes d'apprentissage profond dans l'analyse du sentiment. Plus précisément, elle vise à identifier les facteurs qui contribuent aux variations de ce sentiment dans six disciplines universitaires : les mathématiques, la physique, la biologie, la chimie, l'informatique et la psychologie. La période post-pandémique a été choisie pour étudier la façon dont l'apprentissage à distance est perçu et utilisé dans un environnement où les restrictions sur l'enseignement en personne se sont assouplies et où l'apprentissage à distance est devenu plus courant. Cela permettrait d'obtenir des informations précieuses sur les perceptions actuelles et l'utilisation de l'enseignement à distance dans un contexte post-pandémique.

4.2.3. Système d'analyse du sentiment en temps réel sur les tweets relatifs à l'apprentissage à distance dans l'enseignement supérieur

L'architecture proposée pour l'analyse des sentiments des tweets relatifs à l'enseignement à distance comprend sept étapes, comme le montre la figure 4.1. La première étape est responsable de la collecte des données sur Twitter. La deuxième étape est axée sur le prétraitement des données, suivi de l'annotation des données à l'aide de Text Blob au cours de la troisième étape. L'extraction des caractéristiques à l'aide de GloVe est réalisée au cours de la quatrième étape. La cinquième étape est responsable de la division des données en ensembles d'entraînement et de test afin d'être traitée par le modèle Bi-LSTM basé sur l'auto-attention dans l'étape suivante. Enfin, la dernière étape est l'évaluation du modèle à l'aide de mesures de performance. Une brève description de chacune de ces étapes est donnée dans les sous-sections suivantes.

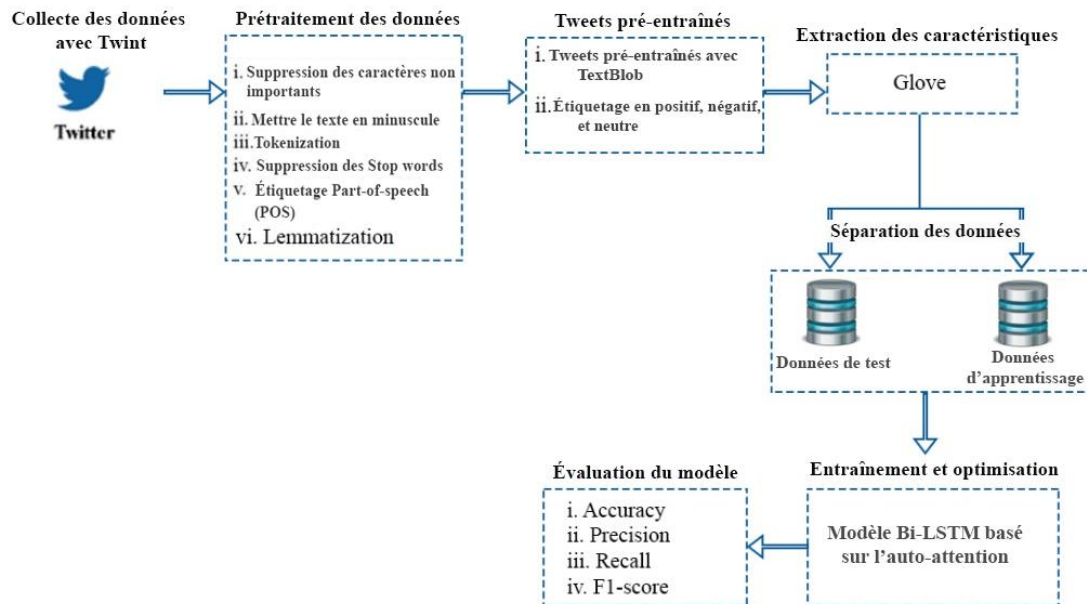


Figure 4.1 – Architecture globale de l'analyse des sentiments.

- *Collecte de données* : pour collecter des tweets sur l'apprentissage à distance, nous avons utilisé Twint, une bibliothèque Python opensource pour le scraping de Twitter. Contrairement à l'interface de programmation d'application (API) de recherche standard qui limite l'accès aux tweets datant de plus d'une semaine. Twint peut extraire presque tous les tweets sans utiliser l'API et l'authentification de Twitter. Un total de 24642 tweets historiques en anglais sur l'apprentissage à distance écrits entre le 20 juillet 2022 et le 6 novembre 2022 ont été collectés sur la base de mots-clés tels que "remote education", "distance learning", "virtual learning", "online learning", "e-learning", "distance education" et "online education". Cette période a été choisie parce qu'elle est postérieure à la pandémie de COVID-19, qui a eu un impact significatif sur le secteur de l'éducation. En choisissant cette période, nous nous attendions à ce que les restrictions liées à la pandémie sur l'enseignement en personne se soient assouplies et que l'enseignement à distance se soit normalisé. Par conséquent, les tweets recueillis au cours de cette période peuvent fournir des informations précieuses sur la façon dont l'apprentissage à distance est perçu et utilisé dans le contexte post-pandémique.
- *Prétraitement des données* : un volume important de données non structurées est généré quotidiennement sur les sites de médias sociaux sous forme de textes, d'images, de vidéos et de sons. Ces données sont souvent bruyantes et non

structurées. Il est donc nécessaire d'éliminer le bruit des données des médias sociaux à l'aide des techniques de prétraitement du texte suivantes afin d'appliquer correctement les algorithmes d'apprentissage profond.

- *Suppression des émoticônes, des URL, de la ponctuation, des chiffres, des noms d'utilisateur, des caractères spéciaux et des hashtags* : dans cette étape, nous avons éliminé la ponctuation, les noms d'utilisateur, les chiffres, les caractères spéciaux et les hashtags du tweet car ils ne jouent aucun rôle dans notre analyse.
- *Minuscules* : la lettre majuscule est remplacée par la lettre minuscule correspondante.
- *Tokenisation* : les phrases sont divisées en jetons avec des caractères d'espace blancs comme délimiteurs.
- *Suppression des mots vides* : nous avons supprimé les mots vides du tweet, qui sont des mots sans valeur sémantique dans le langage naturel, tels que (par, avec, est, le, etc.).
- *L'étiquetage Part-of-speech (POS)* : donne les informations contextuelles d'un mot (verbes, noms, adverbes, conjonctions, etc.) qui sont essentielles dans le processus de lemmatisation.
- *Lemmatisation* : transforme un mot en sa forme racine à l'aide de la lemmatisation NLTK.

Le tableau 4.1 illustre un exemple de mise en œuvre des étapes de prétraitement sur le tweet suivant : "We know that online learning is not only the solution for pandemic times, but also for the uncertainty like this. It is time to bring it back to the table. #onlinelearning #remoteteaching #lms #learningplatform".

- *Pré-entraînement des données avec Text Blob* : après l'étape de prétraitement, chaque tweet est étiqueté comme "positif", "négatif" et "neutre" à l'aide du Text Blob afin d'être transmis au modèle d'apprentissage profond proposé. Le Text Blob est un outil NLP Python à source ouverte qui représente le sentiment émotionnel d'un tweet sur la base du score de polarité, qui se situe entre -1 et 1. Dans notre travail, le sentiment est "positif" si le score de polarité est supérieur à 0.2. Le sentiment est également "neutre" si le score de polarité est compris entre 0 et 0,2. Enfin, le sentiment est "négatif" si le score de polarité est inférieur à 0. L'estimation du score de polarité est illustrée par l'équation (4.1).

$$S_{t_i} = \left\{ \begin{array}{ll} \text{Positive} & P > 0.2 \\ \text{Neutre} & 0 \leq P \leq 0.2 \\ \text{Negative} & P < 0 \end{array} \right\} \quad (4.1)$$

Le résultat de cette étape est le suivant : 6980 tweets positifs, 14048 tweets neutres et 3614 tweets négatifs. La figure 4.2 montre le nombre de tweets par sentiment entre le 20 juillet 2022 et le 6 novembre 2022.

Tableau 4.1 – Exemple de mise en œuvre des étapes de prétraitement sur un tweet.

Étape de prétraitement	Tweet
Après avoir supprimé les émoticônes, les URL, la ponctuation, les chiffres, les noms d'utilisateur, les caractères spéciaux et les hashtags.	We know that online learning is not only the solution for pandemic times, but also for the uncertainty like this. It is time to bring it back to the table
Après le bas de casse	We know that online learning is not only the solution for pandemic times but also for the uncertainty like this it is time to bring it back to the table
Après la tokenisation	['we', 'know', 'that', 'online', 'learning', 'is', 'not', 'only', 'the', 'solution', 'for', 'pandemic', 'times', 'but', 'also', 'for', 'the', 'uncertainty', 'like', 'this', 'it', 'is', 'time', 'to', 'bring', 'it', 'back', 'to', 'the', 'table']
Après avoir supprimé les mots vides	['know', 'online', 'learning', 'solution', 'pandemic', 'times', 'also', 'uncertainty', 'like', 'time', 'bring', 'back', 'table']
Après l'étiquetage de la partie du discours (POS)	[('know', 'VERB'), ('online', 'NOUN'), ('learning', 'VERB'), ('solution', 'NOUN'), ('pandemic', 'ADJ'), ('times', 'NOUN'), ('also', 'ADV'), ('uncertainty', 'NOUN'), ('like', 'ADP'), ('time', 'NOUN'), ('bring', 'VERB'), ('back', 'ADV'), ('table', 'ADJ')]
Après lemmatisation	['know', 'online', 'learning', 'solution', 'pandemic', 'time', 'also', 'uncertainty', 'like', 'time', 'bring', 'back', 'table']

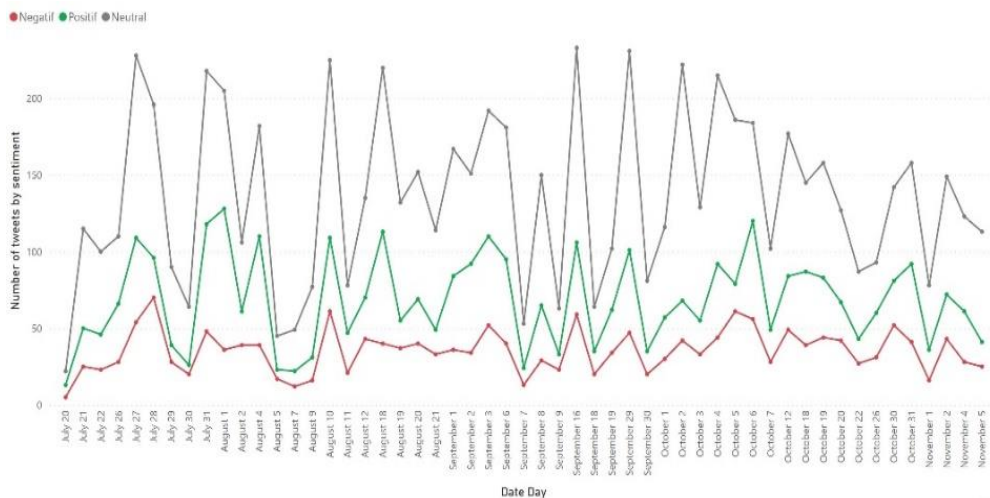


Figure 4.2 – Chronologie des tweets sur l'apprentissage à distance.

- *Extraction des caractéristiques* : il existe de nombreuses techniques d'extraction de caractéristiques telles que Word Embedding, GloVe, et TF-IDF. Dans cette étude, la technique d'intégration de mots GloVe (Global Vectors) est utilisée dans l'approche proposée pour obtenir des caractéristiques sémantiques à partir du tweet. GloVe est un algorithme d'apprentissage non supervisé pour la représentation distribuée des mots du texte. Il permet de capturer les relations sémantiques entre les mots dans l'espace vectoriel. Le but de GloVe est de minimiser la fonction objective indiquée dans l'équation (4.2).

$$J = \sum_{i,j=1}^{|V|} f(X_{ij})(v_i^T v_j' + b_i + b_j' - \log(X_{ij}))^2 \quad (4.2)$$

Où v_i est la représentation du mot w_i , b_i est le biais de v_i , et $f(x)$ est la fonction de pondération décrite dans l'équation (4.3).

$$f(x) = \begin{cases} (x/x_{max})^\alpha, & \text{if } x < x_{max}, \\ 1, & \text{otherwise} \end{cases} \quad (4.3)$$

Dans ce travail, nous avons utilisé les vecteurs de mots Glove [233] pour créer une matrice de cooccurrence X afin d'entraîner le modèle GloVe.

- *Fractionnement des données* : dans notre étude, nous avons appliqué un fractionnement stratifié des données avec un ratio 90:10 où 90% des données sont utilisées pour la formation du modèle, tandis que 10% sont utilisées pour le test. Nous avons ensuite mélangé les données pour que l'ensemble de formation reflète davantage la distribution globale des données. Le nombre de tweets dans l'ensemble de formation est de 22177 et dans l'ensemble de test est de 2465.
- *Classification des sentiments à l'aide du modèle Bi-LSTM basé sur l'auto-attention* : l'architecture générale du modèle proposé comprend la couche d'entrée, la couche Bi-LSTM, le mécanisme d'auto-attention [234], la couche d'abandon, la normalisation des lots et la couche de sortie. La description des composants du modèle est la suivante. La couche d'entrée prend en charge les données sous la forme d'une séquence de jetons. Chaque jeton est converti en un vecteur de mots. Les jetons sont complétés à l'aide de la stratégie de remplissage zéro ou tronqués en fonction de la longueur maximale du modèle. Ils sont ensuite transmis à la couche d'intégration de Glove afin de produire le vecteur d'intégration de mots. Dans notre travail, nous avons utilisé le réseau Bi-LSTM

qui se compose de LSTM avant et arrière pour obtenir des caractéristiques et apprendre leurs contextes. Les LSTM avant et arrière sont définies dans l'équation (4.4) et l'équation (4.5), où m est la longueur maximale des caractéristiques.

$$\vec{h}_{f_{lstm}} = \overrightarrow{LSTM}(p_i), \quad i \in [1, m] \quad (4.4)$$

$$\overleftarrow{h}_{b_{lstm}} = \overleftarrow{LSTM}(p_i), \quad i \in [m, 1] \quad (4.5)$$

h_t , défini dans l'équation (4.6), est le résultat de la concaténation de l'équation (4.4) et de l'équation (4.5).

$$h_t = LSTM[\vec{h}_{f_{lstm}}, \overleftarrow{h}_{b_{lstm}}] \quad (4.6)$$

Étant donné que tous les mots d'un tweet n'ont pas la même signification, nous avons appliqué le mécanisme d'auto-attention aux caractéristiques générées par le Bi-LSTM en donnant aux mots les plus importants un poids plus élevé pour obtenir leur influence sur l'émotion du tweet. L'annotation du mot h_t produite par le Bi-LSTM est introduite dans un MLP perceptron à une couche pour obtenir u_t comme représentation cachée de h_t pour le tweet t^{th} , comme défini dans l'équation (6.7), où W est une matrice de poids du MLP, et b est un vecteur de biais du MLP. Ensuite, nous mesurons l'importance des mots par la similarité entre u_t et le vecteur de contexte v_s , qui est initialisé de manière aléatoire. En outre, nous avons obtenu un poids d'importance normalisé A_{t_i} par le biais d'une fonction Softmax. A_{t_i} le poids normalisé, tel que représenté dans l'équation, est le poids normalisé du mot i^{th} dans le tweet t^{th} , et $\exp(.)$ est la fonction exponentielle. Plus A_{t_i} est grand, plus le mot i^{th} est important pour la représentation émotionnelle. Enfin, le vecteur de phrase s_t défini dans l'équation (6.8) est la somme pondérée des annotations de mots.

$$A_{t_i} = \frac{\exp(u_t * v_s)}{\sum_{i=1}^m \exp(u_t * v_s)} \quad (4.7)$$

$$s_t = \sum_i \alpha_{t_i} h_{t_i} \quad (4.8)$$

Le mécanisme d'auto-attention est suivi d'une couche d'abandon et d'une couche de normalisation par lots pour éviter l'ajustement excessif, puis d'une couche de sortie à trois neurones, de la fonction d'activation Softmax et de l'entropie croisée catégorielle éparse comme fonction de perte pour effectuer la classification des sentiments. La figure 4.3 présente le résumé du modèle Bi-LSTM basé sur l'auto-attention que nous proposons.

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 20, 300)	2842500
bidirectional_1 (Bidirectional)	(None, 20, 32)	40576
seq_self_attention_1 (SeqSelfAttention)	(None, 20, 32)	2113
dropout_1 (Dropout)	(None, 20, 32)	0
batch_normalization_1 (Batch Normalization)	(None, 20, 32)	128
flatten_1 (Flatten)	(None, 640)	0
dense_1 (Dense)	(None, 3)	1923
=====		
Total params: 2,887,240		
Trainable params: 44,676		
Non-trainable params: 2,842,564		

Figure 4.3 – Résumé du modèle Bi-LSTM basé sur l'auto-attention.

- *Évaluation du modèle* : nous avons utilisé quatre mesures de performance pour l'évaluation des modèles proposés : l'accuracy, la précision, le rappel et le score F1, définis dans les équations (2.1), (2.2), (2.3) et (2.4).

4.2.4. Détails de l'implémentation du système proposé

Nous avons réalisé l'expérience sur Windows 10 avec 16 Go de RAM, Intel Core i7, 64 bits, 3.20 GHz dans le CPU, et Nvidia GTX 1070. Le modèle proposé est construit en utilisant Keras et un backend TensorFlow. Pour la formation, l'ensemble de données utilisé dans cette étude est divisé en 90% pour la formation et 10% pour le test. Nous avons utilisé GloVe avec 300 dimensions dans la couche d'entrée pour produire le vecteur d'intégration des mots. La taille de la couche cachée du réseau neuronal Bi-LSTM est de 32 avec la fonction d'activation ReLU. La sortie du Bi-LSTM est transmise au mécanisme d'auto-attention pour obtenir le poids d'attention. Ensuite, pour éviter un surajustement, un dropout de 0.2 est appliqué, suivi d'une normalisation par lots. La couche de sortie comporte trois neurones avec l'entropie croisée catégorielle éparse et la fonction d'activation Softmax. La

taille du lot est de 16, l'optimiseur de modèle est Adam avec un taux d'apprentissage de 0.01 et le nombre d'époques est de 100. Nous avons utilisé la méthode RandomizedSearchCV de la bibliothèque Sklearn [235] pour obtenir les meilleures valeurs d'hyperparamètres.

4.2.5. Résultats expérimentaux

Nous avons entraîné le modèle Bi-LSTM basé sur l'auto-attention sur 22177 tweets et l'avons testé sur 2465 tweets. Nous avons ainsi obtenu une précision de test globale de 95%. Les performances du modèle proposé sont illustrées par la matrice de confusion validée avec un schéma de validation stratifié 90:10%, qui est employé pour identifier la relation entre les classes vraies et prédites, comme illustré à la figure 4.4.

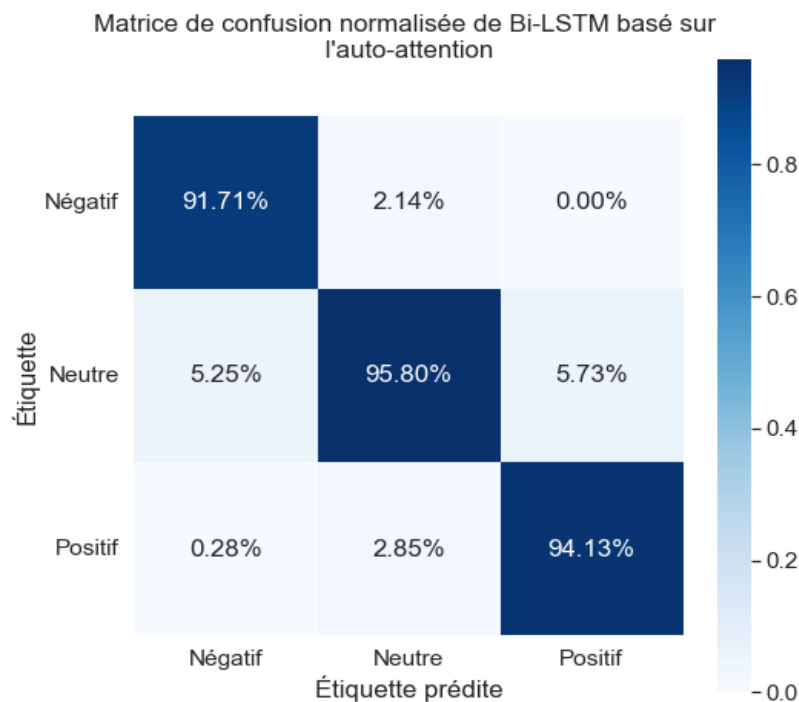


Figure 4.4 – Matrice de confusion normalisée du modèle Bi-LSTM basé sur l'auto-attention.

L'accuracy, la précision, le score F1 et le rappel sont utilisés dans l'étude pour comparer les performances de notre modèle à celles des modèles LSTM, CNN-Bi-LSTM et Bi-LSTM, comme le montrent la figure 4.5 et le tableau 4.2.

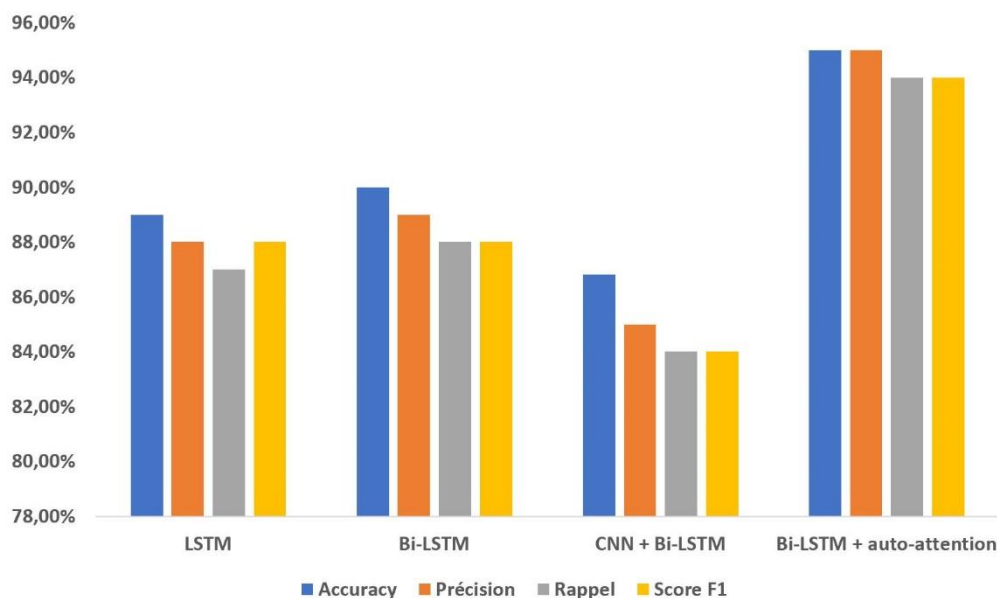


Figure 4.5 – Mesures de performance du modèle proposé avec différents modèles d'apprentissage profond.

Tableau 4.2 – Comparaison des performances du modèle proposé avec différents modèles d'apprentissage profond.

Modèle	Accuracy	Précision	Rappel	Score F1
LSTM	89.00%	88.00%	87.00%	88.00%
Bi-LSTM	90.00%	89.00%	88.00%	88.00%
CNN + Bi-LSTM	86.81%	85.00%	84.00%	84.00%
Bi-LSTM + auto-attention	95.00%	95.00%	94.00%	94.00%

Les architectures de ces trois modèles d'apprentissage profond sont présentées dans les figures 4.6, 4.7 et 4.8. En comparaison, notre modèle d'auto-attention-Bi-LSTM proposé a surpassé les autres modèles avec une exactitude de test globale de 95%. Le modèle Bi-LSTM a quant à lui obtenu une précision de 90%, suivi du modèle LSTM avec une précision de 89%. Toutefois, le modèle CNN-Bi-LSTM a obtenu la plus faible précision de test, soit 86.81%. Cela nous amène à conclure que le modèle que nous proposons atteint la meilleure performance parmi tous les autres modèles.

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 20, 300)	2842500
lstm (LSTM)	(None, 20, 32)	42624
dropout (Dropout)	(None, 20, 32)	0
batch_normalization (Batch Normalization)	(None, 20, 32)	128
flatten (Flatten)	(None, 640)	0
dense (Dense)	(None, 3)	1923

=====
Total params: 2,887,175
Trainable params: 44,611
Non-trainable params: 2,842,564

Figure 4.6 – Résumé du modèle LSTM.

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 20, 300)	2842500
bidirectional (Bidirectional)	(None, 20, 32)	40576
dropout_1 (Dropout)	(None, 20, 32)	0
batch_normalization_1 (Batch Normalization)	(None, 20, 32)	128
flatten_1 (Flatten)	(None, 640)	0
dense_1 (Dense)	(None, 3)	1923

=====
Total params: 2,885,127
Trainable params: 42,563
Non-trainable params: 2,842,564

Figure 4.7 – Résumé du modèle Bi-LSTM.

La segmentation des tweets relatifs à l'apprentissage à distance en fonction de leur sentiment prédit dans les données de test est représentée dans la figure 4.10. On constate que 57% des tweets sont neutres, 28.3% sont positifs et 14.7% sont négatifs. En outre, nous avons classé les tweets de l'ensemble de données de test en fonction du niveau d'enseignement correspondant (enseignement supérieur, collège et lycée, école primaire et maternelle). Si le niveau d'enseignement n'est pas mentionné dans le tweet, celui-ci est classé comme non spécifié. Le tableau 4.3 représente le nombre de tweets concernant chaque niveau d'enseignement et le nombre de sentiments prédits (positifs, négatifs et neutres) pour les données de test.

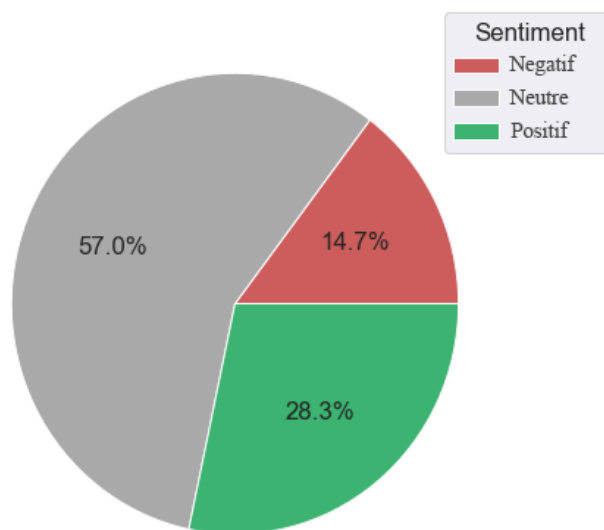


Figure 4.10 – Distribution des sentiments prédits pour les tweets relatifs à l'apprentissage à distance.

Tableau 4.3– Nombre de tweets dans chaque étape éducative par sentiment pour l'ensemble de données de test.

Sentiment	Enseignement supérieur	École intermédiaire et secondaire	École primaire et maternelle	Non spécifié
Positif	108	10	13	569
Neutre	192	13	21	1180
Négaif	26	7	6	320
Nombre total de tweets	326	30	40	2069

La figure 4.11 montre un diagramme à barres horizontal qui représente la comparaison entre les divers niveaux d'enseignement en fonction du nombre de tweets et des sentiments prédits. Il apparaît clairement que c'est au niveau de l'enseignement supérieur que le nombre de tweets et le pourcentage de tweets positifs sur l'apprentissage à distance sont les

plus élevés, suivi de l'école primaire et de l'école maternelle. En revanche, les écoles intermédiaires et secondaires ont le plus faible pourcentage de tweets positifs. En ce qui concerne les tweets neutres, le pourcentage était plus élevé à tous les niveaux d'enseignement.

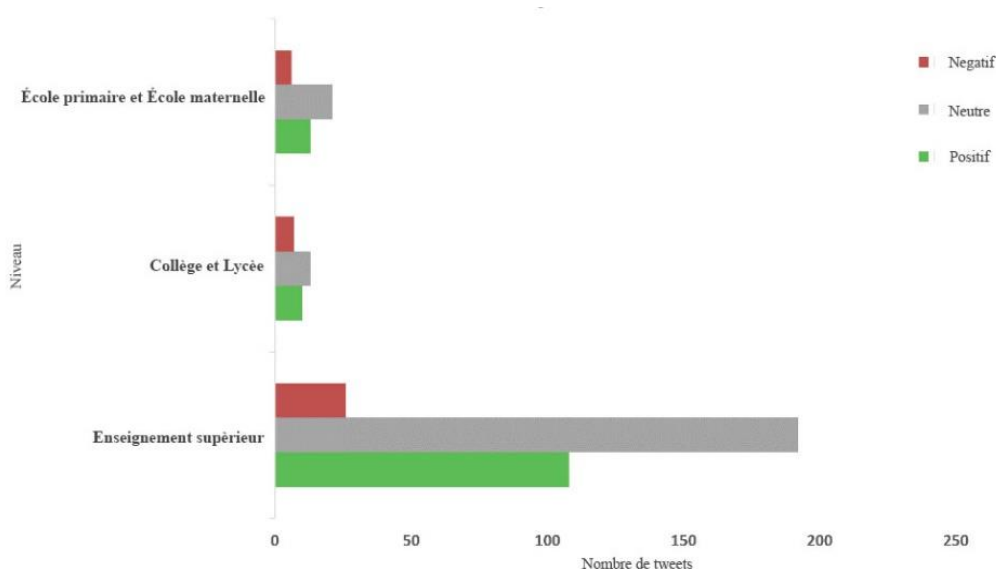


Figure 4.11– Sentiments prédits à l'égard de l'apprentissage à distance à différents niveaux de l'éducation.

Le tableau 4.4 présente un exemple de tweets relatifs à l'enseignement supérieur avec leur sentiment prédit.

Tableau 4.4 – Exemple de tweets relatifs à l'enseignement supérieur avec leur sentiment prédit.

Tweet	Sentiment prédit
taught many people turn learning flexibility learning virtually anywhere team pursue without commute physical campus	Positif
remote student sometimes feel lonely however experience full community building support faculty teaching staff	Négatif
faculty last chance apply virtual exchange training spring virtual exchange provide global collaboration apply	Neutre

Nous avons établi une comparaison entre les disciplines de l'enseignement supérieur. La figure 4.12 présente un diagramme à barres montrant le pourcentage de tweets positifs, négatifs et neutres sur l'apprentissage à distance dans six disciplines universitaires : les mathématiques, la physique, la biologie, la chimie, l'informatique et la psychologie. Nous pouvons constater que la psychologie présente le pourcentage le plus élevé de tweets positifs (41.17%), suivie par l'informatique (34.2%). Les mathématiques ont quant à elles enregistré 23.5 % de tweets positifs. En revanche, la physique, la chimie et la biologie ont le plus faible pourcentage de tweets positifs, soit respectivement 20.8%, 20% et 19%. En revanche, le pourcentage de tweets négatifs est très élevé en biologie, en chimie et en physique, avec respectivement 28%, 26% et 24.6%. En outre, les six disciplines universitaires présentent un pourcentage élevé de tweets neutres, qui varie entre 53% et 67%.

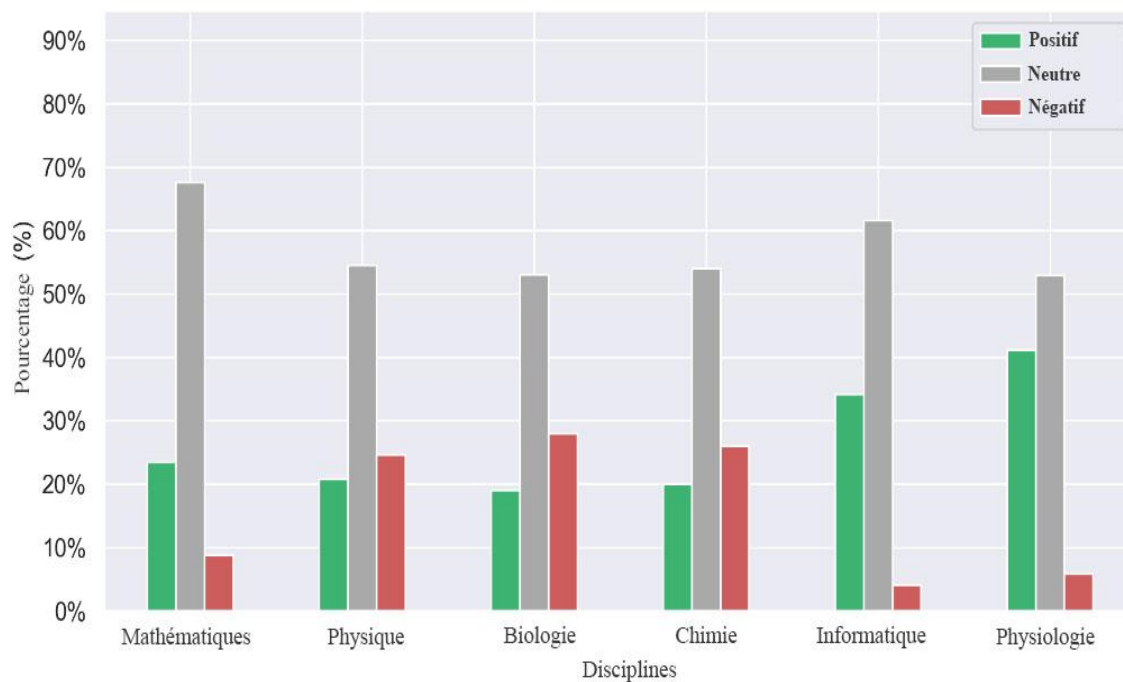


Figure 4.12 – Distribution des sentiments prédits dans les tweets sur l'enseignement à distance par six disciplines académiques.

Comme le montre la figure 4.13, les bigrammes les plus fréquents liés à l'apprentissage à distance dans l'enseignement supérieur sont *understanding speak*, *web development*, *effective communication*, *easy access*, and *remote assistance*. Nous pouvons voir que les gens parlent positivement du développement web en tant qu'application de l'informatique et de la communication efficace en tant que compétence essentielle qu'un étudiant en psychologie doit acquérir au cours de ses études de premier cycle. Cela explique pourquoi le pourcentage de tweets dans les disciplines de la psychologie et de l'informatique est très élevé.

En revanche, les 20 bigrammes les plus fréquents dans les tweets négatifs sur l'enseignement à distance dans l'enseignement supérieur, comme le montre la figure 4.14, sont les suivants : method study, money funds, anatomy accounting, physics essay, following biology, everyone go, et due biology. Il apparaît donc que les principales raisons du sentiment négatif exprimé dans ces tweets sont liées aux fonds, au manque d'engagement et à la nécessité de disposer de laboratoires en personne pour les cours de biologie, de chimie, d'anatomie et de physique, afin que les étudiants puissent voir les concepts en action. Ces raisons expliquent pourquoi le pourcentage de tweets négatifs est très élevé en biologie, en chimie et en physique.

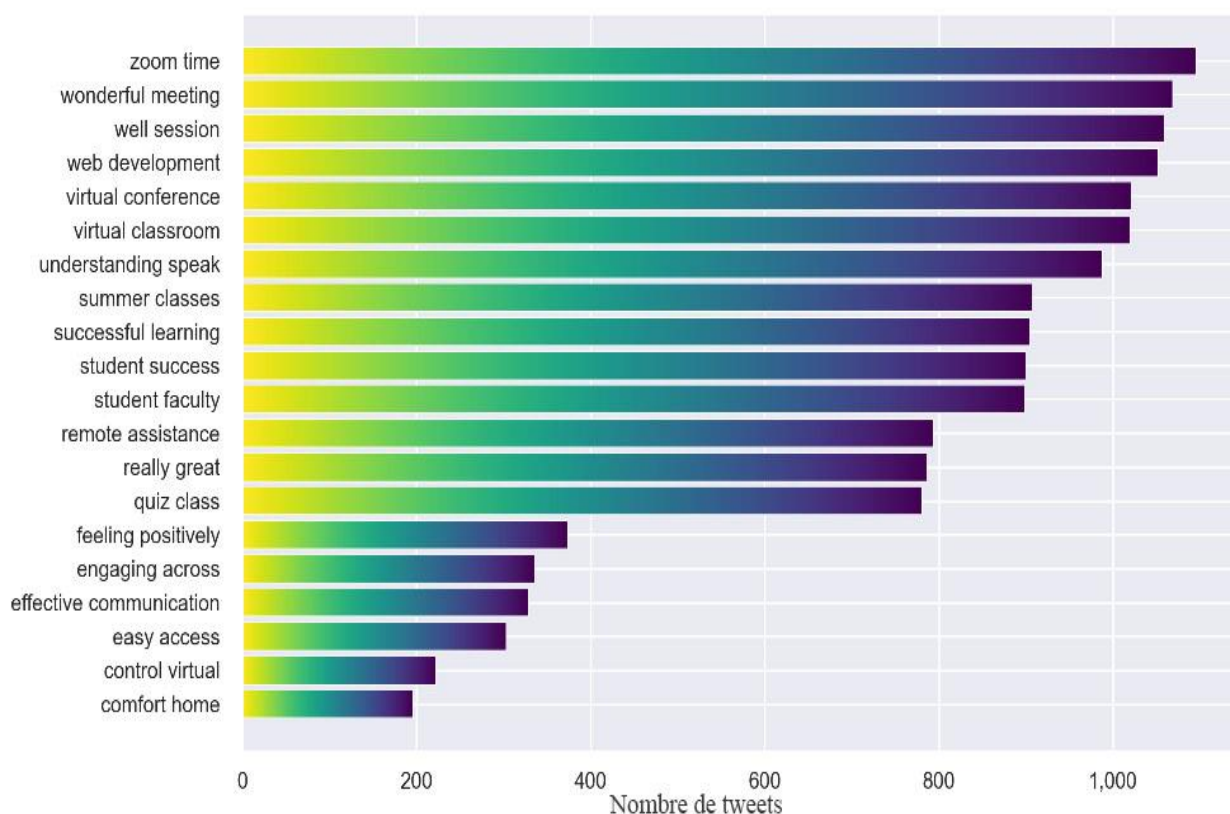


Figure 4.13 – Les 20 bigrammes les plus fréquents des tweets positifs sur l'apprentissage à distance dans l'enseignement supérieur.

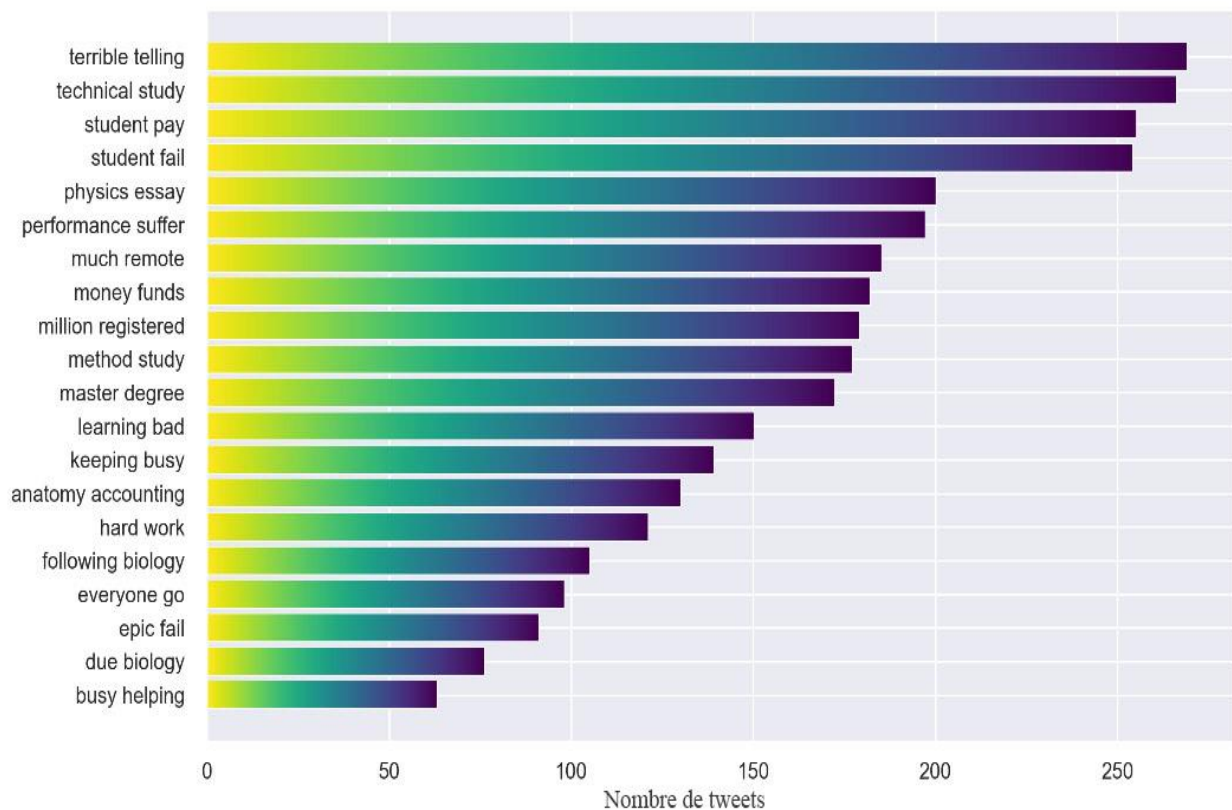


Figure 4.14 – Les 20 bigrammes les plus fréquents des tweets négatifs sur l'apprentissage à distance dans l'enseignement supérieur.

4.3. Analyse de la réputation des universités marocaines en temps réel

4.3.1. Introduction

De nos jours, la croissance des sites de médias sociaux offre aux gens la possibilité de communiquer leurs sentiments, leurs opinions et leurs idées sur un sujet donné. Twitter est considéré comme l'un des sites web les plus utilisés, avec 435 millions d'utilisateurs actifs mensuels [236]. Les données extraites de ce média ont incité les chercheurs à explorer les opinions du public. L'analyse des sentiments sur Twitter est donc devenue un sujet de recherche intéressant. Il identifie si la polarité du sentiment à partir de données textuelles est positive, neutre ou négative en analysant l'opinion. Ce dernier pour aider à améliorer le processus de prise de décision.

Plusieurs études ont montré que l'analyse des sentiments sur Twitter est plus efficace lorsqu'il est utilisé dans certains domaines, notamment la santé [237, 238], le secteur bancaire [239], le marketing [240, 241], le tourisme [242] et la politique [243]. En outre,

L'analyse des sentiments est utilisée dans le secteur de l'éducation pour améliorer la qualité des services des établissements d'enseignement [244, 245], améliorer le processus d'apprentissage en analysant les sentiments, l'orientation des étudiants [246], et l'extraction des informations utiles sur la méthodologie d'enseignement d'un professeur. En particulier, des études récentes ont mis l'accent sur l'analyse des sentiments pour détecter les points forts et les points faibles des cours dans l'enseignement supérieur en analysant les opinions en ligne des étudiants [247] et en mesurant des indicateurs universitaires spécifiques tels que la réputation de l'université à partir des médias sociaux pour construire un mécanisme de classement [248]. L'analyse des tweets sur les universités peut être une source complémentaire essentielle pour les décideurs politiques dans les environnements d'enseignement supérieur pour l'allocation des budgets et l'évaluation des universités.

L'analyse des sentiments peut être réalisée à l'aide de l'apprentissage automatique et de l'approche basée sur les lexiques. Les approches basées sur le lexique reposent sur l'hypothèse que l'orientation sémantique du texte est liée à la polarité des mots et des phrases qui y figurent, tandis que les approches basées sur l'apprentissage automatique reposent sur des modèles construits à partir d'exemples de phrases étiquetées. Cependant, l'analyse des sentiments en temps réel nécessite de puissantes technologies de big data, telles que Apache Spark [249] et Apache Kafka [250] pour le traitement des données. Ensuite, Elasticsearch [251] et Kibana pour le stockage et la visualisation en temps réel de la grande quantité de données.

Jusqu'à présent, aucune étude antérieure n'a été menée pour analyser les tweets liés aux universités marocaines. Par conséquent, cela nous pousse à introduire un nouveau système pour l'analyse des sentiments sur Twitter concernant les universités publiques marocaines en temps réel avec l'apprentissage automatique et le big data. Les résultats prévus permettront d'améliorer les systèmes de classement des universités au Maroc en incluant les sentiments des gens (positifs, négatifs et neutres) concernant ces institutions académiques.

Les principales contributions de cette étude sont les suivantes :

- Introduction d'un système en temps réel pour la prédiction du sentiment des universités publiques marocaines à partir de Twitter en utilisant les technologies du big data.
- Collecte de données de tweets en français sur douze universités publiques marocaines, puis les prétraiter et les étiqueter.
- Mener des études comparatives sur les performances de différents algorithmes

d'apprentissage automatique avec la méthode Term Frequency Inverse Document Frequency (TF-IDF).

- Sélectionner et appliquer le meilleur modèle avec la plus grande précision de test sur des nouveaux tweets non étiquetés collectés en temps réel pour prédire les opinions des gens concernant les universités publiques marocaines.

Dans ce chapitre, nous allons exposer notre système de prédiction du sentiment à partir de Twitter concernant douze universités publiques marocaines en temps réel en utilisant les algorithmes d'apprentissage automatique. Nous allons présenter tout d'abord les travaux connexes (section 4.3.2). Ensuite, nous décrivons le système de reconnaissance des émotions faciales proposé (section 4.3.3). Puis nous allons exposer en détail la mise en œuvre du système dans la section 4.3.4. Enfin, nos résultats expérimentaux, les comparaisons et la discussion sont présentés dans la section 4.3.5. La synthèse du chapitre sera l'objet de la section 4.4.

4.3.2. Travaux connexes

Le vaste développement de l'utilisation des médias sociaux a présenté une opportunité pour chaque organisation, entreprise et institution d'obtenir les avis, commentaires et critiques des utilisateurs sur des questions spécifiques. En janvier 2022, Twitter comptait 435 millions d'utilisateurs actifs mensuels [236]. Les informations de Twitter sont donc précieuses pour l'analyse des sentiments. En conséquence, de nombreuses études antérieures se sont concentrées sur l'analyse des opinions à partir des données de Twitter dans divers domaines, y compris les films, le tourisme et, plus encore, l'éducation.

Parallèlement au grand succès de l'apprentissage automatique dans plusieurs domaines d'application, les algorithmes d'apprentissage automatique sont également appliqués à l'analyse des sentiments. Neethu et Rajasree [252] ont analysé les messages Twitter sur les produits électroniques à l'aide des algorithmes d'apprentissage automatique. Goel et al. [253] ont classé les critiques de films sur Twitter en temps réel en utilisant l'algorithme Naïve bayes (sous-section 1.2.4). Les auteurs ont également expliqué que la précision de Naïve bayes pouvait être améliorée en l'utilisant avec SentiWordNet. Chikersal et al. [254] ont analysé des tweets en utilisant SVM. L'ensemble de données utilisé pour l'apprentissage du modèle SVM comprenait 9418 tweets. Coletta et al. [255] ont utilisé une méthode différente qui combine le SVM avec un ensemble de clusters pour analyser les données Twitter. Elle a montré une meilleure précision de test qu'un SVM autonome. Ensuite, Huq et al. [256] ont utilisé le SVM et l'algorithme de classification des sentiments (SCA) sur les données Twitter. Les performances de SVM et de SCA ont été comparées, et alors SCA

s'est avéré être meilleur que le SVM.

L'analyse des sentiments a également attiré l'attention des chercheurs dans le domaine de l'éducation récemment. De nombreuses études ont appliqué l'analyse des sentiments pour analyser les attitudes des étudiants envers divers aspects en utilisant l'apprentissage automatique et l'apprentissage profond. Altrabsheh et al. [257] ont construit un système de réponse des étudiants pour recueillir des commentaires à partir de Twitter, de cliques et de téléphones mobiles en temps réel, puis ils ont comparé différentes techniques d'analyse des sentiments comme SVM, Naïve bayes et le maximum d'entropie. Mujahid et al. [233] ont analysé les sentiments des gens concernant l'enseignement à distance pendant le COVID-19 en utilisant des techniques d'apprentissage profond et d'apprentissage automatique sur un ensemble de données qui contient 17155 tweets sur l'apprentissage en ligne (e-learning). Waheeb et al. [37] ont proposé une méthode d'analyse des sentiments à propos de l'apprentissage en ligne à partir de Twitter afin d'identifier les fausses nouvelles COVID-19 en utilisant l'autoencodeur Extreme Learning Machine (ELM-AE) avec LSTM.

Dans le contexte de l'enseignement supérieur, de nombreuses études ont été appliquées aux tweets sur les universités en utilisant des techniques d'apprentissage automatique. Balachandran et Kirupananda [258] examinent les commentaires des utilisateurs sur Twitter et Facebook concernant les avec la bibliothèque StanfordCoreNLP. Alruily et Shahin [259] ont proposé une approche d'analyse des sentiments pour analyser les tweets des universités Saoudiennes en utilisant Naïve bayes, SVM, K-Nearest Neighbors classifieur (KNN), Stochastic Gradient Descent (SGD), Random Forest, Sequential Minimal Optimization (SMO), Multilayer Perceptrons, et Sentiment Score Calculation (SSC). L'ensemble de données collectées contient 5882 commentaires positifs et 5882 commentaires négatifs sur 22 universités Saoudiennes. La meilleure précision du test est obtenue en utilisant le SVM. AL-Rubaiee et al. [260] ont introduit un cadre pour l'analyse des tweets en arabe concernant les attitudes des étudiants de l'Université King Abdul-Aziz envers l'apprentissage à distance en utilisant Naïve bayes et SVM.

4.3.3. Système d'analyse du sentiment en temps réel sur les tweets relatifs aux universités marocaines

L'architecture globale du système proposé pour la prédiction en temps réel du sentiment sur les tweets relatifs aux universités marocaines est représentée dans la figure 5.1. Elle comprend deux composants principaux :

1. Un modèle d'analyse de sentiments en utilisant des algorithmes d'apprentissage automatique.

2. Un pipeline d'analyse de sentiments en temps réel en utilisant les technologies de Big data.

4.3.3.1. Phase de création d'un modèle d'analyse de sentiment

Nous avons développé un modèle d'analyse de sentiment en utilisant six algorithmes d'apprentissage automatique avec différentes techniques d'extraction de caractéristiques à appliquer pour la prédiction de sentiments en temps réel. Nous avons utilisé un ensemble de données de tweets français collectés sur 12 universités marocaines pour entraîner les modèles d'apprentissage automatique. La figure 4.15 souligne les étapes de l'analyse des sentiments, comprenant la collecte des données, le prétraitement des données, les tweets pré-entraînés, la séparation des données, l'extraction des caractéristiques, le modèle d'analyse des sentiments et l'évaluation des performances du modèle. Chaque étape est décrite comme suit :

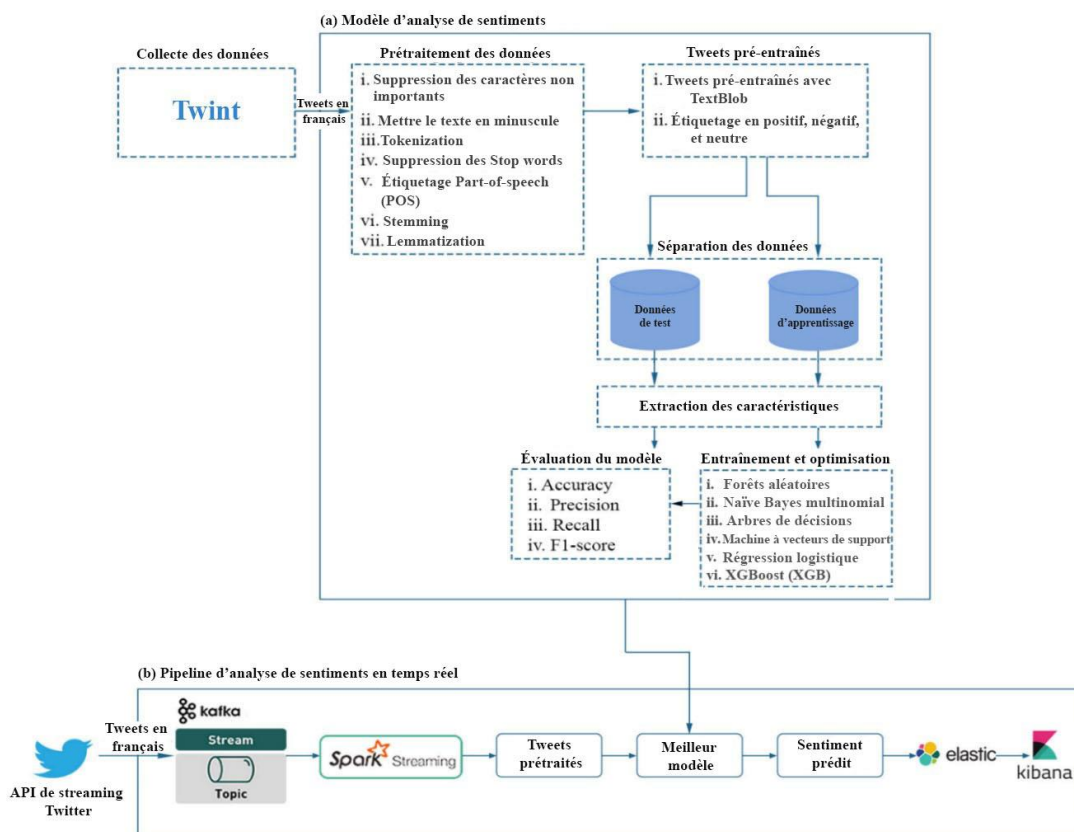


Figure 4.15 – Architecture d'analyse des sentiments en temps réel à partir de Twitter.

- *Collecte de données* : Dans l'étape de collecte des données du modèle d'analyse de sentiment, nous avons collecté 1798 tweets historiques français liés à 12 universités publiques marocaines entre le 11 juin 2009 et le 24 mai 2022. Comme l'API de recherche Twitter (Search API) offre l'accès à des données anciennes remontant juste jusqu'à une semaine auparavant, nous avons utilisé Twint [261], une bibliothèque Python open-source conçue pour extraire les tweets à partir de mots-clés, de dates, de lieux et de hashtags. Le nombre de tweets collectés pour chaque université est indiqué dans le Tableau 4.5.

Tableau 4.5 – Nombre de tweets pour chaque université marocaine.

Université	Nombre de tweets
Université Mohammed V de Rabat	375
Université Sidi Mohamed Ben Abdellah	240
Université Ibn Tofaïl	220
Université Cadi Ayyad	158
Université Hassan 1 ^{er}	154
Université Ibn Zohr	118
Université Abdelmalek Essaâdi	113
Université Chouaib Doukkali	108
Université Hassan II	106
Université Mohammed 1 ^{er}	85
Université Moulay Ismail	61
Université Sultan Moulay Slimane	60

- *Prétraitement des données* : Les données générées sur les sites de médias sociaux sont vastes, bruyantes, distribuées, hétérogènes et non structurées. Il est donc essentiel d'éliminer le bruit des données Twitter avant d'appliquer des algorithmes d'apprentissage automatique ou d'apprentissage profond. Pour y parvenir, nous devons effectuer un prétraitement du texte. Le Tableau 4.6 montre un exemple d'application des étapes de prétraitement mentionnées dans la sous-section 4.2.3 sur un tweet.

Tableau 4.6 – Exemple de mise en œuvre des étapes de prétraitement sur un tweet.

Université	Avant le prétraitement	Après le prétraitement
Université Mohammed V de Rabat	L'Université Mohammed V de Rabat a été classée première au Maghreb par le Center for World University Rankings (CWUR), qui a dévoilé, lundi 25 avril, les résultats de son classement annuel des universités mondiales. #Ma-roc #Enseignementsupérieur @um5rabat	université mohammed v rabat a classée première maghreb center for world university rankings cwur a dévoilé lundi avril résultat classement annuel université mondiales

- *Tweets pré-entraînés* : Nous avons utilisé TextBlob [263], une bibliothèque NLP dans Python à code source ouvert, qui donne la polarité qui est une valeur flottante entre -1 et 1, où -1 caractérise un sentiment négatif, 0 caractérise un sentiment neutre, et 1 caractérise un sentiment positif. Ensuite, la subjectivité qui représente une valeur flottante comprise entre 0 et 1 et qui quantifie le nombre d'émotions, ou d'opinions personnelles dans le texte. La polarité est utilisée dans notre étude, pour étiqueter les tweets collectés afin de les transmettre au modèle d'analyse des sentiments proposé. Le résultat est le suivant : 969 tweets neutres, 721 tweets positifs et 108 tweets négatifs.
- *Séparation des données* : Pour former le modèle d'analyse de sentiments que nous proposons, nous avons utilisé une séparation des données stratifiée avec un ratio 90:10, où 90% de l'ensemble de données pré-entraînés est utilisé pour l'apprentissage et 10% pour le test. Dans l'ensemble d'apprentissage, chaque texte d'entrée est lié au sentiment exact, ce qui permet au modèle d'effectuer le processus d'apprentissage. En outre, l'ensemble de test est composé de données non utilisées pour l'évaluation du modèle.
- *Extraction des caractéristiques* : Il existe plusieurs techniques d'extraction des caractéristiques à partir de différents types de données. Dans cette étude, les techniques Count Vectorizer et TF-IDF sont utilisées dans l'approche proposée pour extraire les caractéristiques sémantiques du tweet.
 - *Count Vectorizer* : C'est une technique qui consiste à compter l'occurrence d'un terme dans un document textuel et utilise cette valeur sous forme d'entiers comme son poids pour convertir une collection de documents textuels en un vecteur de comptage de termes et de tokens et

construire un vocabulaire de termes.

- *Term frequency-inverse document frequency (TF-IDF)* : vise à extraire des caractéristiques du texte. La TF mesure le nombre d'un terme t dans le document d , et l'IDF fournit l'importance du terme. La TF et l'IDF sont définies dans les équations (4.9) et (4.10).

$$TF(t, d) = \frac{\text{Nombre de fois où le terme } t \text{ apparaît dans un document } d}{\text{Nombre total de termes dans un document } d} \quad (4.9)$$

$$IDF(t) = \log_e \frac{\text{Nombre total de documents}}{\text{Nombre total de documents avec le terme } t} \quad (4.10)$$

Ensuite, le TF-IDF calcule le poids d'un terme t dans un document d en multipliant le TF par les valeurs IDF pour chaque terme, comme défini dans l'équation (4.11).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (4.11)$$

TF-IDF est noté entre 0 et 1. Les valeurs les plus élevées correspondent à des mots moins fréquents dans le corpus de documents. La plus petite valeur pour les mots significatifs dans le corpus de documents.

- *Séparation des données* : Pour former le modèle d'analyse de sentiments que nous proposons, nous avons utilisé une séparation des données stratifiée avec un ratio 90:10, où 90% de l'ensemble de données pré-entraînés est utilisé pour l'apprentissage et 10% pour le test. Dans l'ensemble d'apprentissage, chaque texte d'entrée est lié au sentiment exact, ce qui permet au modèle d'effectuer le processus d'apprentissage. D'autre part, l'ensemble de test est constitué de données non utilisées pour l'évaluation du modèle.
- *Classification des données* : Dans cette étude, l'objectif de la tâche de classification est d'assigner un tweet inconnu à une classe de sentiment (positif, négatif ou neutre) sur la base des données d'apprentissage. La méthode de recherche sur grille (ou Grid Search) avec validation croisée 10 fois stratifiée (CV) a été utilisée pour optimiser les hyperparamètres des modèles d'apprentissage automatique utilisés et appliquer les meilleurs hyperparamètres pour prédire le sentiment d'un tweet. Les algorithmes d'apprentissage automatique utilisés et leurs hyperparamètres sont présentés comme suit :
 - *Régression logistique (sous-section 1.2.2)* : Les hyperparamètres du modèle sont : multi_class='multinomial', solver='lbfgs', et C=5e1.

- *Machine à vecteurs de support (SVM) (sous-section 1.2.3)* : Les hyperparamètres du modèle sont : kernel='linear', dual='False', et C=1.
- *Naïve bayes multinomial (sous-section 1.2.4)* : L'hyperparamètre du modèle est force_alpha=True.
- *Arbres de décisions (sous-section 1.2.5)* : Les hyperparamètres du modèle sont : criterion='entropy' et splitter='best'.
- *Forêts aléatoires (sous-section 1.2.6)* : Le nombre d'estimateurs utilisés avec ce modèle a été fixé à 50.
- *Boosting de gradient extrême (XGBoost) (sous-section 1.2.7)* : Le nombre d'estimateurs utilisés avec ce modèle a été fixé à 100.
- *Évaluation des modèles* : Quatre mesures de performance sont utilisées pour l'évaluation des modèles proposés : l'accuracy, la précision, le rappel et le score F1, définis dans les équations (2.1), (2.2), (2.3) et (2.4).

4.3.3.2. Phase de création d'un pipeline d'analyse de sentiments en temps réel en utilisant les technologies de Big data

Le pipeline d'analyse de sentiments prédit le sentiment des tweets en streaming des universités marocaines en effectuant un traitement en temps réel sur les données collectées à l'aide du modèle d'apprentissage automatique déployé. Le tweet est extrait et publié dans un sujet Kafka (Kafka Topic). Le Spark Streaming Job consomme le message du tweet de Kafka, effectue une analyse de sentiment et insère le résultat dans Elasticsearch. Ensuite, les données stockées sur Elasticsearch sont affichées sur le tableau de bord Kibana. L'architecture du pipeline de prédiction en ligne comprend cinq étapes : la collecte des données, l'ingestion des données, l'analyse des sentiments en temps réel, l'exploration des données en temps réel et la visualisation des données.

- *Collecte de données à l'aide de l'API de streaming Twitter* : L'API de streaming Twitter est utilisée pour extraire les tweets de Twitter en utilisant la bibliothèque Tweepy, qui permet à Python de communiquer avec Twitter et d'utiliser son API via des identifiants d'authentification de base tels que (consumer secret, consumer key, access secret, et access key). A partir de cette API, de nombreux champs sont scrappés comme (texte, source, retweets, langue, utilisateur, localisation, etc.). L'API de filtrage (ou filter API) permet de rechercher un flux de tweets correspondant à un hashtag. Un total de 321 tweets en français a été collectés en temps réel pour cette étude sur 12 universités publiques marocaines du 25 mai 2022 au 16 septembre 2022.

- *Ingestion de données à l'aide d'Apache Kafka* : Les données en streaming de Twitter sont ingérées en temps réel dans Kafka Topic. L'ingestion de données vise à importer des données de diverses sources telles que des sites de médias sociaux, des weblogs, des systèmes de gestion de bases de données relationnelles (SGBDR), etc... vers un site cible pour un traitement ultérieur. Apache Kafka est un système de messagerie distribué de type publish-subscribe qui revêt une importance dans l'ingestion de données en temps réel. Il est utilisé pour construire des pipelines de données en temps réel pour transférer de grands volumes de données entre plusieurs applications. Kafka est exécuté en tant que cluster sur de nombreux serveurs appelés Kafka Cluster, qui se composent de plusieurs courtiers (ou brokers). Un courtier possède une partition et chaque partition stocke les flux de données avec ses clés, ses valeurs et ses horodatages dans des catégories appelées sujets. Un producteur envoie le message vers le courtier et les consommateurs reçoivent les messages du courtier. En d'autres termes, le producteur communique avec les courtiers Kafka par le biais du réseau pour écrire des événements sur un sujet, et les consommateurs lisent ces événements. Kafka utilise ZooKeeper [264] pour déterminer quel courtier est le leader d'une partition et d'un sujet donné et pour procéder à l'élection du leader. La figure 4.16 illustre l'architecture d'Apache Kafka.

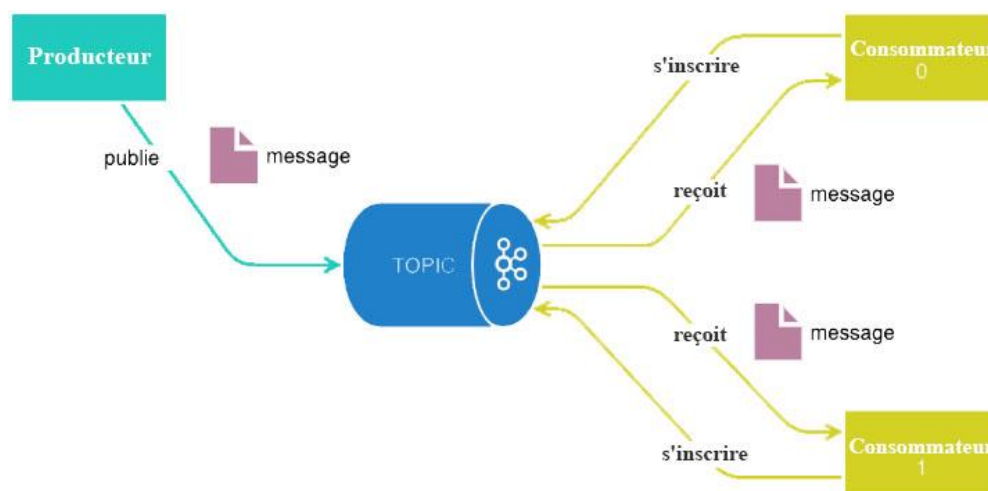


Figure 4.16 – Architecture d'Apache Kafka.

- *Analyse des sentiments en temps réel avec Apache Spark* : Les capacités de streaming et d'apprentissage automatique de Spark sont utilisées pour réaliser le modèle d'analyse des sentiments. Apache Spark est une plateforme de traitement open source distribuée en mémoire, développée pour le calcul rapide.

L'ensemble de données distribuées résilientes (ou Resilient Distributed Dataset RDD) est la principale abstraction de Spark. Il s'agit d'une collection distribuée d'objets répartis sur un ensemble de nœuds d'un cluster, avec la capacité de récupérer les pannes survenues. Il s'agit d'une collection distribuée d'objets répartis sur un ensemble de nœuds d'un cluster, avec la capacité de se rétablir automatiquement en cas de défaillance du système. Spark suit une architecture maître/esclave, qui est une combinaison de SparkContext en tant que maître de l'application Spark, un ensemble de Workers, qui sont acquis par Spark pour exécuter des calculs et stocker des données pour une application. Spark est écrit en Scala, mais il dispose également d'APIs pour Python, R et Java. La figure 4.17 illustre l'architecture d'Apache Spark.

L'écosystème Spark comprend six composants tels que Spark Core, Spark SQL, Spark GraphX, Spark MLlib, Spark Streaming et SparkR. Spark MLlib est une structure d'apprentissage automatique distribuée sur Spark Core. Il dispose de presque tous les algorithmes d'apprentissage automatique populaires, notamment le regroupement, la régression et la classification. Spark Streaming reçoit des flux de données en direct qui sont divisés en petits lots appelés flux discrétisés (ou Discretized Streams (DStreams)). Il s'agit d'une séquence continue de RDDs dans la mémoire de Spark qui peut être collectée à partir de nombreuses sources de données en continu telles que Flume, Kafka et Elasticsearch. Dans notre étude, le streaming Spark prétraite les tweets relatifs aux universités marocaines pour les intégrer dans le modèle d'apprentissage automatique afin de prédire les sentiments correspondants.

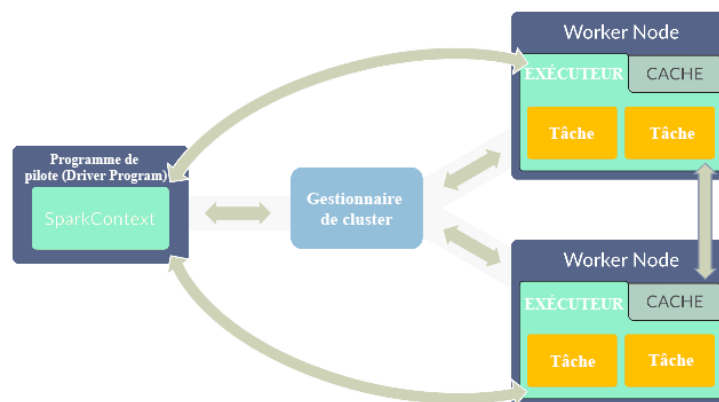


Figure 4.17 – Architecture d'Apache Spark.

- *Exploration des données en temps réel avec Elasticsearch* : Les résultats de la prédiction en ligne obtenus à l'aide de l'API de Spark Streaming sont émis vers Elasticsearch. Ce dernier est un moteur distribué en temps réel construit au-dessus d'Apache Lucene. Elasticsearch stocke les données sous la forme de documents JSON (JavaScript Object Notation). Chaque document possède un ensemble de clés avec leurs valeurs correspondantes. Une fois que les documents sont indexés, Elasticsearch crée une structure de données connue sous le nom d'index inversé conçue pour permettre une récupération très rapide des résultats de recherche. Les données dans Elasticsearch sont distribuées sur plusieurs serveurs appelés nœuds. La collection de nœuds connectés est appelée un Cluster. Elasticsearch peut prendre en charge des clients dans de nombreux langages différents tels que PHP, Java, JavaScript, Python, Go, C# et Ruby.
- *Visualisation des données avec Kibana* : Les données traitées en temps réel stockées sur Elasticsearch sont affichées sur le tableau de bord Kibana pour analyser les sentiments des gens à l'égard des universités marocaines. Kibana est une couche de visualisation de données construite au-dessus d'Elasticsearch utilisée pour analyser et visualiser les données sous la forme d'un graphique à barres, d'un graphique linéaire, d'un graphique circulaire, etc. Kibana fonctionne en synchronisation avec Elasticsearch et Logstash, qui forment ensemble la pile ELK. Les tableaux de bord Kibana affichent les images complètes des données en Elasticsearch et peuvent être personnalisés en fonction des besoins des utilisateurs.

4.3.4. Détails de l'implémentation du système proposé

Le modèle proposé est écrit en PySpark qui est une API Python pour Apache Spark et exécuté dans un ordinateur portable équipé d'une carte graphique Nvidia GTX 1070 et d'un système d'exploitation Windows 10 Professional 64 bits avec 16 Go de RAM et un CPU Intel Core i7 de 3,20 GHz, au sein de notre laboratoire Conception et Systèmes (Electronique, Signaux et Informatique). Lors de la phase d'apprentissage, l'ensemble de données des tweets pré-entraînés utilisé dans cette étude est divisé en ensemble de données d'apprentissage et de test dans un rapport de division stratifié 90-10. Nous avons utilisé la méthode de recherche de grille (Grid Search) pour trouver les meilleures valeurs des hyperparamètres.

4.3.5. Résultats expérimentaux

Dans cette sous-section, nous allons fournir une analyse approfondie des performances de notre modèle d'analyse de sentiment, qui a été développé en utilisant des technologies de Big data. Nous allons exposer en détail les résultats obtenus lors de la création de ce modèle ainsi que les étapes requises pour la mise en œuvre d'un pipeline d'analyse de sentiments en temps réel. L'objectif de cette évaluation est de mettre en évidence l'efficacité de notre modèle et sa capacité à traiter des données en temps réel, tout en fournissant des informations précieuses sur le sentiment des utilisateurs à l'égard des universités marocaines.

Figures 4.18 et 4.19 et le Tableau 4.7 présentent les résultats obtenus en employant six modèles d'apprentissage automatique, notamment les forêts aléatoires (RF), le classificateur Naïve bayes multinomial (MNB), la régression logistique (LR), les arbres de décisions (DT), le classificateur linéaire à vecteur de support (SVC) du SVM et le Boosting de gradient extrême (XGBoost) avec TF-IDF et Count Vectorizer.

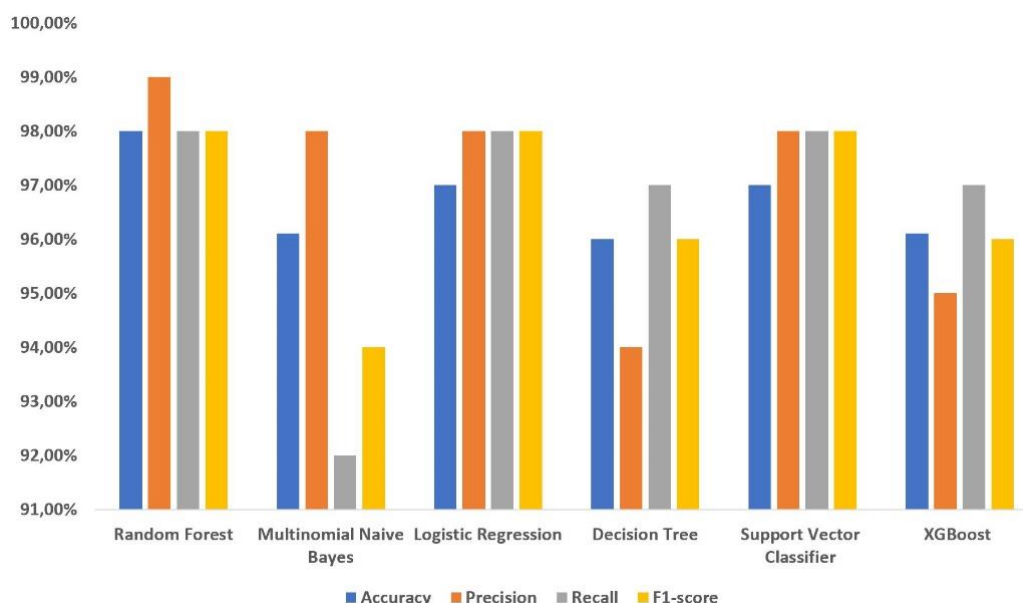


Figure 4.18 – Mesures de performance des six algorithmes d'apprentissage automatique avec TF-IDF.

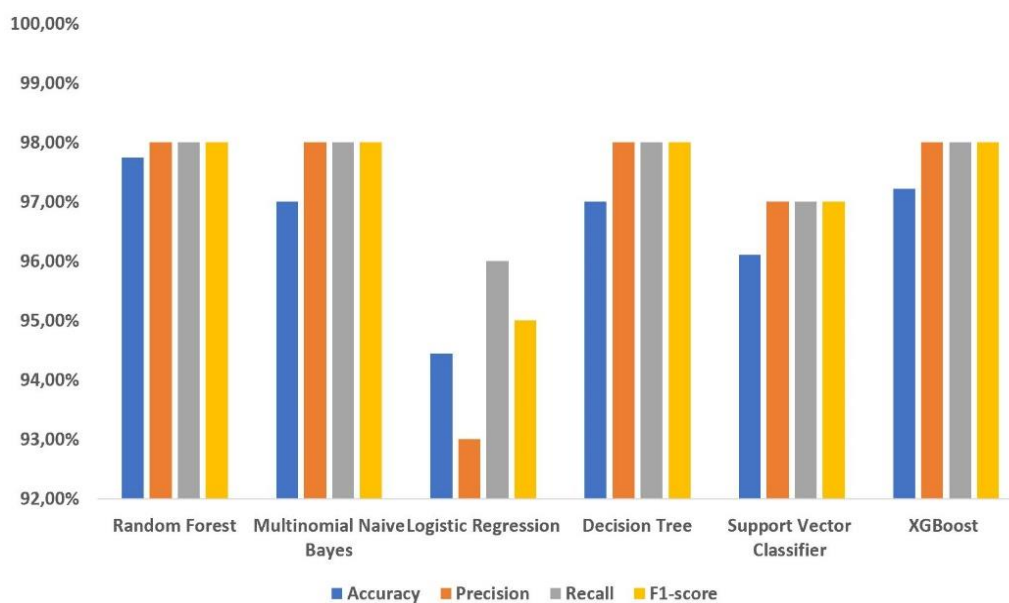


Figure 4.19 – Mesures de performance des six algorithmes d'apprentissage automatique avec Count Vectorizer.

Tableau 4.7 – Résultats des performances des six classificateurs avec TF-IDF et Count Vectorizer.

Extraction des caractéristiques	Modèle	Accuracy	Precision	Recall	F1-score
TF-IDF	RF	98.00%	99.00%	98.00%	98.00%
	MNB	96.11%	98.00%	92.00%	94.00%
	LR	97.00%	98.00%	98.00%	98.00%
	DT	96.00%	94.00%	97.00%	96.00%
	SVC	97.00%	98.00%	98.00%	98.00%
	XGBoost	96.11%	95.00%	97.00%	96.00%
Count Vectorizer	RF	97.74%	98.00%	98.00%	98.00%
	MNB	97.00%	98.00%	98.00%	98.00%
	LR	94.45%	93.00%	96.00%	95.00%
	DT	97.00%	98.00%	98.00%	98.00%
	SVC	96.11%	97.00%	97.00%	97.00%
	XGBoost	97.22%	98.00%	98.00%	98.00%

Les résultats obtenus montrent que l'algorithme des forêts aléatoires (RF) est le plus performant parmi tous les modèles de classification testés. Il a obtenu une précision globale de 98% en utilisant la méthode de vectorisation de texte TF-IDF et une précision de 97.74% avec la méthode Count Vectorizer. Cela signifie que l'algorithme RF a réussi à classer avec précision la plupart des tweets en fonction de leur sentiment, en utilisant les caractéristiques

textuelles extraites par ces méthodes de vectorisation. En revanche, la précision la plus faible a été obtenue par la régression logistique (LR) avec Count Vectorizer, avec une précision de 94.45%. En outre, le modèle d'arbres de décisions (DT) a obtenu la précision de test la plus élevée avec 96% en utilisant la méthode de vectorisation de texte TF-IDF. Bien que cela soit inférieur à la précision globale de l'algorithme RF, cela indique que DT est un modèle viable pour la classification des sentiments des tweets.

En fin de compte, étant donné que l'algorithme des forêts aléatoires (RF) a obtenu les meilleures performances de classification des sentiments des tweets, il a été sélectionné pour être utilisé pour classifier et prédire le sentiment des tweets non étiquetés. La matrice de confusion du modèle RF validé avec un schéma de validation stratifié à 90-10% est présentée dans la figure 4.20.

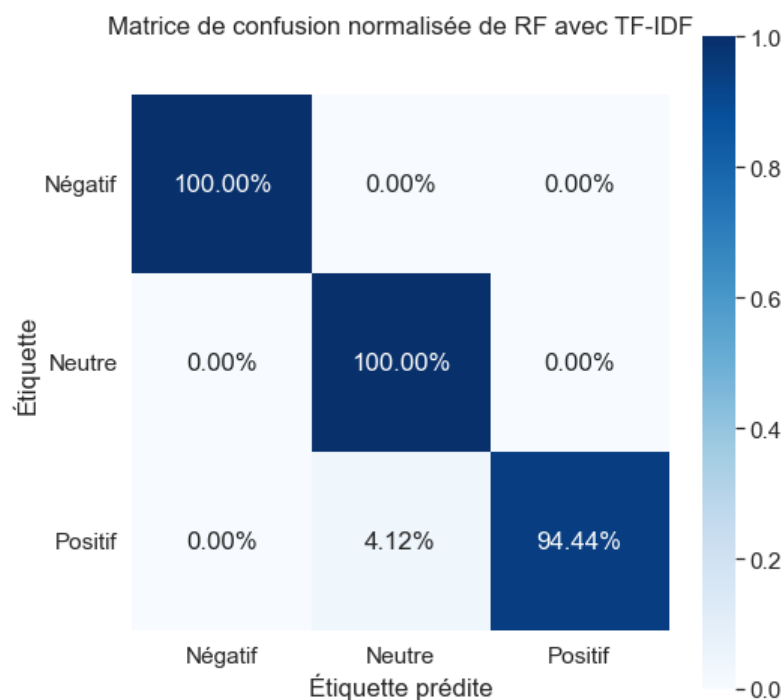


Figure 4.20 – Matrice de confusion normalisée du classificateur Forêts aléatoires (RF) avec TF-IDF.

La figure 4.21 présente un nuage de mots les plus fréquents des tweets relatifs à douze universités marocaines. Le nuage de mots est construit en utilisant les mots les plus fréquents dans les tweets et la dimension de chaque mot est en relation avec sa fréquence d'utilisation. Les mots les plus fréquents qui apparaissent dans la figure incluent des noms d'universités tels que 'Université', 'universite', 'mohammed', 'rabat', 'cadi', 'ayyad', 'maroc'. Les mots clés liés à l'éducation tels que 'faculté', 'science', et 'étudiant'.

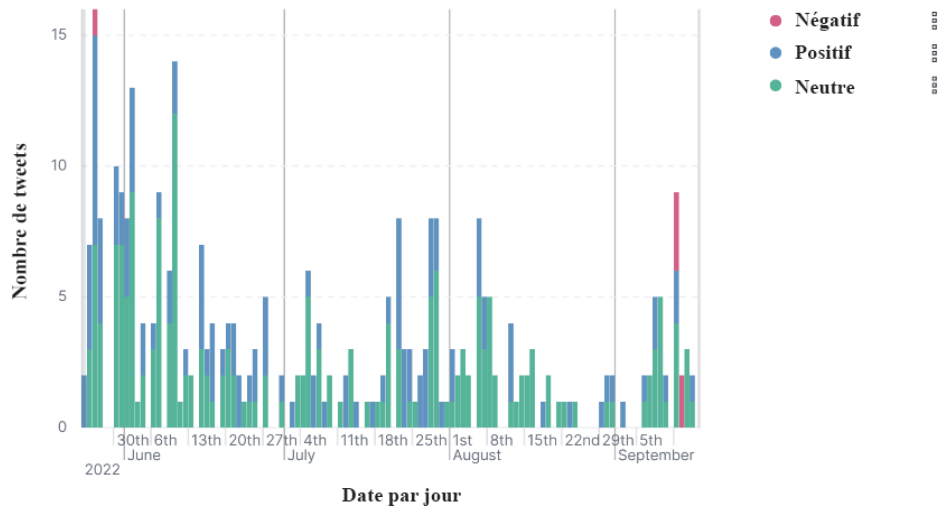


Figure 4.22 – Diagramme à barres montrant le nombre de tweets en temps réel reçus par jour et par sentiment.

Le diagramme en barre de la figure 4.23 représente le sentiment de chaque université publique marocaine. Nous pouvons clairement voir que l'Université Mohammed V de Rabat est celle qui suscite le plus d'intérêt sur Twitter, suivie de l'Université Abdelmalek Essaâdi et de l'Université Cadi Ayyad. Il est intéressant de noter que les sentiments des utilisateurs de Twitter à l'égard des universités marocaines sont généralement positifs et neutres, avec très peu de tweets négatifs. Cela pourrait indiquer un niveau élevé de satisfaction des étudiants et du public en général à l'égard de ces institutions universitaires.

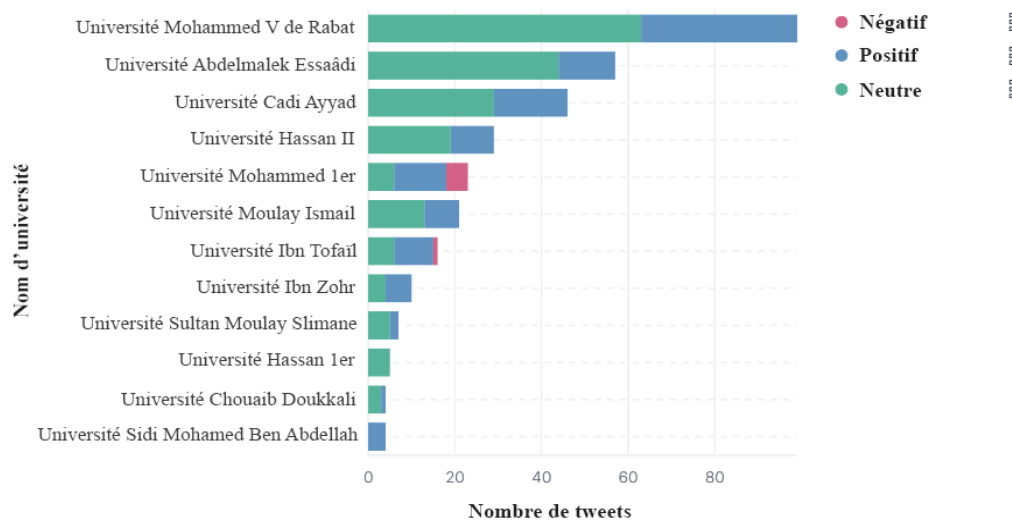


Figure 4.23 – Diagramme à barres horizontales empilées montrant le sentiment prédit des universités Marocaines.

La figure 4.24 présente un diagramme circulaire qui détaille le pourcentage de chaque sentiment des tweets classés en trois catégories : neutres, positifs et négatifs. Les résultats obtenus démontrent que la plupart des tweets collectés étaient neutres, représentant ainsi 61.37% de l'ensemble des tweets. Les tweets positifs sont également nombreux, constituant une part importante avec 36.76%. À l'inverse, les tweets négatifs ne représentent qu'une infime minorité avec seulement 1.87%. Ensuite, l'analyse se focalise davantage sur les tweets positifs. Il est révélé que l'Université Mohammed V de Rabat est l'université la plus appréciée avec un pourcentage de 30.51% des tweets positifs. Elle est suivie de près par l'Université Cadi Ayyad avec 14.41% des tweets positifs, puis l'Université Abdelmalek Essaadi qui occupe la troisième position avec 11.02% des tweets positifs. Cependant, l'Université Chouaib Doukkali a obtenu le plus faible pourcentage de tweets positifs parmi les douze universités étudiées, une information qui est représentée dans la figure 4.25.

En somme, cette analyse démontre que la majorité des tweets collectés étaient neutres, mais qu'il y avait également une proportion significative de tweets positifs. De plus, elle met en évidence que l'Université Mohammed V de Rabat est l'université la plus appréciée parmi les douze universités étudiées. En revanche, pour l'Université Chouaib Doukkali, la majorité des tweets étaient neutres, avec un nombre de tweets positifs inférieur à celui des autres universités étudiées.

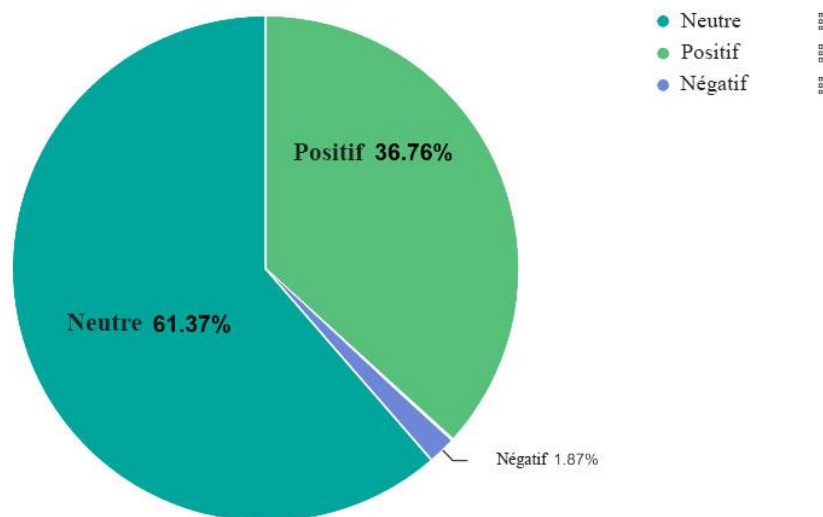


Figure 4.24 – Distribution des sentiments prédits des tweets relatifs aux universités Marocaines.

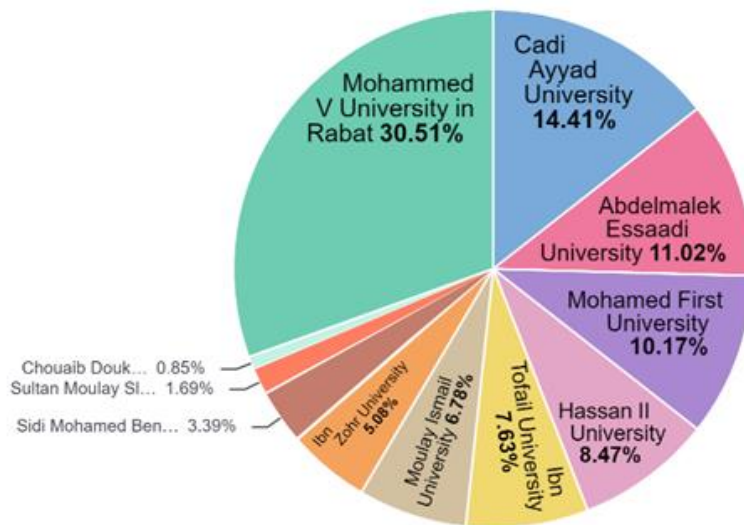


Figure 4.25 – Pourcentage de tweets positifs par université.

Les bigrammes positifs les plus fréquents liés à l'Université Mohammed V sont "recherche scientifique", "qualité education", "education formation", "engagement grand", "um5rabat productivité", "scientifiques partenariats", "recherche innovation", "professeurs université", "meilleurs établissements", et "milieu universitaire". On constate que les gens sont satisfaits de la qualité de l'enseignement, la qualité de la formation, la recherche scientifique, l'engagement, les professeurs, les établissements, l'environnement universitaire, les partenariats scientifiques, la productivité et l'innovation au sein de l'Université Mohammed V de Rabat. Les résultats obtenus sont pratiquement identiques à ceux de l'édition 2019 du Times Higher Education (THE) [265], de l'édition 2022 du Center for World University Rankings (CWUR) [266], et de l'édition 2022 du Centre for Science and Technology Studies (CWTS Leiden) [267].

L'Université Mohammed V de Rabat est classée première parmi les universités marocaines dans le prestigieux classement international 2019 du Times Higher Education selon cinq critères prédéfinis, les citations (28.6), les revenus de l'industrie (40.8), la visibilité internationale (28.9), la recherche (8.9), et l'enseignement (24.9). Elle est également classée première au niveau national et maghrébin et 950ème au niveau mondial selon le classement CWUR. De plus, l'Université Mohammed V de Rabat est la seule université marocaine à figurer dans le classement CWTS Leiden 2022. Les indicateurs ont été calculés par le Centre for Science and Technology Studies (CWTS) de l'Université Leiden à partir de la base de données Scopus.

La figure 4.26 illustre une carte des tweets géolocalisés collectés en temps réel dans notre étude. Chaque tweet est marqué par ses coordonnées de latitude et de longitude, ce qui permet de visualiser précisément leur emplacement sur la carte.



Figure 4.26 – Répartition géographique des tweets en temps réel.

Les résultats montrent que la majorité des tweets se concentrent dans la région du Maroc, avec une densité de tweets plus élevée au nord, notamment à Rabat, Tanger, Casablanca et Kenitra. On observe également une concentration de tweets significative en Espagne, en particulier dans la ville de la Corogne, ainsi qu'aux États-Unis, en particulier dans les États de l'Illinois, du New Jersey et du Texas. En France, la ville de Rennes est un lieu de forte activité pour les tweets. En effet, la géolocalisation des tweets relatifs aux universités marocaines peut avoir une importance considérable pour les décideurs dans l'enseignement supérieur marocain. En analysant la répartition géographique des tweets, il est possible de repérer les zones où les activités des universités sont les plus commentées et partagées sur les réseaux sociaux. Cette démarche peut être bénéfique pour les responsables des universités afin de mieux comprendre les exigences des étudiants et de la communauté éducative dans ces zones géographiques et d'adapter leurs stratégies de communication.

Le tableau de bord présenté dans la figure 4.27 est un outil précieux pour aider les responsables de l'enseignement supérieur marocain à surveiller les commentaires et les opinions des étudiants et du public sur les différents aspects de leurs universités en temps réel. Il comprend cinq graphiques différents qui permettent de visualiser le nombre de tweets reçus par jour et par sentiment, le sentiment prédit des universités marocaines, la distribution des sentiments prédits des tweets relatifs aux universités marocaines, le pourcentage de tweets positifs par université, et la répartition géographique des tweets en temps réel.

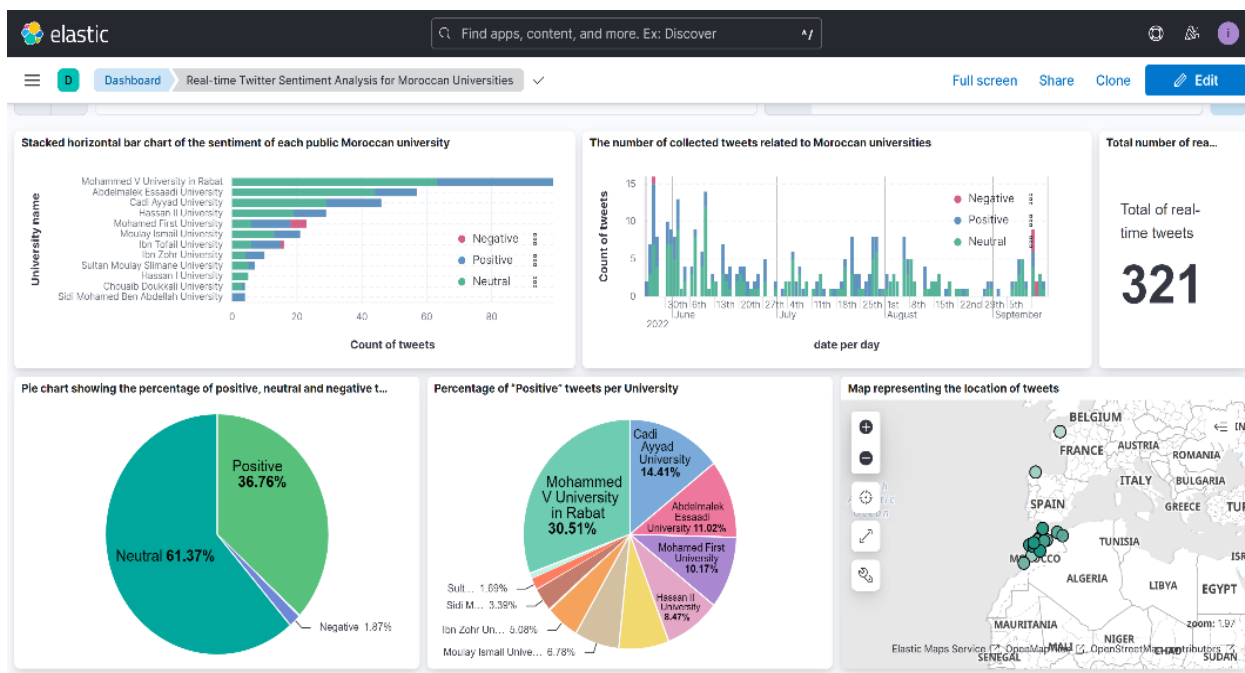


Figure 4.27 – Tableau de bord produit par Kibana pour la visualisation des tweets en temps réel.

Les décideurs peuvent utiliser les données collectées à travers les différents graphiques pour prendre des mesures proactives pour améliorer les aspects qui ont des retours négatifs, tout en renforçant les aspects qui ont des retours positifs. Cela permettra d'améliorer la qualité de l'enseignement supérieur au Maroc. En surveillant les commentaires et les opinions du public sur les universités, en identifiant les domaines qui nécessitent une amélioration et en prenant des mesures proactives pour y remédier, le tableau de bord peut aider à accroître la qualité de l'enseignement supérieur. De plus, en identifiant les domaines de recherche et d'enseignement qui ont reçu des commentaires positifs, les décideurs peuvent utiliser ces informations pour solliciter des financements supplémentaires pour améliorer les infrastructures, recruter des enseignants qualifiés et offrir de meilleurs programmes et équipements.

4.4. Conclusion et discussion

Ce chapitre présente deux études qui utilisent des méthodes fondées sur l'apprentissage automatique et profond pour analyser les sentiments des utilisateurs de Twitter concernant l'enseignement à distance et la réputation des universités marocaines. Dans la première étude, nous avons collecté 24 642 tweets sur l'enseignement à distance entre juillet 2022 et novembre 2022. Les tweets ont été prétraités et étiquetés comme positifs, négatifs ou neutres, puis utilisés pour former un modèle Bi-LSTM basé sur l'auto-attention et Glove pour l'extraction des caractéristiques. Les résultats ont montré que 59% des personnes ont exprimé des sentiments neutres, 33.12% des personnes ont exprimé des sentiments positifs et 7.88% des personnes ont exprimé des sentiments négatifs. Les domaines de psychologie et d'informatique ont enregistré les sentiments positifs les plus élevés, tandis que la biologie et la chimie ont enregistré les sentiments négatifs les plus élevés. Cette étude diffère des études antérieures qui se sont principalement concentrées sur les attitudes négatives envers l'enseignement à distance pendant la pandémie. Les résultats soulignent les avantages potentiels de l'enseignement à distance, tels que la flexibilité et l'accessibilité, tout en soulignant les préoccupations concernant la qualité de l'enseignement à distance dans des domaines pratiques. Dans la deuxième étude, nous avons développé un système de prédiction du sentiment en temps réel basé sur les tweets concernant douze universités publiques marocaines. Nous avons collecté un ensemble de tweets historiques en français et on les a annotés comme positifs, négatifs ou neutres. Les tweets ont été prétraités et utilisés pour entraîner différents modèles d'apprentissage automatique. Les résultats ont montré que 61.37% des tweets en temps réel étaient neutres, 36.76% étaient positifs et 1.87% étaient négatifs. L'université Mohammed V de Rabat a obtenu le plus grand nombre de tweets positifs parmi les douze universités, avec des mentions fréquentes de la qualité de l'enseignement, de la recherche scientifique et de l'environnement universitaire. Les résultats sont cohérents avec les classements internationaux.

Ces études apportent des contributions significatives à la littérature en examinant les opinions du public sur l'apprentissage à distance et la réputation des universités marocaines. Elles soulignent l'importance de l'analyse de sentiment en ligne comme indicateur d'évaluation de la qualité des systèmes éducatifs et de la réputation des universités. Les résultats peuvent être utilisés par les parties prenantes de l'enseignement supérieur pour améliorer l'expérience d'apprentissage à distance et pour prendre des décisions stratégiques en matière d'investissement et d'amélioration des universités.

Conclusion générale

Dans le cadre de ce mémoire, notre objectif principal était d'explorer l'adaptation de l'apprentissage aux apprenants et l'évaluation de la qualité des systèmes éducatifs, en accordant une attention particulière à l'éducation inclusive. Pour ce faire, deux volets de recherche ont été abordés.

Le premier volet visait à améliorer l'évaluation formative en exploitant la reconnaissance des émotions faciales. Cette approche permettait de mieux comprendre les émotions ressenties et le niveau de concentration des apprenants pendant le processus d'apprentissage, offrant ainsi aux enseignants la possibilité d'ajuster leurs méthodes pédagogiques en conséquence. Pour atteindre cet objectif, nous avons développé deux systèmes de reconnaissance des émotions faciales (FER) capables de détecter le visage des étudiants en temps réel, extraire les caractéristiques faciales, reconnaître les émotions à partir de ces caractéristiques, et détecter l'engagement à partir des émotions faciales en se basant sur des algorithmes d'apprentissage profond. Le premier système FER utilise un réseau de neurones convolutif (CNN) Le deuxième système FER utilise un réseau de neurones convolutifs profonds VGG-16 basé sur l'apprentissage par transfert afin de détecter l'engagement des étudiants sourds et malentendants à partir de leurs émotions faciales.

Le deuxième volet de notre recherche s'est concentré sur le développement de deux systèmes d'analyse des sentiments pour évaluer la qualité des systèmes éducatifs. Le premier système utilise l'apprentissage profond pour étudier les opinions des utilisateurs de Twitter sur l'apprentissage à distance dans l'enseignement supérieur. Les disciplines de psychologie (41.17%) et d'informatique (34.2%) affichent les pourcentages les plus élevés de sentiments positifs, tandis que la biologie (28%) et la chimie (24.6%) enregistrent les pourcentages les plus élevés de sentiments négatifs. Cette étude adresse une lacune dans la littérature en analysant les opinions du public sur l'apprentissage à distance après la pandémie et offre des indications précieuses pour améliorer cette expérience dans différentes disciplines de l'enseignement supérieur. L'analyse des sentiments peut également servir d'indicateur d'évaluation de la qualité des systèmes éducatifs, en

fournissant des informations précieuses sur l'expérience des étudiants et en identifiant les domaines nécessitant des améliorations pour optimiser l'apprentissage à distance. Le deuxième système est capable de mesurer la réputation des universités marocaines à partir de Twitter en utilisant les algorithmes d'apprentissage automatique. Les résultats ont révélé que 61.37% des tweets collectés en temps réel, relatifs à douze universités publiques marocaines, étaient neutres, 36.76% étaient positifs et 1.87% étaient négatifs, notamment, l'Université Mohammed V de Rabat a obtenu le plus grand nombre de tweets positifs parmi les 12 universités. D'après les résultats obtenus, on peut constater que l'Université Mohammed V de Rabat a des taux de satisfaction élevés parmi ses étudiants, en particulier parmi les étudiants sourds et malentendants. De plus, une grande proportion d'étudiants sourds et malentendants a été nominalement engagés, ce qui indique que l'université accorde une attention particulière aux besoins des étudiants ayant des besoins spécifiques. Cette satisfaction des étudiants est également en accord avec la réputation de l'université en termes de qualité de la recherche, de la formation, de la visibilité internationale et de la réputation académique, qui est soutenue par des classements internationaux prestigieux.

Dans ce mémoire, le premier chapitre s'est concentré sur les algorithmes d'apprentissage automatique et d'apprentissage profond. Il a examiné leurs avantages, leurs inconvénients et leurs applications dans différents domaines, en accordant une attention particulière à la reconnaissance des émotions faciales. Dans ce contexte, il a présenté les notions de base des émotions, les méthodes de reconnaissance des expressions faciales, et une analyse de l'état de l'art de la reconnaissance des émotions faciales dans l'enseignement supérieur. Le deuxième chapitre a exploré les concepts d'enseignement et d'apprentissage, Il a également examiné des stratégies pour une pédagogie différenciée et inclusive et il a introduit deux approches novatrices utilisant l'apprentissage profond pour reconnaître les émotions faciales et détecter l'engagement des étudiants, dans le but de favoriser une pédagogie différenciée dans des environnements inclusifs. Le troisième chapitre a souligné les bases théoriques de l'analyse de sentiment en tant que méthode potentielle pour évaluer la qualité des systèmes éducatifs. Il a détaillé les différentes tâches, niveaux et étapes du processus d'analyse de sentiment, ainsi que son application actuelle dans l'enseignement supérieur. Le quatrième chapitre a présenté une méthodologie basée sur l'apprentissage profond pour examiner les opinions des utilisateurs de Twitter concernant l'apprentissage à distance dans l'enseignement supérieur après la pandémie de COVID-19, en se concentrant sur six disciplines universitaires spécifiques (psychologie, informatique, mathématiques, physique, chimie et biologie). Ensuite a décrit en détail la mise en place d'une analyse de sentiment en temps réel des tweets portant sur les universités publiques marocaines.

L'intégration de ces deux volets permet une évaluation plus exhaustive de la qualité des services éducatifs offerts. L'évaluation formative, basée sur l'analyse des expressions faciales, mesure l'engagement, la concentration et les émotions des étudiants, fournissant des informations essentielles sur l'efficacité de l'enseignement. Elle sert d'indicateur précieux pour évaluer les méthodes pédagogiques, l'adaptation de l'apprentissage aux besoins des apprenants et l'interaction enseignant-élève. De plus, l'analyse de sentiment, en utilisant les données des réseaux sociaux concernant la réputation et l'enseignement à distance, permet de mesurer la satisfaction globale des apprenants et des autres parties prenantes, jouant un rôle essentiel dans l'évaluation de la qualité du système éducatif. Cette approche offre une vision plus complète de la qualité des services éducatifs et permet d'identifier les domaines d'amélioration pour favoriser un système éducatif de qualité. En prenant en compte l'adaptation de l'apprentissage aux apprenants et l'opinion des utilisateurs, les décideurs, les éducateurs et les institutions peuvent prendre des mesures pour améliorer les méthodes pédagogiques, les ressources et les interactions, contribuant ainsi à une amélioration globale du système éducatif.

Cette thèse a apporté une contribution significative à la recherche sur la reconnaissance des émotions et l'analyse du sentiment dans le domaine de l'éducation en utilisant des algorithmes d'apprentissage automatique, d'apprentissage profond et le Big Data. Ces avancées peuvent contribuer à l'amélioration globale de l'expérience d'apprentissage dans l'enseignement marocain, en créant un environnement inclusif et équitable pour tous les apprenants, y compris ceux ayant des besoins spécifiques. Par conséquent, l'adoption de pratiques pédagogiques plus inclusives pour améliorer la qualité de l'enseignement supérieur au Maroc contribue à l'objectif de développement durable 4 de l'ONU, qui vise à offrir une éducation de qualité inclusive et équitable pour tous. Ainsi, l'amélioration du classement national et international des universités marocaines pourrait favoriser la croissance et le développement durable du pays.

Ces travaux peuvent être considérés comme une base pour les travaux futurs dans cette direction de recherche. Ils ouvrent de vastes perspectives passionnantes et prometteuses pour une multitude d'applications. En utilisant une base de données d'expressions réelles des étudiants, nous pouvons envisager d'étendre la reconnaissance des émotions faciales à d'autres émotions telles que la fierté, la honte et bien d'autres encore. Pour ce faire, il sera important de tenir compte de paramètres tels que l'orientation du visage, la couleur de la peau et l'angle de la caméra afin de garantir une reconnaissance précise et fiable des émotions. En ajoutant le suivi du regard et l'orientation de la tête à la reconnaissance faciale des émotions pour calculer un indice de concentration, il devient possible de mesurer et

d'évaluer l'engagement des étudiants dans différentes situations. Cette capacité pourrait être utilisée dans divers contextes, tels que l'apprentissage en ligne, où il est essentiel de maintenir l'attention et l'intérêt des apprenants.

En dehors du domaine de l'éducation, la reconnaissance faciale des émotions offre de nombreuses perspectives d'application. Dans le domaine de la santé, par exemple, elle pourrait être utilisée pour évaluer l'état émotionnel des patients, notamment pour détecter les signes de détresse, de dépression ou d'anxiété. Cela pourrait contribuer à améliorer les protocoles de diagnostic et de traitement, en offrant des informations supplémentaires sur l'état émotionnel des patients. Les personnes atteintes de troubles émotionnels, tels que l'autisme ou la dépression, pourraient bénéficier de la reconnaissance faciale des émotions comme outil d'assistance. En outre, elle peut aider les psychologues à obtenir une compréhension plus approfondie de leur état émotionnel et à prendre des décisions éclairées concernant le traitement et à suivre l'évolution des émotions de leurs patients au fil du temps afin d'ajuster les interventions thérapeutiques en conséquence. La reconnaissance faciale des émotions peut également être utilisée comme une aide à la thérapie. Par exemple, elle peut être intégrée à des outils de réalité virtuelle pour créer des environnements thérapeutiques interactifs où les patients peuvent interagir avec des avatars virtuels qui réagissent en fonction de leurs expressions faciales. Cela permet aux patients d'explorer et de gérer leurs émotions dans un environnement contrôlé et sécurisé.

Dans le domaine de la robotique, la reconnaissance faciale des émotions permettrait aux robots d'interagir de manière plus intuitive et empathique avec les humains, que ce soit dans un contexte domestique, social ou professionnel. La sécurité est un autre domaine où la reconnaissance faciale des émotions pourrait apporter des avantages significatifs. En détectant les émotions suspectes ou anormales dans les environnements publics, cette technologie pourrait contribuer à améliorer les systèmes de surveillance et de sécurité. Par exemple, elle pourrait être utilisée pour détecter les comportements agressifs ou les situations de stress, permettant ainsi de prévenir ou de réagir rapidement face à des incidents potentiellement dangereux. Un autre domaine où la reconnaissance faciale des émotions peut être appliquée est celui de la sécurité routière. La fatigue du conducteur (driver fatigue) est un facteur de risque majeur dans les accidents de la route, et la capacité de détecter les émotions du conducteur peut contribuer à améliorer la sécurité. En utilisant des caméras ou des capteurs embarqués dans les véhicules, la reconnaissance faciale des émotions peut surveiller les expressions faciales du conducteur pour détecter des signes de fatigue, de distraction ou de frustration. Par exemple, si le système détecte que le conducteur a les paupières lourdes ou les yeux fermés pendant une période prolongée, il peut déclencher une

alerte sonore ou visuelle pour avertir le conducteur du risque de s'endormir au volant. De plus, en détectant des signes de colère ou d'agitation, le système peut ajuster automatiquement les paramètres du véhicule, tels que la musique, la climatisation ou même le niveau d'assistance à la conduite, pour aider à apaiser le conducteur et à maintenir un environnement de conduite plus sûr.

Bibliographie

- [1] K. Vernooij, “Omgaan met verschillen nader bekeken. Wat werkt?”, https://www.mbotaalacademie.nl/upload/files/Artikel_Omgaan_met_verschillen.pdf [Consulté le: 20-mars-2023].
- [2] P. S. George, “A Rationale for Differentiating Instruction in the Regular Classroom”, *Theory Into Practice*, vol. 44, n° 3, 2005, p. 185–193.
- [3] D. D. Ready, et D. L. Wright, “Accuracy and inaccuracy in teachers’ perceptions of young children’s cognitive abilities: the role of child background and classroom context”, *American Educational Research Journal*, vol. 48, 2011, p. 335–360.
- [4] P. Ekman, et W. Friesen, “Constants across cultures in the face and emotion”, *Journal of Personality and Social Psychology*, vol. 17, n° 2, 1971.
- [5] K. Caraway, C. M. Tucker, W. M. Reinke, et C. Hall, “Self-efficacy, goal orientation, and fear of failure as predictors of school engagement with high school students”, *Psychology in the Schools*, vol. 40, 2003, p. 417–427.
- [6] C. Valiente, J. Swanson, et N. Eisenberg, “Linking Students’ Emotions and Academic Achievement: When and Why Emotions Matter”, *Child Dev Perspect.*, vol. 6, n° 2, 2012, p. 129–135.
- [7] L. Hao et C. Yunhuo, “Effects of Academic Emotions on Achievement Among Mainland Chinese Students: A Meta-Analysis”, *Social Behavior and Personality: an international journal*, vol. 44, 2016, p. 1541–1553.
- [8] K. L. Webber, R. B. Krylow, et Q. Zhang, “Does involvement really matter? Indicators of college student success and satisfaction”, *Journal of College Student Development*, vol. 54, n° 6, 2013, p. 591–611.
- [9] UNESCO, “Guide rapide des indicateurs de l’éducation pour l’ODD 4”, https://unesdoc.unesco.org/ark:/48223/pf0000265396_fre, [Consulté le: 20-mars-2023].
- [10] E. Sreehari, et G. S. Ghantasala, “Climate Changes Prediction Using Simple Linear Regression” *Journal of Computational and Theoretical Nanoscience*, vol. 16, p. 655–658, 2019.

-
- [11] R. Karim, M. K. Alam and M. R. Hossain, “Stock Market Analysis Using Linear Regression and Decision Tree Regression,” *1st International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, p. 1–6, Sana’a, Yemen, 2021.
- [12] A. Loganathan, S. Palraj, et V. Deneshkumar, “Estimation of Air Quality Index Using Multiple Linear Regression”, *Applied Ecology and Environmental Sciences*, vol. 10, 2022, p. 717–722.
- [13] Y. Zou, “Forecast of COVID-19 Epidemic Trend Based on Multiple Linear Regression Model”, *FCIS*, vol. 1, n° 3, 2022, p. 65–67.
- [14] D. Haryadi et D. Umi Atmaja, “Identify cholesterol disease risk levels using multiple linear regression algorithms”, *jitk*, vol. 8, n° 1, 2022, p. 10–17.
- [15] K. Berezka, O. Kovalchuk, S. Banakh, S. Zlyvko, R. Hrechaniuk, “A Binary Logistic Regression Model for Support Decision Making in Criminal Justice”, *Folia Oeconomica Stetinensia*, vol. 22, n° 1, 2022, p. 1–17.
- [16] L. Biwei, “Factors Affecting the Punctuality of Logistics Services Using Binary Logistic Regression” *BCP Business & Management*, vol. 34, 2022, p. 704–712.
- [17] H. Zhang, H. Xing, D. Yao, et al., “The multiple logistic regression recognition model for mine water inrush source based on cluster analysis”, *Environ Earth Sci*, vol. 78, 2019, p. 612.
- [18] S. M. C. Vidal E Silva et al., “Factors associated with preventable infant death: a multiple logistic regression”, *Rev Saude Publica.*, 2018, p. 52–32.
- [19] A. Risfian, S. Nurrohmah, et A. Sarini. “Multinomial Logistic Regression and Spline Regression for Credit Risk Modelling”, *Journal of Physics: Conference Series.*, vol. 1108, 2018, p. 012019.
- [20] H. Soufi, et A. Esfahanipour, “Modeling Financial Resilience in Commercial Banks using Multinomial Logistic Regression”, 2021.
- [21] S. Santoshkumar, “Multinomial Logistic Regression Model to Identify Factors Associated with Food Insecurity in Rural Households in Nepal”, *Nepalese Journal of Statistics.*, vol. 4., 2020, p. 17–32.
- [22] V. Vapnik, et A. Lerner, “A Pattern Recognition Using Generalized Portrait”, *Automation and Remote Control*, vol. 24, 1963, p. 6.
- [23] B. T. Smith et B. Eng, “Lagrange Multipliers Tutorial in the Context of Support Vector Machines”, *Memorial University of Newfoundland St. John’s*, vol. 16, 2004, p. 17.
- [24] J. Mercer, “Functions of Positive and Negative Type, and their Connection with the Theory of Integral Equations”, *Philosophical Transactions of the Royal Society of London Series A*, vol. 209, 1909, p. 415–446.

- [25] T. Joachims, “Text Categorization with Support Vector Machines,” dans *Proceedings of the European Conference Machine Learning (ECML'98)*, 1998.
- [26] R. Romero, E. L. Iglesias, et L. Borrajo, “A linear-RBF multikernel SVM to classify big text corpora”, *BioMed research international*, 2015, p. 878291.
- [27] L. Hoang, L. Thai, T. Son, T. Hai, et T. Nguyen, “Image Classification using Support Vector Machine and Artificial Neural Network”, *International Journal of Information Technology and Computer Science*, vol. 4., 2012.
- [28] K. Aida-zade, A. Xocayev, et S. Rustamov, “Speech recognition using Support Vector Machines,” *IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)*, p. 1–4, 2016.
- [29] H. Supreeth, “An Adaptive SVM Technique for Object Tracking”, *International Journal of Pure and Applied Mathematics*, vol. 118, 2018.
- [30] S. Bhatia, et J. Malhotra, “Naïve Bayes Classifier for Predicting the Novel Coronavirus,” *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, p. 880–883, 2021.
- [31] W. Wei, S. Visweswaran, et G. F. Cooper, “The application of naive Bayes model averaging to predict Alzheimer’s disease from genome-wide data”, *Journal of the American Medical Informatics Association : JAMIA*, vol. 18, n° 4, 2011, p. 370–375.
- [32] V. Sai Krishna Reddy et al., “Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers”, *J. Phys.: Conf. Ser.*, vol. 2161, 2022, p. 012015.
- [33] M. Abbas et al., “Multinomial Naive Bayes Classification Model for Sentiment Analysis”, *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, n° 3, 2019.
- [34] L. Hong et al. “Decision tree model for predicting in-hospital cardiac arrest among patients admitted with acute coronary syndrome”, *Clinical cardiology*, vol. 42, n° 11, 2019, p. 1087–1093.
- [35] P. Jayanthi et al., “Decision Trees for Predicting Brain Tumors: A Case Study in Health Care,” *Third International Congress on Information and Communication Technology: ICICT 2018*, p. 921–928, London, Springer Singapore, 2019.
- [36] I. Lasri, A. RiadSolh, et M. ElBelkacemi, “Toward an Effective Analysis of COVID-19 Moroccan Business Survey Data using Machine Learning Techniques,” *13th International Conference on Machine Learning and Computing ICMLC*, p. 50–58, China, 2021.
- [37] L. Breiman, “Random Forests”, *Machine Learning*, vol.45, n° 1, 2001, p. 5–32.

-
- [38] A. Sarica, A. Cerasa, et A. Quattrone, “Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer’s Disease: A Systematic Review”, *Front. Aging Neurosci.*, vol. 9, 2017, p. 329.
- [39] X. Gao, J. Wen, et C. Zhang, “An Improved Random Forest Algorithm for Predicting Employee Turnover”, *Mathematical Problems in Engineering*, 2019, p. 1–12.
- [40] C. G. Villegas-Mier, J. Rodriguez-Resendiz, J. M. Álvarez-Alvarado, H. Jiménez-Hernández, et Á. Odry, “Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours”, *Micromachines*, vol. 13, n° 9, 2022, p. 1406.
- [41] Z. Fang et al., “Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study”, *BMJ Open*, vol 12, 2022, p. e056685.
- [42] J. Yang, et J. Guan, “A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm”, *Information*, vol. 13, 2022, p. 475.
- [43] B. Noh et al., “XGBoost based machine learning approach to predict the risk of fall in older adults using gait outcomes”, *Scientific reports*, vol. 11, n° 1, 2021, p. 12183.
- [44] L. Gan, “XGBoost-Based E-Commerce Customer Loss Prediction”, *Computational intelligence and neuroscience*, vol. 2022, p. 1858300.
- [45] K. K. Mohbey, M. Z. Khan, et A. Indian, “Credit Card Fraud Prediction Using XGBoost: An Ensemble Learning Approach”, *Int. J. Inf. Retr. Res.*, vol. 12, n° 2, 2022, p. 1–17.
- [46] W. S. McCulloch et W. Pitts, “A logical calculus of the ideas immanent in nervous activity”. *The bulletin of mathematical biophysics*, vol. 5, n° 4, 1943, p. 115–133.
- [47] F. Rosenblatt, “The Perceptron A Probabilistic Model for Information Storage and Organization in the Brain”, *Psychological Review*, vol. 65, 1958, p. 386.
- [48] H. J. Kelley, “Gradient theory of optimal flight paths”, *Ars Journal*, vol. 30, n° 10, 1960, p. 947–954.
- [49] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, et J. Schmidhuber, “LSTM: A Search Space Odyssey”, *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, n° 10, 2017, p. 2222–2232.
- [50] F. A. Gers, J. Schmidhuber et F. Cummins, “Learning to forget: continual prediction with LSTM,” *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, vol. 2, p. 850–855, Edinburgh, UK, 1999.
- [51] F. Rodrigues, I. Markou, F. C. Pereira, “Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach”, *Information Fusion*, vol. 49, 2019, p. 120–129.
- [52] M. Z. Uddin, “A wearable sensor-based activity prediction system to facilitate edge computing in smart healthcare system”, *Journal of Parallel and Distributed Computing*, vol. 123, 2019, p. 46–53.

- [53] S. Ryu, S. Kim, J. Choi, H. Yu, et G. G. Lee, “Neural sentence embedding using only in-domain sentences for out-of-domain sentence detection in dialog systems”, *Pattern Recognition Letters*, vol. 88, 2017, p. 26–32.
- [54] Y. Sun, Z. Ji, L. Lin, D. Tang, et X. Wang, “Entity disambiguation with decomposable neural networks”, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, n° 5, 2017, p. e1215.
- [55] O. Melamud, J. Goldberger, I. Dagan, S. Riezler, et Y. Goldberg, “context2vec: Learning Generic Context Embedding with Bidirectional LSTM,” dans *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, Association for Computational Linguistics (ACL)*, p. 51–61 Berlin, Germany, August 2016.
- [56] M. Arbane, R. Benlamri, Y. Brik, et A. D. Alahmar, “Social media-based COVID-19 sentiment classification model using Bi-LSTM”, *Expert Systems with Applications*, vol. 212, 2023, p. 118710.
- [57] M. Mujahid et al., “Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network”, *Diagnostics*, vol. 12, 2022, p. 1280.
- [58] K. Yeon-Jae et al., “Deep convolutional neural network for classification of thyroid nodules on ultrasound: Comparison of the diagnostic performance with that of radiologists”, *European Journal of Radiology*, vol. 152, 2022, p. 110335.
- [59] M. K. Bohmrah, et H. Kaur, “Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model”, *Global Transitions Proceedings*, vol 2, n° 2, 2021, p. 476–483.
- [60] K. Simonyan, et A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, 2015, arXiv:1409.1556.
- [61] C. Szegedy et al., “Going deeper with convolutions,” dans *Proceedings of 2015 IEEE conference on computer vision and pattern recognition (cvpr)*, 2015.
- [62] F. Chollet, “Xception: Deep learning with depthwise separable convolutions”, 2017, arXiv:1610.02357.
- [63] K. He, X. Zhang, S. Ren, et J. Sun, “Deep residual learning for image recognition,” dans *Proceedings of 2016 IEEE conference on computer vision and pattern recognition (cvpr)*, p.770–778, Las Vegas NV, USA: IEEE, 2016.
- [64] A. Howard et al., “Mobilenets: Efficient convolutional neural networks for mobile vision applications”, 2017, arXiv:1704.04861.
- [65] G. Huang, Z. Liu, Z., L. Van Der Maaten, et K. Weinberger, “Densely connected convolutional networks,” dans *Proceedings of 2017 IEEE conference on computer vision and pattern recognition (cvpr)*, pp. 2261–2269, Honolulu, HI, USA: IEEE, 2017.

- [66] K. R. Scherer, “Emotions are emergent processes: they require a dynamic computational architecture”, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 364, n° 1535, 2009, p. 3459–3474.
- [67] P. Ekman, R. J. Davidson, et W. V. Friesen, “The Duchenne smile: Emotional expression and brain physiology: II”, *Journal of Personality and Social Psychology*, vol. 58, n° 2, 1990, p. 342–353.
- [68] P. Viola et M. Jones, “Rapid object detection using a boosted cascade of simple features,” dans *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, p. I–I, Kauai, HI, USA, 2001.
- [69] X. Xie, G. Cheng, J. Wang, X. Yao, et J. Han, “Oriented R-CNN for object detection,” dans *Proceedings of the IEEE/CVF International Conference on Computer Vision*, p. 3520–3529, 2021.
- [70] H. Nakahara, H. Yonekawa et S. Sato, “An object detector based on multiscale sliding window search using a fully pipelined binarized CNN on an FPGA,” *2017 International Conference on Field Programmable Technology (ICFPT)*, p. 168–175, Melbourne, VIC, Australia, 2017.
- [71] T. Haryanto, H. Suhartanto, A. M. Arymurthy, et K. Kusmardi, “Conditional sliding windows: An approach for handling data limitation in colorectal histopathology image classification”, *Informatics in Medicine Unlocked*, vol 23, 2021, p. 100565.
- [72] W. Liu, D. Anguelov, D., Erhan, C. Szegedy, S., Reed, C. Y. Fu, et A. C. Berg, “Ssd: Single shot multibox detector,” dans *Computer Vision—ECCV 2016: 14th European Conference*, p. 21–37, Amsterdam, The Netherlands, 11–14 octobre 2016, Springer International Publishing.
- [73] A. L. Yuille, “Deformable templates for face recognition, *Journal of cognitive neuroscience*, vol. 3, n° 1, 1991, p. 59–70.
- [74] M. Çarıkçı, et F. Özen, “A Face Recognition System Based on Eigenfaces Method”, *Procedia Technology*, vol. 1, 2012, p. 118–123.
- [75] B. Suman et R. Kumar, “Face recognition by linear discriminant analysis”, *International Journal of Communication Network Security*, vol. 2, 2013, p. 31–35.
- [76] P. Corcoran et C. Iancu, “Hidden Markov Models in Automatic Face Recognition - A Review”, *Reviews, Refinements and New Ideas in Face Recognition*, 2011.
- [77] T. F. Cootes, C. J. Taylor, D. H. Cooper, et J. Graham, “Active shape models - their training and application”. *Computer Vision and Image Understanding*, vol. 61, 1995, p. 38–59.
- [78] T. F. Cootes, G. J. Edwards, et C. J. Taylor, “Active appearance models”, dans *Computer Vision — ECCV’98. ECCV 1998. Lecture Notes in Computer Science*, H. Burkhardt, B. Neumann, Éd. Springer, Berlin, Heidelberg, vol. 1407, 1998.

- [79] W. Burger, et M. J. Burge, “Scale-Invariant Feature Transform (SIFT)”, dans *Digital Image Processing. Texts in Computer Science*, Springer, London, 2016.
- [80] W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, et H. Lin, “Facial expression recognition using radial encoding of local Gabor features and classifier synthesis”, *Pattern Recognition*, vol. 45, n° 1, 2012, p. 80–91.
- [81] J. Chen, Z. Chen, Z. Chi, et H. Fu, “Facial expression recognition based on facial components detection and hog features,” dans *International Workshops on Electrical and Computer Engineering Subfields*, p. 884–888, 2014.
- [82] C. Shan, S. Gong, et P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study”, *Image and Vision Computing*, vol. 27, n° 6, 2009, p. 803–816.
- [83] S. L. Happy, A. George, et A. Routray, “A real time facial expression classification system using Local Binary Patterns,” dans *Proceedings of the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, décembre 2012.
- [84] W. Wei, Q. Jia, et G. Chen, “Real-time facial expression recognition for affective computing based on Kinect,” dans *Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, juin 2016.
- [85] R. Vedantham, et E. S. Reddy, “A robust feature extraction with optimized DBN-SMO for facial expression recognition”, *Multimedia Tools and Applications*, vol. 79, 2020, p. 21487–21512.
- [86] L. Zhao, Z. Wang, X. Wang, et Q. Liu, “Driver drowsiness detection using facial dynamic fusion information and a DBN”, *IET Intelligent Transport Systems*, vol. 12, n° 2, 2018, p. 127–133.
- [87] S. Jaiswal, et M. Valstar, “Deep learning the dynamic appearance and shape of facial action units,” dans *Proc. IEEE Int. Winter Conf. Applications of Computer Vision (WACV)*, p. 1–8, Lake Placid, NY, USA, 2016.
- [88] G. Hegde, M. Seetha, et N. Hegde, “Kernel locality preserving symmetrical weighted fisher discriminant analysis based subspace approach for expression recognition”, *Engineering Science and Technology, an International Journal*, vol. 19, n° 3, 2016, p. 1321–1333.
- [89] H. Wang et al., “Emotion detection via discriminative kernel method,” dans *Proceedings of the 3rd international conference on pervasive technologies related to assistive environments*, p. 1–7, juin 2010.
- [90] H. Zhao, Z. Wang, et J. Men, “Facial complex expression recognition based on fuzzy kernel clustering and support vector machines,” dans *Third international conference on natural computation (ICNC 2007)*, vol. 1, p. 562–566, Août 2007.

-
- [91] K. Anastasios, et I. F. Dimitrios, “Image processing and machine learning techniques for facial expression recognition”, dans *Handbook of research on advanced techniques in diagnostics imaging and biomedical applications*, T. P. Exarchos, A. Papadopoulos, D. I. Fotiadis, Éd. IGI Global, p. 247–262, 2009.
- [92] G. Yijun, “Facial expression recognition using convolutional neural network,” dans *Proc. Int. Conf. Vision, Image and Signal Processing (ICVISIP)*, Las Vegas, NV, USA, Août 2018.
- [93] H. Yang, et L. Yin, “CNN based 3D facial expression recognition using masking and landmark features,” dans *Proc. Int. Conf. Affective Computing and Intelligent Interaction (ACII)*, p. 556–560, San Antonio, TX, 2017.
- [94] J. Donahue et al., “Long-term recurrent convolutional networks for visual recognition and description”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, 2017, p. 677–691.
- [95] Y. Wu, et W. Qiu, “Facial expression recognition based on improved deep belief networks,” dans *AIP Conference Proceedings*, vol. 1864, n° 1, p. 020130, AIP Publishing LLC, Août 2017.
- [96] T. Teixeira T, É. Granger, et A. L. Koerich, “Continuous Emotion Recognition with Spatiotemporal Convolutional Neural Networks”, *Applied Sciences*, vol. 11, n° 24, 2021, p. 11738.
- [97] E. Tran, M. B. Mayhew, H. Kim, P. Karande et A. D. Kaplan, “Facial Expression Recognition Using a Large Out-of-Context Dataset,” *2018 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, p. 52–59, Lake Tahoe, NV, USA, 2018.
- [98] J. Tan, J. Mao, Y. Jiang, et M. Gao, “The Influence of Academic Emotions on Learning Effects: A Systematic Review”, *International journal of environmental research and public health*, vol. 18, n° 18, 2021, p. 9678.
- [99] E. Elshareif, et E. A. Mohamed, “The Effects of E-Learning on Students' Motivation to Learn in Higher Education”, *Online Learning*, vol. 25, n° 3, 2021, p. 128–143.
- [100] C. Li, M. Yang, Y. Zhang, et K. W. Lai, “An Intelligent Mental Health Identification Method for College Students: A Mixed-Method Study”, *Int. J. Environ. Res. Public Health*, vol. 19, 2022, p. 14976.
- [101] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, et S. A. Velastin, “Facial expression recognition of instructor using deep features and extreme learning machine”, *Computational Intelligence and Neuroscience*, 2021, p. 1–17.
- [102] I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Facial emotion recognition of students using convolutional neural network,” dans *2019 third international conference on intelligent computing in data sciences (ICDS)*, p. 1–6, IEEE, octobre 2019.

- [103] I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning”, *Educ Inf Technol*, 2022.
- [104] M. Altet, “Les pédagogies de l’apprentissage”, Éd. PUF, Paris, 1997.
- [105] A. Prost, “Éloge des pédagogues”, Éd. Éditions du Seuil, Paris, 1985.
- [106] Y. Chevallard, “The Didactic Transposition—Of Learned Knowledge to Knowledge Taught”, dans *La pensée wild*, Éd. Grenoble, Second Enlarged Edition, 1991.
- [107] J. Houssaye, “Le triangle pédagogique”, dans *Berne*, P. Lang, 1988.
- [108] J. C Bertin, “L’ergonomie didactique : une approche de la recherche dans le domaine de l’apprentissage médiatisé par ordinateur”, dans *L’anglais de spécialité en France. Mélanges en l’honneur de Michel Perrin*, M. Mémet, et M. Petit, Éd. GERAS, Bordeaux, p. 237–254, 2001.
- [109] D. Paquelin, “Analyse d’applications multimédias pour un usage pédagogique. À la recherche de l’intentionnalité partagée”, *Alsic*, vol. 5, n° 1, 2002.
- [110] M. Mastafi, “Obstacles à l’intégration des technologies de l’information et de la communication (TIC) dans le système éducatif marocain Barriers to integration of information and communication technology (ICT) in Moroccan education system”, 2014.
- [111] E. E. Villiot-Leclercq, “Modèle de soutien à l’élaboration et à la réutilisation de scénarios pédagogiques”, Université Joseph-Fourier-Grenoble I, 2007.
- [112] J. Watson, “Le béhaviorisme”, Éd. Cepi, Paris, 1972.
- [113] J. Piaget, “L’équilibration des structures cognitives”, Éd. PUF, Paris, 1975.
- [114] L. S. Vygotsky, “Mind in society : The development of higher psychological processes”, Éd. Harvard university press, 1980.
- [115] J. S. Bruner, et Y. Bonin, “L’éducation, entrée dans la culture : les problèmes de l’école à la lumière de la psychologie culturelle”, Éd. Retz, 1996.
- [116] G. Siemens, “Connectivism : A learning theory for the digital age”, *International journal of instructional technology and distance learning*, vol. 2, n° 1, 2005, p. 3–10.
- [117] “La pédagogie différenciée : enfermement ou ouverture ?”, <http://www.meirieu.com/ARTICLES/pedadif.pdf> [Consulté le: 26-mars-2023].
- [118] R. Dottrens, “L’Enseignement individualisé”, Éd. Neuchâtel et Paris : Delachaux et Niestlé.
- [119] J. Gal, et C. Freinet, “La construction d’une pédagogie populaire et d’un mouvement d’éducateurs engagé”, *Le Télémaque*, 1996, p. 111–120.
- [120] B. S. Bloom, “New views of the learner: Implications for instruction and curriculum. Childhood Education”, vol. 56, n° 1, 1979, p. 4–11.
- [121] C. Tomlinson, et A. Demirsky, “Leadership for differentiating schools and classrooms”, dans *Alexandria, VA : ASCD*, 2000.

- [122] <https://albayane.press.ma/reforme-du-prescolaire-et-si-la-pedagogie-differenciee-etait-lissue.html> [Consulté le: 26-mars-2023].
- [123] H. Przesmycki et al. “La pédagogie différenciée”, Éd. Hachette éducation, Paris, 2004.
- [124] C. R. Rogers, “Client-centered therapy; its current practice, implications, and theory”, Éd. Houghton Mifflin, 1951.
- [125] U. Bronfenbrenner, “The experimental ecology of education”, *Educational Researcher*, vol. 5, n° 5, 1976, p. 5–15.
- [126] “Convention des Nations Unies relative au droit des personnes handicapées”, <https://www.un.org/disabilities/documents/convention/convoptprot-f.pdf> [Consulté le: 28-mars-2023].
- [127] “Global Education Monitoring Report: Inclusion and Education UNESCO 2020”, <https://www.unesco.org/gem-report/en/inclusion> [Consulté le: 28-mars-2023].
- [128] “Évaluation du modèle d’éducation des enfants en situation de handicap au Maroc : vers une éducation inclusive Conseil Supérieur de l’Éducation, de la Formation et de la Recherche Scientifique 2019”, <https://www.csefrs.ma/wp-content/uploads/2019/05/Rapport-e%CC%81ducation-inclusive-Fr.pdf> [Consulté le: 28-mars-2023].
- [129] P. Vuilleumier, “How brains beware: neural mechanisms of emotional attention”, *Trends in cognitive sciences*, vol. 9, no 12, 2005, p. 585–594.
- [130] E. A. Phelps, “Human emotion and memory: interactions of the amygdala and hippocampal complex”, *Curr. Opin. Neurobiol.*, vol. 14, 2004, p. 198–202.
- [131] A. M. Isen, K. A. Daubman, et G. P. Nowicki, “Positive affect facilitates creative problem solving”, *J. Pers. Soc. Psychol.*, vol. 52, 1987, p. 1122–1131.
- [132] N. Jung, C. Wranke, K. Hamburger, et M. Knauff, “How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety”, *Front. Psychol.*, 2014, p. 5–570.
- [133] H. T. Schupp, J. Stockburger, M. Codispoti, M. Junghöfer, A. I. Weike, et A. O. Hamm, “Selective visual attention to emotion”, *J. Neurosci.*, vol. 27, 2007, p. 1082–1089.
- [134] S. D’Mello, B. Lehman, R. Pekrun, et A. Graesser, “Confusion can be beneficial for learning”, *Learn. Instr.*, vol. 29, 2014, p.153–170.
- [135] C. Tang, P. Xu, Z. Luo, G. Zhao, et T. Zou, “Automatic Facial Expression Analysis of Students in Teaching Environments,” dans *Biometric Recognition*, J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, vol. 9428, 2015, p. 439–447.

- [136] A. Savva, V. Stylianou, K. Kyriacou, et F. Domenach, “Recognizing student facial expressions: A web application,” dans *2018 IEEE Global Engineering Education Conference (EDUCON)*, Tenerife, 2018, p.1459–1462.
- [137] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions,” *IEEE Transactions on Affective Computing*, vol. 5, n° 1, p. 86–98, janvier 2014.
- [138] N. Bosch et al., “Automatic Detection of Learning-Centered Affective States in the Wild,” dans *Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI '15*, p. 379–388, Atlanta, Georgia, USA, 2015.
- [139] L. B. Krithika et G. G. L. Priya, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric”, *Procedia Computer Science*, vol. 85, 2016, p. 767–776.
- [140] U. Ayvaz, H. Gürüler, et M. O. Devrim, “Use of facial emotion recognition in e-learning systems”, *Information Technologies and Learning Tools*, vol. 60, n° 4, 2017, p. 95,
- [141] Y. Kim, T. Soyata, et R. F. Behnagh, “Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education”, *IEEE Access*, vol. 6, 2018, p. 5308–5331.
- [142] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, et A. Elchouemi, “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment”, *Procedia Computer Science*, vol. 125, 2018, p. 2–10.
- [143] G. Bradski, “The OpenCV Library. Dr. Dobb’s Journal of Software Tools”, vol. 12, 2000, p. 122–125.
- [144] Y. Freund et R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, *Journal of Computer and System Sciences*, vol. 55, n° 1, 1997, p. 119–139.
- [145] F. Chollet, “Keras: the python deep learning library”, 2015, <https://keras.io>. [Consulté le: 20-mar-2019].
- [146] H. Robbins, et S. Monro, “A stochastic approximation method”, *Annals of Mathematical Statistics*, vol. 22, n° 3, 1951, p. 400–407.
- [147] I. J. Goodfellow et al., “Challenges in Representation Learning: A report on three machine learning contests”, juillet 2013, arXiv:1307.0414 [cs, stat].
- [148] H. Ellaban, et E. Elsaed, “A real-time system for facial expression recognition using support vector machines and k-nearest neighbor classifier”, *International Journal of Computer Applications*, vol. 159, n° 8, 2017, p. 23–29.

-
- [149] C. Thomas, et D. Jayagopi, “Predicting student engagement in classrooms using facial behavioral cues,” dans *Proceedings of the 1st ACM sigchi international workshop on multimodal interaction for education (mie)*, p. 33–40, Glasgow Scotland, UK: ACM, 2017.
- [150] S. Aslan et al., “Investigating the impact of a real-time, multi-modal student engagement analytics technology in authentic classrooms,” dans *Proceedings of the 2019 conference on human factors in computing systems (chi)*, pp. 1–12, Glasgow Scotland, UK: ACM, 2019.
- [151] C. Lee, C. Shih, W. Lai, et P. Lin, “An improved boosting algorithm and its application to facial emotion recognition”, *Journal of Ambient Intelligence and Humanized Computing*, vol. 3, n° 1, 2012, p. 11–17.
- [152] C. Liew, et T. Yairi, “Facial expression recognition and analysis: A comparison study of feature descriptors”, *IPSP Transactions on Computer Vision and Applications*, vol. 7, 2015, p. 104–120.
- [153] S. Eng, H. H. Ali, A. Cheah, et Y. Chong, “Facial expression recognition in JAFFE and KDEF datasets using histogram of oriented gradients and support vector machine”, *IOP Conference Series: Materials Science and Engineering*, vol. 705, n° 1, 2019, p. 012031.
- [154] R. Holder, et J. Tapamo, “Improved gradient local ternary patterns for facial expression recognition”, *EURASIP Journal on Image and Video Processing*, 2017, p. 42.
- [155] X. Zhao, X. Shi, et S. Zhang, “Facial expression recognition via deep learning”, *IETE Technical Review*, vol. 32, n° 5, 2015, p. 347–355.
- [156] P. Liu, S. Han, Z. Meng, et Y. Tong, “Facial expression recognition via a boosted deep belief network,” dans *Proceedings of 2014 IEEE conference on computer vision and pattern recognition*, p. 1805–1812, Columbus, OH, USA: IEEE, 2014.
- [157] M. Sari, A. Moussaoui, et A. Hadid, “A simple yet effective convolutional neural network model to classify facial expressions,” dans *Lecture notes in networks and systems*, S. Chikhi, A. Amine, A. Chaoui, D. Saidouni, et M. Kholadi, Éd. Springer, vol. 156, p. 188–202, 2021.
- [158] D. Hamster, P., Barros, et S. Wermter, “Face expression recognition with a 2-channel convolutional neural network,” dans *Proceedings of 2015 international joint conference on neural networks (ijcnn)*, p. 1–8, Killarney, Ireland: IEEE, 2015.
- [159] N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, et M. Zareapoor, “Hybrid deep neural networks for face emotion recognition”, *Pattern Recognition Letters*, vol. 115, 2018, p. 101–106.
- [160] S. Ayouni, F. Hajjej, M. Maddeh, et S. Al-Otaibi, “A new ml-based approach to enhance student engagement in online environment”, *PLoS ONE*, vol. 16, n° 11, 2021, p. 0258788.

- [161] J. Shen, H. Yang, et J. Li, “Assessing learning engagement based on facial expression recognition in mooc’s scenario”, *Multimedia Systems*, vol. 28, 2022, p. 469–478.
- [162] Y. Nesterov, “A method of solving a convex programming problem with convergence rate $o(1/k^2)$ ”, *Soviet Mathematics Doklady*, vol. 27, n° 2, 1983, p. 372–376.
- [163] P. Sharma et al., “Student engagement detection using emotion analysis, eye tracking and head movement with machine learning,” dans *Technology and Innovation in Learning, Teaching and Education: Third International Conference, TECH-EDU 2022*, p. 52–68, Lisbon, Portugal, Cham: Springer Nature Switzerland, 2022.
- [164] M. Abadi et al., “Tensorflow: A system for large-scale machine learning,” dans *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, p. 265–283, 2016.
- [165] M. Lyons, S. Akamatsu, M. Kamachi, et J. Gyoba, “Coding facial expressions with gabor wavelets,” dans *Proceedings of 3rd IEEE international conference on automatic face and gesture recognition*, p. 200–205, Nara, Japan: IEEE, 1998.
- [166] M. Calvo, et D. Lundqvist, “Facial expressions of emotion (KDEF): Identification under different display-duration conditions”, *Behavior Research Methods*, vol. 40, n° 1, 2008, p. 109–115.
- [167] N. Qian, “On the momentum term in gradient descent learning algorithms”, *Neural Networks : The Official Journal of the International Neural Network Society*, vol. 12, n° 1, 1999, p. 145–151.
- [168] D. Kingma, et J. Ba, “Adam: a method for stochastic optimization”, 2014, arXiv:1412.6980v9.
- [169] J. Duchi, E. Hazan, et Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization”, *Journal of Machine Learning Research*, vol. 12, 2011, p. 2121–2159.
- [170] D. Zeiler, “Adadelta: an adaptive learning rate method”, 2012, arXiv:1212.5701.
- [171] C. Haelermans, M. Jacobs, R. van der Velden, L. van Vugt, et S. van Wetten, “Inequality in the Effects of Primary School Closures Due to the COVID-19 Pandemic: Evidence from the Netherlands”, *AEA Papers and Proceedings*, vol. 112, 2022, p. 303–07.
- [172] D. Papanagnou, R. Jaffe, et D. Ziring, “Highlighting a curricular need: Uncertainty, COVID-19, and health systems science”, *Health Science Reports*, vol. 4, n° 3, 2021.
- [173] J. Bengtson, “The quality of education at the beginning of the 21st Century”, dans *EFA Global Monitoring Report*, 2005.
- [174] J. Scheerens, “Perspectives on education quality education indicators and benchmarking”. *European Educational Research Journal*, vol. 3, no 1, 2004, p. 115–138.

- [175] T. Kellaghan, et V. Greaney, “Using assessment to improve the quality of education”, Paris: IIEP, UNESCO, 2001.
- [176] UNESCO, “Understanding education quality”, dans *EFA Global Monitoring Report*. 2005.
- [177] R. Hofman, W. Hofman, J. Gray, et P. Daly, “Institutional context of education systems in Europe: A cross-country comparison on quality and equity”, 2005.
- [178] B. Pang, L. Lee, et S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” dans *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, p. 79–86, Morristown, NJ, USA, Association for Computational Linguistics, juillet 2002.
- [179] B. Liu, “Sentiment Analysis and Subjectivity”, dans *Handbook of Natural Language Processing, Second Edition*, N. Indurkha et F. J. Damerau, Éd. 2010.
- [180] B. Pang et L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” dans *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL’04)*, p. 271–278, 2004.
- [181] H. Nishikawa, T. Hasegawa, Y. Matsuo, et G. Kikui, “Opinion summarization with integer linear programming formulation for sentence extraction and ordering,” dans *Coling 2010: Posters*, p. 910–918, Août 2010.
- [182] P. C. Carvalho, L. Sarmiento, M. J. Silva, et E. deOliveira, “Clues for Detecting Irony in User-Generated Contents: Oh...!! It’s “so easy” ;-),” dans *Text Sentiment Analysis (TSA’09)*, Hong Kong, China, 2009.
- [183] D. I. H. Farias, et P. Rosso, “Chapter 7 - Irony, Sarcasm, and Sentiment Analysis”, dans *Sentiment Analysis in Social Networks*, F. A. Pozzi, E. Fersini, E. Messina, B. Liu, Éd. Morgan Kaufmann, p. 113–128, 2017.
- [184] A. Reyes, P. Rosso, et T. Veale, “A multidimensional approach for detecting irony in Twitter”, *Language Resources and Evaluation*, vol. 47, n° 1, p. 239–268, 2013.
- [185] N. Jindal et B. Liu, “Opinion spam and analysis,” dans *Proceedings of the international conference on Web search and web data mining, WSDM ’08*, p. 219–230, New York, NY, USA. ACM, 2008.
- [186] Y. Dang, Y. Zhang, et H. Chen, “A lexicon-enhanced method for sentiment classification: An experiment on online product reviews”, *IEEE Intelligent Systems*, vol. 25, n° 4, p. 46–53, 2010.
- [187] J. Bollen, H. Mao, et X. Zeng, “Twitter mood predicts the stock market”, *Journal of computational science*, vol. 2, n° 1, 2011, p. 1–8.
- [188] A. Moreo, M. Romero, J. Castro, et J. M. Zurita, “Lexicon-based comments-oriented news sentiment analyzer system”, *Expert Systems with Applications*, vol. 39, n° 10, 2012, p. 9166–9180.

- [189] A. Abbasi, H. Chen, et A. Salem, “Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums”, *ACM Transactions on Information Systems (TOIS)*, vol. 26, n° 3, 2008, p. 12.
- [190] E. Boiy et M. F Moens, “A machine learning approach to sentiment analysis in multilingual web texts”, *Information retrieval*, vol. 12, n° 5, 2009, p. 526–558.
- [191] L. Garcia-Moya, H. Anaya-Sanchez, et R. Berlanga-Llavori, “Retrieving product features and opinions from customer reviews”, *IEEE Intelligent Systems*, vol. 28, n° 3, 2013, p. 19–27.
- [192] C. Quan, et F. Ren, “Unsupervised product feature extraction for feature-oriented opinion determination”, *Information Sciences*, vol. 272, 2014, p. 16–28.
- [193] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, et M. Ma, “Aspect-based opinion polling from customer reviews”, *IEEE Transactions on Affective Computing*, vol. 2, n° 1, 2011, p. 37–49.
- [194] E. Cambria, B. Schuller, Y. Xia, et C. Havasi, “New avenues in opinion mining and sentiment analysis”, *IEEE Intelligent Systems*, vol. 28, n° 2, 2013, p. 15–21.
- [195] S. Poria, E. Cambria, G. Winterstein, et G. B. Huang, “Sentic patterns: Dependency-based rules for concept-level sentiment analysis”, *Knowledge-Based Systems*, vol. 69, 2014, p. 45–63.
- [196] A. Tsai, R. T. H. Tsai, et J. Y. J. Hsu, “Building a concept-level sentiment dictionary based on commonsense knowledge”, *IEEE Intelligent Systems*, 2013, p. 1.
- [197] G. Salton et C. Buckley, “Term weighting approaches in automatic text retrieval,” *Technical report*, Cornell University, Ithaca, NY, USA, 1989.
- [198] J. Pennington, R. Socher, et C. D. Manning, “Glove: Global vectors for word representation,” dans *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, p. 1532–1543, octobre 2014.
- [199] H. S. Hota, D. K. Sharma, et N. Verma, “Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad”, *Data Science for COVID-19*, 2021, p. 275–295.
- [200] Z. Zhou, X. Zhang, et M. Sanderson, “Sentiment Analysis on Twitter through Topic-Based Lexicon Expansion”, dans *Databases Theory and Applications ADC 2014. Lecture Notes in Computer Science*, H. Wang, M. A. Sharaf, Éd. Springer, Cham, vol. 8506, 2014.
- [201] M. Ahmad, S. Aftab, et I. Ali, “Sentiment analysis of tweets using svm”, *Int. J. Comput. Appl.*, vol. 177, n° 5, 2017, p. 25–29.

- [202] M. R. Saleh, M. T., Martín-Valdivia, A. Montejo-Ráez, et L. Ureña-López, “Experiments with svm to classify opinions in different domains”, *Expert Systems with Applications*, vol. 38, n° 12, 2011, p. 14799–14804.
- [203] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, et K. Filippova, “Multi-lingual opinion mining on youtube”, *Information Processing & Management*, vol. 52, n° 1, 2016, p. 46–60.
- [204] S. Ruder, P. Ghaffari, et J. G. Breslin, “Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis”, *arXiv preprint*, 2016, arXiv:1609.02748.
- [205] O. Appel, F. Chiclana, J. Carter, et H. Fujita, “A hybrid approach to the sentiment analysis problem at the sentence level”, *Knowledge-Based Systems*, vol. 108, 2016, p. 110–124.
- [206] M. Ghiassi, J. Skinner, et D. Zimbra, “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network”, *Expert Systems with applications*, vol. 40, n° 16, 2013, p. 6266–6282.
- [207] N. A. S. Remali, M. R. Shamsuddin et S. Abdul-Rahman, “Sentiment Analysis on Online Learning for Higher Education During Covid-19,” *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, IPOH, Malaysia, 2022, p. 142–147.
- [208] R. Baragash, et H. Aldowah, “Sentiment analysis in higher education: a systematic mapping review,” dans *Journal of Physics: Conference Series*, vol. 1860, n° 1, 2021, p. 012002, IOP Publishing.
- [209] M. Oliveira, A. Penedo, et V. Pereira, “Distance education: advantages and disadvantages of the point of view of education and society”, *Dialogia*, 2018 , p. 139–152.
- [210] M. Pakdaman, M. Moghadam, et H. R. Dehghan, A. Dehghani, S. M. Namayandeh, “Evaluation of the Cost-Effectiveness of Virtual and Traditional Education Models in Higher Education: A Systematic Review”, *Health Technology Assessment in Action*, vol.3, n°1, 2019.
- [211] J. S. Barrot, I. I. Llenares, et L. S. Del Rosario, “Students’ online learning challenges during the pandemic and how they cope with them: The case of the Philippines”, *Education and information technologies*, vol. 26, n°6, 2021, p. 7321–7338.
- [212] P. Vagos, et L. Carvalhais, “Online Versus Classroom Teaching: Impact on Teacher and Student Relationship Quality and Quality of Life”, *Front. Psychol.*, vol. 13, 2022, p. 828774.
- [213] E. R. Pelikan, et al, “Distance learning in higher education during COVID-19: The role of basic psychological needs and intrinsic motivation for persistence and procrastination—a multi-country study”, *PLoS One*, vol.16, n°10, 2021, p. e0257346.
- [214] <https://higheredpartners.co.uk/why-online-learning-in-higher-education-is-here-to-stay-a-trends-assessment/> [Consulté le: 15-mars-2022].

- [215] M. Jebbour, “The unexpected transition to distance learning at Moroccan universities amid COVID-19: A qualitative study on faculty experience”, *Social Sciences & Humanities Open*, vol. 5, n°1, 2022, p. 100253.
- [216] M. Niu, “Classification of Learning Sentiments of College Students Based on Topic Discussion Texts of Online Learning Platforms”, *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, n°24, 2022, p. 42–56.
- [217] J. Sun, et X. Zhang, “Exploring Chinese college students’ emotions as they engage in online learning during a pandemic”, *Asia Pacific Journal of Education*, 2021, p. 1–12.
- [218] Y. Liang, “Intelligent Emotion Evaluation Method of Classroom Teaching Based on Expression Recognition”, *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, n°04, 2019, p. 127–141.
- [219] H. J. So, et E. Oh, “The effects of COVID-19 on online learning based on sentiment analysis”, *The Journal of Educational Technology Systems*, vol. 49, n°1, 2020, p. 5–22.
- [220] J. Kim, , Y. J. Kim, H. W. Lee, et D. Cho, “Analyzing learners’ tweets and perceptions in a massive open online course”, *Computers in Human Behavior*, vol. 51, 2015 , p. 1293–1302.
- [221] G. C. Kessler, et J. DeCarlo, “ Sentiment Analysis in Education: Opportunities, Challenges, and Future Directions”, *Educational Researcher*, vol. 46, n°9, 2017, p. 492–499.
- [222] S. Hochreiter, et J. Schmidhuber, “Long short-term memory”, *Neural Comput.*, vol. 9, n°8, 1997, p. 1735–1780.
- [223] M. Schuster, et K. K. Paliwal, “Bidirectional recurrent neural networks”, *IEEE Trans. Signal Process*, vol. 45, n°11, 1997, p. 2673–2681.
- [224] V. A. Kharde, et S. S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques”, *International Journal of Computer Applications*, vol. 139, n°11, 2016, p. 5–15.
- [225] K. I. Senadhira, R. A. H. M. Rupasingha, et B. T. G. S. Kumara, “Sentiment Analysis on Twitter Data Related to Online Learning During the Covid-19 Pandemic,” dans *International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 2022, p. 131–136.
- [226] C. Cortes, et V. Vapnik, “Support-vector networks”, *Machine Learning*, vol. 20, n°3, 1995, p. 273–297.
- [227] S. Das et al., “ Sentiment Dynamics Detection of Online Learning Impact using Hybrid Approach”, *Special Education*, vol. 1, n°43, 2022, p. 1225–1236.

- [228] S. H. Sahir et al., “Online learning sentiment analysis during the covid-19 Indonesia pandemic using twitter data,” *IOP Conference Series: Materials Science and Engineering*, vol. 1156, n°1, 2021, p. 012011.
- [229] J. Almalki, “A machine learning-based approach for sentiment analysis on distance learning from Arabic Tweets”, *PeerJ Comput Sci.*, vol.8, 2022, p. e1047.
- [230] M. Aljabri et al., “Sentiment Analysis of Arabic Tweets Regarding Distance Learning in Saudi Arabia during the COVID-19 Pandemic”, *Sensors*, vol. 21, 2021, p. 5431.
- [231] A. Althagafi, G. Althobaiti, H. Alhakami, et T. Alsubait, “Arabic Tweets Sentiment Analysis about Online Learning during COVID-19 in Saudi Arabia”, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, n°3, 2021, p. 620–625.
- [232] M. Mujahid et al., “Sentiment analysis and topic modelling on tweets about online education during COVID-19”, *Appl. Sci.*, vol. 11, no 18, 2021, p. 8438.
- [233] <https://nlp.stanford.edu/projects/glove/> [Consulté le: March 24, 2022].
- [234] A. Vaswani et al., “Attention is All you Need”, *arXiv*, 2017, p. 11.w
- [235] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, vol. 12, n° Oct, 2011, p. 2825–2830.
- [236] “Statista”, <https://www.statista.com/forecasts/1146722/twitter-users-in-the-world> [Consulté le: 20-may-2022].
- [237] H. Htet, S. S. Khaing, et Y. Y. Myint, “Tweets sentiment analysis for healthcare on big data processing and IoT architecture using maximum entropy classifier”, *Big Data Analysis and Deep Learning Applications*, vol. 744, 2019, p. 28–38.
- [238] C. R. Machuca, C. Gallardo, et R. M. Toasa, “Twitter sentiment analysis on Coronavirus: machine learning approach”, *J. Phys. Conf. Ser.*, vol. 1828, n° 1, 2021, p. 012104.
- [239] A. Riadsolh, I. Lasri, et M. ElBelkacemi, “Cloud-based sentiment analysis for measuring customer satisfaction in the Moroccan banking sector using Naive Bayes and Stanford NLP”, *J. Autom. Mob. Robot. Intell. Syst.*, vol. 14, n° 4, 2020, p. 64–71.
- [240] N. A. Vidya, M. I. Fanany, et I. Budi, “Twitter sentiment to analyze net brand reputation of mobile phone providers,” dans *3th Information Systems International Conference*, Surabaya, Indonesia, p. 519–526, 2015.
- [241] M. F. Omar, N. H. Mahathir, M. N. M. Nawati, et F. Zulhumadi, “Prototype development and pre-commercialization strategies for mobile based property analytics”, *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 13, n° 10, 2019, p. 198–204.

- [242] M. Paolanti et al., “Tourism destination management using sentiment analysis and geo-location information: a deep learning approach”, *Inf. Technol. Tour.*, vol. 23, n° 2, 2021, p. 241–264.
- [243] M. Z. Ansari, M. B. Aziz, M. O. Siddiqui, H. Mehra, et K. P. Singh, “Analysis of political sentiment orientations on Twitter,” dans *International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*, vol. 167, 2020, p. 1821–1828.
- [244] Z. E. Aydin, Z. K. Ozturk, et Z. I. E. Cicek, “Turkish sentiment analysis for open and distance education systems”, *Turk. Online J. Distance Educ.*, 2021, p. 124–138.
- [245] N. Z. Dina, R. T. Yunardi, A. A. Firdaus, et N. Juniarta, “Measuring user satisfaction of educational service applications using text mining and multicriteria decision-making approach”, *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, n° 17, 2021, p. 76–88.
- [246] F. Ouatik, M. Erritali, F. Ouatik, et M. Jourhmane, “Students’ orientation using machine learning and big data”, *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no 01, 2021, p. 111–119.
- [247] S. Ulfa, R. Bringula, C. Kurniawan, et M. Fadhli, “Student feedback on online learning by using sentiment analysis: a literature review,” dans *6th International Conference on Education and Technology (ICET)*, Malang, Indonesia, p. 53–58, 2020.
- [248] Z. K. Ozturk, Z. E. Cicek, et Z. Ergul, “Sentiment analysis: an application to Anadolu university”, *Acta Phys. Pol. A*, vol. 132, no 3, 2017, p. 753–755.
- [249] M. Zaharia et al., “Apache Spark: a unified engine for big data processing”, *Commun ACM*, vol. 59, no 11, 2016, p. 56–65.
- [250] J. Kreps, N. Narkhede, et J. Rao, “Kafka: a distributed messaging system for log processing,” dans *6th International Workshop on Networking Meets Databases (NetDB 2011)*, 2011, p. 7.
- [251] C. Gormley, et Z. Tong, “Elasticsearch: the definitive guide”, 2015.
- [252] M. S. Neethu, et R. Rajasree, “Sentiment analysis in Twitter using machine learning techniques,” dans *4th International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 2013, p. 1–5.
- [253] A. Goel, J. Gautam, et S. Kumar, “Real time sentiment analysis of tweets using Naïve Bayes,” dans *2nd International Conference on Next Generation Computing Technologies (NGCT)*, p. 257–261, 2016.
- [254] P. Chikersal, S. Poria, et E. Cambria, “SeNTU: sentiment analysis of tweets by combining a rule-based classifier with supervised learning,” dans *9th International Workshop on Semantic Evaluation (SemEval 2015)*, p. 647–651, 2015.

- [255] L. F. S. Coletta, N. F. F. Da Silva, et E. R. Hruschka, “Combining classification and clustering for tweet sentiment analysis,” dans *Brazilian Conference on Intelligent Systems*, Sao Paulo, Brazil, p. 210–215, 2014.
- [256] M. R. Huq, A. Ali, et A. Rahman, “Sentiment analysis on Twitter data using KNN and SVM”, *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no 6, 2017, p. 19–25.
- [257] N. Altrabsheh, M. Cocea, et S. Fallahkhair, “Learning sentiment from students feedback for real-time interventions in classrooms,” dans *Adaptive and Intelligent Systems*, Springer International Publishing, vol. 8779, 2014, p. 40–49.
- [258] L. Balachandran, et A. Kirupananda, “Online reviews evaluation system for higher education institution: an aspect based sentiment analysis tool,” dans *11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2017)*, p. 1–7, Malabe, Sri Lanka, 2017.
- [259] M. Alruily, et O. R. Shahin, “Sentiment analysis of Twitter data for Saudi universities”, *Int. J. Mach. Learn. Comput.*, vol. 10, no 1, 2021, p. 18–24.
- [260] H. AL-Rubaiee, R. Qiu, K. Alomar, et D. Li, “Sentiment analysis of Arabic tweets in e-learning”, *J. Comput. Sci.*, vol. 12, no 11, 2016, p. 553–563.
- [261] “Twint”, <https://github.com/twintproject/twint> [Consulté le: 14-avril-2022].
- [262] S. Bird, E. Klein, et E. Loper, “Natural language processing with Python: analyzing text with the natural language toolkit”, *OReilly Media Inc*, 2009.
- [263] S. Loria, “Textblob Documentation. Release 0.15, 2”, 2018.
- [264] P. Hunt, M. Konar, F. P. Junqueira, et B. Reed, “ZooKeeper: wait-free coordination for internet-scale systems”, 2010, p. 14.
- [265] “Times Higher Education World University Rankings 2019”, <https://www.timeshighereducation.com/world-university-rankings/2019/world-ranking> [Consulté le: 25-may-2022].
- [266] “Center for World University Rankings (CWUR) 2022-23 Edition”, <https://cwur.org/2022-23.php> [Consulté le: 25-may-2022].
- [267] “CWTS Leiden Ranking 2022”, <https://www.leidenranking.com/ranking/2022/list> [Consulté le: 25-may-2022].

Liste des publications

Articles principaux

- I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Facial emotion recognition of deaf and hard-of-hearing students for engagement detection using deep learning”, *Education and Information Technologies*, vol. 28, n° 11, 2022, p. 1–24.
- I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Real-time Twitter Sentiment Analysis for Moroccan Universities using Machine Learning and Big Data Technologies”, *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, n° 5, 2023, p. 42.
- I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Facial emotion recognition of students using convolutional neural network,” dans *2019 third international conference on intelligent computing in data sciences (ICDS)*, p. 1–6, IEEE, octobre 2019.
- I. Lasri, A. Riadsolh, et M. Elbelkacemi, “ Self-Attention-Based Bi-LSTM Model for Sentiment Analysis on Tweets about Distance Learning in Higher Education”, *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, n° 12, 2023, p. 119.
- I. Lasri, A. Riadsolh, et M. Elbelkacemi, “Toward an Effective Analysis of COVID-19 Moroccan Business Survey Data using Machine Learning Techniques,” *13th International Conference on Machine Learning and Computing ICMLC*, p. 50–58, China, 2021.

Collaboration avec d’autres auteurs

- A. Riadsolh, I. Lasri, et M. Elbelkacemi, “Cloud-based sentiment analysis for measuring customer satisfaction in the Moroccan banking sector using Naive Bayes and Stanford NLP”, *J. Autom. Mob. Robot. Intell. Syst.*, vol. 14, n° 4, 2020, p. 64–71.
- C. F. Choukhan, I. Lasri, R. Elhatimi, M. R. Lemnaouar, et M. Esghir, “SARS-CoV-2 Prediction Strategy Based on Classification Algorithms from a Full Blood Examination”, *The Scientific World Journal*, 2023, p. 3248192.

Résumé

À l'ère actuelle, l'adaptation de l'apprentissage aux besoins des apprenants est essentielle pour évaluer la qualité des systèmes éducatifs. C'est dans cet esprit, conscient de la diversité des apprenants, que notre thèse se focalise sur l'importance de l'adaptation de l'apprentissage, étroitement liée à l'évaluation de la qualité des systèmes éducatifs, tout en accordant une attention particulière à l'éducation inclusive. Dans ce contexte, deux volets ont été explorés. Le premier volet se focalise sur l'amélioration de l'évaluation formative en utilisant deux systèmes de reconnaissance des émotions faciales (FER). Le premier utilise un réseau de neurones convolutifs (CNN) pour reconnaître les émotions faciales des étudiants en temps réel. Le deuxième système FER repose sur un réseau de neurones convolutifs profonds VGG-16, basé sur l'apprentissage par transfert, pour détecter l'engagement des étudiants sourds et malentendants en se basant sur leurs expressions faciales. Les systèmes ont été évalués auprès d'étudiants en master et en licence fondamentale d'excellence à la Faculté des Sciences de Rabat (FSR). Le deuxième volet de notre recherche se concentre sur le développement de systèmes d'analyse des sentiments en utilisant l'apprentissage automatique et les technologies des Big Data pour évaluer la qualité des systèmes éducatifs. Nous avons exploré différentes approches, y compris l'analyse des tweets pour mesurer la réputation des universités marocaines, ainsi que l'opinion des utilisateurs de Twitter sur l'apprentissage à distance dans l'enseignement supérieur. Les résultats de ces systèmes fournissent des indications précieuses pour évaluer en profondeur la qualité des systèmes éducatifs.

Mots-clefs : Reconnaissance des émotions faciales, Éducation inclusive, Analyse des sentiments, Intelligence artificielle, Big data.

Abstract

In today's era, the adaptation of learning to the needs of learners is essential for evaluating the quality of educational systems. It is in this spirit, aware of the diversity of learners, that our thesis focuses on the importance of adapting learning, which is closely linked to the evaluation of the quality of educational systems, while paying particular attention to inclusive education. In this context, two strands have been explored. The first strand focuses on improving formative assessment using two facial emotion recognition (FER) systems. The first one uses a convolutional neural network (CNN) to recognize students' facial emotions in real time. The second FER system relies on a VGG-16 deep convolutional neural network, based on transfer learning, to detect the engagement of deaf and hard-of-hearing students based on their facial expressions. The systems were evaluated with students in the Master's and Bachelor's degree programs of excellence at the Faculty of Sciences of Rabat (FSR). The second part of our research focuses on the development of sentiment analysis systems using machine learning and Big Data technologies to evaluate the quality of educational systems. We explored different approaches, including analyzing tweets to measure the reputation of Moroccan universities, as well as Twitter users' opinions on distance learning in higher education. The results of these systems provide valuable insights for in-depth evaluation of the quality of educational systems.

Keywords : Facial emotion recognition, Inclusive education, Sentiment analysis, Artificial intelligence, Big data.