جامعة سيدي محمد بن عبد الله بفاس
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵙⵉⴷⵉ ⵎⵓⵃⵎⴷ ⴱⵏ ⵄⴱⴷⵍⵍⴰⵀ ⴱ ⴼⴰⵙ
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

كلية العلوم ظهر المهراز فاس
ⵜⴰⵖⵉⵡⵉⵜ ⵏ ⵜⵣⵔⴼⵜ ⵉⴼⵓⵙ ⵏ ⵥⵀⵔ ⵍⵎⵀⵔⴰⵣ ⴼⴰⵙ
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

**N° d'ordre : 26 / 2024**

Centre des Études : ***Sciences et Techniques et Sciences Médicales***

Formation Doctorale : ***Sciences et Technologies de l'Information et de Communication (STIC)***

Discipline : ***Informatique***

Spécialité : ***Informatique***

***Laboratoire d'Informatique, Signaux, Automatique et Cognitivisme (LISAC)***

# THÈSE DE DOCTORAT

# Contribution à l'apprentissage intelligent : analyse de données et recommandations personnalisées de ressources pédagogiques vidéo

Présentée par :

**Mohamed TIMMI**

Soutenue le 12 juin 2024 devant le jury :

| | | | |
|---|---|---|---|
| **Pr. NFAOUI El Habib** | PES | FSDM, USMBA, Fès | **Président** |
| **Pr. AGHOUTANE Badraddine** | PES | FS, UMI, Meknès | **Rapporteur & examinateur** |
| **Pr. RADOUANE Abdelhay** | MCH | CRMEF Fès-Meknès | **Rapporteur & examinateur** |
| **Pr. TAIRI Hamid** | PES | FSDM, USMBA, Fès | **Rapporteur & examinateur** |
| **Pr. RIFFI Jamal** | MCH | FSDM, USMBA, Fès | **Examinateur** |
| **Pr. EL AKKAD Nabil** | MCH | ENSA, USMBA, Fès | **Examinateur** |
| **Pr. JEGHAL Adil** | MCH | ENSA, USMBA, Fès | **Examinateur** |
| **Pr. EL GAROUANI Said** | PES | FSDM, USMBA, Fès | **Directeur de thèse** |
| **Pr. YAHYAOUY Ali** | PES | FSDM, USMBA, Fès | **Co-Directeur de thèse** |

**Année Universitaire : 2023-2024**

# Remerciements

Au cours de ma période de recherche au sein du Laboratoire d'Informatique, Signaux, Automatique et Cognitivisme (LISAC) de la faculté des sciences Dhar El Mehraz à Fès, j'ai ressenti un sentiment d'appartenance à une famille bien plus qu'à une simple équipe de recherche. À l'issue de cette thèse, je tiens à exprimer ma profonde gratitude à l'ensemble des membres de notre laboratoire, en particulier au directeur du laboratoire, le Professeur **TAIRI Hamid**, Ainsi je tiens à remercier Professeur **NFAOUI El Habib**, de m'avoir fait l'honneur d'être le président de ma soutenance.

Je tiens également à remercier chaleureusement mon directeur de thèse, le Professeur **Said El Garouani**, pour son accompagnement, son soutien, le temps qu'il m'a consacré, sa bienveillance, ses conseils avisés, sa compréhension et ses directives précieuses. Son mentorat a été une source d'apprentissage inestimable tout au long de ces années.

Je tiens à exprimer ma profonde gratitude à mon co-encadrant de thèse, le Professeur **Ali Yahyaouy**, pour son soutien académique, ses conseils précieux, et son amitié tout au long de ce parcours. Sa présence bienveillante et son expertise ont été des atouts inestimables dans l'achèvement de ma thèse. Merci pour tout.

Mes remerciements s'étendent aussi au Professeur **Adil JEGHAL** de l'école nationale des sciences appliquées de Fès, pour son suivi attentif, son encadrement continu et ses conseils judicieux. Sa réactivité et son efficacité face à mes interrogations ont fait de lui plus qu'un mentor, un véritable frère.

Je suis également reconnaissant envers les rapporteurs de cette thèse, les Professeurs **AGHOUTANE Badraddine**, **RADOUANE Abdelhay**, et **TAIRI Hamid,** pour l'attention qu'ils ont portée à mon travail.

Je souhaite également remercier les Professeurs **RIFFI Jamal**, **EL AKKAD Nabil** et **JEGHAL Adil** pour avoir accepté d'examiner ma thèse.

Je tiens à exprimer ma gratitude à l'égard de tous ceux qui ont contribué, de près ou de loin, à la réalisation de cette thèse. Je leur adresse toute ma reconnaissance et mes sincères remerciements pour leur soutien et leur confiance.

Enfin, je suis profondément redevable à **ma famille**, proche et élargie, qui m'a apporté soutien et encouragement dans les moments de doute, qui a été un pilier dans toutes mes décisions, et qui m'a enseigné la persévérance face aux obstacles. Leur présence constante, leur confiance

sans faille et leur soutien indéfectible ont été pour moi une source d'encouragement continue. Merci également à mes collègues du **CRMEF Fès-Meknès** et à **mes amis**, pour leur solidarité et leur motivation tout au long de mes études.

# Résumé

La présente thèse se concentre sur le développement de systèmes de recommandation optimisés pour la sélection de ressources pédagogiques vidéo dans l'enseignement, répondant ainsi à une problématique centrale : comment améliorer l'accès et la personnalisation des ressources éducatives auprès des apprenants à travers des recommandations précises et adaptées ? Cette question soulève le défi d'intégrer efficacement les préférences individuelles et les interactions sociales dans les algorithmes de recommandation, de façon à surmonter les limites des méthodes existantes, telles que la rareté des données et le démarrage à froid, tout en améliorant l'expérience d'apprentissage en ligne.

La contribution principale de cette recherche réside dans la proposition et l'évaluation de deux innovations majeures : l'approche "global local trust with Matrix Factorization", qui intègre les dimensions globales et locales des relations de confiance entre utilisateurs pour améliorer les prédictions des recommandations, et une approche hybride combinant le filtrage basé sur le contenu et le filtrage collaboratif pour les vidéos éducatives. Ces approches complémentaires exploitent à la fois les évaluations des utilisateurs et les informations contextuelles pour générer des recommandations plus pertinentes et personnalisées, visant de surcroît à enrichir l'expérience éducative des apprenants.

Les résultats obtenus à l'issue de nos expérimentations démontrent l'efficacité remarquable des méthodes proposées, avec une amélioration significative de la précision des recommandations par rapport aux approches traditionnelles. La mise en pratique de ces techniques sur des données concrètes, telles que les vidéos YouTube, a révélé leur capacité à guider efficacement les utilisateurs vers des ressources pédagogiques adaptées à leurs besoins et préférences individuels. Ces résultats soulignent le potentiel des systèmes de recommandation intelligents pour optimiser l'accès aux ressources éducatives, offrant ainsi de nouvelles perspectives pour la personnalisation et l'amélioration de l'apprentissage en ligne.



**Mots clés :** Apprentissage Intelligent, Système de Recommandation, Ressources Pédagogiques Vidéo, Filtrage Collaboratif, Filtrage Basé sur le Contenu, Filtrage Hybride, Factorisation Matricielle, Analyse de l'Apprentissage, Data Mining Éducatif, Personnalisation de l'Apprentissage, Technologie Éducative.

## Abstract

This thesis focuses on the development of optimized recommendation systems for the selection of video educational resources in teaching, addressing a central dilemma on how to improve access and personalization of educational resources for learners through precise and tailore suggestions. This delicate concern raises the challenge of effectively incorporating individual preferences and social interactions into algorithm-based recommendations, to overcome the limitations of existing methods, such as data scarcity and cold start, while enhancing the online learning experience. The main contribution of this research lies in the proposal and evaluation of two major innovations: the “global local trust with Matrix Factorization” approach, which integrates the global and local dimensions of trust relationships among users to enhance recommendation predictions, and a hybrid approach combining content-based filtering and collaborative filtering for educational videos. These approaches leverage both user ratings and contextual information to generate more relevant and personalized recommendations, aiming to enrich learners' educational experience. The results demonstrate the effectiveness of the proposed methods, with a significant improvement in recommendation precision compared to traditional approaches. The empirical application of these techniques on concrete data, such as YouTube videos, has revealed their ability to effectively guide users towards educational resources that meet their individual needs and preferences. These results highlight the potential of intelligent recommendation systems to transform access to educational resources, thus offering new perspectives for the personalization and enhancement of online learning.



**Keywords:** Intelligent Learning, Recommendation System, Video Educational Resources, Collaborative Filtering, Content-Based Filtering, Hybrid Filtering, Matrix Factorization, Learning Analysis, Educational Data Mining, Learning Personalization, Educational Technologies.

# Tables des matières

*Remerciements* ..... *II*

*Résumé* ..... *IV*

*Abstract* ..... *V*

*Tables des matières* ..... *VI*

*Liste des figures* ..... *X*

*Liste des tableaux* ..... *XII*

*Liste des abréviations* ..... *XIII*

*Introduction générale* ..... *2*

*Chapitre 1* ..... *8*

*État de l'art sur l'apprentissage intelligent* ..... *8*

**1.1. Introduction** ..... **8**

**1.2. Systèmes de recommandation** ..... **9**

*1.2.1.* Historique des systèmes de recommandation ..... 10

*1.2.2.* Problèmes et limites des systèmes de recommandation ..... 11

*1.2.3.* Systèmes de recommandation sociale ..... 12

**1.3. Réseaux sociaux en ligne** ..... **14**

1.3.1. Types de systèmes de recommandation sociale ..... 14

1.3.2. Systèmes de recommandation dans les environnements informatiques pour l'apprentissage humain 15

**1.4. Travaux connexes** ..... **16**

1.4.1. Travaux connexes : sur l'intégration des systèmes de recommandation dans les environnements informatiques pour l'apprentissage humain ..... 16

1.4.2. Travaux connexes : Systèmes de recommandation éducatifs ..... 17

1.4.3. Travaux connexes : Systèmes de recommandation de vidéos éducatives ..... 19

1.4.4. Synthèse ..... 20

**1.5. Évaluation des systèmes de recommandation** ..... **21**

1.5.1. Évaluation hors ligne ..... 22
1.5.2. Évaluation en ligne ..... 23
**1.6. Apprentissage intelligent ..... 24**
1.6.1. Fondements théoriques de l'apprentissage intelligent ..... 24
1.6.2. Technologies sous-jacentes à l'apprentissage intelligent ..... 25
**1.7. Conclusion ..... 26**
***Chapitre 2 ..... 29***
***Data mining et analyse de l'apprentissage : révolutionner l'éducation à l'ère numérique ..... 29***
***2.1.* Introduction ..... 29**
***2.2.* Data mining dans les environnements éducatifs ..... 30**
2.2.1. Introduction au data mining appliqué aux contextes éducatifs ..... 30
2.2.2. Contexte et principe du data mining éducatif ..... 32
2.2.3. Caractéristiques et modèles d'apprentissage des apprenants ..... 33
2.2.4. Synergie entre le data mining éducatif et les systèmes de recommandation ..... 35
2.2.5. Défis et préoccupations ..... 38
***2.3.* Analyse de l'apprentissage ..... 40**
2.3.1. Analyse de l'apprentissage et son rôle dans l'éducation moderne ..... 40
2.3.2. Définition et contexte historique ..... 41
2.3.3. Objectifs et importance de l'analyse de l'apprentissage ..... 43
2.3.4. Méthodes et techniques employées ..... 44
2.3.5. Applications pratiques ..... 46
2.3.6. Défis et préoccupations éthiques ..... 47
2.3.7. Synthèse finale ..... 48
***2.4.* Conclusion ..... 49**
***Chapitre 3 ..... 52***
***Système de recommandation utilisant la confiance et les évaluations des utilisateurs ..... 52***
***3.1.* Introduction ..... 52**
***3.2.* Différentes approches existantes ..... 54**

*3.3.* **Approche proposée ..... 55**

3.3.1. Description du problème ..... 55
3.3.2. Modèle de factorisation matricielle ..... 56
3.3.3. Étapes de la construction du modèle proposé ..... 57
3.3.4. Apprentissage du modèle proposé ..... 65

*3.4.* **Évaluation expérimentale ..... 68**

3.4.1. Présentation de la base de données ..... 68
3.4.2. Analyse de la distribution des donneurs de confiance et receveurs de confiance ..... 69
3.4.3. Évaluation du modèle ..... 71
3.4.4. Paramètres expérimentaux et évaluation ..... 72

*3.5.* **Résultats et discussion ..... 74**

3.5.1. Analyses et interprétations ..... 74
3.5.2. Évaluation de la performance du modèle proposé selon différents niveaux de confiance 76

*3.6.* **Conclusion ..... 79**

***Chapitre 4 ..... 82***

***Optimisation des recommandations de vidéos éducatives : filtrage hybride combinant le filtrage basé sur le contenu et le filtrage collaborative pour une personnalisation avancée ..... 82***

**4.1. Introduction ..... 82**

**4.2. Quelques exemples de systèmes de recommandation ..... 83**

4.2.1. Système de recommandation pour vidéos médicales ..... 83
4.2.2. Système hybride de recommandation de vidéos YouTube ..... 83
4.2.3. Système de recommandation pour des vidéos éducatives ouvertes basé sur les compétences requises 84
4.2.4. Système de recommandation pour YouTube basé sur son réseau de critiques ..... 84
4.2.5. Système de résumé de contenu vidéo pour recommandation ..... 85
4.2.6. Comparatif des systèmes de recommandation ..... 86

**4.3. Approche proposée ..... 86**

4.3.1. Aperçu ..... 87
4.3.2. Extraction de vidéo ..... 90
4.3.3. Prétraitement et classification des vidéos ..... 92
4.3.4. Algorithme du système de recommandation et résultats ..... 93

4.3.5. Algorithme de filtrage hybride de l'approche proposée ..... 98
4.3.6. Collecte des feedbacks ..... 99
4.3.7. Interface du système proposé ..... 99

**4.4. Résultat et discussion ..... 100**

4.4.1. Évaluation ..... 100
4.4.2. Discussion ..... 107

**4.5. Conclusion ..... 108**

***Conclusion générale et perspectives ..... 110***

***Publications ..... 112***

***Bibliographie ..... 113***

# Liste des figures

*Figure 1: Data mining éducatif (Romero and Ventura 2013)* ..... *31*

*Figure 2 : Étapes du processus de data mining.* ..... *32*

*Figure 3: Conception d'un schéma intégré pour l'analyse de données éducatives, l'analyse de l'apprentissage et les systèmes de recommandation dans l'environnement de l'apprentissage numérique* ..... *36*

*Figure 4: Intersection des domaines d'analyse en éducation* ..... *43*

*Figure 5: Principe de fonctionnement de la technique de factorisation matricielle* ..... *57*

*Figure 6: Modèle graphique de la fusion de la confiance globale et locale par la factorisation matricielle pour des recommandations personnalisées.* ..... *59*

*Figure 7: Organigramme d'Algorithme d'apprentissage du modèle proposé.* ..... *67*

*Figure 8: Répartition des administrateurs par rapport aux personnes qui leur accordent leur confiance* ..... *70*

*Figure 9: Comparaison des performances selon divers degrés de confiance des utilisateurs (MAE en fonction de degré de confiance avec d=5)* ..... *77*

*Figure 10: Comparaison des performances selon divers degrés de confiance des utilisateurs (RMSE en fonction de degré de confiance avec d=5)* ..... *77*

*Figure 11: Comparaison des performances selon divers degrés de confiance des utilisateurs (MAE en fonction de degré de confiance avec d=10)* ..... *78*

*Figure 12: Comparaison des performances selon divers degrés de confiance des utilisateurs (RMSE en fonction de degré de confiance avec d=10)* ..... *78*

*Figure 13: Architecture du système VCSR [108]* ..... *85*

*Figure 14: Aperçu du système de recommandation.* ..... *88*

*Figure 15: Étapes suivies par la solution.* ..... *89*

*Figure 16: Détails des vidéos.* ..... *91*

*Figure 17: Organigramme de l'algorithme combiné pour la proposition de suggestions de vidéos.* ..... *94*

*Figure 18: Page d'accueil de l'Interface du système proposé* ..... *100*

*Figure 19: Fréquence d'utilisation de YouTube. ..... 101*

*Figure 20: Utilisation de YouTube à des fins éducatives..... 102*

# Liste des tableaux

*Tableau 1: Caractéristiques principales de la base de données Epinions ..... 69*

*Tableau 2: Comparaison des performances des différentes méthodes ..... 72*

*Tableau 3: Comparaisons des performances des recommandations sur la base de données Epinions, dans le cas de "Tous les utilisateurs" ..... 74*

*Tableau 4: Comparaisons des performances des recommandations sur la base de données d'Epinions, dans le cas des "utilisateurs ayant démarré à froid" ..... 74*

*Tableau 5: Calcul de la moyenne pondérée ..... 97*

*Tableau 6: Analyse du Taux de rejet et des opinions des utilisateurs sur différentes listes de recommandations vidéo ..... 106*

# Liste des abréviations

| | |
|---|---|
| **API** | Application Programming Interface |
| **AUC** | Area Under the Curve |
| **EDM** | Educational Data Mining |
| **EIAH** | Environnements Informatique Pour l'Apprentissage Humain |
| **FBC** | Filtrage Basé Sur Le Contenu |
| **FC** | Filtrage Collaboratif |
| **GLoTMF** | **G**lobal **Lo**cal **T**rust with **M**atrix **F**actorization |
| **IA** | Intelligence Artificielle |
| **IoT** | Internet of Things |
| **ITS** | Intelligent Tutoring System (Systèmes Tutoriels Intelligents) |
| **LA** | Learning Analytics (Analyse De L'apprentissage) |
| **LMS** | Learning Management System |
| **MAE** | Mean Absolute Error) |
| **ML** | Machine Learning (Apprentissage Automatique) |
| **MOOC** | Massine Online Open Course |
| **OCR** | Optical Character Recognition |
| **RMSE** | Root Mean Square Error |
| **SPOC** | System for Small Private Online Courses |
| **SR** | Système De Recommandation |
| **VCSR** | Video Content Summarization for Recommendation |
| **YRN** | YouTube Recommendation Network |

# INTRODUCTION GÉNÉRALE

# Introduction générale

L'apprentissage intelligent est un domaine fascinant et en constante évolution, rendu possible par les avancées récentes en matière de technologies de l'information et de la communication [1]. Au cœur de ce domaine se trouve le potentiel de transformer fondamentalement la façon dont les individus apprennent et interagissent avec le contenu éducatif. La thèse qui sera développée dans les pages suivantes se concentre sur l'exploration de cette frontière, en particulier l'utilisation des outils de data mining pédagogiques et des analyses d'apprentissage pour améliorer le processus d'apprentissage grâce à un système de recommandation intelligent.

Le système éducatif actuel, bien que riche en ressources, est souvent marqué par des inégalités dans la qualité de l'apprentissage et l'accès à l'information. De plus, la nature « taille unique » des systèmes traditionnels ne tient pas compte des différences individuelles en termes de style d'apprentissage, de rythme et de préférences [2]. Par conséquent, la possibilité de personnaliser l'expérience d'apprentissage grâce à une approche basée sur les données pourrait fournir des solutions à ces problèmes persistants.

Cette thèse explore le développement et l'application de processus d'apprentissage intelligents, qui reposent sur l'exploration de données pédagogiques et l'analyse d'apprentissage pour générer des recommandations personnalisées. L'objectif est d'explorer comment ces technologies peuvent être utilisées pour créer des environnements d'apprentissage qui s'adaptent et répondent aux besoins individuels des apprenants, améliorant ainsi l'efficacité de l'apprentissage et l'engagement des apprenants.

Dans cette optique, cette thèse est structurée en plusieurs parties. Dans un premier temps, une revue exhaustive de la littérature existante permettra d'établir un cadre conceptuel autour de l'apprentissage intelligent, de l'exploration de données pédagogiques et de l'analyse d'apprentissage. Ensuite, nous proposerons un modèle de système de recommandation basé sur ces concepts, en décrivant son architecture et sa mise en œuvre. Nous discuterons ensuite des résultats obtenus lors de l'application de ce système dans divers contextes d'apprentissage. Enfin, nous conclurons en résumant les conclusions clés et en proposant des orientations pour les recherches futures.

Cette thèse vise à contribuer au corps croissant de connaissances dans le domaine de l'apprentissage intelligent et à éclairer la voie pour de nouvelles approches qui pourraient rendre l'éducation plus inclusive, engageante et efficace pour tous.

## 1. Contexte de la thèse

L'évolution rapide des technologies numériques a profondément transformé le domaine de l'éducation. Les méthodes traditionnelles d'enseignement et d'apprentissage sont constamment repensées et redéfinies, offrant de nouvelles opportunités et de nouveaux défis. De plus, la quantité croissante de données générées par les plateformes d'apprentissage en ligne offre une possibilité sans précédent de comprendre et d'améliorer les processus d'apprentissage. Dans ce contexte, cette thèse se concentre sur l'exploitation de ces données pour améliorer l'apprentissage par le biais de systèmes de recommandation intelligents.

Plus spécifiquement, la thèse se situe à l'intersection de plusieurs domaines de recherche : l'apprentissage intelligent, l'exploration de données pédagogiques, l'analyse de l'apprentissage et les systèmes de recommandation. Le potentiel de ces domaines pour améliorer l'apprentissage est largement reconnu, mais leur intégration efficace reste un défi.

En ce qui concerne les ressources pédagogiques vidéo, elles sont de plus en plus utilisées en raison de leur capacité à présenter des informations de manière visuelle et engageante. Cependant, étant donné l'énorme quantité de vidéos éducatives disponibles en ligne, il peut être difficile pour les apprenants de trouver les ressources les plus pertinentes et utiles pour leurs besoins spécifiques.

Par conséquent, le développement d'un système de recommandation intelligent capable de proposer des ressources vidéo pédagogiques personnalisées en fonction des besoins individuels de chaque apprenant est d'une grande importance. Une telle innovation pourrait contribuer à améliorer l'efficacité de l'apprentissage, l'engagement des apprenants et, en fin de compte, leurs résultats.

## 2. Problématique scientifique

La problématique scientifique de cette thèse, centrée sur le système de recommandation de ressources pédagogiques vidéo, peut être formulée comme suit :

"Comment développer un système de recommandation intelligent et efficace qui facilite la découverte de ressources pédagogiques vidéo pertinentes et personnalisées pour améliorer l'expérience et les résultats d'apprentissage des apprenants ?"

Cette problématique soulève plusieurs défis importants :

Comment recueillir et analyser des données pertinentes (comme les préférences des apprenants, leur historique de visionnage, leurs performances d'apprentissage, les caractéristiques des vidéos) pour soutenir un système de recommandation ?

Comment concevoir des algorithmes de recommandation qui peuvent exploiter efficacement ces données pour fournir des suggestions de vidéos pédagogiques qui sont à la fois pertinentes et adaptées aux besoins individuels de chaque apprenant ?

Comment ces recommandations peuvent-elles être présentées de manière à encourager l'engagement des apprenants et à faciliter leur apprentissage ?

Comment évaluer l'efficacité de ces recommandations sur l'expérience d'apprentissage des apprenants et sur leurs résultats ?

Comment le système de recommandation peut-il s'adapter et évoluer avec le temps pour répondre aux besoins changeants des apprenants et aux développements de nouvelles ressources pédagogiques vidéo ?

Ce sujet invite à une exploration en profondeur des possibilités offertes par l'exploration des données pédagogiques, l'analyse de l'apprentissage et les technologies d'intelligence artificielle pour améliorer l'accès à l'utilisation de ressources pédagogiques vidéo dans le cadre de l'apprentissage.

## 3. Démarche méthodologique et organisation de la thèse

La démarche méthodologique de cette thèse s'articule autour de trois étapes principales : une revue de la littérature, l'élaboration d'un système de recommandation, et enfin, l'évaluation et l'amélioration du système.

Dans un premier temps, une revue exhaustive de la littérature permettra de situer le travail dans le contexte des recherches actuelles sur l'apprentissage intelligent, l'exploration de données

pédagogiques, l'analyse de l'apprentissage et les systèmes de recommandation. Cela fournira les bases théoriques nécessaires et aidera à identifier les lacunes actuelles dans ces domaines.

Ensuite, un système de recommandation de ressources pédagogiques vidéo sera développé en utilisant des techniques d'exploration de données et d'analyse de l'apprentissage. Ce système sera conçu pour être adaptatif et pour répondre aux besoins individuels des apprenants. Ce travail comprendra le recueil et l'analyse de données, la conception et l'implémentation d'algorithmes de recommandation, et le développement d'une interface utilisateur intuitive.

Enfin, le système de recommandation sera évalué pour déterminer son efficacité. Les métriques d'évaluation pourraient inclure la satisfaction des utilisateurs, l'engagement des apprenants, l'amélioration des résultats d'apprentissage et la capacité du système à s'adapter aux besoins changeants des apprenants. Sur la base de ces évaluations, des améliorations seront apportées au système pour augmenter son efficacité et son utilité.

L'organisation de la thèse pourrait se présenter comme suit :

- **Introduction générale** - Établissement le cadre pour notre étude approfondie sur l'amélioration de l'accès et de la personnalisation des ressources pédagogiques vidéo à travers des systèmes de recommandation intelligents.
- **Chapitre 1 : État de l'art sur l'apprentissage intelligent -** Exploration l'évolution des systèmes de recommandation, en mettant en évidence les développements actuels et les lacunes dans ces domaines.
- **Chapitre 2 : Data mining et analyse de l'apprentissage : révolutionner l'éducation à l'ère numérique -** Discussion des méthodologies et techniques de data mining et d'analyse de l'apprentissage spécifiques à l'éducation.
- **Chapitre 3 : Système de recommandation utilisant la confiance et les évaluations des utilisateurs** - Présentation de notre première contribution majeure, en décrivant comment la confiance et les évaluations des utilisateurs peuvent être intégrées pour améliorer la précision des recommandations.
- **Chapitre 4 : Optimisation des recommandations de vidéos éducatives : filtrage hybride combinant le filtrage basé sur le contenu et le filtrage collaborative pour une personnalisation avancée -** Exposition de notre seconde contribution, détaillant comment les approches de filtrage hybride peuvent être utilisées pour une personnalisation plus poussée des recommandations.

- **Conclusion générale et perspectives** - Résumé des contributions principales, discussion de l'impact de notre travail, et exploration des directions futures possibles pour la recherche dans ce domaine.

# CHAPITRE 1

# ÉTAT DE L'ART SUR L'APPRENTISSAGE INTELLIGENT

# Chapitre 1

# État de l'art sur l'apprentissage intelligent

## 1.1. Introduction

A l'ère numérique actuelle, où la technologie imprègne chaque aspect de notre vie quotidienne, l'importance des systèmes de recommandation, des réseaux sociaux en ligne et des méthodes d'analyse de l'apprentissage dans les environnements éducatifs est devenue plus prépondérante que jamais. Ces systèmes, initialement conçus pour optimiser l'expérience utilisateur dans divers domaines tels que le commerce électronique et les médias sociaux, se sont progressivement imposés comme des outils cruciaux dans le secteur de l'éducation, façonnant la manière dont l'information est filtrée, présentée et consommée.

Ce chapitre se propose d'explorer l'évolution et l'impact significatif de ces technologies avancées. Nous débuterons par une analyse des systèmes de recommandation, en mettant en lumière leur développement historique, les différents types et leur application croissante dans des domaines variés, y compris celui de l'éducation. Ensuite, nous plongerons dans le monde des réseaux sociaux en ligne, pour tenter d'examiner comment ces plateformes ont remodelé la communication, la diffusion d'informations, et les interactions sociales dans notre société numérisée.

Poursuivant notre exploration, nous aborderons les systèmes de recommandation sociale, une fusion entre les systèmes de recommandation traditionnels et les éléments sociaux, qui vise à affiner davantage les recommandations à travers l'analyse des relations et interactions sociales des utilisateurs. Cette section soulignera l'importance croissante de la dimension sociale dans la personnalisation des expériences utilisateur.

Enfin, nous examinerons comment l'analyse de l'apprentissage et le data mining éducatif révolutionnent le paysage de l'éducation. Nous discuterons de la manière dont ces approches basées sur les données permettent une compréhension plus profonde des besoins et comportements des apprenants, conduisant à des expériences éducatives plus personnalisées et efficaces. Cette analyse sera complétée par une discussion sur les défis éthiques et pratiques associés à la collecte et à l'utilisation de données dans un contexte éducatif.

## 1.2. Systèmes de recommandation

Les systèmes de recommandation sont des outils logiciels et techniques conçus pour suggérer des produits, des services ou des informations à utiliser par les utilisateurs en fonction de leurs préférences et besoins [3]. Le principal objectif est de fournir des suggestions personnalisées et pertinentes aux utilisateurs [4]. Ces systèmes sont devenus des composantes essentielles de nombreux services en ligne, tels que les plateformes de streaming vidéo comme Netflix pour suggérer des films et des séries télévisées, les sites de commerce électronique comme Amazon pour recommander des produits et les réseaux sociaux comme Facebook pour recommander des publications, des amis et des publicités [5].

Il existe principalement trois types de systèmes de recommandation : le filtrage collaboratif, le filtrage basé sur le contenu et les méthodes hybrides.

a. **Filtrage collaboratif** (FC) : Cette méthode fait des prédictions sur les intérêts d'un utilisateur en collectant les préférences de nombreux utilisateurs. L'hypothèse sous-jacente est que si deux utilisateurs sont d'accord sur un sujet, ils sont susceptibles de s'accorder sur d'autres sujets [6, 7].
b. **Filtrage basé sur le contenu** (FBC) : Cette approche recommande des éléments en comparant le profil de l'utilisateur à une description de l'élément. En d'autres termes, l'utilisateur se voit recommander des éléments similaires à ceux qu'il a aimés dans le passé [6, 7].
c. **Méthodes hybrides** : Ces systèmes combinent les approches du filtrage collaboratif basé sur le contenu pour surmonter leurs limitations respectives [6]. Par exemple, ils peuvent utiliser le filtrage collaboratif pour faire des prédictions initiales, puis utiliser le filtrage basé sur le contenu pour affiner ces prédictions [8].

Les systèmes de recommandation ont connu un développement considérable au cours des dernières décennies. Avec l'avancement de l'intelligence artificielle et de l'apprentissage automatique, de nouveaux algorithmes et techniques sont continuellement développés pour améliorer la précision et l'efficacité des recommandations. Cependant, il reste encore de nombreux défis à relever, tels que la gestion de l'échelle des données, le traitement du démarrage à froid et l'amélioration de la transparence des recommandations.

### *1.2.1.* Historique des systèmes de recommandation

Les systèmes de recommandation sont devenus une partie intégrante de notre vie numérique quotidienne, en guidant nos choix dans une variété de domaines allant de l'achat de produits en ligne au visionnage de films et à l'écoute de musique. Pour comprendre comment nous en sommes arrivés là, il est utile de jeter un coup d'œil à l'histoire des systèmes de recommandation.

a. **Années 1990** : Les premiers systèmes de recommandation ont vu le jour dans les années 1990. Le système GroupLens [9], introduit en 1992, est souvent considéré comme l'un des premiers systèmes de recommandation. Il utilisait la technologie des filtres collaboratifs pour recommander des items de nouvelles en ligne en se basant sur les notes des utilisateurs.
b. **Années 2000** : La popularité des systèmes de recommandation a explosé avec l'avènement des sites de commerce électronique comme Amazon et des sites de streaming comme Netflix [10]. Amazon a commencé à utiliser des systèmes de recommandation pour suggérer des produits en se basant sur les achats antérieurs des utilisateurs et les produits consultés. En 2006, Netflix a lancé le Netflix Prize, un concours avec un prix d'un million de dollars pour améliorer la précision de son système de recommandation [11].
c. **Années 2010** : L'essor des réseaux sociaux a introduit une nouvelle dimension dans les systèmes de recommandation, permettant d'incorporer les informations sociales des utilisateurs dans le processus de recommandation. Les algorithmes de recommandation sont également devenus plus sophistiqués, en intégrant des techniques d'apprentissage automatique et d'intelligence artificielle pour améliorer la précision des recommandations [12].
d. **Années 2020 et au-delà** : À l'heure actuelle, les systèmes de recommandation continuent d'évoluer. Les tendances récentes incluent l'utilisation de l'apprentissage profond pour les recommandations, l'exploration des systèmes de recommandation multicritères et l'intégration des préférences explicites et implicites des utilisateurs [13, 14]. De plus, il y a une prise de conscience croissante de l'importance de la transparence et de l'équité dans les systèmes de recommandation.

Dans le contexte éducatif, les systèmes de recommandation sont utilisés pour proposer des ressources d'apprentissage pertinentes aux apprenants. Ces systèmes prennent en compte les

préférences, les compétences et les objectifs d'apprentissage des utilisateurs pour faire des recommandations personnalisées, aidant ainsi les apprenants à naviguer dans la vaste quantité de ressources éducatives disponibles en ligne.

### *1.2.2.* Problèmes et limites des systèmes de recommandation

Bien que les systèmes de recommandation aient apporté d'énormes avantages en facilitant la découverte de produits et de services pertinents parmi une large gamme d'options, ils présentent également un certain nombre de défis et de limitations :

a. **Problème du démarrage à froid :** Lorsqu'un nouvel utilisateur ou un nouvel élément entre dans le système, il n'y a pas assez d'informations pour faire des recommandations précises. C'est ce qu'on appelle le problème du démarrage à froid [15]. Pour les nouveaux utilisateurs, on parle parfois de problème de démarrage à froid pour l'utilisateur, et pour les nouveaux éléments, de problème de démarrage à froid pour l'élément.
b. **Problème de la popularité :** Les systèmes de recommandation ont tendance à recommander des éléments déjà populaires, ce qui peut entraîner un manque de diversité et une sous-estimation des éléments moins connus ou nouveaux qui pourraient être pertinents pour l'utilisateur [15, 16].
c. **Problème de l'échelle :** Avec l'augmentation du nombre d'utilisateurs et d'éléments, fournir des recommandations précises et en temps réel peut être un défi en termes de calcul et de stockage des données.
d. **Problème de la transparence :** Les recommandations produites par des systèmes complexes (par exemple, ceux utilisant l'apprentissage profond) peuvent être difficiles à comprendre pour les utilisateurs [15, 17]. Cette opacité peut conduire à une méfiance à l'égard du système.
e. **Problème de la vie privée :** Les systèmes de recommandation nécessitent souvent de recueillir des informations détaillées sur les utilisateurs pour fonctionner efficacement, ce qui soulève des préoccupations en matière de vie privée [17].
f. **Problème de l'évaluation :** Évaluer la qualité des recommandations peut être difficile, car les véritables préférences de l'utilisateur ne sont pas toujours connues, et les préférences peuvent changer avec le temps.

Ces problèmes constituent des défis importants pour la recherche en matière de systèmes de recommandation, et de nombreux chercheurs travaillent à l'élaboration de méthodes pour les surmonter.

Après avoir exploré la nature et l'importance des systèmes de recommandation, leur évolution et leurs diverses méthodologies, nous abordons maintenant une facette plus ciblée de cette technologie : les systèmes de recommandation sociale. Ces systèmes enrichissent l'approche traditionnelle en tirant parti des interactions et des relations sociales pour affiner encore davantage les suggestions proposées aux utilisateurs. L'intégration de ces dimensions sociales représente une avancée significative, reflétant l'évolution des comportements en ligne et la synergie entre les technologies de recommandation et les réseaux sociaux. Cette prochaine section dévoilera comment les systèmes de recommandation sociale tirent profit de l'interconnectivité des utilisateurs pour offrir des recommandations qui ne sont pas seulement personnalisées, mais également contextualisées en fonction des réseaux et des communautés d'utilisateurs.

### *1.2.3.* Systèmes de recommandation sociale

Les systèmes de recommandation sociale sont une évolution des systèmes de recommandation traditionnels. Ils tirent parti des informations sociales et des relations d'un utilisateur pour générer des recommandations. Ces systèmes supposent que les amis ou les pairs d'un utilisateur ont des goûts similaires et utilisent cette information pour influencer les recommandations.

Dans un système de recommandation sociale, les informations sociales peuvent provenir de diverses sources, telles que les réseaux sociaux, les évaluations et les commentaires des utilisateurs et d'autres formes d'interactions sociales.

Il existe principalement deux approches pour intégrer les informations sociales dans les systèmes de recommandation :

a. **Filtrage collaboratif social** : Cette approche est similaire au filtrage collaboratif traditionnel, mais elle accorde plus de poids aux opinions des amis ou des pairs d'un utilisateur lors de la création de recommandations [18]. Par exemple, si vos amis sur un site de médias sociaux ont aimé un certain film, ce film pourrait vous être recommandé.

b. **Filtrage basé sur le contenu social** : Cette approche utilise les informations sociales pour enrichir le profil d'un utilisateur [19]. Par exemple, en plus de vos propres évaluations, les évaluations et les commentaires de vos amis pourraient être utilisés pour comprendre vos préférences.

L'un des principaux avantages des systèmes de recommandation sociale est qu'ils peuvent aider à résoudre certains des problèmes des systèmes de recommandation traditionnels, tels que le problème du démarrage à froid. En utilisant les informations sociales, ces systèmes peuvent faire des recommandations même pour les nouveaux utilisateurs qui n'ont pas encore d'historique d'interaction avec le système.

Cependant, les systèmes de recommandation sociale posent également des défis, notamment en matière de confidentialité et de manipulation sociale. Par exemple, ils doivent être conçus de manière à protéger la confidentialité des utilisateurs tout en exploitant leurs informations sociales pour faire des recommandations. De plus, ils doivent être robustes face aux utilisateurs qui tentent de manipuler le système pour promouvoir certains éléments.

Dans la section précédente, nous avons examiné les systèmes de recommandation sociale, qui intègrent des informations provenant des réseaux de contacts et des interactions des utilisateurs pour affiner et personnaliser les suggestions de contenu. Ces systèmes tirent parti de la puissance des connexions et des activités sociales pour offrir une expérience utilisateur améliorée, en s'appuyant sur l'hypothèse que les préférences d'un individu sont souvent influencées par son cercle social.

Nous allons maintenant faire la transition vers les "Réseaux sociaux en ligne", le terrain fertile sur lequel les systèmes de recommandation sociale prospèrent. Ces plateformes de médias sociaux ne sont pas seulement des espaces pour l'échange d'informations et la communication, mais aussi de vastes réservoirs de données utilisateurs et d'interactions qui alimentent les algorithmes de recommandation. En explorant les fonctionnalités, les utilisations et l'impact des plateformes de médias sociaux en ligne, nous allons comprendre comment elles constituent une composante essentielle de l'écosystème numérique moderne, en influençant les comportements des utilisateurs et en servant de catalyseur pour les systèmes de recommandation sociale.

## 1.3. Réseaux sociaux en ligne

Les réseaux sociaux en ligne sont des plateformes numériques qui facilitent la création et le partage de contenus [20], ainsi que la formation de réseaux sociaux avec d'autres utilisateurs. Ils sont devenus une partie importante de la vie quotidienne de nombreuses personnes, influençant la façon dont nous communiquons, travaillons, apprenons et nous divertissons.

Les réseaux sociaux offrent de nombreuses fonctionnalités, notamment :

a. **Partage de contenu** : Les utilisateurs peuvent créer et partager une variété de contenus, y compris des textes, des photos, des vidéos, et des liens vers d'autres sites Web [20].

b. **Interaction sociale** : Les utilisateurs peuvent interagir avec d'autres en aimant, en commentant, en partageant du contenu, et en envoyant des messages privés [21, 22].

c. **Formation de réseaux** : Les utilisateurs peuvent se connecter avec d'autres en les suivant, en devenant leurs amis, ou en se joignant à des groupes d'intérêt commun [22].

d. **Personnalisation** : Les utilisateurs peuvent personnaliser leur profil, leur flux d'actualités, et leurs notifications en fonction de leurs préférences et de leurs intérêts [23].

Certains des réseaux sociaux les plus populaires comprennent Facebook, Instagram, Twitter, LinkedIn, et Snapchat. Chaque réseau social a sa propre orientation et ses propres caractéristiques. Par exemple, LinkedIn est axé sur les réseaux professionnels et le développement de carrière, tandis qu'Instagram est principalement utilisé pour partager des photos et des vidéos.

Malgré leurs nombreux avantages, les réseaux sociaux en ligne ont également soulevé des préoccupations, notamment en ce qui concerne la vie privée, la désinformation, le cyberharcèlement et l'impact sur la santé mentale. De plus, la manière dont les informations sont collectées et utilisées par ces plateformes pour cibler les utilisateurs avec du contenu et des publicités personnalisés a également fait l'objet de débats.

### 1.3.1. Types de systèmes de recommandation sociale

Les systèmes de recommandation sociale sont conçus pour utiliser les informations provenant des réseaux sociaux ou des interactions sociales pour améliorer la qualité des recommandations

[24]. Ils visent à exploiter les relations sociales et les interactions entre les utilisateurs pour générer des recommandations plus pertinentes et précises. Voici trois types principaux de systèmes de recommandation sociale :

a. **Systèmes de recommandation sociale basés sur la confiance** : Ces systèmes supposent que les utilisateurs font plus confiance aux recommandations de leurs amis ou de leurs connexions de confiance dans un réseau social [25]. Ils pondèrent ou filtrent les recommandations en fonction de la proximité sociale et de la confiance entre les utilisateurs [26].

b. **Systèmes de recommandation sociale basés sur l'influence** : Ces systèmes tiennent compte de l'influence d'un utilisateur dans le réseau social[27]. Par exemple, un utilisateur qui a beaucoup de followers ou qui reçoit beaucoup de likes ou de partages peut être considéré comme ayant une grande influence [28]. Les recommandations de ces utilisateurs peuvent être considérées comme plus importantes.

c. **Systèmes de recommandation sociale basés sur les groupes** : Ces systèmes génèrent des recommandations en se basant sur les préférences et les comportements de groupes d'utilisateurs similaires [29]. Par exemple, si les membres d'un groupe de discussion sur les films aiment tous un certain film, ce film pourrait être recommandé à d'autres membres du groupe.

Il est important de noter que ces catégories ne sont pas mutuellement exclusives et qu'un système de recommandation sociale peut utiliser une combinaison de ces approches pour améliorer ses recommandations. Par ailleurs, la conception et l'implémentation de systèmes de recommandation sociale efficaces posent plusieurs défis, notamment en termes de respect de la vie privée, de manipulation et de biais dans les recommandations.

### 1.3.2. Systèmes de recommandation dans les environnements informatiques pour l'apprentissage humain

Les Environnements Informatiques pour l'Apprentissage Humain (EIAH), visent à utiliser la technologie pour faciliter et améliorer l'apprentissage [30]. Les systèmes de recommandation jouent un rôle important dans les EIAH en aidant les apprenants à naviguer dans le vaste éventail de ressources pédagogiques disponibles.

Voici quelques applications typiques des systèmes de recommandation dans les EIAH :

a. **Recommandation de ressources pédagogiques** : Les systèmes de recommandation peuvent aider les apprenants à trouver les ressources pédagogiques les plus pertinentes, telles que les livres, les articles, les cours en ligne, les vidéos, etc. Ces recommandations peuvent être basées sur les préférences de l'apprenant, son niveau de compétence, ses objectifs d'apprentissage, et d'autres facteurs [31].

b. **Recommandation d'activités d'apprentissage** : Les systèmes de recommandation peuvent suggérer des activités d'apprentissage qui sont susceptibles d'aider l'apprenant à atteindre ses objectifs [32]. Par exemple, ils peuvent recommander des exercices pratiques, des projets, des jeux éducatifs, etc.

c. **Recommandation de pairs pour l'apprentissage collaboratif** : Dans un contexte d'apprentissage collaboratif, les systèmes de recommandation peuvent aider à trouver des partenaires d'apprentissage ou des groupes d'étude qui ont des intérêts ou des objectifs d'apprentissage similaires [33].

d. **Recommandation de parcours d'apprentissage** : Pour les apprenants qui cherchent à acquérir une compétence ou une connaissance spécifique, les systèmes de recommandation peuvent proposer des parcours d'apprentissage personnalisés. Par exemple, ils peuvent recommander une séquence de cours pour apprendre une nouvelle langue ou maîtriser une compétence technique [34].

Tout comme les autres systèmes de recommandation, ceux utilisés dans les EIAH doivent surmonter des défis tels que le problème du démarrage à froid, la nécessité d'adapter les recommandations au fil du temps pour refléter l'évolution des besoins de l'apprenant, et le respect de la vie privée de l'apprenant.

## 1.4. Travaux connexes

### 1.4.1. Travaux connexes : sur l'intégration des systèmes de recommandation dans les environnements informatiques pour l'apprentissage humain

De nombreux travaux ont été réalisés sur l'intégration des systèmes de recommandation dans les Environnements Informatiques pour l'Apprentissage Humain (EIAH). Ces travaux visent

généralement à aider les apprenants à trouver des ressources pédagogiques pertinentes, à proposer des activités d'apprentissage adaptées, ou à faciliter l'apprentissage collaboratif. Voici quelques exemples de travaux dans ce domaine :

a. **GraphSAGE :** est un système de recommandation pour l'apprentissage en ligne qui utilise les graphes de connaissances pour générer des recommandations. Il peut recommander des cours en ligne ou des ressources d'apprentissage en fonction de la structure du graphe de connaissances et des intérêts de l'apprenant [35].

b. **ITS :** Les Systèmes Tutoriels Intelligents (ITS) sont des systèmes qui utilisent l'intelligence artificielle pour fournir un apprentissage personnalisé. Ils peuvent adapter le contenu d'apprentissage, le rythme et la séquence en fonction des besoins spécifiques de chaque apprenant. De nombreux ITS intègrent des systèmes de recommandation pour proposer des ressources ou des activités d'apprentissage [36].

Ces travaux démontrent la diversité et le potentiel des systèmes de recommandation pour améliorer l'efficacité et l'efficience de l'apprentissage dans les EIAH. Cependant, de nombreux défis restent à relever, tels que l'adaptation des recommandations à l'évolution des besoins des apprenants, le respect de la vie privée et la gestion de la grande variété de ressources d'apprentissage disponibles.

### 1.4.2. Travaux connexes : Systèmes de recommandation éducatifs

Les systèmes de recommandation éducatifs sont un sous-domaine en pleine expansion, avec de nombreuses recherches en cours pour explorer comment ces systèmes peuvent être utilisés pour améliorer les résultats d'apprentissage. Voici quelques exemples de travaux pertinents dans ce domaine :

a. De Medio et ses collègues [31] proposent une solution visant à simplifier la recherche et la sélection de ressources d'apprentissage dans le domaine de l'éducation en ligne. Leur article met en lumière le défi d'accéder à plusieurs dépôts de ressources éducatives en ligne et présente MoodleRec, un système de recommandation hybride intégré au système de gestion de l'apprentissage Moodle. MoodleRec permet aux enseignants d'explorer ces dépôts de manière transparente et de recevoir des suggestions de ressources adaptées à leurs besoins spécifiques. En utilisant des stratégies de recommandation basées sur des mots-clés et des caractéristiques générées socialement,

MoodleRec vise à améliorer l'efficacité de la recherche de ressources d'apprentissage en ligne. L'article comprend également une étude expérimentale pour valider cette approche et discute de son importance dans le contexte de l'enseignement en ligne.

**b.** Monsalve-Pulido et al. [37] proposent une architecture d'un système de recommandation intelligent et autonome destiné à être implémenté dans tout environnement d'apprentissage virtuel, dans le but de fournir des recommandations efficaces de ressources numériques. Le document détaille l'architecture des dimensions intelligentes et autonomes du système de recommandation, présentant un modèle hybride de recommandation qui coordonne et gère les informations disponibles ainsi que les besoins spécifiques en recommandations pour déterminer les algorithmes appropriés à utiliser. Cette approche hybride permet l'intégration de diverses méthodes basées sur le filtrage collaboratif, le contenu ou les connaissances. Les données sont extraites de quatre sources principales : le contexte, les étudiants, le cours et les ressources numériques, en identifiant des variables telles que les styles d'apprentissage individuels, les données socio-économiques, les paramètres de connexion, et la localisation, entre autres. Des expérimentations ont été menées dans le cadre de la création d'un cours académique afin d'évaluer les capacités intelligentes et autonomes de cette architecture.

**c.** Li Bifeng et leurs collaborateurs [38] explorent une approche de recommandation personnalisée adaptée aux Massive Open Online Courses (MOOC), en combinant les technologies de deep learning et de big data. En utilisant le modèle Bidirectional Encoder Representations from Transformers (BERT), ils développent des stratégies visant à améliorer la précision du système de recommandation. Ils commencent par décrire l'acquisition et le prétraitement des données ouvertes. Ensuite, ils élaborent un cadre de modèle de recommandation en exploitant BERT et en intégrant un mécanisme d'auto-attention. Enfin, pour mieux appréhender les caractéristiques des cours, ils proposent une méthode d'apprentissage des différences de caractéristiques de domaine. Les résultats de leurs expériences mettent en évidence l'efficacité du modèle proposé par rapport à d'autres approches de recommandation.

Ces travaux mettent en évidence l'importance croissante des systèmes de recommandation éducatifs dans le paysage de l'apprentissage et de l'enseignement. Cependant, il reste encore

beaucoup à faire pour comprendre comment ces systèmes peuvent être conçus et utilisés de manière plus efficace pour soutenir l'apprentissage personnalisé et adaptatif.

### 1.4.3. Travaux connexes : Systèmes de recommandation de vidéos éducatives

Les systèmes de recommandation de vidéos éducatives sont une application spécifique des systèmes de recommandation éducatifs. Leur but est de recommander aux apprenants des vidéos qui correspondent à leurs besoins et préférences d'apprentissage. Voici quelques exemples de travaux de recherche dans ce domaine :

a. **Improving Video Ranking on E-Learning Platforms Driven by Syllabus : An Approach Using Fuzzy Logic**: Ce travail de P. Rishad et all (2021), aborde les défis liés à l'évaluation de la valeur éducative et au classement des vidéos sue les plateformes d'apprentissage en ligne basées sur les programmes. Il présente un système de recommandation Acharya, une plate-forme permettant aux utilisateurs de lier des vidéos de diverses plates-formes à des sujets de programme. Cependant, le défi réside dans le classement de ces vidéos en fonction de paramètres tels que les vues, les commentaires et la pertinence par rapport au programme. La solution proposée implique un système de classement flou prenant en compte des facteurs tels que le temps de création de la vidéo, la durée de visionnage, la durée, les commentaires, les évaluations du site Web parent, le temps d'intégration et l'analyse des sentiments. Cette approche globale vise à fournir un classement significatif des vidéos sur la plateforme E-learning [39].

b. **Learning Path Recommendation in MOOCs**: Ce travail de CL Tang and all (2020), explore comment recommander des chemins d'apprentissage personnalisés dans les cours en ligne ouverts et massifs (MOOCs). Le système de recommandation suggère une séquence de vidéos d'apprentissage qui est adaptée aux objectifs et aux préférences de l'apprenant [40].

c. **Enhancing Learning Experiences through a Hybrid Video Recommender System for Small Private Online Courses (SPOCs)**: Ce système de recommandation, développé par N Belarbi et all (2019), présente un système de recommandation vidéo conçu pour un petit cours privé en ligne (SPOC). Pour améliorer les expériences d'apprentissage, le système utilise des techniques d'adaptabilité, de personnalisation et de recommandation. Il utilise une approche de recommandation hybride, analysant le

comportement vidéo des utilisateurs lors de l'inscription au SPOC pour estimer leurs intérêts. Le processus implique la construction de profils d'utilisateurs basés sur des inférences implicites à partir des données de parcours, l'application de l'algorithme de clustering K-Means pour regrouper les utilisateurs ayant un comportement vidéo similaire et la recommandation de vidéos provenant de profils similaires pour améliorer l'engagement et la satisfaction des utilisateurs dans l'environnement d'apprentissage [41].

Ces travaux soulignent l'importance des systèmes de recommandation de vidéos éducatives dans l'amélioration de l'expérience d'apprentissage des utilisateurs. Cependant, de nombreux défis demeurent, tels que la personnalisation des recommandations pour répondre aux besoins individuels des apprenants, la prise en compte du contexte d'apprentissage, et l'assurance de la qualité des vidéos recommandées.

### 1.4.4. Synthèse

Les systèmes de recommandation ont été largement étudiés et utilisés dans de nombreux domaines, y compris l'éducation. Dans le contexte de l'apprentissage et de l'enseignement, les systèmes de recommandation visent à améliorer l'expérience d'apprentissage en proposant des ressources d'apprentissage, des activités, des pairs pour l'apprentissage collaboratif et des parcours d'apprentissage personnalisés qui sont pertinents pour chaque apprenant individuel.

Dans les Environnements Informatiques pour l'Apprentissage Humain (EIAH), des systèmes de recommandation ont été intégrés pour aider les apprenants à naviguer parmi les nombreuses ressources pédagogiques disponibles. Des travaux tels que GraphSAGE et les Systèmes Tutoriels Intelligents (ITS) ont démontré la variété et le potentiel des systèmes de recommandation dans ce contexte.

En ce qui concerne les vidéos éducatives, plusieurs travaux ont exploré comment les systèmes de recommandation peuvent aider à recommander des vidéos pertinentes aux apprenants. Par exemple, le projet EdutubePlus a développé un système de recommandation de vidéos éducatives en utilisant une combinaison de techniques de recommandation. De plus, des travaux ont été réalisés pour recommander des chemins d'apprentissage personnalisés dans les MOOCs, qui peuvent comprendre une séquence de vidéos d'apprentissage.

En somme, les travaux connexes montrent que les systèmes de recommandation peuvent jouer un rôle important dans l'amélioration de l'apprentissage et de l'enseignement. Cependant, il reste encore de nombreux défis à relever, notamment en ce qui concerne la personnalisation des recommandations, la prise en compte du contexte d'apprentissage, l'assurance de la qualité des ressources recommandées, et la protection de la vie privée des apprenants.

Après avoir exploré la conception et l'application des systèmes de recommandation, nous allons maintenant aborder une étape cruciale qui garantit leur pertinence et efficacité : l'évaluation de ces systèmes. Cette section fera la lumière sur les mécanismes et critères utilisés pour mesurer la performance des recommandations, assurant qu'elles répondent aux attentes et besoins des utilisateurs. Nous examinerons les diverses méthodologies d'évaluation, les métriques clés, ainsi que les défis inhérents à l'évaluation précise de tels systèmes complexes.

## 1.5. Évaluation des systèmes de recommandation

L'évaluation des systèmes de recommandation est une étape cruciale pour garantir leur efficacité et leur pertinence. Elle consiste généralement à examiner les aspects suivants :

a. **Pertinence des recommandations :** Une mesure clé de l'efficacité d'un système de recommandation est la pertinence des recommandations qu'il produit. Cela peut être évalué en mesurant la satisfaction des utilisateurs avec les recommandations, ou en comparant les recommandations à une liste de "bonnes" recommandations préétablies [42].

b. **Précision :** La précision est souvent mesurée en utilisant des méthodes comme la précision à K (precision@K), le rappel à K (recall@K), le score F1, ou l'AUC (aire sous la courbe ROC) [43]. Ces mesures comparent les recommandations faites par le système à un ensemble de référence de recommandations "correctes".

c. **Diversité et nouveauté :** Un bon système de recommandation doit être capable de proposer une variété de recommandations pour éviter de trop se focaliser sur un sous-ensemble limité d'items [3, 42, 44]. De plus, il doit être capable de recommander de nouveaux items que l'utilisateur n'a pas encore découverts.

d. **Robustesse :** Un système de recommandation doit être robuste face à différentes formes de manipulation [45, 46], comme le bourrage de popularité (où certains items sont artificiellement rendus plus populaires).

e. **Évolutivité :** Un système de recommandation doit être capable de gérer un grand nombre d'utilisateurs et d'items sans perdre en performance [47].

f. **Transparence et explicabilité :** Les utilisateurs sont généralement plus à l'aise avec un système de recommandation s'ils comprennent comment il fonctionne et pourquoi il a fait certaines recommandations [48, 49]. Par conséquent, l'évaluation de la transparence et de l'explicabilité du système peut être un aspect important.

g. **Satisfaction des utilisateurs :** En fin de compte, la mesure la plus importante de l'efficacité d'un système de recommandation peut être la satisfaction globale des utilisateurs [50]. Cela peut être mesuré par des enquêtes auprès des utilisateurs, ou par l'observation du comportement des utilisateurs (par exemple, combien de temps ils passent sur le site, combien de recommandations ils suivent, etc.).

L'évaluation des systèmes de recommandation peut se faire de deux manières : hors ligne (offline) et en ligne (online). Chacune de ces approches à ses propres avantages et inconvénients, et elles sont souvent utilisées ensemble pour obtenir une image complète de la performance d'un système.

### 1.5.1. Évaluation hors ligne

L'évaluation hors ligne est réalisée à l'aide de données historiques collectées auprès des utilisateurs [51]. Ces données sont généralement divisées en un ensemble d'apprentissage pour entraîner le système de recommandation, et un ensemble de test pour évaluer sa performance. L'évaluation hors ligne peut mesurer des aspects tels que la précision des recommandations (par exemple, en utilisant la précision@k, le rappel@k, l'AUC) [52], la diversité des recommandations, et l'évolutivité du système.

**Avantages :** L'évaluation hors ligne peut être réalisée rapidement et à moindre coût, car elle n'exige pas d'interaction en temps réel avec les utilisateurs.

**Inconvénients :** L'évaluation hors ligne ne peut pas mesurer directement la satisfaction des utilisateurs, ni comment les utilisateurs interagissent avec le système en temps réel. En outre,

les données historiques peuvent ne pas refléter fidèlement les préférences actuelles des utilisateurs.

### 1.5.2. Évaluation en ligne

L'évaluation en ligne est réalisée en testant le système de recommandation en temps réel avec des utilisateurs réels. Cette évaluation peut se faire par le biais de tests A/B, où différents utilisateurs ou groupes d'utilisateurs reçoivent des recommandations de différentes versions du système, et leurs interactions sont ensuite comparées [53]. L'évaluation en ligne peut mesurer des aspects tels que la satisfaction des utilisateurs, le taux de clics sur les recommandations, et le temps passé sur le site.

**Avantages :** L'évaluation en ligne peut fournir des informations précieuses sur le comportement réel des utilisateurs et leur satisfaction. Elle permet également de tester le système dans des conditions réelles, y compris l'impact de facteurs imprévus ou inconnus.

**Inconvénients :** L'évaluation en ligne peut être coûteuse et prendre beaucoup de temps. Elle nécessite également un nombre suffisant d'utilisateurs pour obtenir des résultats significatifs. En outre, elle peut présenter des risques si les utilisateurs réagissent négativement à certaines versions du système.

Alors que les sections précédentes ont mis en lumière les systèmes de recommandation sociale et les méthodes d'évaluation de ces systèmes, nous nous apprêtons à franchir une frontière technologique en abordant le domaine de "L'apprentissage intelligent". Cette transition représente le passage d'une focalisation sur les mécanismes de suggestion et d'évaluation externes vers une introspection dans les méthodes d'apprentissage renforcées par la technologie.

Dans les systèmes de recommandation sociale, nous avons vu comment l'interaction humaine et les données comportementales sont exploitées pour améliorer la pertinence des suggestions fournies aux utilisateurs. L'évaluation de ces systèmes nous a permis de mesurer leur efficacité et d'en déduire des implications pour le développement futur.

En progressant vers "L'apprentissage intelligent", nous entrons dans une ère où les technologies avancées telles que l'intelligence artificielle et l'analyse de données massives sont utilisées pour personnaliser et améliorer le processus d'apprentissage lui-même. Ici, l'accent est mis sur l'adaptabilité et la personnalisation de l'expérience éducative, exploitant des techniques

sophistiquées pour répondre aux besoins d'apprentissage individuels et optimiser le parcours pédagogique.

L'apprentissage intelligent fait donc écho aux concepts précédemment discutés, mais avec une application directe dans l'optimisation de l'expérience d'apprentissage. Alors que les systèmes de recommandation peuvent diriger un utilisateur vers un contenu éducatif, l'apprentissage intelligent vise à rendre l'expérience d'apprentissage elle-même plus intuitive, efficace et enrichissante. C'est une évolution naturelle qui s'appuie sur les fondements de la recommandation et de l'évaluation pour élaborer des stratégies d'apprentissage plus sophistiquées et centrées sur l'apprenant.

## 1.6. Apprentissage intelligent

L'apprentissage intelligent est un concept qui se réfère à l'utilisation de technologies avancées pour faciliter et améliorer l'apprentissage [54, 55]. Nous examinerons les recherches existantes sur les fondements théoriques de l'apprentissage intelligent, y compris les principes de la personnalisation, de l'adaptabilité et de la contextualisation. Ensuite, nous aborderons l'étude de différentes technologies qui sous-tendent l'apprentissage intelligent, tels que l'intelligence artificielle, l'apprentissage machine, l'internet des objets et le cloud computing [56].

### 1.6.1. Fondements théoriques de l'apprentissage intelligent

L'apprentissage intelligent repose sur un certain nombre de principes clés. Le premier est la personnalisation [57, 58], qui vise à adapter l'expérience d'apprentissage aux besoins et aux préférences individuels de chaque apprenant. Cela peut impliquer des ajustements en termes de rythme, de style d'apprentissage, de contenu et d'évaluation. La littérature a largement documenté l'impact positif de la personnalisation sur l'engagement et la réussite des apprenants.

Le second principe est l'adaptabilité [57, 58], qui se réfère à la capacité du système d'apprentissage à évoluer en fonction des retours et des performances des apprenants. Les systèmes d'apprentissage adaptatifs utilisent des données en temps réel pour ajuster continuellement l'expérience d'apprentissage, offrant ainsi un parcours personnalisé et flexible.

Le troisième principe est la contextualisation, qui implique l'intégration du contexte d'apprentissage dans le processus d'enseignement [57, 58]. Cela peut signifier l'incorporation d'éléments de la vie réelle, l'adaptation à la culture ou aux environnements spécifiques de l'apprenant, ou la connexion avec des expériences d'apprentissage antérieures.

### 1.6.2. Technologies sous-jacentes à l'apprentissage intelligent

Une gamme de technologies avancées soutient l'apprentissage intelligent, y compris l'intelligence artificielle (IA), Le machine learning ou apprentissage automatique (ML), l'Internet des objets (IoT) et le cloud computing [56].

L'IA et le ML sont essentiels pour permettre la personnalisation et l'adaptabilité. Par exemple, les algorithmes ML peuvent être utilisés pour modéliser le comportement d'apprentissage, prévoir les performances et générer des recommandations personnalisées. L'IA peut également être utilisée pour créer des tuteurs virtuels et des agents conversationnels qui peuvent faciliter l'apprentissage interactif.

L'IoT offre des possibilités intéressantes pour la contextualisation. Les capteurs et les dispositifs IoT peuvent collecter des données contextuelles qui peuvent être utilisées pour améliorer l'expérience d'apprentissage. Par exemple, les données de localisation peuvent être utilisées pour fournir des informations d'apprentissage pertinentes en fonction du lieu de l'apprenant.

Le cloud computing fournit l'infrastructure nécessaire pour stocker et traiter les grandes quantités de données générées par les systèmes d'apprentissage intelligent. Il offre également la flexibilité et l'évolutivité nécessaires pour soutenir des expériences d'apprentissage personnalisées et adaptatives.

En somme, l'apprentissage intelligent est une approche prometteuse qui utilise des technologies avancées pour créer des expériences d'apprentissage plus personnalisées, adaptatives et contextuelles. Dans les sections suivantes, nous explorerons comment ces principes et technologies peuvent être appliqués à l'exploration des données pédagogiques et à l'analyse de l'apprentissage.

Tandis que la section "L'apprentissage intelligent" a mis en exergue la convergence de la technologie avancée et des méthodes pédagogiques pour créer des expériences d'apprentissage personnalisées et adaptatives, nous nous acheminons à présent vers une exploration plus approfondie de ces concepts au sein des "Environnements éducatifs soutenus par le data mining". L'intersection entre ces deux domaines est essentielle : l'apprentissage intelligent forge le cadre dans lequel le data mining éducatif opère, en fournissant les données nécessaires pour affiner et personnaliser l'enseignement.

Le data mining éducatif représente la machinerie analytique qui permet d'exploiter les vastes bases de données générées dans les environnements d'apprentissage numériques. C'est la clé qui déverrouille le potentiel de l'apprentissage intelligent, en fournissant les insights nécessaires pour adapter l'expérience éducative aux besoins individuels des apprenants. En ce sens, le data mining n'est pas simplement un outil d'analyse, mais un catalyseur qui rend l'apprentissage intelligent opérationnel et mesurable.

Alors que nous orientons vers le "Data mining dans les environnements éducatifs", nous allons examiner comment les techniques de fouille de données sont spécifiquement appliquées pour décortiquer, comprendre et améliorer les processus d'apprentissage. Nous verrons comment ces méthodologies transforment des données brutes en connaissances pédagogiques précieuses et comment elles peuvent être utilisées pour soutenir des décisions éducatives éclairées, en alignant la technologie avec les théories de l'apprentissage pour réaliser une éducation plus efficace et engagée.

## 1.7. Conclusion

À travers ce chapitre, nous avons entrepris un voyage à travers le paysage en constante évolution des systèmes de recommandation, des réseaux sociaux en ligne, et de l'analyse de l'apprentissage dans les contextes éducatifs. Notre exploration a révélé non seulement la complexité technique et théorique de ces domaines, mais aussi leur impact profond et croissant sur notre société numérique.

Nous avons constaté que les systèmes de recommandation, au-delà de leur utilisation initiale dans le commerce électronique et les plateformes de streaming, jouent un rôle crucial dans l'éducation en personnalisant et en améliorant l'expérience d'apprentissage. Les systèmes de recommandation sociale, en intégrant les dimensions sociales dans les algorithmes de recommandation, ouvrent de nouvelles voies pour des suggestions plus pertinentes et contextualisées.

Parallèlement, les réseaux sociaux en ligne continuent de transformer radicalement nos modes de communication, d'interaction et de partage d'informations, jouant un rôle central dans la diffusion de connaissances et la formation de communautés. Toutefois, ces plateformes soulèvent des questions importantes en termes de vie privée, de sécurité des données, et d'éthique, qui nécessitent une attention et une régulation continues.

Dans le domaine de l'éducation, l'analyse de l'apprentissage et le data mining éducatif se sont révélés être des outils puissants pour comprendre et répondre aux besoins des apprenants. Ils permettent une personnalisation poussée de l'enseignement, mais soulèvent également des défis en termes d'interprétation des données, de respect de la vie privée, et d'équité dans l'éducation.

Ce chapitre souligne la nécessité d'une approche équilibrée qui reconnaît le potentiel immense de ces technologies tout en étant attentif à leurs implications éthiques et sociales. Alors que nous continuons à naviguer dans l'ère numérique, il est impératif de développer et d'appliquer ces technologies de manière responsable, en veillant à ce qu'elles servent l'intérêt collectif et contribuent à une société plus informée, équitable et connectée.

En somme, les systèmes de recommandation, les réseaux sociaux en ligne, et l'analyse de l'apprentissage dans les environnements éducatifs représentent des domaines d'innovation et de croissance continus. Leur intégration judicieuse dans nos systèmes éducatifs et sociaux promet non seulement d'enrichir l'expérience d'apprentissage, mais aussi de façonner la façon dont nous interagissons, partageons et nous développons dans un monde interconnecté

# CHAPITRE 2 :

## DATA MINING ET ANALYSE DE L'APPRENTISSAGE : RÉVOLUTIONNER L'ÉDUCATION À L'ÈRE NUMÉRIQUE

# Chapitre 2

# Data mining et analyse de l'apprentissage : révolutionner l'éducation à l'ère numérique

## *2.1.* Introduction

Dans un monde numérique en rapide évolution, l'éducation se trouve à un carrefour technologique. Les environnements d'apprentissage génèrent chaque jour une quantité phénoménale de données, allant des interactions en ligne des apprenants à leurs performances académiques. L'exploitation de ces données à travers le data mining éducatif et l'analyse de l'apprentissage promet de révolutionner les méthodes pédagogiques et de transformer l'expérience éducative elle-même.

Ce chapitre explore comment le data mining éducatif et l'analyse de l'apprentissage, bien que distincts, convergent pour améliorer l'efficacité et la personnalisation de l'éducation. Le data mining éducatif utilise des techniques avancées pour découvrir des patterns et des insights cachés dans les données éducatives, permettant ainsi une meilleure compréhension des comportements d'apprentissage et des résultats scolaires. Parallèlement, l'analyse de l'apprentissage applique ces découvertes pour optimiser les décisions pédagogiques et améliorer les résultats des apprenants.

Nous examinerons les méthodologies spécifiques employées dans chaque domaine, de la modélisation prédictive au clustering, et discuterons de leur rôle dans la détection des apprenants à risque, la personnalisation des parcours d'apprentissage, et la conception de systèmes de recommandation éducatifs. Des exemples concrets illustreront la mise en œuvre de ces techniques dans des contextes réels, mettant en lumière leur potentiel pour enrichir l'éducation.

De plus, ce chapitre traitera des défis éthiques et techniques associés à l'utilisation de ces technologies. Questions de confidentialité, de consentement éclairé, et d'équité seront abordées pour souligner l'importance d'une approche éthique et responsable dans l'exploitation des données éducatives.

En somme, nous offrirons une vue d'ensemble sur la manière dont le data mining éducatif et l'analyse de l'apprentissage peuvent travailler de concert pour non seulement comprendre mais aussi améliorer la manière dont l'apprentissage est facilité et perçu dans l'ère numérique.

## *2.2.* Data mining dans les environnements éducatifs

Data mining éducatif (EDM) est une méthode qui permet d'extraire des informations précieuses à partir de grandes quantités de données éducatives [59]. Dans ce chapitre, nous passerons en revue les travaux existants sur le data mining éducatif, en mettant l'accent sur les techniques utilisées, comme le clustering, la classification, la régression, l'association et la détection d'anomalies. De plus, nous examinerons également les applications de l'exploration des données pédagogiques dans différents contextes d'apprentissage.

### 2.2.1. Introduction au data mining appliqué aux contextes éducatifs

Avec la montée en puissance de l'éducation numérique, une quantité phénoménale de données est générée chaque jour par des apprenants du monde entier [59, 60]. Ces données, bien qu'abondantes, restent souvent inexploitées. Cependant, elles renferment des trésors d'informations qui, si analysés correctement, peuvent révolutionner la manière dont l'éducation est dispensée et perçue. C'est ici que le data mining éducatif entre en jeu, apportant des méthodologies et des outils pour extraire des informations pertinentes de ces vastes bases de données.

Le data mining éducatif, à la croisée des chemins entre l'informatique, la statistique et l'éducation, a gagné en importance ces dernières années [59], comme illustré dans la figure 1. Il permet non seulement de comprendre les comportements d'apprentissage des apprenants, mais aussi d'identifier des tendances, des schémas et des domaines nécessitant une intervention. Dans un monde où l'éducation s'efforce d'être de plus en plus centrée sur l'apprenant, ces insights sont inestimables [60].

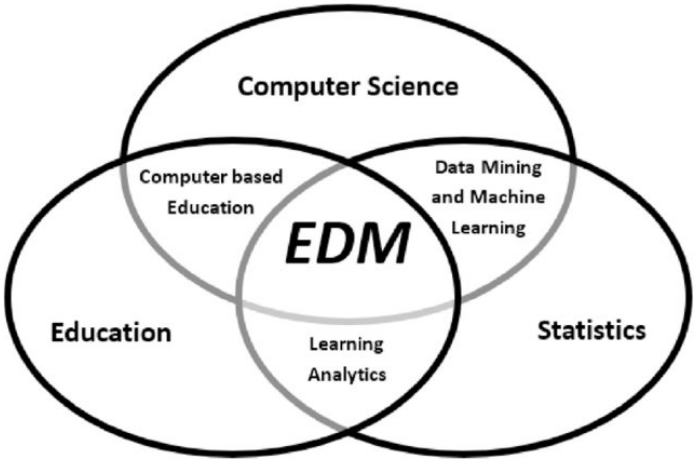


*Figure 1: Data mining éducatif (Romero and Ventura 2013)*

Parallèlement, les systèmes de recommandation, bien connus dans des domaines tels que le commerce électronique et le streaming de médias, trouvent également leur place dans le paysage éducatif. Au lieu de suggérer un produit ou un film, ils recommandent des ressources pédagogiques, des cours ou des méthodes d'apprentissage. Dans un environnement éducatif, ces systèmes visent à personnaliser l'expérience d'apprentissage, en veillant à ce que chaque apprenant reçoive du contenu adapté à ses besoins, ses préférences et son niveau de compétence.

L'intersection du data mining éducatif et des systèmes de recommandation ouvre la voie à une synergie prometteuse. Si nous pouvons correctement identifier les modèles d'apprentissage des apprenants, nous sommes alors mieux équipés pour recommander des ressources qui sont non seulement pertinentes mais aussi bénéfiques pour leur progression académique.

L'objectif de ce chapitre est d'explorer cette interaction. Nous chercherons à comprendre comment le data mining éducatif, en identifiant les caractéristiques et les modèles d'apprentissage des apprenants, peut fournir des données essentielles pour alimenter et affiner les systèmes de recommandation. À travers cette exploration, nous visons à démontrer comment une intégration judicieuse de ces deux domaines peut améliorer l'expérience d'apprentissage, la rendant plus personnalisée, plus efficace et, en fin de compte, plus enrichissante pour l'apprenant.

### 2.2.2. Contexte et principe du data mining éducatif

L'évolution rapide des technologies de l'information et de la communication a profondément transformé le paysage éducatif mondial [60-62]. Avec l'augmentation des cours en ligne, des plateformes d'apprentissage électronique et des outils éducatifs numériques, une immense quantité de données est produite quotidiennement. Le data mining éducatif se présente comme une solution prometteuse pour naviguer dans ce déluge d'informations et en extraire des connaissances utiles.

#### 2.2.2.1. Définition et portée du data mining éducatif

Le data mining éducatif peut être défini comme le processus d'application de techniques de data mining pour analyser et interpréter les vastes bases de données produites dans les contextes éducatifs [61, 63] tel que représenté dans la figure 2. Alors que le data mining traditionnel se concentre sur la conversion de données brutes en informations utiles, le data mining éducatif vise spécifiquement à améliorer les résultats et les pratiques d'apprentissage.




*Figure 2 : Étapes du processus de data mining.*

La portée de ce domaine est vaste. Elle s'étend de la compréhension des comportements d'apprentissage des apprenants et de la prédiction de leurs performances futures à l'identification des domaines d'intervention nécessaires pour les éducateurs.

#### 2.2.2.2. Méthodes couramment utilisées

a. **Clustering** : Cette méthode cherche à regrouper les apprenants en fonction de leurs comportements ou caractéristiques similaires, permettant aux éducateurs de comprendre les différents types ou groupes d'apprenants présents dans leur classe [64, 65].

b. **Classification** : Elle vise à prédire une catégorie ou un résultat pour les apprenants. Par exemple, un modèle de classification pourrait être formé pour prédire si un apprenant

est susceptible de réussir ou d'échouer dans un cours donné, en se basant sur son comportement d'apprentissage ou ses performances antérieures [64, 65].

c. **Règles d'association :** Ces règles identifient des relations entre différentes variables. Par exemple, elles peuvent révéler que les apprenants qui participent activement aux forums de discussion ont tendance à obtenir de meilleures notes dans les tests [65].

#### 2.2.2.3. Avantages potentiels pour les éducateurs et les apprenants

a. **Pour les éducateurs [61, 66] :**

- Une meilleure compréhension des besoins et des comportements des apprenants.
- La possibilité d'intervenir tôt pour les apprenants à risque.
- Des insights pour améliorer le contenu et la méthodologie d'enseignement.

b. **Pour les apprenants [61, 66] :**

- Une expérience d'apprentissage plus adaptée à leurs besoins.
- Des retours d'information plus précis et opportuns sur leurs performances.
- Un environnement d'apprentissage qui évolue en fonction de leurs comportements et de leurs besoins.

En somme, le data mining éducatif, par son approche centrée sur la donnée, offre un éclairage précieux pour optimiser l'expérience éducative. En comprenant et en répondant mieux aux besoins des apprenants, on peut espérer une amélioration des résultats d'apprentissage et une plus grande satisfaction tant pour les éducateurs que pour les apprenants.

### 2.2.3. Caractéristiques et modèles d'apprentissage des apprenants

Le cœur de l'éducation réside dans la compréhension des apprenants. Chaque apprenant est unique, avec ses propres méthodes, rythmes et préférences d'apprentissage [67]. Dans le contexte numérique, où l'interaction face à face est souvent limitée, comprendre ces nuances devient un défi. Toutefois, les environnements d'apprentissage en ligne génèrent une richesse de données qui, si elles sont correctement analysées, peuvent offrir des insights précieux sur les apprenants.

#### 2.2.3.1. Types de données collectées dans les environnements d'apprentissage en ligne

a. **Données comportementales** : Elles capturent les actions des apprenants [68], comme les clics, la navigation, le temps passé sur une page, les participations aux forums, etc.

b. **Données d'évaluation :** Cela inclut les scores des tests, les soumissions de devoirs, les feedbacks des pairs, etc… [69].

c. **Données d'interaction :** Ces données proviennent des discussions en ligne, des commentaires, des chats, et d'autres formes d'interactions entre les apprenants ou entre les apprenants et les enseignants [68].

d. **Données démographiques et d'arrière-plan** : Ces informations peuvent inclure l'âge, le sexe, la formation antérieure, le lieu, etc… [68].

#### 2.2.3.2. Comment le data mining peut identifier des modèles d'apprentissage

En exploitant les techniques de data mining, il est possible d'analyser d'immenses bases de données éducatives pour identifier des tendances, des modèles de comportement et des caractéristiques distinctives parmi les apprenants. Le clustering permet de dévoiler des groupes d'apprenants partageant des habitudes d'apprentissage similaires, tels que ceux qui privilégient les supports visuels face aux apprenants orientés vers le texte. La classification aide à prédire la catégorie à laquelle un apprenant pourrait appartenir, tel qu'un potentiel décrocheur ou un apprenant réussissant, en se basant sur ses interactions et comportements. De même, les règles d'association révèlent les liens entre certaines actions ou caractéristiques et des résultats académiques spécifiques, comme la corrélation entre une participation assidue aux forums et une meilleure performance académique.

#### 2.2.3.3. Exemples de caractéristiques d'apprentissage

a. **Styles d'apprentissage :** Certains apprenants peuvent être des apprenants visuels, auditifs ou kinesthésiques.

b. **Préférences de contenu** : Tandis que certains apprenants pourraient graviter autour de vidéos ou d'animations, d'autres pourraient préférer des articles, des études de cas ou des podcasts.

c. **Rythmes d'étude :** Certains apprenants peuvent être des lève-tôt, apprenant principalement le matin, tandis que d'autres peuvent être plus actifs la nuit.

d. **Niveaux d'interaction :** Il peut s'agir d'apprenants très engagés qui participent activement aux forums et aux discussions, ou d'apprenants plus passifs qui consomment principalement du contenu.

En reconnaissant et en comprenant ces caractéristiques et modèles d'apprentissage, les éducateurs et les plateformes d'apprentissage en ligne peuvent mieux répondre aux besoins individuels des apprenants, conduisant à une expérience d'apprentissage plus enrichissante et efficace.

### 2.2.4. Synergie entre le data mining éducatif et les systèmes de recommandation

L'éducation numérique d'aujourd'hui s'appuie fortement sur des algorithmes et des systèmes sophistiqués pour offrir des expériences d'apprentissage optimales. En combinant les puissantes capacités d'analyse du data mining éducatif avec les méthodes adaptatives des systèmes de recommandation [70], il est possible de créer une expérience d'apprentissage profondément personnalisée et centrée sur l'apprenant.

La figure 3 illustre l'architecture d'un système avancé d'apprentissage numérique qui intègre trois composantes essentielles : le data mining éducatif, l'analyse de l'apprentissage et le système de recommandation. Conçu pour optimiser l'expérience éducative dans un environnement numérique, ce modèle représente l'interaction dynamique entre la collecte et le traitement des données, l'analyse approfondie des comportements et des styles d'apprentissage des apprenants, et la génération de recommandations personnalisées pour améliorer l'efficacité pédagogique. Chaque étape du processus est soigneusement détaillée, depuis la collecte initiale des données jusqu'à la boucle de feedback qui permet une amélioration continue. Ce schéma offre un aperçu complet de la manière dont les technologies modernes peuvent être harmonieusement combinées pour enrichir l'apprentissage numérique, mettant en lumière les interactions complexes entre les différents éléments du système pour créer une expérience d'apprentissage enrichissante et personnalisée.

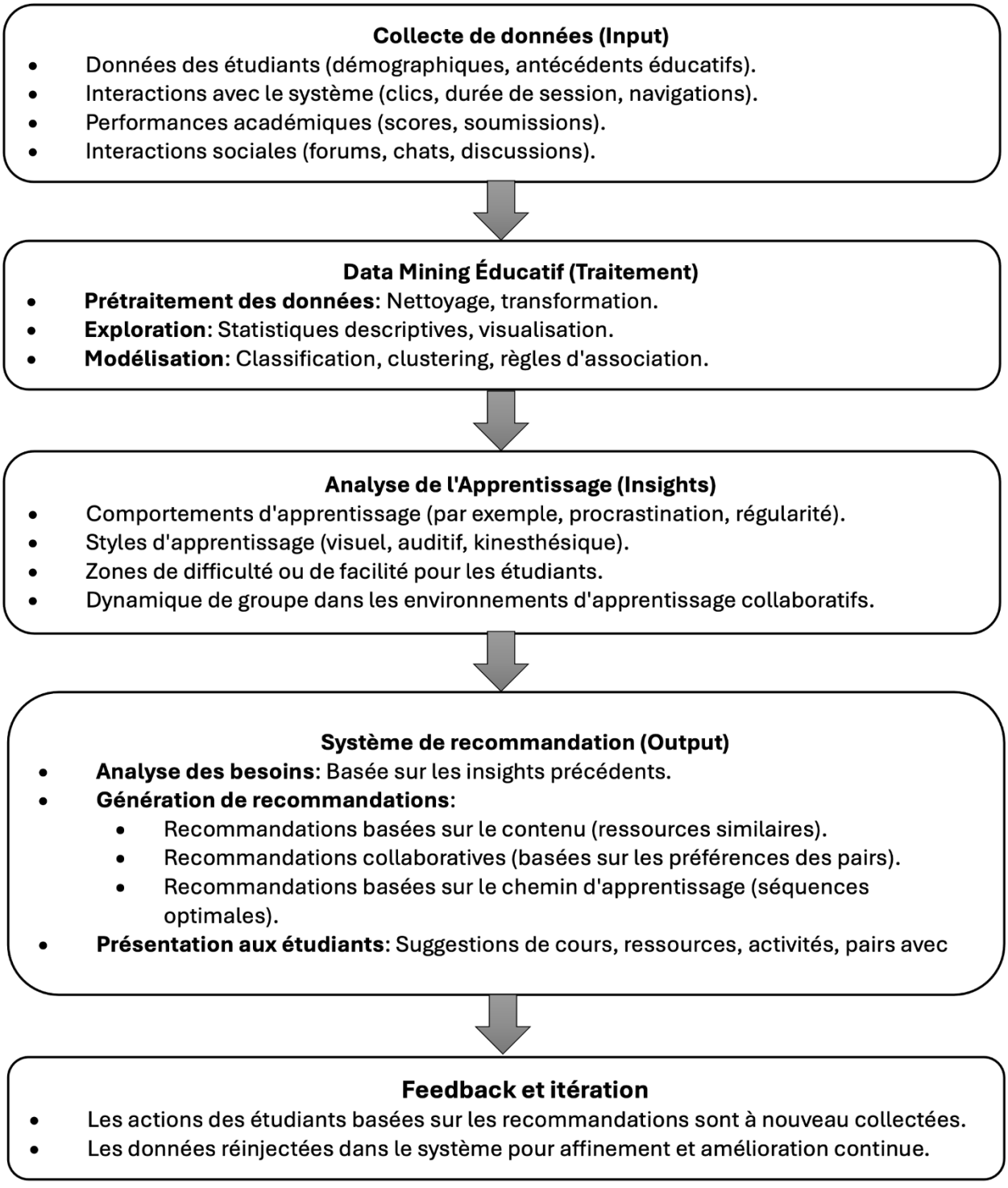


*Figure 3: Conception d'un schéma intégré pour l'analyse de données éducatives, l'analyse de l'apprentissage et les systèmes de recommandation dans l'environnement de l'apprentissage numérique*

#### 2.2.4.1. Comment les insights du data mining peuvent renforcer les recommandations

Les analyses réalisées via le data mining éducatif peuvent fournir une multitude d'informations précieuses sur les apprenants [71] :

- Identification des styles d'apprentissage.
- Compréhension des sujets qui captivent le plus les apprenants.
- Reconnaissance des points où les apprenants rencontrent généralement des difficultés.

Lorsque ces informations sont intégrées dans les systèmes de recommandation, elles permettent [38] de/d':

- Cibler les recommandations en fonction des besoins spécifiques et des préférences de chaque apprenant.
- Ajuster dynamiquement les contenus proposés en fonction des progrès et des interactions de l'apprenant.
- Prévoir et de prévenir les éventuels décrochages ou difficultés en recommandant des ressources ou des interventions appropriées.

#### 2.2.4.2. Exemples d'institutions ou de plateformes qui ont intégré le data mining éducatif dans leurs systèmes de recommandation

a. **Coursera** : Cette plateforme de MOOC utilise le data mining pour analyser les comportements des apprenants, tels que les taux de clics, les temps de visionnage et les interactions sur les forums. En utilisant ces données, Coursera adapte ses recommandations pour suggérer des cours qui correspondent le mieux aux intérêts et aux besoins des apprenants [72].

b. **Khan Academy** : En utilisant des analyses pour suivre la progression des apprenants, Khan Academy recommande des vidéos et des exercices basés sur les performances passées et actuelles des apprenants. Cela garantit que les apprenants reçoivent des ressources adaptées à leur niveau de compétence [73].

c. **DreamBox Learning** : Cette plateforme d'apprentissage des mathématiques ajuste dynamiquement son contenu en fonction du niveau de compétence de l'apprenant,

utilisant le data mining pour analyser les performances des apprenants en temps réel et adapter le contenu en conséquence [74].

#### 2.2.4.3. Avantages potentiels pour les apprenants

a. **Recommandations plus précises** : En comprenant les besoins, les préférences et les progrès de chaque apprenant, les recommandations deviennent plus ciblées et pertinentes [68].

b. **Parcours d'apprentissage personnalisé** : Les apprenants ne sont plus soumis à une approche uniforme. Au lieu de cela, ils reçoivent un parcours d'apprentissage adapté à leurs propres rythmes et besoins [68].

c. **Meilleur engagement** : Lorsque les apprenants se voient proposer du contenu qui résonne avec eux et qui répond à leurs besoins, ils sont plus susceptibles de rester engagés et motivés [68].

En somme, la fusion du data mining éducatif avec les systèmes de recommandation offre une opportunité inestimable d'améliorer l'éducation numérique, en plaçant l'apprenant au cœur de l'expérience d'apprentissage.

### 2.2.5. Défis et préoccupations

Tandis que le data mining éducatif et les systèmes de recommandation offrent de formidables opportunités pour améliorer l'éducation, ils apportent aussi avec eux un ensemble de défis et de préoccupations qui doivent être abordés pour assurer une utilisation éthique et efficace.

#### 2.2.5.1. Les questions éthiques et de confidentialité liées à l'utilisation du data mining en éducation

a. **Consentement informé** : Les apprenants et leurs tuteurs doivent être conscients de la manière dont leurs données sont collectées, analysées et utilisées. Le consentement doit être donné librement et peut être révoqué à tout moment.

b. **Protection des données** : Avec la collecte de grandes quantités de données sur les apprenants, il est impératif de garantir que ces données soient stockées en toute sécurité et protégées contre tout accès non autorisé.

c. **Utilisation des données** : Il est crucial d'assurer que les données recueillies soient utilisées uniquement à des fins éducatives et ne soient pas vendues ou utilisées à des fins commerciales sans le consentement explicite de l'individu.

#### 2.2.5.2. La nécessité d'un équilibre entre personnalisation et sur-adaptation (risque de créer des "bulles d'apprentissage")

a. **Isolation** : Une personnalisation excessive peut conduire à ce que les apprenants soient enfermés dans des "bulles" où ils ne sont exposés qu'à des contenus qu'ils connaissent déjà ou qu'ils aiment, limitant ainsi leur exposition à de nouveaux concepts ou perspectives.

b. **Dépendance** : La sur-reliance sur les systèmes de recommandation peut également rendre les apprenants dépendants des suggestions, diminuant leur capacité à explorer et à chercher activement de nouvelles connaissances par eux-mêmes.

#### 2.2.5.3. Limitations techniques et nécessité d'une mise à jour constante

a. **Précision des modèles** : Aucun modèle ou algorithme n'est parfait. Il peut y avoir des occasions où les recommandations ne sont pas exactes ou pertinentes, ce qui peut frustrer ou dérouter les apprenants.

b. **Évolution des besoins** : Les besoins et les comportements des apprenants évoluent avec le temps. Les systèmes doivent être régulièrement mis à jour pour s'assurer qu'ils restent pertinents.

c. **Complexité technique** : L'intégration du data mining éducatif dans les systèmes de recommandation peut être techniquement complexe, nécessitant des compétences spécialisées et une maintenance continue.

En conclusion, bien que le data mining éducatif et les systèmes de recommandation aient le potentiel de transformer positivement l'éducation numérique, il est essentiel de naviguer avec précaution, en tenant compte des préoccupations éthiques, des défis techniques et des risques pour garantir que les avantages l'emportent sur les inconvénients.

Dans la section sur le data mining, nous avons exploré comment les données brutes sur les comportements et les performances des apprenants sont recueillies et analysées pour révéler des tendances et des informations cachées. Ce processus est essentiel pour poser les fondations

d'une analyse plus approfondie - il s'agit de la phase de diagnostic où les données sont triées, nettoyées et préparées pour l'examen.

En progressant vers la section le learning analytics (LA), ou «analyse de l'apprentissage» , nous nous engageons dans une démarche proactive d'application de ces connaissances. Ici, l'accent est mis sur l'usage des informations obtenues à partir du data mining pour améliorer les pratiques pédagogiques, personnaliser l'apprentissage pour répondre aux besoins individuels des apprenants, et, en fin de compte, pour améliorer les résultats éducatifs. L'analyse de l'apprentissage vise à interpréter les données de manière à informer les décisions et les actions - c'est une transition de l'observation à l'action.

## *2.3.* Analyse de l'apprentissage

Après avoir exploré les méthodologies du data mining éducatif, nous comprenons maintenant comment extraire des informations précieuses des vastes bases de données éducatives. Cependant, extraire des informations n'est que la première étape. Dans cette section, nous nous concentrerons sur l'application de ces informations pour améliorer l'apprentissage - un domaine connu sous le nom d'analyse de l'apprentissage. Alors que le data mining éducatif met l'accent sur l'extraction, l'analyse de l'apprentissage se concentre sur l'interprétation et l'application des connaissances pour améliorer l'expérience éducative.

### 2.3.1. Analyse de l'apprentissage et son rôle dans l'éducation moderne

À mesure que le monde de l'éducation évolue, la capacité à comprendre en profondeur le processus d'apprentissage des apprenants devient non seulement souhaitable, mais essentielle. L'ère numérique a apporté avec elle une abondance de données - une trace digitale de chaque clic, réponse, interaction et discussion dans les environnements d'apprentissage en ligne. Mais que faisons-nous de toutes ces informations ? Comment les exploitons-nous pour améliorer l'éducation ? C'est ici qu'intervient l'analyse de l'apprentissage.

L'analyse de l'apprentissage, ou "Learning Analytics", est une approche systématique de l'interprétation des vastes quantités de données générées par les apprenants dans le cadre de leur parcours éducatif [75]. Bien plus que de simples statistiques ou graphiques, l'analyse de l'apprentissage vise à produire des insights significatifs qui peuvent influencer positivement le processus d'enseignement et d'apprentissage [61].

Dans le contexte de l'éducation moderne, où la personnalisation et l'efficacité deviennent des mots d'ordre, l'analyse de l'apprentissage offre une promesse : celle d'une éducation plus réactive, qui s'adapte aux besoins individuels de chaque apprenant. Que ce soit pour identifier un apprenant qui pourrait avoir besoin d'une aide supplémentaire, pour comprendre quels éléments d'un cours en ligne sont les plus engageants, ou pour prédire les tendances et les comportements futurs, l'analyse de l'apprentissage se positionne comme un outil clé pour les éducateurs du 21e siècle [61].

L'objectif de cette partie est de plonger au cœur de l'analyse de l'apprentissage. Nous explorerons son essence, les méthodes qu'elle emploie, les diverses applications où elle peut apporter une valeur ajoutée, ainsi que les défis et dilemmes éthiques qu'elle présente. En naviguant à travers ces pages, nous espérons offrir un aperçu complet et éclairant de ce domaine passionnant, tout en mettant en lumière son potentiel révolutionnaire dans l'amélioration de l'éducation pour tous.

### 2.3.2. Définition et contexte historique

#### 2.3.2.1. Origines et évolution de l'analyse de l'apprentissage

L'analyse de l'apprentissage, bien qu'elle puisse sembler être un concept récent, trouve ses racines dans des décennies de recherche en psychologie de l'éducation, en statistiques et, plus récemment, en informatique [61, 76]. Avec l'essor des technologies éducatives dans les années 2000, la nécessité de comprendre et d'analyser les comportements des apprenants dans les environnements d'apprentissage numériques est devenue primordiale [77].

Les premiers systèmes de gestion de l'apprentissage (LMS) ont fourni une plateforme pour collecter des données sur les interactions des apprenants, mais c'est avec l'explosion des MOOCs (Massive Open Online Courses) et de l'apprentissage en ligne à grande échelle que l'analyse de l'apprentissage a vraiment gagné en traction [78]. Ces vastes bases de données, combinés à des avancées dans les domaines du big data et de l'analytique, ont permis aux éducateurs et aux chercheurs de mieux comprendre les dynamiques d'apprentissage à une échelle sans précédent.

#### 2.3.2.2. Distinction entre l'analyse de l'apprentissage, le data mining éducatif et d'autres disciplines connexes

a. **Analyse de l'apprentissage :** Cette discipline se concentre principalement sur l'application de méthodes statistiques et analytiques pour comprendre et améliorer le processus d'apprentissage des apprenants [61]. Elle s'appuie sur des données générées par les apprenants au cours de leurs interactions avec les ressources et les environnements d'apprentissage.

b. **Data Mining Éducatif** : Bien que similaire dans son objectif de comprendre les processus d'apprentissage, le data mining éducatif tend à utiliser des techniques de machine learning et d'exploration de données pour découvrir des motifs et des modèles dans de grands bases de données éducatives. Il s'agit plus d'une approche orientée vers la découverte de nouveaux patterns que d'une analyse prescriptive [61].

c. **Analyse académique** : Cette discipline se penche sur la performance académique des apprenants, en évaluant les résultats et en cherchant des moyens d'améliorer les programmes et les politiques éducatives [79]. Elle est plus centrée sur les résultats institutionnels et les politiques que sur les comportements individuels des apprenants.

En somme, bien que ces disciplines partagent des objectifs communs, elles diffèrent par leurs méthodes, leurs foci et leurs applications, comme illustré dans la figure 4. L'analyse de l'apprentissage est devenue un pont essentiel entre la technologie éducative et la pédagogie, fournissant des insights qui permettent d'affiner l'enseignement et l'apprentissage pour le bénéfice des apprenants dans l'éducation moderne.

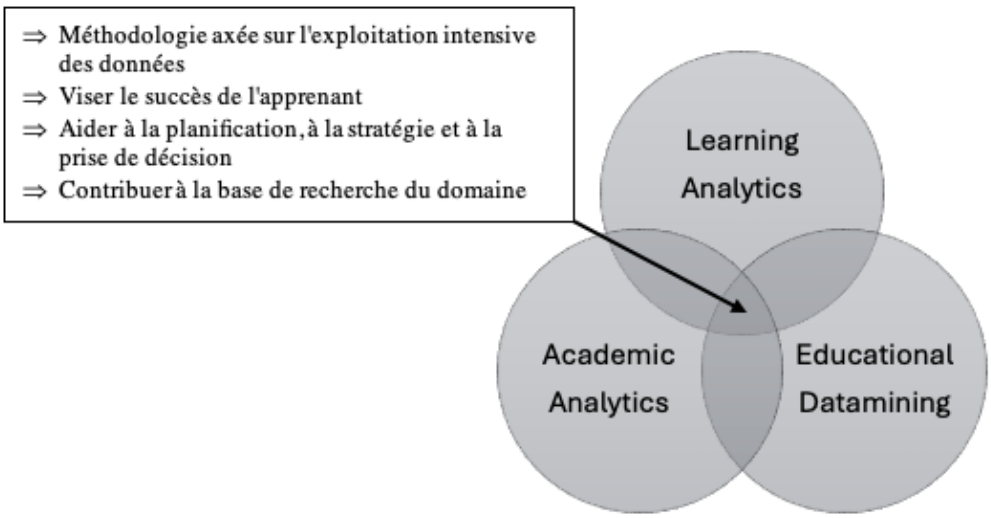


*Figure 4: Intersection des domaines d'analyse en éducation*

### 2.3.3. Objectifs et importance de l'analyse de l'apprentissage

#### 2.3.3.1. Amélioration de l'expérience de l'apprenant

L'analyse de l'apprentissage offre une occasion unique de comprendre comment chaque apprenant interagit avec le contenu, les plateformes et ses pairs. En apprenant ces interactions, il est possible de repérer des tendances, des préférences et des défis qui peuvent être adressés de manière proactive [80]. Par exemple, si un apprenant passe un temps disproportionné sur une section particulière d'un cours, cela peut indiquer une difficulté de compréhension, ou peut-être une fascination pour le sujet. Dans les deux cas, une intervention adaptée - que ce soit une ressource supplémentaire ou un encouragement - peut améliorer considérablement l'expérience de cet apprenant.

#### 2.3.3.2. Détection précoce des apprenants à risque

Un des avantages les plus tangibles de l'analyse de l'apprentissage est sa capacité à identifier rapidement les apprenants qui peuvent être en difficulté [81]. En surveillant des indicateurs clés, comme la fréquence de connexion, l'achèvement des devoirs ou l'interaction avec les ressources en ligne, il est possible de repérer des signes d'un éventuel décrochage ou d'une détresse académique. Une intervention précoce, basée sur ces signaux, peut fournir un soutien crucial aux apprenants avant que les problèmes ne s'aggravent.

#### 2.3.3.3. Optimisation des environnements et ressources d'apprentissage

L'analyse de l'apprentissage ne se limite pas à la compréhension des apprenants ; elle peut aussi éclairer la façon dont les ressources pédagogiques et les environnements d'apprentissage sont utilisés [82, 83]. Cela permet aux éducateurs de repérer les contenus qui fonctionnent bien, ceux qui nécessitent des améliorations ou ceux qui sont largement ignorés. En conséquence, les contenus et les plateformes peuvent être continuellement affinés pour mieux répondre aux besoins des apprenants.

#### 2.3.3.4. Informer et éclairer les décisions pédagogiques

Avec des données concrètes sur la manière dont les apprenants apprennent, les éducateurs et les administrateurs sont mieux équipés pour prendre des décisions pédagogiques éclairées [61, 84]. Que ce soit pour ajuster la structure d'un cours, introduire de nouvelles méthodes d'enseignement ou investir dans de nouvelles technologies, l'analyse de l'apprentissage fournit une base solide de données pour guider ces choix. Elle garantit que les décisions prises soient en phase avec les besoins réels des apprenants, conduisant à des résultats d'apprentissage améliorés.

En conclusion, l'analyse de l'apprentissage sert de boussole dans le paysage souvent complexe de l'éducation moderne. Elle guide, informe et illumine, garantissant que chaque apprenant ait l'opportunité d'une éducation qui est non seulement de haute qualité, mais aussi profondément personnalisée pour ses besoins uniques.

### 2.3.4. Méthodes et techniques employées

#### 2.3.4.1. Introduction aux méthodes quantitatives et qualitatives

L'analyse de l'apprentissage utilise un éventail de méthodes pour déchiffrer les données d'apprentissage. Ces méthodes peuvent être largement catégorisées en deux types :

a. **Méthodes quantitatives** : Ces méthodes se concentrent sur l'analyse de données numériques pour obtenir des informations statistiquement significatives. Elles permettent d'extraire des tendances, des corrélations et des patterns à partir de grandes bases de données [85].

b. **Méthodes qualitatives** : Ces méthodes analysent des données non numériques, telles que le texte ou les enregistrements audio/vidéo, pour obtenir des insights qualitatifs.

Elles sont souvent utilisées pour comprendre les perceptions, les opinions et les motivations des apprenants [85].

#### 2.3.4.2. Modélisation prédictive

La modélisation prédictive utilise des algorithmes pour prédire des résultats futurs en fonction des données historiques. Dans le contexte de l'analyse de l'apprentissage, elle peut être utilisée pour prévoir les performances académiques d'un apprenant, identifier les apprenants à risque de décrochage ou prédire l'efficacité d'une ressource pédagogique [86].

#### 2.3.4.3. Analyse de réseau social

L'analyse de réseau social examine les relations et les interactions entre les apprenants pour comprendre les dynamiques de groupe et les structures communautaires. Elle peut révéler des informations sur les leaders d'opinion, les groupes d'apprentissage collaboratif ou identifier les apprenants isolés qui pourraient bénéficier d'un soutien supplémentaire [87].

#### 2.3.4.4. Analyse du sentiment

Cette technique évalue le ton et l'émotion des communications des apprenants, que ce soit dans les forums, les chats ou les évaluations [88]. Elle permet de comprendre le sentiment général des apprenants vis-à-vis d'un cours, d'une ressource ou d'un sujet, offrant des insights sur leur engagement et leur satisfaction.

#### 2.3.4.5. Visualisation des données

La visualisation transforme les données complexes en représentations graphiques intuitives, facilitant l'interprétation et la prise de décision [89]. Les tableaux de bord, les graphiques, les cartographies de chaleur et d'autres représentations visuelles peuvent aider les éducateurs à rapidement saisir des patterns, des tendances ou des domaines préoccupants.

En somme, l'analyse de l'apprentissage s'appuie sur une combinaison de méthodes et de techniques pour offrir une vue complète et nuancée de l'apprentissage des apprenants. Ces outils, lorsqu'ils sont utilisés judicieusement, ont le potentiel de transformer radicalement la manière dont l'éducation est dispensée et perçue, créant des environnements d'apprentissage plus réactifs, inclusifs et efficaces.

### 2.3.5. Applications pratiques

#### 2.3.5.1. Tableaux de bord et systèmes de feedback pour les apprenants

Les tableaux de bord personnalisés offrent aux apprenants une vision claire de leurs progrès, de leurs réalisations et de leurs domaines à améliorer [90]. Ces interfaces, souvent intégrées aux systèmes de gestion de l'apprentissage (LMS), permettent aux apprenants de suivre leurs performances en temps réel, de comparer leurs résultats à ceux des normes de classe ou à leurs propres objectifs, et de recevoir des feedbacks instantanés. Cela encourage la responsabilité et la motivation de l'apprenant, tout en leur fournissant les outils nécessaires pour orienter de manière autonome leur parcours d'apprentissage.

#### 2.3.5.2. Outils d'alerte et de monitoring pour les enseignants

En exploitant les données des apprenants, les enseignants peuvent recevoir des alertes en temps réel concernant les apprenants qui peuvent être en difficulté ou qui présentent des signes de décrochage. Ces outils peuvent également mettre en évidence des domaines spécifiques du curriculum qui posent des problèmes à un grand nombre d'apprenants, suggérant ainsi des opportunités d'ajustement ou de renforcement. Cette surveillance proactive permet aux enseignants d'intervenir rapidement, offrant un soutien ciblé lorsque cela est le plus nécessaire [91].

#### 2.3.5.3. Analyse des contenus et des parcours pédagogiques

L'analyse de l'apprentissage peut révéler des informations précieuses sur l'efficacité de certains contenus ou méthodes pédagogiques. Par exemple, si un module particulier de cours voit un taux d'échec élevé ou une faible interaction, il peut nécessiter une révision. De même, l'analyse peut identifier des chemins d'apprentissage optimaux, suggérant des séquences de cours ou de ressources qui fonctionnent le mieux pour différents types d'apprenants.

#### 2.3.5.4. Impact sur la conception et la livraison des cours

Avec une compréhension profonde des besoins, des comportements et des préférences des apprenants, les concepteurs de cours peuvent créer des expériences d'apprentissage plus alignées sur les besoins des apprenants. Cela peut se traduire par l'adoption de méthodes pédagogiques innovantes, l'intégration de technologies émergentes ou la création de parcours d'apprentissage adaptatifs qui s'ajustent en fonction des progrès et des préférences de chaque

apprenant. De plus, la livraison du cours peut être ajustée en temps réel en fonction des feedbacks des apprenants, assurant ainsi que l'enseignement demeure pertinent et engageant.

En conclusion, les applications pratiques de l'analyse de l'apprentissage ne se limitent pas à la simple collecte de données, mais s'étendent à la transformation concrète de l'expérience éducative. Qu'il s'agisse d'outils de monitoring pour les enseignants ou de feedbacks en temps réel pour les apprenants, l'objectif final reste le même : créer un environnement d'apprentissage optimal, éclairé par des données et centré sur l'apprenant.

### 2.3.6. Défis et préoccupations éthiques

#### 2.3.6.1. Problèmes liés à la vie privée et à la protection des données

Avec la collecte et l'analyse de données sur les comportements, les interactions et les performances des apprenants, la question de la vie privée et de la sécurité des données devient primordiale. Qui a accès à ces données ? Comment sont-elles stockées et protégées ? Les apprenants ont-ils la possibilité de refuser la collecte de leurs données ou de demander leur suppression ? Le non-respect de la vie privée ou des fuites de données peuvent avoir des conséquences graves, tant pour les institutions que pour les individus concernés.

#### 2.3.6.2. Risques d'interprétation erronée ou de mésusage des données

Les données, si elles ne sont pas correctement interprétées, peuvent mener à des conclusions erronées. Un taux de clics élevé sur un module de cours, par exemple, peut être interprété comme un signe d'engagement, alors qu'il pourrait en réalité refléter une confusion ou un contenu mal conçu. De plus, il y a un danger potentiel dans la stigmatisation ou le profilage d'apprenants basé uniquement sur des analyses de données, sans prendre en compte le contexte plus large.

#### 2.3.6.3. Nécessité d'une infrastructure adéquate et de compétences spécialisées

La mise en œuvre efficace de l'analyse de l'apprentissage nécessite une infrastructure robuste pour la collecte, le stockage et l'analyse des données [92]. De plus, l'interprétation et l'application des insights tirés de ces données nécessitent des compétences spécialisées. Non seulement les éducateurs doivent-ils être formés pour utiliser ces outils, mais ils doivent aussi avoir une compréhension approfondie des implications pédagogiques des données.

#### 2.3.6.4. Questions éthiques et implications pour les parties prenantes

Au-delà de la vie privée et de la sécurité, il existe d'autres questions éthiques associées à l'analyse de l'apprentissage. Par exemple, est-il éthique de catégoriser ou de profiler les apprenants sur la base de leurs données ? Les apprenants sont-ils informés de la manière dont leurs données sont utilisées et des avantages qu'ils peuvent en tirer ? Et comment garantir que l'analyse de l'apprentissage ne renforce pas les inégalités existantes, mais contribue plutôt à les atténuer ?

En conclusion, bien que l'analyse de l'apprentissage offre un potentiel considérable pour améliorer l'expérience éducative, elle doit être abordée avec précaution, réflexion et intégrité. Assurer une mise en œuvre éthique et responsable de l'analyse de l'apprentissage est essentiel pour maintenir la confiance des apprenants et garantir que les avantages l'emportent sur les risques.

### 2.3.7. Synthèse finale

L'analyse de l'apprentissage, bien que relativement nouvelle, a déjà laissé une empreinte indélébile sur le paysage éducatif. Elle a ouvert une fenêtre sur les mécanismes complexes et souvent mystérieux de l'apprentissage, offrant aux éducateurs, institutions et plateformes d'apprentissage des outils pour créer des expériences éducatives plus réactives, efficaces et centrées sur l'apprenant.

Au cours de ce chapitre, nous avons exploré les origines, les méthodes et les applications de l'analyse de l'apprentissage. De sa définition et de ses racines historiques à ses applications pratiques et ses implications éthiques, nous avons tenté de peindre un tableau complet de ce domaine en plein essor.

L'avenir de l'analyse de l'apprentissage est riche en opportunités. Avec l'évolution rapide de la technologie, on peut s'attendre à ce que l'analyse devienne encore plus précise, avec des modèles prédictifs plus affinés et des visualisations plus intuitives. Toutefois, cette voie est également semée de défis, en particulier en ce qui concerne les questions éthiques et la protection des données des apprenants.

Il est donc essentiel que, à mesure que nous avançons, nous adoptions une approche centrée sur l'apprenant, en veillant toujours à ce que les méthodes et techniques utilisées servent l'intérêt supérieur de l'apprentissage. Il faut veiller à ce que les données soient utilisées de manière

responsable, en respectant la vie privée des apprenants et en garantissant que les insights tirés de ces données soient utilisés pour améliorer, et non compromettre, l'expérience éducative.

En fin de compte, l'analyse de l'apprentissage est un outil, et comme tout outil, son véritable potentiel est déterminé par la manière dont il est utilisé. Avec une mise en œuvre judicieuse, elle a le potentiel de transformer l'éducation, en créant des expériences d'apprentissage plus personnalisées, inclusives et efficaces pour tous.

## *2.4.* Conclusion

Au terme de ce chapitre, nous avons exploré en profondeur le rôle transformateur du data mining éducatif et de l'analyse de l'apprentissage dans le paysage éducatif moderne. Nous avons vu comment ces techniques peuvent décrypter d'immenses volumes de données pour en extraire des insights précieux qui améliorent l'expérience d'apprentissage. Ces outils ne se contentent pas de révéler les tendances et les modèles de comportement des apprenants, mais ils aident également à façonner des environnements éducatifs plus réactifs et personnalisés.

En regardant vers l'avenir, le prochain chapitre, "Système de recommandation utilisant la confiance et les évaluations des utilisateurs", s'appuiera sur ces fondations pour examiner comment le data mining et l'analyse de l'apprentissage peuvent enrichir les systèmes de recommandation. Nous explorerons des techniques avancées où la confiance et les évaluations des utilisateurs sont intégrées pour affiner la pertinence et l'efficacité des recommandations, assurant ainsi que chaque apprenant reçoit des suggestions adaptées à ses besoins uniques et à ses préférences personnelles.

Puis, dans le chapitre suivant, "Optimisation des recommandations de vidéos éducatives : filtrage hybride combinant FBC et FC pour une personnalisation avancée", nous irons encore plus loin en détaillant comment combiner différents types de filtrage pour optimiser la recommandation de contenus éducatifs, spécifiquement les vidéos. Ce chapitre démontrera comment le data mining éducatif et l'analyse de l'apprentissage peuvent être exploités pour créer un système de recommandation hybride qui non seulement répond aux intérêts des apprenants mais anticipe également leurs besoins d'apprentissage futurs.

Ainsi, en progressant dans ces chapitres, nous continuerons à construire sur le cadre établi ici pour déployer des technologies éducatives qui ne sont pas seulement innovantes, mais également centrées sur l'apprenant, adaptatives et, surtout, efficaces. Le voyage à travers le data

mining éducatif et l'analyse de l'apprentissage est un témoignage de l'engagement continu envers l'amélioration de l'éducation par la technologie, garantissant que chaque étape de l'apprentissage est aussi enrichissante et personnalisée que possible.

# CHAPITRE 3

## SYSTÈME DE RECOMMANDATION UTILISANT LA CONFIANCE ET LES ÉVALUATIONS DES UTILISATEURS.

# Chapitre 3

# Système de recommandation utilisant la confiance et les évaluations des utilisateurs

## *3.1.* Introduction

Les plateformes regorgeant d'items (vidéos, livres, …) se doivent d'orienter les utilisateurs vers des contenus adaptés à leurs besoins et préférences. Face à cet enjeu, les systèmes de recommandation se distinguent comme des outils essentiels, proposant des contenus personnalisés grâce à des méthodologies et logiciels spécifiques visant à améliorer l'expérience utilisateur [93]. Parmi les diverses approches existantes pour élaborer ces recommandations, le filtrage collaboratif est largement privilégié [94, 95]. Cependant, cette méthode, malgré sa popularité, présente des limitations, notamment en termes de précision des prédictions dues à des lacunes telles que la rareté des données ou le problème de démarrage à froid [96, 97]. Afin de pallier ces défis et d'affiner les recommandations, des données complémentaires, comme celles issues des réseaux sociaux (relations d'amitié, groupes d'appartenance, niveaux de confiance), peuvent être intégrées au filtrage collaboratif. En effet, les liens basés sur la confiance jouent un rôle majeur, renforçant la pertinence des suggestions en mettant en évidence les similarités entre utilisateurs [98, 99].

De nombreuses recherches ont souligné l'efficacité supérieure du filtrage collaboratif en termes de précision prédictive par rapport aux méthodes de recommandation classiques [100]. Le filtrage collaboratif, basé principalement sur l'historique des interactions des utilisateurs, est largement plébiscité en raison de sa simplicité d'application dans divers contextes réels. Cette méthode découle de l'observation selon laquelle les individus ont souvent tendance à se fier aux avis de leurs pairs. Par exemple, avant d'acheter un produit, une personne se tournera souvent vers les avis de ceux qui ont déjà fait cet achat. Ce phénomène est courant dans diverses situations du quotidien, qu'il s'agisse du choix d'un film, d'un livre, etc. Plusieurs systèmes ont tenté d'intégrer la dimension de la confiance dans leurs modèles pour pallier les défis tels que la rareté des données ou le démarrage à froid, en combinant l'influence de la confiance sociale à leurs stratégies [101].

Cependant, les tentatives antérieures dans ce domaine ont rencontré des obstacles dans la modélisation complète des interactions réciproques entre les utilisateurs. En effet, il n'a pas été clairement identifié comment les opinions des autres influencent un utilisateur, et vice versa, en prenant en compte que les préférences d'une personne en tant que donneur ou receveur de confiance peuvent différer [102]. Il semble donc judicieux, lors de la prédiction des préférences d'un utilisateur pour un item, de considérer à la fois ses propres préférences et la manière dont il est perçu par les autres. Or, les approches antérieures ont souvent simplifié cette complexité en traitant les utilisateurs de manière unidimensionnelle ou en séparant les deux rôles. Une véritable distinction entre le rôle d'un utilisateur en tant que donneur de confiance et en tant que receveur semble donc nécessaire pour améliorer la génération des recommandations.

Indépendamment de la méthode d'apprentissage utilisée dans les systèmes de recommandation, une historique d'interactions entre les utilisateurs et les éléments du système est essentiel. Ces interactions antérieures, souvent désignées sous le terme de "feedback", peuvent être soit explicites, soit implicites. De nombreuses approches s'appuient principalement sur un feedback explicite pour établir un lien de confiance entre les utilisateurs, où ils manifestent leurs préférences clairement, contrairement à ceux qui interagissent de manière plus subtile ou implicite. Bien que les techniques actuelles modélisent efficacement les relations explicites, elles tendent à négliger les interactions implicites entre utilisateurs ayant des similitudes mais non directement connectés dans un réseau de confiance. La dimension locale des relations sociales met en avant les interactions entre un utilisateur et son entourage immédiat, alors que la dimension globale met en lumière sa réputation au sein d'un réseau plus large [103]. Alors que la tendance naturelle des utilisateurs est de chercher des recommandations auprès de leurs proches, ils peuvent aussi être influencés par les opinions de figures plus influentes ou réputées au sein du réseau. Cela suggère que l'exploitation conjointe des opinions locales et globales pourrait optimiser la pertinence des systèmes de recommandation.

Dans ce chapitre, nous introduisons une méthode de recommandation qui intègre à la fois les dimensions globales et locales des relations de confiance (**G**lobal **Lo**cal **T**rust with **M**atrix **F**actorization "GLoTMF"). Cette approche innove en fusionnant les données d'évaluation et de confiance, en mettant l'accent sur la manière dont les évaluations sont influencées par les comportements de confiance au sein des réseaux sociaux. Contrairement à d'autres études qui tendaient à simplement juxtaposer différents types de données, notre méthode vise à représenter plus fidèlement les influences réciproques entre utilisateurs.

Notre modèle emploie une factorisation matricielle pour déchiffrer les préférences des utilisateurs basées sur la confiance. Elle appréhende distinctement les préférences des deux catégories d'utilisateurs en analysant à la fois leurs interactions explicites et implicites. Plus concrètement, elle évalue les préférences des utilisateurs en se basant à la fois sur les évaluations et les relations de confiance explicites, contrairement aux méthodes traditionnelles qui se fondent uniquement sur les liens établis. La structure du réseau de confiance est également exploitée pour décrypter les liens entre les différents acteurs, permettant ainsi de modéliser les interactions subtiles entre les utilisateurs.

Les tests effectués sur des données réelles attestent de la performance de GLoTMF en matière de précision prédictive. Nos résultats suggèrent que cette méthode offre des avantages significatifs, notamment face aux défis des données de démarrage à froid et de rareté, comparativement à d'autres approches existantes.

## *3.2.* Différentes approches existantes

La combinaison des méthodes de filtrage collaboratif basées sur des modèles avec les relations de confiance pour renforcer la justesse des recommandations a gagné en popularité ces dernières années. Notamment, l'utilisation de la factorisation matricielle se démarque en raison de sa précision élevée et de sa capacité à adresser efficacement le défi de la rareté des données, comparée à d'autres techniques [104]. Plusieurs spécialistes se sont penchés sur cette méthode afin de déceler les propriétés cachées des utilisateurs et des éléments évalués, combinant les liens sociaux entre utilisateurs avec les données d'évaluation via diverses approches.

Une contribution notable est le modèle "SoRec", qui intègre des données issues de réseaux sociaux dans une approche de factorisation matricielle probabiliste. Cette approche analyse conjointement la matrice d'évaluation et la matrice de confiance sociale, en mutualisant les propriétés latentes des utilisateurs [100]. Bien que leurs tests aient démontré une performance supérieure par rapport à la factorisation matricielle standard et à d'autres approches exploitant les relations de confiance, ce modèle ne prend pas véritablement en compte les mécanismes réels de recommandation. Dans cette optique, ces mêmes chercheurs ont développé le modèle "RSTE" [105], envisageant la décision d'évaluation d'un utilisateur comme un compromis entre ses préférences personnelles et celles de ses contacts directs. Bien que leurs résultats soient prometteurs, ce modèle laisse de côté l'influence réelle des contacts directs sur les propriétés latentes d'un utilisateur et ne considère pas la propagation de la confiance à travers le réseau.

Les scientifiques ont amélioré ce modèle en intégrant la propagation de la confiance, aboutissant à "SocialMF" [106]. Ce modèle tient compte de l'impact social en reliant les propriétés cachées d'un utilisateur à celles de ses contacts directs au sein du réseau social. Par ailleurs, face à l'absence d'informations de confiance, un autre modèle, "SoReg", a été introduit [107, 108]. Celui-ci réalise une factorisation matricielle tout en exploitant la régularisation sociale basée à la fois sur les matrices utilisateur-item et sur les liens sociaux affirmés. Ce qui le distingue des recherches antérieures en matière de recommandation basée sur la confiance est sa capacité à distinguer entre une relation de confiance et une simple relation d'amitié, encourageant ainsi les préférences d'un utilisateur à converger vers celles de ses amis sur le réseau.

Les approches "SoReg", "SocialMF", "STE" et "SoRec" poursuivent essentiellement le même dessein [109, 110]. Cependant, "TrustMF", qui est particulièrement pertinent pour notre étude, met l'accent sur l'apprentissage distinct des préférences selon les différents contextes pour anticiper les évaluations. Selon ce modèle, lors de la génération d'une évaluation, il semble plus judicieux de prendre en compte simultanément les préférences de celui qui fait confiance et celles du bénéficiaire de cette confiance. Néanmoins, "TrustMF" peine à percevoir les liens sous-jacents entre l'utilisateur faisant confiance et celui bénéficiant de cette confiance lorsqu'ils ne sont pas directement connectés socialement. De plus, il néglige la confiance collective, une lacune que nous souhaitons combler avec notre nouveau modèle[111].

## *3.3.* Approche proposée

### 3.3.1. Description du problème

Nous supposons qu'il existe un système de recommandation qui implique *m* utilisateurs et *n* items, nous introduisons quelques notations utilisées pour modéliser le problème de recommandation dans cette étude. Considérons $U = \{u_1, u_2, \ldots, u_n\}$ comme l'ensemble des utilisateurs et $V = \{v_1, v_2, \ldots, v_m\}$ comme l'ensemble des items, où *n* représente le total des utilisateurs et *m* le nombre d'items. *R*, une matrice de dimension *nxm*, symbolise les évaluations données par les utilisateurs aux items. L'élément $R_{i,j}$ indique la notation attribuée à l'item $v_j$ par l'utilisateur $u_i$. Chaque utilisateur note un ensemble spécifique d'items en se basant sur une échelle déterminée par le système de recommandation. Les coordonnées pour lesquelles une évaluation a été donnée dans la matrice *R* sont représentées par $\Omega = \{(\boldsymbol{i}, \boldsymbol{j}): \boldsymbol{R_{i,j}} \neq \boldsymbol{0}\}$.

$T \in R^{nxn}$ Représente la matrice des liens de confiance entre utilisateurs, où $T_{i,k}$ est un coefficient compris entre 0 et 1 qui évoque la force du lien de confiance entre les utilisateurs $u_i$ et $u_k$. Les coordonnées au sein de la matrice *T* où une relation de confiance est présente sont symbolisées par $\psi = \{(i,k): T_{i,k} \neq 0\}$. Dans cette étude, nous utilisons une technique de factorisation matricielle. Ainsi, $B_i \in R^k$ désigne le vecteur de préférence (de dimension *K*) pour l'utilisateur $u_i$ en tant qu'émetteur de confiance, et $E_i \in R^k$ représente le vecteur de préférence (aussi de dimension *K*) pour l'utilisateur $u_i$ en tant que récepteur de confiance. $V_i \in R^k$ , quant à lui, est le vecteur caractéristique (de dimension *k*) de l'élément $v_j$. L'objectif de cette recherche est donc le suivant : étant donné un ensemble d'évaluations d'éléments *R* par des utilisateurs et un ensemble de valeurs de confiance *T* entre utilisateurs, comment peut-on utiliser la factorisation matricielle pour appréhender les préférences spécifiques des utilisateurs et les attributs des éléments afin de prédire avec précision l'évaluation de l'élément $v_j$ par l'utilisateur $u_i$.

### 3.3.2. Modèle de factorisation matricielle

Le modèle de factorisation matricielle part de l'idée qu'il existe des facteurs latents qui modulent les évaluations d'un utilisateur. Ces facteurs façonnent le vecteur de préférences de l'utilisateur, qui dépend de l'interaction de ces facteurs avec ledit utilisateur. Cette hypothèse permet de découvrir les évaluations manquantes dans la matrice d'évaluation à partir des évaluations connues. Cette technique décompose la matrice d'évaluation *R* en deux matrices d'ordre inférieur *K* qui sont la matrice des caractéristiques latentes de l'utilisateur U et la matrice des caractéristiques latentes de l'élément *V*, c'est-à-dire $R \approx U^T V$ (comme le montre la figure 5), où les matrices *U* et *V* de faible dimension sont inconnues et doivent être prédites. L'objectif de la technique de factorisation matricielle est donc d'apprendre les matrices des caractéristiques latentes *U* et *V* et de les utiliser par la suite pour prédire les évaluations manquantes en résolvant le problème d'optimisation suivant :

$$\min_{U,V} \sum_{i=1}^{n} \sum_{j=1}^{m} W_{i,j} \left(R_{i,j} - U_i^T V_j\right)^2 + \lambda(\|U\|_F^2 + \|V\|_F^2) \qquad (1)$$

$U_i \in R^k$ est le vecteur représentant les caractéristiques latentes des préférences de l'utilisateur $u_i$, tandis que $V_j \in R^k$ symbolise le vecteur des caractéristiques latentes associées à l'item $v_j$. *K* quantifie le nombre de ces caractéristiques latentes. Pour éviter le surajustement, un paramètre de régulation, $\lambda$, est utilisé. Ce paramètre module la complexité du modèle en ajoutant

le terme $\lambda(\|U\|_F^2 + \|V\|_F^2)$ où $\|.\|_F^2$ représente la "norme de Frobenius". Parallèlement, $W \in R^{nxm}$ définit la matrice des poids, avec $W_{i,j}$ *r*eprésentant l'importance de l'évaluation donnée par l'utilisateur $u_i$ pour l'item $v_j$. Habituellement, $W_{i,j}$ est défini comme 1 si une évaluation est présente $R_{i,j} \neq 0$. Cependant, $W$ peut également encapsuler d'autres types d'informations, comme des opinions implicites ou des mesures de similarité entre utilisateurs et items. L'illustration de cette factorisation est dépeinte dans la figure 5 qui suit.

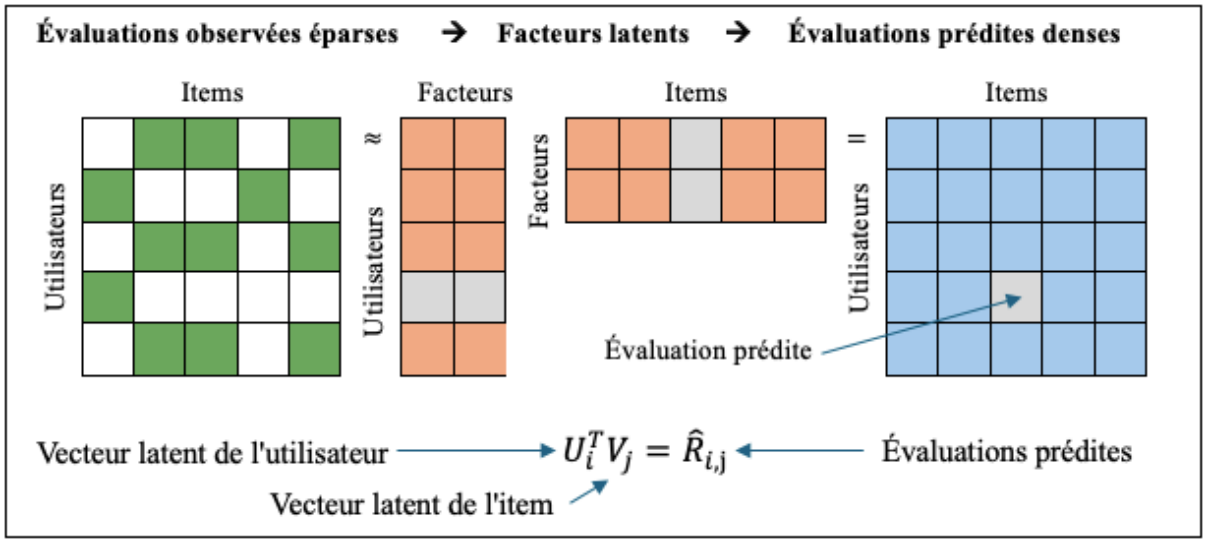


*Figure 5: Principe de fonctionnement de la technique de factorisation matricielle*

### 3.3.3. Étapes de la construction du modèle proposé

#### 3.3.3.1. Modélisation de l'information globale dans les réseaux sociaux

Au sein des réseaux sociaux, les informations disponibles peuvent servir de baromètre à la réputation d'un utilisateur. Cette réputation est un critère particulièrement intéressant car elle octroie aux systèmes de recommandation une source d'information supplémentaire, amplifiant ainsi leur potentiel. Plusieurs algorithmes ont été conçus pour estimer cette valeur de réputation, parmi lesquels "PageRank" est l'un des plus notoires. Ces algorithmes permettent de hiérarchiser les utilisateurs en s'appuyant sur une vue d'ensemble des interactions au sein des réseaux sociaux. Dans cette optique, la valeur $r_i \in \{1,2,\dots,\mathrm{N}\}$ représente le rang de réputation d'un utilisateur $u_i$ de sorte que $r_i = 1$ signifie que l'utilisateur $u_i$ est perçu comme le plus influent ou le plus réputé au sein du réseau. La relation entre le poids et le rang peut être exprimée par :

$$w_i = f(r_i) = \frac{1}{1 + log(r_i)} \qquad (2)$$

où " $f$ " est une fonction définissant comment le rang se traduit en poids ou importance dans le système de recommandation.

Dès lors que la fonction $(f)$ assure que la valeur de la réputation $w_i$ se situe entre [0, 1], elle est en décroissance vis-à-vis de $r_i$. Autrement dit, plus le rang d'un utilisateur est élevé, plus sa valeur de réputation est conséquente. Par conséquent, pour intégrer l'information des réseaux sociaux, les valeurs de réputation des utilisateurs peuvent servir à pondérer la pertinence de leurs recommandations. Ainsi, l'équation (3) peut être reformulée comme suit:

$$\min_{B,E,V} \sum_{(i,j)\in\Omega} w_i \left(R_{i,j} - g(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j)\right)^2 + \lambda(\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2) \qquad (3)$$

Dans cette factorisation matricielle, une valeur élevée de $w_i$ (qui signifie une forte réputation de $u_i$) poussera le terme $(\alpha B_i^T V_j + (1-\alpha) E_i^T V_j)$ à s'ajuster étroitement à l'évaluation $R_{i,j}$. Inversement, une valeur faible de $w_i$ permettra à ce même terme de n'offrir qu'une approximation de $R_{i,j}$.

#### 3.3.3.2. Modélisation de l'information locale dans les réseaux sociaux

L'information locale reflète les inclinations propres à l'utilisateur dans ses doubles rôles de donneur et de receveur de confiance, déterminées à travers les interactions directes avec d'autres utilisateurs. Les préférences du donneur de confiance et celles du receveur de confiance sont formulées en intégrant à la fois les évaluations explicites et les relations de confiance avérées. La structure des liens au sein du réseau de confiance local offre une évaluation des connexions entre les donneurs et les receveurs de confiance. Cette structure sert de cadre pour appréhender les interactions subtiles de l'utilisateur avec ses pairs, en tenant compte de leurs inclinations en tant que donneurs et receveurs de confiance.

##### 3.3.3.2.1. Modélisation des interactions explicites entre les utilisateurs

Dans cette section, nous expliquons comment produire des évaluations et des relations basées sur la confiance à partir de différentes perspectives d'utilisateurs, comme illustré à la figure 6. Ici, l'influence sociale sur les évaluations d'un utilisateur est bidirectionnelle. Cela signifie que l'évaluation d'un utilisateur n'est pas seulement façonnée par ceux qui lui font confiance, mais

aussi par ceux en qui il a confiance. Des études ont validé cette observation, suggérant que l'impact des personnes ayant confiance en un utilisateur dans la prédiction de l'évaluation peut être aussi significatif que celui de ceux en qui l'utilisateur a confiance, enrichissant ainsi la précision de la prédiction des évaluations.

La figure 6 illustre le modèle de confiance suggéré qui décrit comment les évaluations d'un utilisateur, $U_i$, sont influencées par d'autres utilisateurs en qu'il a confiance, à travers $B_i^T V_j$. De plus, cette figure met en avant comment les opinions d'un utilisateur, $U_i$, peuvent influencer les décisions des utilisateurs qui lui accordent leur confiance, via $E_i^T V_j$.

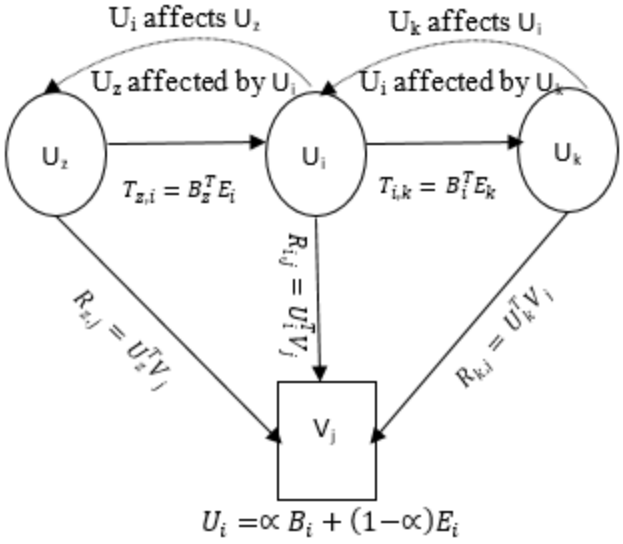


*Figure 6: Modèle graphique de la fusion de la confiance globale et locale par la factorisation matricielle pour des recommandations personnalisées.*

### i. Modélisation de la confiance

En plus des évaluations, un large réseau de confiance, créé par les utilisateurs, est accessible sur les plateformes en ligne de produits. Les préférences des utilisateurs sont influencées par celles des donneurs de confiance via des activités telles que la navigation et les commentaires (par exemple, en affichant les notations et critiques de ceux en qui ils ont confiance pour certains produits) et, à leur tour, elles influencent les receveurs de confiance. Au-delà de la relation entre les préférences des utilisateurs et la confiance, il est également possible de modéliser les préférences des utilisateurs à partir des données de confiance existantes en utilisant une technique de factorisation matricielle, comme mentionné dans les références [112, 113]. Quand les utilisateurs sont représentés dans un espace vectoriel commun à k dimensions,

chaque valeur de confiance connue, $T_{i,k}$, peut être modélisée comme le produit scalaire de $B_i$ et $E_k$. C'est-à-dire, l'intensité prévue de la relation de confiance entre $u_i$ et $u_k$ est représentée par :

$$\hat{T}_{i,k} = B_i^T E_k \qquad (4)$$

En se basant uniquement sur les données de confiance, il est possible d'apprendre les matrices de caractéristiques $B \in R^{k\times n}$ et $E \in R^{k\times n}$ en traitant le problème d'optimisation suivant :

$$\min_{B,E} \sum_{(i,k)\in\psi} \left(T_{i,k} - g\left(\hat{T}_{i,k}\right)\right)^2 + \lambda(\|B\|_F^2 + \|E\|_F^2) \qquad (5)$$

Par conséquent, en réalisant cette factorisation, on peut déduire les préférences de l'utilisateur à travers les vecteurs de caractéristiques latentes des donneurs de confiance et des receveurs de confiance, à partir des données de confiance disponibles.

### ii. Modélisation de l'évaluation

En prenant en compte l'idée que différents rôles d'utilisateurs influencent de manière distincte la formation des évaluations [99, 114], chaque évaluation existante devrait être influencée à la fois par les préférences de l'utilisateur donneur de confiance et par celles de l'utilisateur receveur de confiance. Ainsi, la prédiction de l'évaluation faite par l'utilisateur $u_i$ pour l'objet $v_i$ peut être formulée comme suit :

$$\hat{R}_{i,j} = \propto B_i^T V_j + (1-\propto) E_i^T V_j \qquad (6)$$

α est le paramètre qui ajuste l'importance relative des deux rôles d'utilisateurs dans l'évaluation. Donc, en se concentrant uniquement sur les données d'évaluation, nous pouvons déduire les matrices de caractéristiques $B \in R^{k\times n}$, $E \in R^{k\times n}$ , $V \in R^{k\times m}$ en résolvant le problème d'optimisation ci-dessous :

$$\min_{B,E,V} \sum_{(i,j)\in\Omega} \left(R_{i,j} - g\left(\hat{R}_{i,j}\right)\right)^2 + \lambda F \qquad (7)$$

où $F = (\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2)$ représente la "norme de Frobenius".

La fonction $g(x)$, connue sous le nom de fonction logistique, est utilisée pour mappier le produit interne des vecteurs de caractéristiques latentes dans l'intervalle [0,1]. Elle est définie

par l'expression $g(x) = 1/(1 + e^{-X})$. Afin de faciliter l'apprentissage des paramètres, nous normalisons les évaluations brutes $R_{i,j}$ à l'intervalle [0, 1] à l'aide de cette fonction logistique, en prenant en compte $R_{max}$, qui est la plus haute évaluation possible dans le système de recommandation. Une fois le modèle formé et les matrices de caractéristiques déduites, la prédiction de l'évaluation donnée par l'utilisateur $u_i$ pour l'item $v_i$ est calculée par la formule $g(\hat{R}_{i,j})$ x $R_{max}$. Ainsi, en employant cette méthode de factorisation matricielle, il est possible d'appréhender les préférences des utilisateurs en fonction de leur rôle, soit en tant que donneur, soit en tant que receveur de confiance, à partir des évaluations existantes.

#### 3.3.3.2.2. Modélisation des interactions implicites entre les utilisateurs

Dans le segment suivant, nous expliquons la méthode pour représenter les interactions sous-jacentes entre deux utilisateurs - qu'ils soient donneurs ou receveurs de confiance - en intégrant la structure de liaison du réseau de confiance local afin d'affiner notre fonction objective.

##### i. L'influence implicite des receveurs de confiance

Dans cette section, nous expliquons comment deux utilisateurs ayant un rôle de receveur de confiance peuvent être considérés comme similaires s'ils ont de nombreux liens sortants communs dans un réseau de confiance. Autrement dit, s'ils interagissent tous les deux avec un ensemble similaire d'utilisateurs donneurs de confiance. En examinant l'ensemble des liens de confiance d'un utilisateur avec d'autres, plutôt que de se baser uniquement sur un lien spécifique, on peut établir une relation plus robuste et précise entre deux receveurs de confiance, même s'ils n'ont pas de lien direct. Ainsi, afin de déterminer la similarité entre deux utilisateurs receveurs de confiance, $u_i$ et $u_k$, basée sur la structure des liens émergents, nous utilisons la formule de "similarité cosinus" [115] comme suit :

$$S_{i,k}^{B} = \frac{\sum_{f=1}^{n} T_{i,f} T_{k,f}}{\sqrt{(\sum_{f=1}^{n} T_{i,f}^{2})\ x\ (\sum_{f=1}^{n} T_{k,f}^{2})}} \qquad (8)$$

Dans notre étude, nous avons adopté une approche binaire pour représenter la confiance, c'est-à-dire que $T_{i,k}$ peut prendre la valeur 0 ou 1. Ici, *B* symbolise la similarité entre deux receveurs de confiance. En exploitant cette notion de similarité, nous sommes en mesure de modéliser l'influence implicite des receveurs de confiance. Pour ce faire, nous cherchons à minimiser

l'écart entre les caractéristiques des receveurs de confiance similaires en utilisant l'expression suivante :

$$\sum_{i=1}^{n}\sum_{k=1}^{n} S_{i,k}^{B} \|B_i - B_k\|_F^2 \qquad (9)$$

Une forte valeur de $S_{i,k}^{B}$ suggère que les utilisateurs $u_i$ et $u_k$, en tant que receveurs de confiance, ont plusieurs liens communs. Par conséquent, il est attendu que leurs vecteurs de préférence soient similaires. À l'inverse, une faible valeur de $S_{i,k}^{B}$ signifie qu'il y a peu de similitudes entre les liens de $u_i$ et $u_k$, impliquant que leurs vecteurs de préférence devraient être distincts. Ainsi, en intégrant cette similarité dérivée de la structure, les vecteurs de préférence des receveurs de confiance sont interdépendants lors de la phase d'apprentissage.

### ii. L'impact implicite des donneurs de confiance

Deux utilisateurs, en tant que donneurs de confiance, sont considérés comme semblables s'ils reçoivent de la confiance de plusieurs utilisateurs communs au sein du réseau de confiance. Autrement dit, ils sont tous deux perçus comme fiables par un ensemble similaire de déposants de confiance. Ainsi, la similarité entre deux donneurs de confiance $u_i$ et $u_k$, basée sur la structure des liens qui leur sont associés, peut être quantifiée par la formule de similarité cosinus suivante :

$$S_{i,k}^{E} = \frac{\sum_{f=1}^{n} T_{f,i}.T_{f,k}}{\sqrt{(\sum_{f=1}^{n} T_{f,i}^{2})\ x\ (\sum_{f=1}^{n} T_{f,k}^{2})}} \qquad (10)$$

Lorsque E symbolise la similarité entre deux donneurs de confiance, tout comme nous avons modelé l'effet implicite pour les receveurs de confiance, nous intégrons également l'effet implicite pour les donneurs de confiance. Ce modèle cherche à minimiser la distance entre les vecteurs de préférence de donneurs de confiance similaires, comme illustré par l'expression suivante :

$$\sum_{i=1}^{n}\sum_{k=1}^{n} S_{i,k}^{E} \|E_i - E_k\|_F^2 \qquad (11)$$

Ici, une valeur élevée de $S_{i,k}^{E}$ suggère que les donneurs de confiance $u_i$ et $u_k$, ont une confiance partagée provenant de nombreux déposants communs. Par conséquent, il est souhaitable que leurs vecteurs de préférence soient rapprochés. À l'inverse, une faible valeur pour $S_{i,k}^{E}$ implique

que les deux donneurs de confiance ont des profils de confiance distincts, et leurs vecteurs de préférence devraient donc être éloignés.

#### 3.3.3.3. Modèle proposé unifié

Avec toutes les informations et modèles précédents, nous construisons maintenant un cadre unifié pour la recommandation basée sur la confiance. Ce cadre englobe à la fois les données d'évaluation explicite et les informations de confiance, tout en capturant les effets sociaux implicites et explicites parmi les utilisateurs.

##### 3.3.3.3.1. Factorisation matricielle basée sur l'évaluation :

$$\min_{B,E,V} \sum\nolimits_{(i,j)\in\Omega} \left(R_{i,j} - g\left(\hat{R}_{i,j}\right)\right)^2 + \lambda(\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2)$$

##### 3.3.3.3.2. Incorporation des relations de confiance :

$$\min_{B,E} \sum\nolimits_{(i,k)\in\psi} \left(T_{i,k} - g\left(\hat{T}_{i,k}\right)\right)^2 + \lambda(\|B\|_F^2 + \|E\|_F^2)$$

##### 3.3.3.3.3. Modélisation des interactions implicites :

- Pour les donneurs de confiance :

$$\sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{n} S_{i,k}^{E} \|E_i - E_k\|_F^2$$

- Pour les receveurs de confiance :

$$\sum\nolimits_{i=1}^{n} \sum\nolimits_{k=1}^{n} S_{i,k}^{B} \|B_i - B_k\|_F^2$$

En utilisant l'approche de descente de gradient stochastique ou une autre méthode d'optimisation, nous cherchons à résoudre le problème d'optimisation global résultant, qui vise à minimiser l'erreur entre les évaluations prédites et réelles, tout en incorporant les effets sociaux implicites et explicites des relations de confiance.

Notre modèle normalisé se base sur la combinaison des évaluations explicites des utilisateurs, leurs interactions sociales et la structure des réseaux de confiance pour offrir des recommandations pertinentes. Il prend en compte non seulement les opinions directes et

explicites des utilisateurs, mais aussi la manière dont ces opinions sont influencées et formées à travers le réseau de confiance.

Cette approche intégrative est essentielle pour améliorer la précision des recommandations, en particulier dans les scénarios où les données sont rares ou où les nouvelles interactions des utilisateurs doivent être rapidement intégrées pour offrir des recommandations mises à jour. En fin de compte, ce cadre est conçu pour être flexible et adaptable, permettant d'incorporer d'autres sources d'information ou de s'ajuster à différentes dynamiques de réseau selon les besoins spécifiques de l'application.

Notre approche complète dans ce travail est d'intégrer les interactions explicites, les interactions implicites, et la réputation des utilisateurs au sein des réseaux de confiance dans un modèle unifié de recommandation. En fusionnant ces diverses sources d'information, nous visons à améliorer la précision et la pertinence de nos recommandations.

Le modèle unifié que nous présentons est mathématiquement décrit par la fonction objective $\mathcal{L}$ donnée dans l'équation (12). Cette fonction vise à minimiser la différence entre les évaluations réelles et prédites des utilisateurs tout en prenant en compte la structure du réseau de confiance.

$$\mathcal{L} = \frac{1}{2}\Bigg(\sum_{(i,j)\in\Omega} w_i \left(R_{i,j} - g\left(\hat{R}_{i,j}\right)\right)^2 + \sum_{(i,k)\in\psi} \left(T_{i,k} - g\left(\hat{T}_{i,k}\right)\right)^2 + \lambda_B \sum_{i=1}^{n}\sum_{k=1}^{n} S_{i,k}^{B} \|B_i - B_k\|_F^2 + \lambda_E \sum_{i=1}^{n}\sum_{k=1}^{n} S_{i,k}^{E} \|E_i - E_k\|_F^2 + \lambda \mathrm{F}\Bigg) \quad (12)$$

Les composants de cette fonction objective sont :

**Terme d'interaction explicite** : qui capture la différence entre les évaluations réelles et prédites, pondérée par la réputation de l'utilisateur $w_i$.

$$\sum_{(i,j)\in\Omega} w_i \left(R_{i,j} - g\left(\hat{R}_{i,j}\right)\right)^2$$

**Terme d'interaction de confiance :** qui mesure l'écart entre la confiance réelle et prédite entre les utilisateurs.

$$\sum_{(i,k)\in\psi} \left(T_{i,k} - g\left(\hat{T}_{i,k}\right)\right)^2$$

**Termes d'interactions implicites :** Ces termes cherchent à rapprocher les vecteurs de caractéristiques des utilisateurs ayant des profils similaires dans le réseau de confiance.

- $\sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k}^{E} \|E_i - E_k\|_F^2$ pour les donneurs de confiance.
- $\sum_{i=1}^{n} \sum_{k=1}^{n} S_{i,k}^{B} \|B_i - B_k\|_F^2$ pour les récepteurs de confiance.

**Terme de régularisation :** est utilisé pour éviter le surajustement en limitant la magnitude des vecteurs de caractéristiques.

$$\lambda \mathrm{F} \text{ où } \mathrm{F} = (\|B\|_F^2 + \|E\|_F^2 + \|V\|_F^2)$$

Les paramètres $\lambda_B$ et $\lambda_E$ sont utilisés pour contrôler la contribution des termes d'interaction implicite. Dans notre expérimentation, en fixant $\lambda_B = \lambda_E$, nous assurons que les donneurs et les receveurs de confiance ont une influence égale sur la formation des recommandations.

En résolvant cette fonction objective, nous pouvons obtenir des matrices de caractéristiques latentes optimales qui capturent à la fois les évaluations explicites et la structure du réseau de confiance, permettant ainsi des recommandations plus précises et contextuelles pour les utilisateurs.

### 3.3.4. Apprentissage du modèle proposé

Pour minimiser la fonction objective décrite précédemment et apprendre les matrices de caractéristiques *V, E* et *B* de manière à prédire efficacement les évaluations inconnues, nous employons la méthode de descente de gradient stochastique. Cette méthode est largement reconnue pour son efficacité dans les systèmes de recommandation.

#### 3.3.4.1. Algorithme d'apprentissage du modèle proposé

***Algorithme : Algorithme d'apprentissage du modèle proposé***

***Début***

***Entrées :***

*Matrice d'évaluation R*

*Matrice de confiance T*

*Paramètres de régulation $\lambda_E$, $\lambda_B$ et $\lambda$*

*Contribution des différents cas d'utilisateur $\alpha$*

*Taux d'apprentissage $\mu$*

*Nombre de facteurs latents K*

***Procédure :***

*Initialiser aléatoirement les matrices V, E, et B avec de petites valeurs.*

*Répéter jusqu'à la convergence de la fonction objective L :*

*Pour chaque triplet (i,j,k) existant dans R et T :*

*Calculer l'erreur pour le triplet actuel en utilisant la fonction objective.*

*Mettre à jour les matrices V, E, et B en utilisant les dérivées par rapport à chaque variable et le taux d'apprentissage $\mu$.*

*Une fois l'entraînement terminé, utiliser les matrices V, E, et B pour faire des prédictions.*

***Sorties** : Vecteurs de caractéristiques latentes optimisés V, E, et B.*

***Fin***

Cette approche garantit que nous incorporons à la fois les évaluations explicites fournies par les utilisateurs et la structure implicite du réseau de confiance lors de la génération de recommandations. L'utilisation de la descente de gradient stochastique garantit que l'algorithme converge rapidement vers une solution qui minimise l'erreur entre les évaluations prédites et les véritables évaluations des utilisateurs.

### 3.3.4.2. Organigramme de l'algorithme du modèle proposé

La Figure 7 expose un organigramme qui décrit le processus d'un algorithme d'apprentissage pour un modèle de recommandation, où diverses matrices et hyperparamètres sont utilisés pour ajuster les caractéristiques des utilisateurs et des éléments en vue de générer des recommandations.

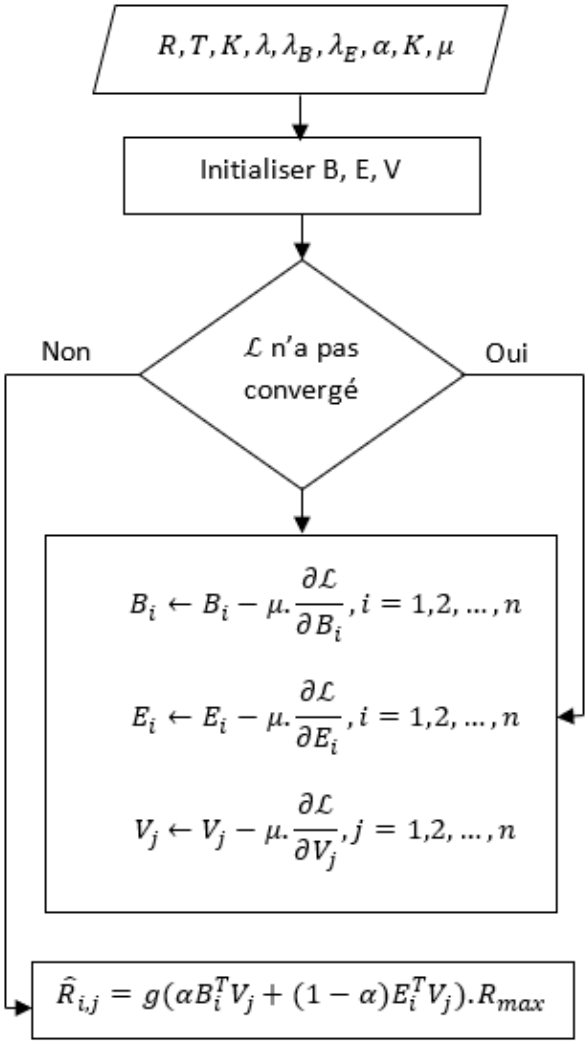


*Figure 7: Organigramme d'Algorithme d'apprentissage du modèle proposé.*

Où *:*

$R$ : Matrice d'évaluation,

$T$ : matrice sociale,

$\lambda, \lambda_B, \lambda_E, \alpha$ et $k$ : hyperparamètres

$\mu$ : Taux d'apprentissage

*B* et *E* : Matrices de caractéristiques pour les utilisateurs ayant le rôle de truster (donneur de confiance) et de trustee (receveur de confiance),

V : Matrice des caractéristiques pour les éléments avec rétroactions implicites et explicites.

Voici une description de cet algorithme :

a. **Initialisation** : Les matrices *B, E* et V sont initialisées. *B* et *E* correspondent aux matrices de caractéristiques pour les utilisateurs ayant les rôles de truster (celui qui accorde la confiance) et de trustee (celui qui reçoit la confiance), tandis que *V* est la matrice des caractéristiques pour les éléments, incluant les rétroactions implicites et explicites.
b. **Convergence** : L'algorithme entre dans une boucle où il vérifie la convergence, un état où les changements des caractéristiques deviennent négligeables, indiquant que le modèle a appris de façon optimale les préférences et les comportements des utilisateurs.
c. **Mise à jour des Caractéristiques** : Si le modèle n'a pas convergé, il procède à la mise à jour des matrices de caractéristiques *B, E* et *V* en utilisant les dérivées partielles de la fonction de coût par rapport à chaque caractéristique et le taux d'apprentissage $\mu$. Ces mises à jour sont réalisées dans le but de minimiser la fonction de coût et d'améliorer la précision du modèle.
d. **Calcul de la Recommandation** : Une fois que la convergence est atteinte, l'algorithme calcule la recommandation pour un utilisateur et un élément donné en utilisant une fonction qui prend en compte les caractéristiques mises à jour des utilisateurs et des éléments, pondérées par les hyperparamètres et le maximum de la matrice d'évaluation *R*.

## *3.4.* Évaluation expérimentale

### 3.4.1. Présentation de la base de données

Epinions est une base de données largement reconnu, issu du monde réel et librement accessible, fréquemment utilisé comme référence pour tester et évaluer les systèmes de recommandation dans les recherches académiques. Cette base de données comprend non seulement les évaluations des utilisateurs concernant divers items, mais également des relations

explicites de confiance et de méfiance entre les utilisateurs. Ces éléments font d'Epinions un choix idéal pour tester notre modèle.

La base de données "Epinions" que nous avons utilisé pour nos tests a été collecté par Paolo Massa lors d'un processus de capture de données étendu sur une période de 5 semaines (d'octobre à décembre 2003) depuis le site Epinions.com.

Le Tableau 1 fournit un résumé des attributs clés de cette base de données :

*Tableau 1: Caractéristiques principales de la base de données Epinions*

| Caractéristique | Effectif |
|---|---|
| Items | 139,738 |
| Utilisateurs | 49,290 |
| Utilisateurs ayant évalué un item | 40,163 |
| Évaluations | 664,824 |
| Relations de confiance | 487,183 |
| Utilisateurs ayant des relations de confiance | 33,960 |
| Plage d'évaluations | [1-5] |

Ces statistiques fournissent un aperçu complet de la diversité et de la densité de la base de données Epinions, le rendant approprié pour des tests robustes et complets de notre modèle de recommandation.

### 3.4.2. Analyse de la distribution des donneurs de confiance et receveurs de confiance

Nos analyses ont révélé que la distribution des donneurs de confiance (trusters) et des receveurs de confiance (trustees) dans la base de données "Epinions" suit une distribution en loi de puissance (Power-Law), un schéma couramment observé dans de nombreux réseaux sociaux.

La figure 8 ci-dessous le démontre clairement. On y observe que seulement quelques donneurs de confiance ont un grand nombre de receveurs de confiance. En d'autres termes, un petit nombre d'utilisateurs (donneurs de confiance) sont extrêmement populaires et bénéficient de la confiance d'un grand nombre d'autres utilisateurs. En revanche, la majorité des donneurs de confiance n'ont que quelques receveurs de confiance. Ce phénomène de concentration élevée pour un petit groupe, avec une longue traîne pour la majorité, est caractéristique des distributions en loi de puissance.

Cette observation soulève également un point important : il existe une grande disparité dans les données de confiance fournies par les utilisateurs. Cela pourrait être dû à la nature du site, où certains utilisateurs sont considérés comme des influenceurs ou des experts dans certains domaines et sont donc davantage suivis et font l'objet d'une plus grande confiance. Alors que la plupart des autres utilisateurs, qui sont peut-être des participants plus occasionnels, n'accumulent pas autant de relations de confiance. Cette caractéristique doit être prise en compte lors de la construction et de l'évaluation de modèles basés sur ces données.

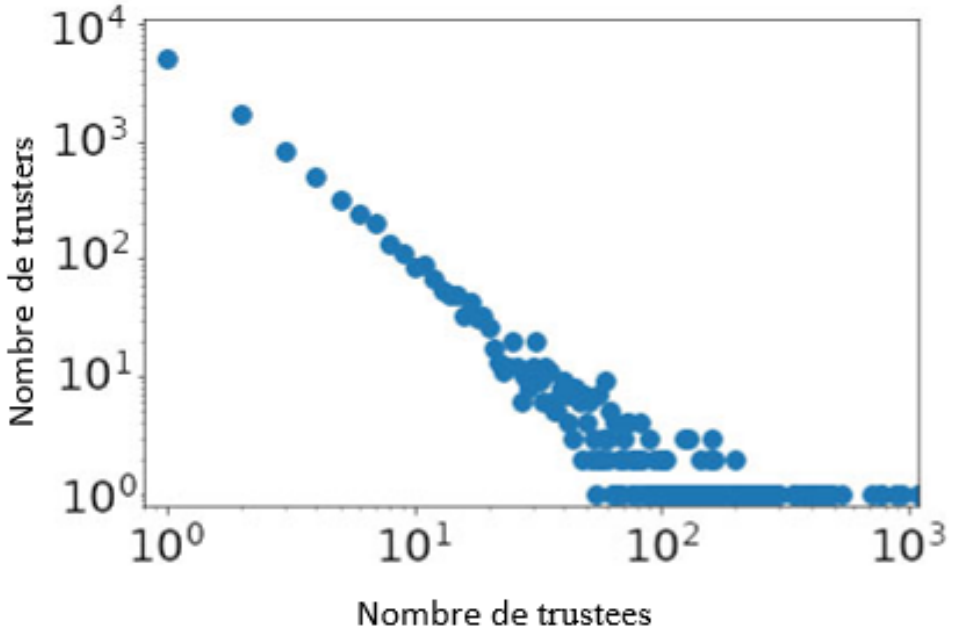


*Figure 8: Répartition des administrateurs par rapport aux personnes qui leur accordent leur confiance*

### 3.4.3. Évaluation du modèle

#### 3.4.3.1. Métriques d'évaluation

Pour évaluer l'efficacité de notre méthode proposée, nous la comparons avec le filtrage collaboratif traditionnel et d'autres techniques basées sur la confiance en utilisant deux métriques clés : la MAE (Mean Absolute Error) et la RMSE (Root Mean Square Error).

**MAE (Erreur Absolue Moyenne)** : La MAE mesure la différence moyenne absolue entre les notations réelles attribuées par l'utilisateur et les notations prédites par le modèle de recommandation, comme indiqué par l'équation (13) :

$$MAE = \frac{1}{N}\sum_{(u_i,v_j)\in TestSet}\left|\hat{R}_{i,j} - R_{i,j}\right| \qquad (13)$$

Où *N* est le total des notations dans le jeu de test. Une MAE plus faible indique une prédiction plus précise.

**RMSE (Root Mean Square Error)** : La RMSE mesure l'écart quadratique moyen entre les notations réelles et prédites, offrant une vue plus pénalisée des erreurs importantes, comme indiqué par l'équation (14) :

$$RMSE = \sqrt{\frac{1}{N}\sum_{(u_i,v_j)\in TestSet}(\hat{R}_{i,j} - R_{i,j})^2} \qquad (14)$$

Une RMSE plus faible signifie également une meilleure précision des prédictions.

#### 3.4.3.2. Protocole d'évaluation

Nous adoptons la technique de validation croisée à 5 plis pour l'entraînement et la validation de notre modèle. La base de données est divisée en cinq sous-ensembles distincts. Pour chaque itération, un sous-ensemble est réservé pour le test, tandis que les quatre autres servent à l'entraînement. Ce processus est répété cinq fois, en alternant le sous-ensemble de test à chaque fois.

Dans nos tests, **80%** des données sont allouées à l'apprentissage et **20%** au test. Nous utilisons une version randomisée de la base de données "Epinions" pour ce processus, garantissant que chaque sous-ensemble de test contient une représentation équilibrée des données.

#### 3.4.3.3. Analyse et interprétation

La performance du modèle sera ensuite analysée en fonction des valeurs obtenues pour les métriques MAE et RMSE. Une comparaison détaillée avec d'autres méthodes existantes offrira des insights sur la précision et l'efficacité de notre approche proposée. Une analyse plus approfondie des résultats nous permettra d'ajuster et d'optimiser notre modèle pour de meilleures performances et une plus grande précision dans les prédictions futures.

### 3.4.4. Paramètres expérimentaux et évaluation

Nous évaluons notre méthode, que nous appelons GLoTMF, par rapport à une série de méthodes populaires dans le domaine des systèmes de recommandation. Le tableau 2 décrit brièvement chaque méthode et présente leurs paramètres optimaux déterminés à partir des références et de nos expériences.

*Tableau 2: Comparaison des performances des différentes méthodes*

| **Méthodes** | **Description** | **Paramètres optimaux** |
|---|---|---|
| PMF | Une méthode de factorisation matricielle qui utilise seulement la matrice d'évaluation[116]. | $\lambda_u = 0.001$<br>$\lambda_v = 0.001$ |
| SoRec | Intègre à la fois la matrice d'évaluation et la matrice de confiance sociale[117]. | $\lambda_c = 1$<br>$\lambda_u = 0.001$<br>$\lambda_v = 0.001$<br>$\lambda_z = 0.001$ |
| RSTE | Pèse les évaluations en fonction des préférences propres et celles de la tutelle[118]. | $\alpha = 0.4$<br>$\lambda_u = 0.001$<br>$\lambda_v = 0.001$ |
| SocialMF | Modèle où la propagation de la confiance est prise en compte[119]. | $\lambda_t = 1$<br>$\lambda_u = 0.001$<br>$\lambda_v = 0.001$ |

| | | |
|---|---|---|
| SoReg | Minimise la distance entre les caractéristiques latentes des personnes de confiance[120]. | $\beta = 0.001$<br>$\lambda_1 = 0.001$<br>$\lambda_2 = 0.001$ |
| TrustMF | Modélise les utilisateurs dans deux espaces de caractéristiques latentes, basé sur le réseau de confiance[101]. | $\lambda_t = 1$<br>$\lambda = 0.001$ |
| GLoTMF | Notre proposition exploitant le contexte social à la fois général et local des relations. | $\alpha = 0.6$<br>$\lambda_b = 0.1$<br>$\lambda_e = 0.1$<br>$\lambda = 0.001$ |

Pour évaluer les performances, toutes les méthodes sont testées avec une configuration standard. En particulier :

**a. Initialisation** : Toutes les matrices de caractéristiques latentes sont initialisées aléatoirement en suivant une distribution uniforme dans l'intervalle [0,1].

**b. Dimensionnalité** : Nous testons deux dimensions pour l'espace des caractéristiques latentes, à savoir K=5 et K=10. Cette variabilité permet de vérifier l'impact de la dimensionnalité sur la qualité des recommandations.

**c. Sous-ensembles d'utilisateurs** : Pour mesurer la performance des méthodes, deux points de vue d'utilisateurs sont considérés :

- ✓ **Tous les utilisateurs** : Cela concerne tous les utilisateurs ayant au moins une évaluation.
- ✓ **Utilisateurs du démarrage à froid** : Ce sont les utilisateurs qui ont donné moins de 5 évaluations. C'est un scénario plus difficile car ces utilisateurs ont un historique limité, rendant la prédiction plus complexe.

L'objectif est de vérifier comment chaque méthode se comporte dans différentes situations, en particulier dans les cas difficiles, comme les utilisateurs avec peu d'historique.

## *3.5.* Résultats et discussion

### 3.5.1. Analyses et interprétations

Dans cette section, nous exposons les résultats obtenus à partir de nos tests sur la base de données "Epinions", en mettant l'accent sur les groupes d'utilisateurs "Tous les utilisateurs" et "Utilisateurs démarrant à froid". Les mesures "MAE" et "RMSE" ont été employées pour l'évaluation, en tenant compte des différentes dimensions de l'espace des facteurs latents (k=5, k=10). Les performances de la méthode que nous avons développée sont mises en évidence à travers des tableaux 3 et 4, où elles sont comparées à d'autres approches existantes.

*Tableau 3: Comparaisons des performances des recommandations sur la base de données Epinions, dans le cas de "Tous les utilisateurs"*

| Métriques | K | PMF | RSTE | SoRec | SoReg | SocialMF | TrustMF | GLoTMF |
|---|---|---|---|---|---|---|---|---|
| **MAE** | 5 | 0.979 | 0.950 | 0.882 | 0.994 | 0.825 | 0.818 | 0.804 |
| | 10 | 0.909 | 0.958 | 0.884 | 0.932 | 0.826 | 0.820 | 0.805 |
| **RMSE** | 5 | 1.290 | 1.196 | 1.114 | 1.315 | 1.070 | 1.069 | 1.043 |
| | 10 | 1.197 | 1.278 | 1.142 | 1.232 | 1.082 | 1.095 | 1.044 |

*Tableau 4: Comparaisons des performances des recommandations sur la base de données d'Epinions, dans le cas des "utilisateurs ayant démarré à froid"*

| Métriques | K | PMF | RSTE | SoRec | SoReg | SocialMF | TrustMF | GLoTMF |
|---|---|---|---|---|---|---|---|---|
| **MAE** | 5 | 1.451 | 1.051 | 0.892 | 1.398 | 0.884 | 0.891 | 0.868 |
| | 10 | 1.153 | 0.981 | 0.846 | 1.139 | 0.857 | 0.853 | 0.868 |
| **RMSE** | 5 | 1.770 | 1.266 | 1.138 | 1.735 | 1.133 | 1.125 | 1.105 |
| | 10 | 1.432 | 1.313 | 1.180 | 1.437 | 1.152 | 1.176 | 1.108 |

Sur la base des expériences réalisées sur la base de données "Epinions" et des résultats présentés, voici une série d'observations et d'analyses :

a. **Importance de la confiance** : Il est évident que l'intégration de la confiance améliore considérablement la performance des recommandations. Cela est manifeste par la faible performance de PMF par rapport à d'autres méthodes qui utilisent des informations de confiance.

b. **Efficacité de GLoTMF** : Notre modèle, se démarque par ses performances exceptionnelles dans les deux scénarios - "Tous les utilisateurs" et "Utilisateurs démarrant à froid". En particulier, dans le scénario "Tous les utilisateurs", notre modèle dépasse le TrustMF en termes de MAE et de RMSE. Cela montre l'efficacité de notre méthode, qui tient compte à la fois des informations locales et globales du contexte social.

c. **Performance en Cold Start** : Le scenario "Cold Start" est un défi majeur pour les systèmes de recommandation. Ici, notre modèle a montré une robustesse, en particulier en termes de RMSE. Même si d'autres méthodes peuvent avoir de légères avances sur MAE, la faible valeur RMSE pour notre modèle indique que nos prédictions sont généralement proches des vraies évaluations, surtout lorsqu'il s'agit d'erreurs importantes.

d. **Influence des caractéristiques latentes** : Les différentes dimensions de l'espace des caractéristiques latentes (K = 5, K = 10) ont montré une influence sur les performances, mais notre modèle a maintenu une performance solide indépendamment de la dimensionnalité.

e. **Impact des liens implicites** : La meilleure performance de notre modèle met en évidence l'importance de considérer à la fois les liens explicites et implicites. En utilisant ces deux types de liens, notre modèle peut capturer des nuances et des informations qui pourraient être négligées par d'autres méthodes.

f. **Signification de petites améliorations** : Bien que nos améliorations soient relativement petites en pourcentage, elles sont significatives dans le domaine des systèmes de recommandation. Comme le suggèrent les références, même de légères améliorations peuvent avoir un impact considérable sur l'expérience utilisateur.

g. **Validation de l'approche** : L'écart notable en performance entre notre méthode et d'autres méthodes conventionnelles renforce la validité de notre approche, qui utilise une combinaison d'informations explicites, implicites, locales et globales pour améliorer la qualité des recommandations.

En conclusion, les résultats démontrent clairement l'efficacité de notre modèle dans l'exploitation du contexte social dans les systèmes de recommandation. Il serait également intéressant d'explorer d'autres informations contextuelles et de voir comment elles peuvent être intégrées pour améliorer davantage les recommandations.

### 3.5.2. Évaluation de la performance du modèle proposé selon différents niveaux de confiance

Nous avons entrepris une série supplémentaire d'expériences pour évaluer l'efficacité de notre modèle pour les utilisateurs présentant divers niveaux de relations de confiance. L'objectif était de mettre en opposition notre approche avec d'autres techniques basées sur la confiance et d'évaluer la capacité de ces méthodes à capitaliser sur les informations de confiance pour formuler des recommandations. Les niveaux de confiance sont considérés comme étant l'agrégation des relations directes de l'utilisateur avec d'autres, plus ses propres relations.

Nous avons divisé les degrés de confiance en sept catégories : 1-5, 6-10, 11-20, 21-40, 41-100, 101-500, >500, en suivant une démarche similaire à celle adoptée par d'autres chercheurs. Nous avons ensuite évalué l'erreur prédictive pour chaque catégorie, en utilisant les mesures "MAE" et "RMSE", pour des ensembles de caractéristiques latentes de 5 et 10. Les graphiques ci-joints présentent ces résultats.

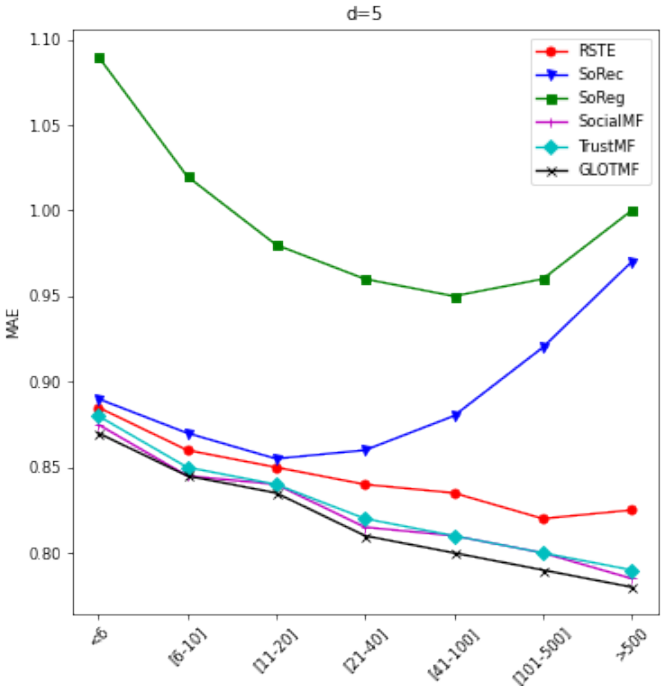


*Figure 9: Comparaison des performances selon divers degrés de confiance des utilisateurs (MAE en fonction de degré de confiance avec d=5)*

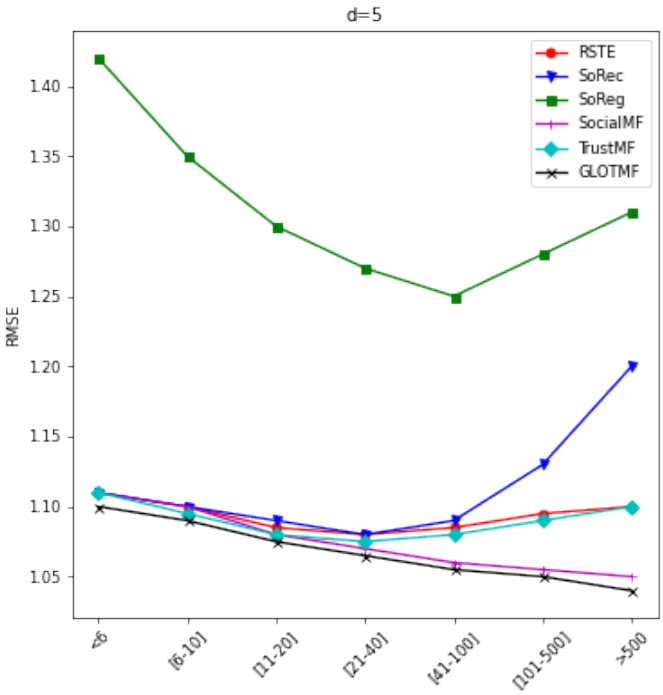


*Figure 10: Comparaison des performances selon divers degrés de confiance des utilisateurs (RMSE en fonction de degré de confiance avec d=5)*

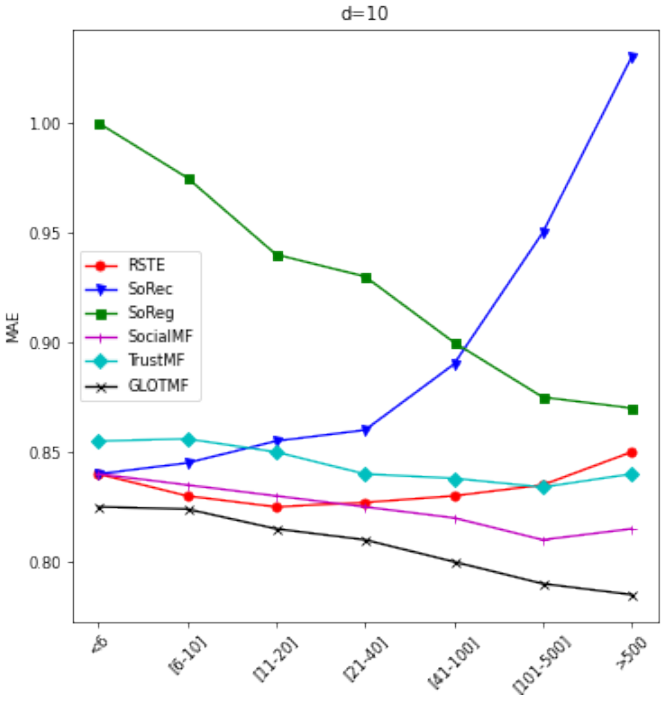


*Figure 11: Comparaison des performances selon divers degrés de confiance des utilisateurs (MAE en fonction de degré de confiance avec d=10)*

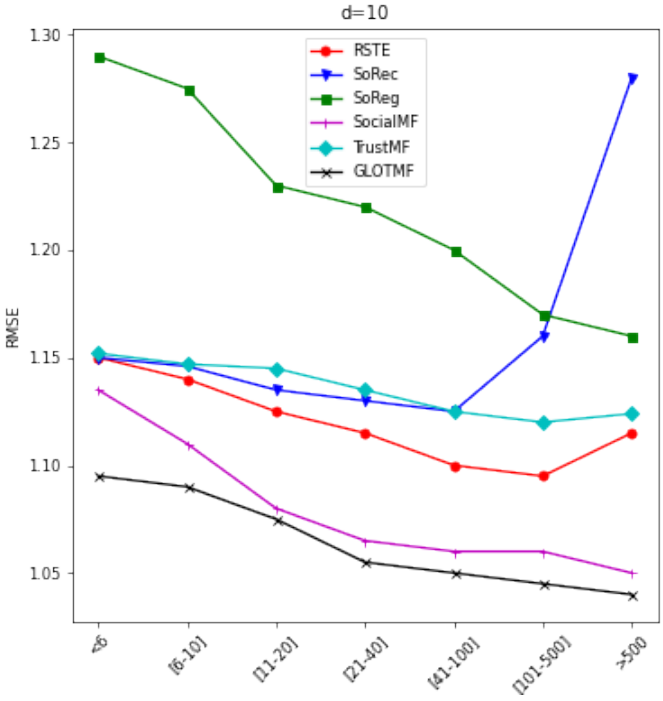


*Figure 12: Comparaison des performances selon divers degrés de confiance des utilisateurs (RMSE en fonction de degré de confiance avec d=10)*

Les données révèlent que l'efficacité des six méthodes en comparaison fluctue légèrement en fonction des différents groupes de confiance. Cependant, notre modèle "GLoTMF" démontre une stabilité et surpasse les autres en matière de précision pour la majorité des catégories, en particulier pour celles avec un maximum de 5 relations de confiance (qui englobent environ 63,4% des utilisateurs) et pour la catégorie de 6 à 10 relations (représentant approximativement 11,6% des utilisateurs). Ces observations sont manifestes pour les mesures MAE et RMSE, et ce, pour les deux dimensions de 5 et 10. Il en ressort que notre modèle est particulièrement apte à tirer profit des données de confiance éparses, plus efficacement que d'autres techniques de recommandation collaboratives fondées sur la confiance.

## *3.6.* Conclusion

Dans ce chapitre, nous avons détaillé une approche innovante visant à renforcer la précision des recommandations dans les techniques de filtrage collaboratif, fondées sur un modèle en capitalisant sur les relations de confiance établies entre les utilisateurs. Notre démarche introduit une manière novatrice de combiner les données d'évaluation et de confiance pour mieux refléter l'influence réciproque des utilisateurs sur leurs opinions. Ainsi, nous parvenons à prédire avec une meilleure justesse et efficience les évaluations inconnues des utilisateurs sur différents items. Dans notre méthode, les préférences de l'utilisateur sont discernées à travers deux rôles distincts : en tant que donneur et en tant que receveur de confiance. Ces préférences sont extrapolées en observant les interactions tant explicites qu'implicites entre les utilisateurs. Parallèlement, nous intégrons la notion de réputation de l'utilisateur au sein du réseau social.

Nous avons validé notre méthode sur une base de données authentique, "Epinions", en observant sous deux angles différents : celui de "Tous les utilisateurs" et celui des "Utilisateurs débutants". Nos expérimentations ont démontré que notre proposition offre une amélioration tangible en matière de précision prédictive. De plus, elle s'avère particulièrement efficiente pour pallier les défis liés à la rareté des données et aux problèmes de démarrage à froid, surtout lorsqu'elle comparée à d'autres méthodes fondées sur des modèles qui exploitent également les relations de confiance.

Cependant, il est crucial de reconnaître que notre approche repose actuellement sur la confiance explicite manifestée par les utilisateurs. Or, dans certains contextes, les utilisateurs peuvent se montrer réticents à partager ou à divulguer de telles informations, notamment pour des raisons de confidentialité. D'ailleurs, nombre des bases de données existantes ne renferment que des

informations de confiance explicites. Il s'agit donc d'une piste d'exploration future : dériver des valeurs de confiance implicites à partir des comportements des utilisateurs pour encore améliorer la portée et l'efficacité de la méthode que nous avons mise en avant.

# CHAPITRE 4

## OPTIMISATION DES RECOMMANDATIONS DE VIDÉOS ÉDUCATIVES : FILTRAGE HYBRIDE COMBINANT LE FILTRAGE BASÉ SUR LE CONTENU ET LE FILTRAGE COLLABORATIVE POUR UNE PERSONNALISATION AVANCÉE.

# Chapitre 4

# Optimisation des recommandations de vidéos éducatives : filtrage hybride combinant le filtrage basé sur le contenu et le filtrage collaborative pour une personnalisation avancée

## 4.1. Introduction

La transformation numérique et la prolifération des technologies de l'information ont redéfini le paysage de l'éducation au XXIe siècle. Au-delà des salles de classe traditionnelles, une nouvelle forme d'apprentissage s'est affirmée : l'éducation informelle [121]. Elle repose sur le libre accès à une multitude de ressources éducatives offertes sur diverses plateformes en ligne, adaptées aux besoins changeants des apprenants modernes.

Les chercheurs, enseignants et apprenants ont aujourd'hui à leur disposition un éventail de dépôts contenant des informations éducatives allant des articles académiques aux tutoriels vidéo, podcasts, et autres forums de discussion. Cette abondance de ressources, bien qu'avantageuse, présente également un défi : comment s'y retrouver dans cette masse d'informations et identifier ce qui est réellement pertinent ?

Parallèlement à cette explosion de contenus, la technologie de l'information a connu des avancées sans précédent, avec le développement de plateformes sociales, forums, blogs et autres outils interactifs [122]. Ces innovations ont conduit à une accumulation rapide de données variées - vidéos, images, courriels, publications sur les réseaux sociaux et bien plus encore.

Face à cette surabondance d'informations, une question prédomine : comment filtrer et personnaliser le contenu pour répondre aux besoins spécifiques des utilisateurs ? C'est là que les systèmes de recommandation entrent en jeu. En utilisant des techniques analytiques avancées et en puisant dans les historiques des interactions des utilisateurs, ces systèmes sont conçus pour prédire et présenter les ressources les plus pertinentes pour chaque individu ou groupe [38, 123].

Cependant, bien que ces systèmes existent depuis un certain temps, il reste des lacunes et des opportunités d'amélioration. Dans ce contexte, notre recherche vise à explorer une combinaison synergique des algorithmes FBC (Filtrage Basé sur le Contenu) et FC (Filtrage Collaboratif). L'objectif est d'affiner la précision des recommandations, en veillant à ce que chaque apprenant, enseignant ou chercheur trouve rapidement et efficacement les ressources éducatives qui correspondent le mieux à ses besoins et intérêts spécifiques.

Cette étude aspire donc à jeter les bases d'une nouvelle ère dans la recommandation de ressources éducatives, en mettant l'accent sur la pertinence, l'efficacité et l'alignement avec les intérêts individuels des utilisateurs.

## 4.2. Quelques exemples de systèmes de recommandation

Dans ce chapitre, nous présenterons plusieurs propositions concernant les systèmes de recommandation en lien avec le sujet abordé. Nous décrirons certaines études pertinentes qui ont servi de fondement à cette recherche, tandis que d'autres sont envisagées comme des tendances actuelles pour le système discuté ici.

### 4.2.1. Système de recommandation pour vidéos médicales

Rivero Rodriguez et al. ont développé un système visant à recommander des vidéos médicales pertinentes, s'appuyant sur des sources provenant de chaînes vidéo de santé reconnues. Ils ont utilisé des vidéos disponibles sur YouTube ainsi que le service "Medline Plus", fourni par la Bibliothèque nationale de médecine des États-Unis. En testant quatre méthodes de recommandation, ils ont trouvé qu'une combinaison de deux techniques préexistantes était la plus efficace. Cependant, ils estiment que l'enrichissement des métadonnées doit être amélioré pour perfectionner la qualité des recommandations proposées [124].

### 4.2.2. Système hybride de recommandation de vidéos YouTube

Abbas et al. ont élaboré une méthode hybride pour suggérer des vidéos sur YouTube. Néanmoins, un inconvénient majeur de cette méthode est que les vidéos suggérées ne correspondent pas toujours au contexte actuel de l'utilisateur. Étant donné qu'un utilisateur peut avoir des intérêts variés selon la situation, ils ont introduit un système pour YouTube qui cible ces intérêts changeants, ne proposant des vidéos que basées sur le contexte immédiat de l'utilisateur [125].

### 4.2.3. Système de recommandation pour des vidéos éducatives ouvertes basé sur les compétences requises

M. Tavakoli et al. ont proposé une nouvelle méthode pour aider les apprenants à trouver des vidéos éducatives ouvertes, pertinentes, afin d'acquérir des compétences recherchées sur le marché du travail. Ils ont élaboré un prototype qui : 1) utilise des méthodes de classification de texte et d'exploration de données sur les annonces d'offres d'emploi pour associer des postes et leurs compétences requises ; 2) prévoit la qualité des vidéos ; et 3) met en place un système de recommandation de vidéos éducatives ouvertes pour suggérer un contenu d'apprentissage personnalisé aux apprenants. Pour la première évaluation de ce prototype, ils se sont concentrés sur le domaine des métiers liés à la science des données. Leur prototype a été évalué par des entretiens semi-structurés approfondis. 15 experts en la matière ont donné leur avis pour évaluer la performance du prototype en termes d'objectifs, de logique et de contribution à l'apprentissage. Plus de 250 vidéos ont été recommandées, et 82,8% de ces recommandations ont été jugées utiles par les personnes interrogées. De plus, les entretiens ont montré que le système de recommandation vidéo personnalisé a le potentiel d'enrichir l'expérience d'apprentissage [126].

### 4.2.4. Système de recommandation pour YouTube basé sur son réseau de critiques

Qin et al. [127] ont introduit SR, une approche classant les vidéos YouTube selon les données recueillies des réseaux sociaux des internautes, mettant l'accent sur les commentaires et critiques vidéo. Les utilisateurs s'expriment sur les contenus qui captivent leur attention. Ainsi, SR offre un éventail plus large de suggestions par rapport à YouTube [128], qui se base principalement sur les vidéos visionnées et les tags associés à l'intérêt de l'utilisateur. Nous avons fait appel à l'API YouTube afin d'ajouter des options comme la recherche de vidéos et la collecte d'informations sur celles-ci, ainsi que leurs visionneurs, pour enrichir notre plateforme.

L'intérêt majeur de cette démarche réside dans sa capacité à cerner les préférences des utilisateurs, suggérant non seulement des vidéos sur les thématiques visionnées, mais aussi sur des sujets susceptibles de les captiver. YRN ne prend pas en compte la qualité intrinsèque des vidéos, mais vise uniquement à présenter des contenus qui pourraient séduire son audience. Par exemple, lorsqu'un utilisateur effectue une recherche sur YouTube avec le terme " le **conjugué** ", la plateforme propose des vidéos populaires basées sur le nombre de vues, sans

garantie quant à la qualité didactique, sonore ou visuelle. YRN, de son côté, mise principalement sur des recommandations pertinentes et accessibles.

### 4.2.5. Système de résumé de contenu vidéo pour recommandation

Dans "An Automatic Multimedia Content Summarization System for Video Recommendation" par Yang et al[129], une méthode est suggérée pour identifier le sous-titre d'une vidéo en utilisant un module de Reconnaissance Optique de Caractères (OCR) afin de résumer son contenu. L'idée est de permettre l'ajout automatique de résumés de vidéos aux collections tout en recommandant des vidéos basées sur leur adéquation au profil de l'utilisateur. La Figure 13 illustre l'architecture du système de résumé de contenu vidéo pour recommandation VCSR (Video Content Summarization for Recommendation).

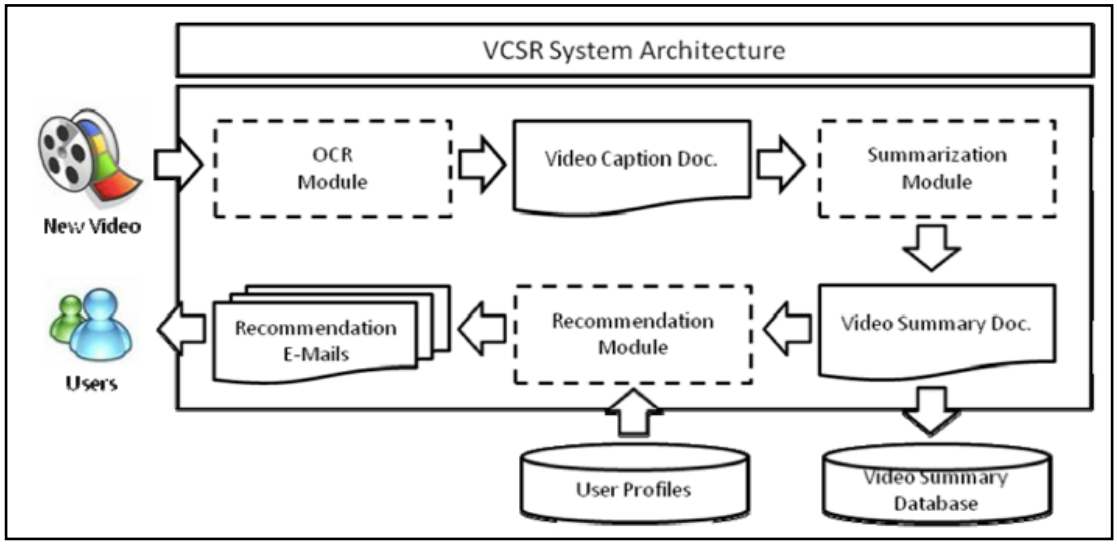


*Figure 13: Architecture du système VCSR [108]*

Lorsqu'une nouvelle vidéo est uploadée, le module OCR identifie les sous-titres et crée un document texte. Ces documents sont ensuite traités par un module de réduction qui emploie l'extraction de mots-clés pour produire des résumés conceptuels des vidéos. Finalement, le module crée des suggestions de vidéos par e-mail et détermine la pertinence selon le profil de chaque utilisateur. En fusionnant ces trois modules, le système est capable de produire des recommandations de façon automatique et d'envoyer des e-mails de suggestion dès qu'une nouvelle vidéo est ajoutée. Autrement dit, le processus du système VCSR est entièrement automatisé et ne requiert aucune intervention manuelle.

La démarche proposée dans ce chapitre se distingue du système VCSR sur plusieurs points :

• Le VCSR récapitule les vidéos et propose des recommandations par e-mail, cependant, il semble y avoir une déconnection notable entre les vidéos.

• L'ambition de ce chapitre est d'exploiter l'API YouTube pour extraire les fichiers de sous-titres.

### 4.2.6. Comparatif des systèmes de recommandation

On peut observer que même si le VCSR se base sur des recommandations vidéo, la technique qu'il emploie pour extraire les sous-titres fait appel à la reconnaissance de texte à partir d'images. Cette étape devient superflue avec l'avènement des plateformes vidéo modernes qui offrent un accès direct aux sous-titres. Même pour les vidéos dépourvues de sous-titres, il est désormais possible de les obtenir grâce à des fonctionnalités de reconnaissance vocale intégrées, comme celle présente sur YouTube. D'ailleurs, YouTube est largement plébiscité par de nombreuses chaînes éducatives. De nombreux apprenants y ont recours pour approfondir leurs connaissances ou lever des ambiguïtés. Ainsi, le projet en question exploite la capacité d'extraction vidéo de YouTube.

Toutes les études existantes s'appuient sur une base de données préexistante remplie de contenu éducatif. Cependant, la démarche proposée ici se distingue par son recours à des algorithmes pour déterminer la qualité d'une vidéo.

Chaque proposition a sa propre pertinence selon son domaine d'application, mais le travail présenté ici offre une perspective et une mise en œuvre plus actuelles, adaptées à une audience mondiale étant donné l'omniprésence de YouTube. Ce projet a aussi l'atout de s'appuyer sur un algorithme de recommandation hybride, fusionnant les atouts des recommandations axées sur le contenu avec ceux des recommandations basées sur les retours d'utilisateurs.

## 4.3. Approche proposée

Le cadre général de cette approche sert de guide pour son exécution, englobant l'extraction des données depuis YouTube, leur classification et le prétraitement, étapes cruciales pour produire des recommandations de premier ordre. Il convient de souligner que cette approche intègre des techniques spécifiques, comme le filtrage orienté contenu et les méthodes basées sur les retours

des utilisateurs. De surcroît, les modalités de sa mise en place, ainsi que ses points forts et faibles, sont exposés. En outre, différents cas d'usage sont minutieusement décrits afin de mettre en lumière la pertinence de la solution face aux besoins préalablement identifiés.

### 4.3.1. Aperçu

Étant donné que le but principal du système proposé est de suggérer des vidéos pédagogiques pertinentes et de garantir leur pertinence éducative, il est crucial de tenir compte de ses dimensions éducatives et techniques. Bien que ces aspects soient développés plus en détail dans ce chapitre, il est d'ores et déjà possible d'obtenir un aperçu global du fonctionnement du système sans plonger dans les détails internes du mécanisme de recommandation. L'interface utilisateur se présente sous forme d'une application web équipée d'une barre de recherche, comme le montre la Figure 14 qui offre une vue d'ensemble du système proposé.

**Étape A :** L'utilisateur lance une requête, par exemple, "conjugue maths" dans le but de comprendre comment rationaliser le dénominateur d'une fraction en éliminant les racines carrées. Cette demande est alors transmise au serveur de l'application pour traitement.

**Étape B :** Le système sonde YouTube à la recherche de vidéos pertinentes liées au thème pour en extraire et stocker les informations pour des analyses ultérieures.

**Étape C :** Le système collecte les retours des utilisateurs, trie et prépare les données recueillies à partir des vidéos.

**Étape D :** Grâce à des méthodes de recommandation avancées, le système intègre les informations collectées lors des étapes B et C pour formuler des suggestions pertinentes de vidéos éducatives en adéquation avec la demande initiale de l'utilisateur.

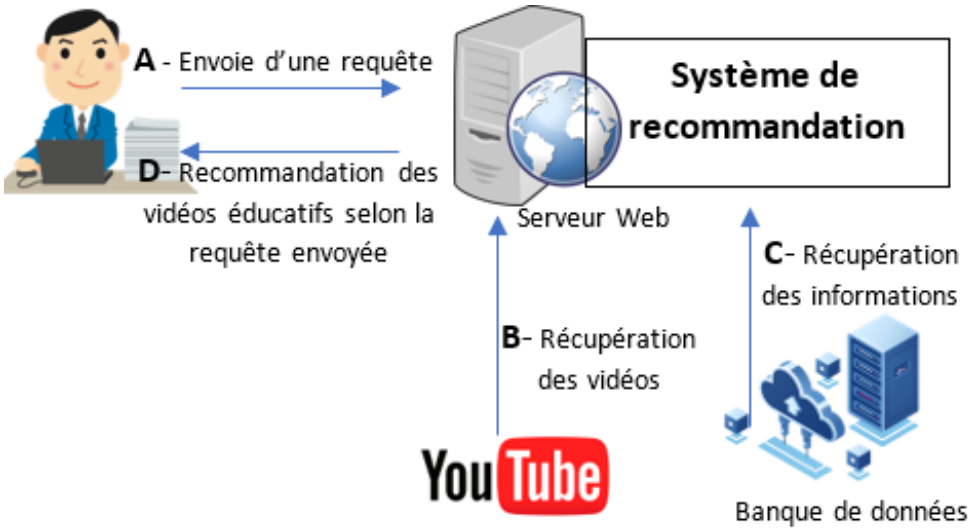


*Figure 14: Aperçu du système de recommandation.*

La solution proposée est un système en ligne doté d'une application qui trie les vidéos YouTube selon des thématiques précises, et qui utilise ces données pour proposer des vidéos associées à des contenus éducatifs susceptibles d'intéresser l'utilisateur. En essence, les suggestions formulées guident les utilisateurs vers des contenus éducatifs de haute qualité, réduisant ainsi la durée de recherche de ressources adéquates. Les bénéfices de cette solution s'étendent non seulement aux apprenants, mais aussi aux enseignants en quête de ressources pédagogiques complémentaires.

La solution mise en place exploite le moteur de recherche de YouTube, offrant aux utilisateurs une expérience similaire à celle du site web de YouTube. Ils peuvent effectuer les mêmes actions, telles que lire ou mettre en pause des vidéos, avancer ou revenir en arrière, et chercher des vidéos à l'aide de mots-clés ou de phrases. Lorsqu'un utilisateur saisit un mot-clé relatif à un sujet, le processus de suggestion de vidéos éducatives s'enclenche. Suite à cela, l'application développée orchestre une série d'opérations via ses différents modules, comme illustré dans la Figure 15.

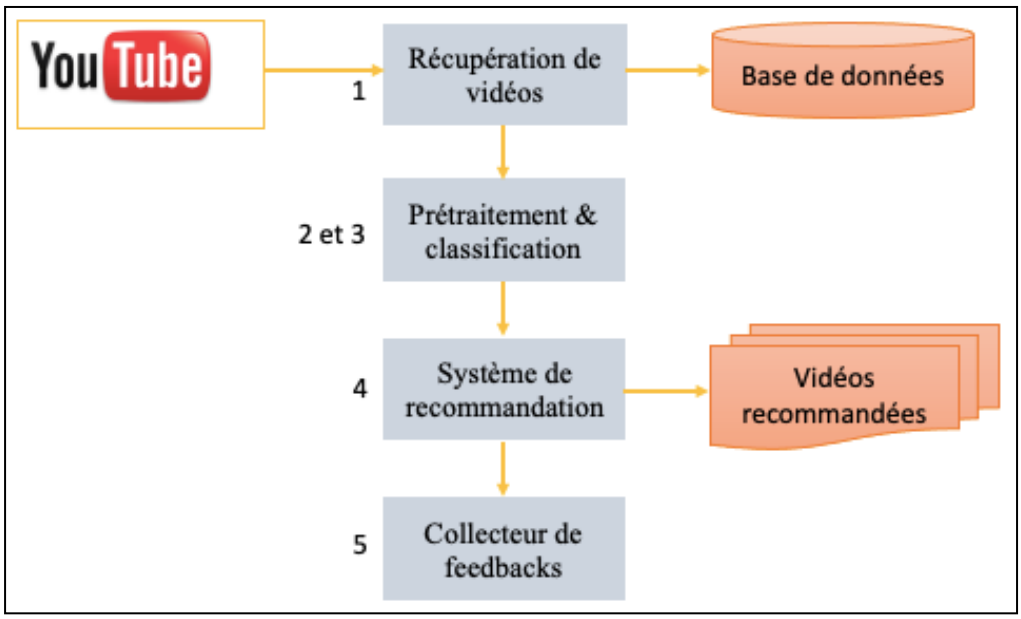


*Figure 15: Étapes suivies par la solution.*

Des détails supplémentaires concernant les modules présentés dans la Figure 15 seront détaillés dans les sections qui suivent. Voici une description sommaire des modules et de leurs fonctionnalités :

1. **Module de récupération de vidéo** : Il extrait des informations relatives aux vidéos YouTube en fonction des termes et mots-clés renseignés par l'utilisateur. Ces données sont ensuite enregistrées dans une base de données pour des traitements ultérieurs.

2. **Module de prétraitement** : Il est dédié au traitement des données recueillies précédemment. Agissant comme un filtre, il se focalise sur les mots et informations pertinents, afin de préparer les données pour les phases de classification et de recommandation.

3. **Module de classification** : En utilisant les données extraites de YouTube, ce module classe les vidéos en se basant sur divers attributs pertinents, définis par les algorithmes d'apprentissage.

4. **Système de recommandation** : Il calcule les correspondances entre le sujet recherché par l'utilisateur et les vidéos éducatives précédemment classées. De plus, il met en œuvre des algorithmes de recommandation tenant compte des préférences et intérêts de l'utilisateur, afin d'initier le processus de recommandation de vidéos.

5. **Collecteur de feedbacks** : Son rôle est de recueillir les avis et commentaires des utilisateurs sur les recommandations proposées.

La prise en compte des retours des utilisateurs est cruciale pour affiner et optimiser l'algorithme d'apprentissage utilisé dans le module de classification. Ainsi, la solution peut progressivement se perfectionner, devenant un outil éducatif de plus en plus efficace, enrichissant continuellement l'expérience d'apprentissage des utilisateurs.

### 4.3.2. Extraction de vidéo

La phase initiale du processus est axée sur l'extraction d'informations sur les vidéos depuis YouTube et leur sauvegarde dans une base de données locale pour des étapes ultérieures de classification et de prétraitement. Le mécanisme de recherche de vidéos est assez direct : le système met à disposition une interface dotée d'une barre de recherche, où l'utilisateur est invité à saisir des mots-clés pertinents pour mener à bien la recherche.

Ce système capitalise sur l'API de Google pour YouTube, offrant la capacité d'intégrer des fonctionnalités propres à YouTube dans un site web tiers ou une application distincte. Afin de dialoguer avec cette API, il est impératif d'obtenir un identifiant d'authentification. Ce dernier peut être acquis via la Console des développeurs Google, accessible via le lien https://console.developers.google.com/ [130]. Il existe trois catégories d'identifiants. Pour les besoins de ce projet, le choix s'est porté sur "API Key".

L'API offre une variété de ressources récupérables pour les utilisateurs. Dans le cadre de ce projet, la méthode SEARCH a été privilégiée, car elle fournit des résultats englobant divers éléments tels que les chaînes, les vidéos et les playlists. Cependant, pour exploiter ces ressources, certaines méthodes préliminaires doivent être mises en œuvre.

La méthode LIST a été sélectionnée pour ce projet. Elle retourne un ensemble de résultats basés sur les critères spécifiés dans la requête d'API [130, 131]. Bien que par défaut, elle identifie les vidéos, les chaînes et les éléments de playlist qui correspondent à la recherche, il est possible pour l'utilisateur d'ajuster la requête pour ne récupérer que des types d'éléments spécifiques.

Lors de l'envoi d'une requête HTTP pour interagir avec l'API de YouTube, on utilise la méthode GET avec l'URL suivante : https://www.googleapis.com/youtube/v3/search. Il est impératif de spécifier un paramètre nommé PART, qui détermine quelles propriétés de la ressource de

recherche seront renvoyées dans la réponse de l'API. Les options disponibles pour ce paramètre sont ID et SNIPPET [131, 132].

Chaque vidéo sur YouTube possède un identifiant unique, constitué de onze caractères, désigné sous le nom "ID". En utilisant le paramètre FIELDS, on peut définir les informations spécifiques à extraire pour chaque vidéo, telles que le titre, la description, l'image par défaut et le nom de la chaîne associée à la vidéo. Il est également essentiel de définir le paramètre MAX RESULTS, qui détermine le nombre maximal de vidéos qui seront renvoyées en réponse à la requête [133].

L'intégration de la clé créée dans le paramètre KEY est un aspect essentiel lors de l'interaction avec l'API de YouTube. La raison principale de stocker les informations récupérées de YouTube dans une base de données dédiée réside dans la nécessité de préserver l'intégrité des données face aux éventuels changements qui pourraient survenir sur la plateforme de YouTube. De plus, le traitement et le prétraitement des données sont cruciaux pour déduire des informations pertinentes et significatives.

Toutefois, il est important de souligner que la méthode de recherche "SEARCH" ne renvoie pas toutes les informations nécessaires concernant une vidéo. Par conséquent, pour obtenir des données complémentaires sur une vidéo spécifique, un appel supplémentaire à l'API est requis pour chaque vidéo. Cela nécessite l'utilisation de l'ID de la vidéo comme paramètre et la spécification des données requises dans le champ PART, comme illustré dans la Figure 16.

```
vidId = item['id']['videoId']
r = youtube.videos().list(
    part="statistics,contentDetails",
    id=vidId,
    fields="items(statistics," + \
                "contentDetails(duration))"
).execute()
try:
    duration = r['items'][0]['contentDetails']['duration']
    views = r['items'][0]['statistics']['viewCount']
    likes = r['items'][0]['statistics']['likeCount']
    favorites = r['items'][0]['statistics']['favoriteCount']
    comments = r['items'][0]['statistics']['commentCount']
    cats_info['id'].append(vidId)
    cats_info['duration'].append(duration)
    cats_info['views'].append(views)
    cats_info['likes'].append(likes)
    cats_info['favorites'].append(favorites)
    cats_info['comments'].append(comments)
except:
    pass
```

*Figure 16: Détails des vidéos.*

Ainsi, nous avons pu obtenir des détails supplémentaires comme la date à laquelle la vidéo a été mise en ligne, l'identifiant de la chaîne, l'identifiant de la catégorie, les étiquettes, la durée, la présence de sous-titres, le nombre total de vues, ainsi que le nombre de mentions "likes" et "dislikes" [134, 135].

**a.** La date de mise en ligne revêt une importance car une vidéo ancienne pourrait contenir des données qui ne sont plus actuelles, et le système pourrait s'en rendre compte, évitant ainsi de suggérer cette vidéo.

**b.** Des détails tels que la chaîne et la catégorie sont aussi essentiels. Par exemple, la catégorie pourrait permettre de cibler uniquement les vidéos éducatives, et certaines chaînes, en raison de leur réputation, pourraient être privilégiées comme sources fiables.

**c.** L'étiquette "TAG" est un élément-clé souvent utilisé dans divers systèmes de recommandation pour déterminer la ressemblance entre les vidéos.

**d.** La durée d'une vidéo peut donner une idée sur la profondeur de traitement d'un sujet, indiquant s'il s'agit d'un aperçu rapide ou d'une explication approfondie [135].

**e.** Le signe "CAPTION" indique la présence de sous-titres sur la vidéo, une caractéristique qui sera exploitée dans des projets ultérieurs pour convertir les sous-titres en mots-clés, améliorant ainsi la pertinence des vidéos suggérées par rapport au sujet d'intérêt.

**f.** Le nombre de vues peut suggérer le degré d'intérêt que la vidéo a suscité. Pour confirmer cette supposition, le nombre de mentions "likes" et le nombre de mentions "dislikes" apportent des éclaircissements [135].

L'objectif principal ici est de recueillir des informations pertinentes des vidéos sur YouTube et de les archiver dans une base de données. Bien que certaines données, comme le nombre de mentions "likes" ou de vues, puissent évoluer avec le temps, des mesures peuvent être envisagées pour actualiser ces données dans la base de données locale lors de futures étapes.

### 4.3.3. Prétraitement et classification des vidéos

Suite à la collecte des données mentionnée précédemment, celles-ci ont été enregistrées dans une base de données locale tel qu'elles ont été fournies par l'API YouTube, sans modification préalable. Toutefois, pour optimiser la classification et la recommandation des vidéos, il est

impératif de prétraiter ces données de façon ciblée. Par exemple, pour établir une corrélation entre un sujet de recherche et une vidéo, il est nécessaire de retravailler le titre en supprimant des éléments superflus comme les points, les espaces et autres caractères spéciaux, ainsi que des mots non pertinents. Ce processus minutieux a abouti à la conservation uniquement des termes pertinents dans une table spécifique.

Un autre aspect à considérer concerne le traitement d'informations spécifiques, comme la durée de la vidéo notée sous le format "PT21M14S". Dans ce format, "P" désigne une période (année, mois, semaine, jour) et "T" précise la durée, soit ici 21 minutes et 14 secondes. Cette manière de représenter la durée n'est pas propice à une classification efficace car elle combine deux informations distinctes [136]. Ainsi, lors du prétraitement, les durées ont été isolées et réparties dans des catégories pour avoir des données plus structurées. De plus, les mots-clés ou "tags" des vidéos, fournis comme des séquences de mots séparés par des virgules, ont été traités individuellement, étant donné leur récurrence à travers les vidéos. L'objectif étant de mieux comprendre leur emploi pour affiner les calculs de correspondance entre les vidéos.

Enfin, une fois les données préparées, c'est au tour de l'algorithme de machine learning de déterminer si une vidéo est à caractère pédagogique et si elle est qualitativement fiable. Pour ce faire, une base de cent vidéos a été établie, comprenant des avis de spécialistes et d'apprenants familiers avec le sujet "expression conjuguée en mathématiques". Il est aussi possible de se référer à la catégorie renvoyée par l'API YouTube. Parmi les 15 catégories définies [137], l'une d'elles, "Éducatif", certifie que le contenu est strictement pédagogique.

### 4.3.4. Algorithme du système de recommandation et résultats

Après avoir écarté les vidéos considérées non pertinentes, soit à cause de leur non-conformité avec l'aspect éducatif, soit en raison de leur qualité insuffisante, les vidéos retenues sont alors transmises à notre système de recommandation. Ce dernier est chargé de prendre en compte une liste de vidéos validées et de les trier en fonction de leur adéquation avec les besoins d'un utilisateur spécifique. Dans notre démarche, nous avons choisi de combiner deux techniques, FBC et FC, pour optimiser les suggestions et compenser les limites inhérentes à l'utilisation d'une seule méthode. Selon plusieurs experts [138], la meilleure classification pour cette démarche est une méthode hybride, où les deux techniques sont combinées dans une liste unifiée.

L'algorithme mis en œuvre pour élaborer les suggestions est qualifié « d'hybride ». Il s'agit du seul algorithme de cette nature qui repose sur deux techniques, FBC et FC, fonctionnant simultanément, comme illustré dans la Figure 17.

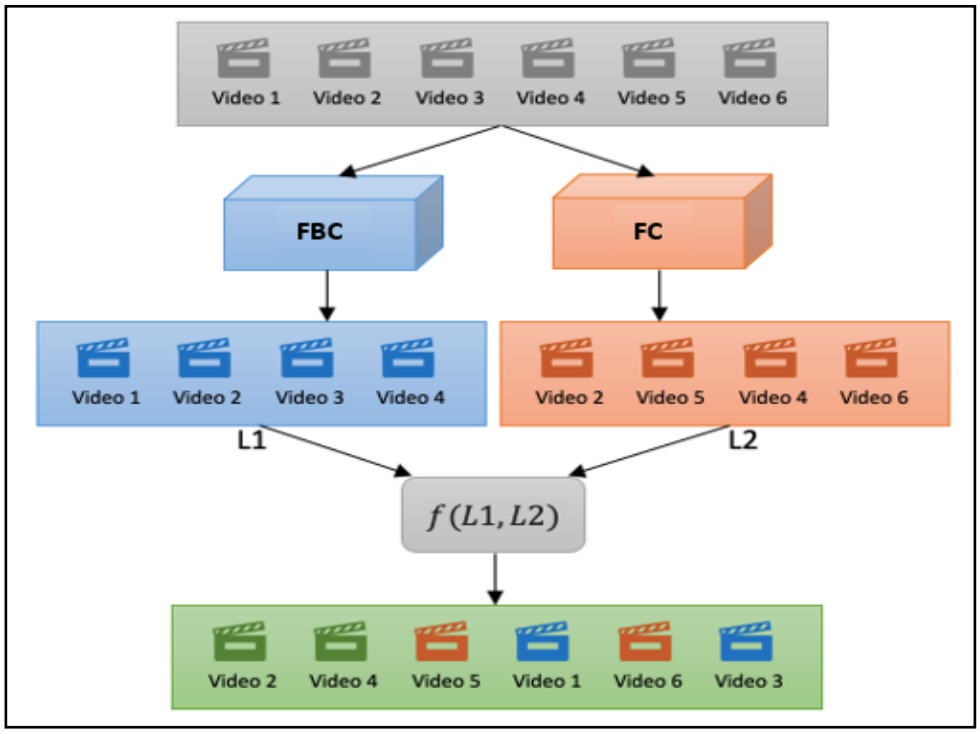


*Figure 17: Organigramme de l'algorithme combiné pour la proposition de suggestions de vidéos.*

Lorsqu'une vidéo est soumise au système de recommandation, elle est simultanément analysée par les algorithmes FBC et FC. Chacun de ces algorithmes produit sa propre liste de recommandations. Toutefois, les vidéos qui apparaissent sur les deux listes sont privilégiées lors de l'établissement de la liste finale. Ainsi, chaque vidéo des deux listes reçoit un score basé sur la somme de ses positions respectives dans les listes initiales.

La liste définitive est ensuite classée par ordre croissant de scores. Ainsi, une vidéo avec un score plus faible sera positionnée en tête de liste. Quant aux vidéos qui n'apparaissent que dans une des deux listes, elles sont ajoutées à la fin de la liste consolidée. Le processus d'ajout s'effectue alternativement : une vidéo de la liste du filtrage collaboratif est ajoutée, suivie d'une vidéo de la liste du filtrage basé sur le contenu, en respectant leur ordre d'apparition dans leurs listes respectives.

Sur la Figure 17 illustrée précédemment, l'organisation des vidéos en fonction de leur score combiné est mise en évidence. Prenons "Vidéo 2" comme exemple : elle occupe la 2e place sur

la liste FBC et la 1ère place sur la liste FC, lui donnant un score cumulé de 3 (2 + 1). Par conséquent, elle est positionnée en tête de la liste finale (LF) car elle possède le score le plus faible parmi toutes les vidéos. Ensuite, nous avons "Vidéo 4", qui avec un score total de 7 (4 + 3), se place juste après "Vidéo 2" sur la liste finale. Ces deux vidéos étaient les seules à apparaître sur les deux listes FBC et FC.

Après avoir traité les vidéos présentes sur les deux listes, le système procède à l'ajout des vidéos exclusives à chaque liste. Il commence par "Vidéo 5" de la liste FBC, suivie de "Vidéo 1" de la liste FC, puisque cette dernière est positionnée plus à gauche. Le processus continue ainsi, en alternant entre les vidéos des deux listes, en fonction de leur placement relatif sur leurs listes respectives.

La technologie de filtrage collaboratif s'appuie sur le partage d'opinions entre les utilisateurs pour élaborer la liste de recommandation *L2*. Cette approche s'inspire du traditionnel bouche-à-oreille que les individus ont toujours employé pour se forger une opinion sur des vidéos. La mise en œuvre de cette technique se décompose en plusieurs étapes :

**a.** Recueillir les avis et évaluations des utilisateurs concernant les vidéos qu'ils ont visionnées.

**b.** Incorporer ces retours dans le profil utilisateur.

**c.** Utiliser ce profil enrichi pour guider les utilisateurs vers de nouvelles découvertes vidéo.

Cependant, cette méthode présente quelques défis. L'un d'entre eux est le problème du "démarrage à froid" (cold start). C'est une situation où il est difficile de fournir des recommandations pour les nouveaux utilisateurs qui n'ont pas encore laissé d'avis ou d'évaluations. Sans ces données initiales, il est compliqué de les comparer à d'autres utilisateurs et de déduire leurs préférences. Heureusement, en s'appuyant sur le vaste référentiel de YouTube, il est possible d'obtenir des données préexistantes qui peuvent aider à générer des recommandations même en l'absence d'évaluations préalables de l'utilisateur.

YouTube, fréquenté par des millions d'internautes évaluant les vidéos, collecte des données comme le nombre de vues, de likes et de dislikes pour obtenir des renseignements utiles. Cette démarche vise principalement à recueillir des informations démographiques essentielles. Si les évaluations proviennent d'un échantillon d'utilisateurs homogène ou sont trop peu nombreuses,

il est conseillé d'exploiter ces informations issues des vidéos YouTube. La formule ci-après a été employée à cet effet :

$$\text{Weighted Average} = [(\%\ \text{Likes}) * 1 + (\%\ \text{Views}) * 2] / 3 \qquad (15)$$

Il arrive que les vidéos sur YouTube enregistrent un grand nombre de vues mais peu de "likes". Ce phénomène peut s'expliquer si le contenu ou la qualité n'a pas satisfait les spectateurs. Il est donc essentiel d'analyser le nombre total d'évaluations d'une vidéo pour déterminer la proportion de "likes" par rapport à l'ensemble des appréciations, positives et négatives. De plus, on calcule le pourcentage de vues d'une vidéo en relation avec le total des vues des vidéos. Par la suite, une moyenne pondérée est établie, accordant un poids de 1/3 au pourcentage de likes et 2/3 au pourcentage de vues, comme illustré dans l'équation 15 précédente.

L'approche décrite ci-dessus pour évaluer et classer les vidéos sur YouTube à l'aide d'une moyenne pondérée est intéressante et tente de résoudre un problème courant sur les plateformes de contenu généré par les utilisateurs. En combinant le nombre de vues et le pourcentage de mentions "likes", nous prenons en compte à la fois la popularité d'une vidéo et la façon dont elle est perçue par ceux qui la regardent. Voici quelques points de réflexion sur cette méthode :

a. **Nombre de vues** : C'est un indicateur de la popularité d'une vidéo. Cependant, comme nous l'avons souligné, ce n'est pas nécessairement un indicateur de qualité. Une vidéo peut avoir beaucoup de vues parce qu'elle a été bien commercialisée ou parce qu'elle a un titre accrocheur, mais elle peut ne pas être appréciée par les spectateurs.

b. **Mentions "**like**"** et **"**dislike**"** : Elles offrent une perspective directe sur la manière dont la vidéo est perçue par ceux qui la regardent. En examinant le pourcentage de mentions "likes" par rapport au total des mentions "likes" et "dislikes", nous obtenons une mesure de l'appréciation réelle de la vidéo par le public.

c. **Moyenne pondérée** : En utilisant la formule 15 ci-dessus, vous donnez plus de poids au nombre de vues qu'aux mentions "likes". Cela pourrait être particulièrement utile pour les nouvelles vidéos qui n'ont pas encore reçu beaucoup de feedback sous forme de mentions "likes" ou "dislikes".

d. **Considérations supplémentaires** :

- Il pourrait être utile d'introduire un seuil minimal de vues pour qu'une vidéo soit prise en compte. Cela éviterait que des vidéos avec un faible nombre de vues mais un pourcentage élevé de mentions "likes" ne soient pas classées trop haut.

- La durée de la vidéo pourrait également être un facteur à considérer. Par exemple, une vidéo de 30 secondes qui reçoit beaucoup de vues pourrait ne pas avoir autant de "valeur" qu'une vidéo de 10 minutes avec un nombre similaire de vues.

En conclusion, notre approche cherche à offrir une évaluation équilibrée des vidéos en combinant la popularité et l'appréciation. C'est une méthode pertinente pour évaluer le contenu sur une plateforme comme YouTube.

Il a été déterminé que le nombre de vues devrait être privilégié par rapport au nombre de "likes", en particulier pour les vidéos récemment mises en ligne qui pourraient ne pas avoir accumulé de nombreuses appréciations positives. Par exemple, le Tableau 5 illustre que la vidéo V3 a été vue seulement 7 fois, mais a reçu 5 "like". Pour pallier les évaluations insuffisantes, une approche consiste à ajuster les coefficients de similarité avec un "poids", réduisant ainsi la moyenne lorsque le nombre d'évaluations est faible.

*Tableau 5: Calcul de la moyenne pondérée*

| **Vidéos** | **V1** | **V2** | **V3** | **V4** |
|---|---|---|---|---|
| Nombre de "likes" | 2530 | 5142 | 5 | 17 |
| Nombre de "dislikes" | 12 | 19 | 0 | 3 |
| Nombre de Vues | 55641 | 96840 | 7 | 45 |
| Pourcentage de "likes" | 0,99527931 | 0,99631854 | 1 | 0,85 |
| Pourcentage de Vues | 0,36478008 | 0,63487901 | 0,00004589 | 0,00029502 |
| Moyenne Pondérée | 0,57494649 | 0,75535885 | 0,33336393 | 0,28353001 |

Les méthodes de filtrage basé sur le contenu, employant des techniques telles que l'extraction de caractéristiques, l'analyse textuelle et la mesure de similarité, ont été déployées pour élaborer les recommandations de la liste *L1*. Ces approches permettent d'examiner votre contenu pour y détecter des éléments semblables ou associés. Jusqu'à présent, nous nous sommes concentrés sur les titres et descriptions des vidéos. Cependant, dans le futur, nous envisageons d'intégrer

le contenu des sous-titres, visant ainsi une meilleure pertinence en mettant l'accent sur les termes les plus récurrents au sein d'une vidéo.

### 4.3.5. Algorithme de filtrage hybride de l'approche proposée

L'algorithme conçu pour le système de recommandation hybride combine les forces du filtrage collaboratif et du filtrage basé sur le contenu afin d'offrir des suggestions personnalisées aux utilisateurs. Au cœur de cette approche se trouve la méthode de calcul des similarités entre utilisateurs et entre éléments, qui permet de capturer les préférences et les comportements des utilisateurs. Les recommandations issues du filtrage collaboratif sont obtenues en identifiant les utilisateurs similaires et en extrapolant leurs préférences pour faire des suggestions. Parallèlement, le filtrage basé sur le contenu se concentre sur les caractéristiques des éléments eux-mêmes, recommandant des éléments similaires à ceux que l'utilisateur a appréciés par le passé. La fusion de ces deux ensembles de recommandations se fait via une fonction qui calcule une moyenne pondérée basée sur le pourcentage de "likes" et de vues des éléments, assurant ainsi que les recommandations finales représentent un équilibre entre les deux méthodologies. Ce processus hybride vise à améliorer la pertinence des recommandations et à enrichir l'expérience utilisateur en tenant compte de diverses dimensions de l'interaction utilisateur-contenu.

**Algorithm:** hybrid filtering in recommendation system

**Begin**

```
# Collaborative Filtering
user_sim = cal_user_sim () # Calculate similarities between users
user_pref = get_user_pref () # Get user preferences # Content-Based
Filtering item_sim = cal_item_sim () # Calculate similarities between
items
# Hybrid Filtering
def hybrid_rec(user_id):
    # Collaborative Filtering recommendations
    cf_rec = col_fil_rec (user_id, user_sim, user_pref) # Content-
    Based Filtering recommendations
    cb_rec = con_based_fil_rec (user_id, item_sim) #Combine
    recommendations
    hybrid_rec = combine_rec (cf_rec, cb_rec) return hybrid_rec
```

```
#Collaborative Filtering recommendations
def col_fil_rec (user_id, user_sim, user_pref):
      sim_users = find_sim_users (user_id, user_sim)
      rec = gen_rec_from_sim_users (sim_users, user_pref)
      return rec
# Content-Based Filtering recommendations
def con_based_fil_rec (user_id, item_sim):
      user_pref = get_user_pref(user_id)
      rec = gen_rec_based_on_pref (user_pref, item_sim)
      return rec
# Combine recommendations
def combine_rec (cf_rec, cb_rec):
      # Apply weights or other techniques
      # to combine the recommendations
      combined_rec = cf_rec + cb_rec
      return combined_rec
# Usage
Print (hybrid_rec (123))
End.
```

### 4.3.6. Collecte des feedbacks

L'obtention des feedbacks d'utilisateurs constitue l'une des phases finales du système de recommandation que nous proposons. Après la présentation des vidéos avec une explication sur la raison de leur suggestion, vient le moment d'évaluer la pertinence de ces recommandations. Les utilisateurs peuvent attribuer une note à la vidéo, allant d'une étoile à cinq étoiles. Ces évaluations sont ensuite prises en compte pour affiner les suggestions, en vérifiant si les recommandations correspondent aux attentes des utilisateurs.

### 4.3.7. Interface du système proposé

L'interface de la page d'accueil, illustrée par la figure 18, est conçue pour fournir une navigation aisée et une expérience sur mesure aux utilisateurs, avec une barre de recherche permettant de trouver du contenu spécifique, une section de recommandations qui s'adapte aux préférences de visionnage, et une mise en avant des vidéos les plus populaires du moment. Sur la page de

vidéo, les utilisateurs bénéficient d'un lecteur vidéo intégré, d'une description détaillée pour chaque vidéo, d'une option pour évaluer le contenu, d'une explication des raisons de la recommandation, et d'un espace de commentaires pour interagir avec la communauté. Le profil de l'utilisateur recense l'historique de visionnage et permet de personnaliser les préférences de recommandation. Enfin, la page des paramètres est essentielle pour la gestion des notifications et la protection de la vie privée, donnant aux utilisateurs le contrôle sur leurs données et la façon dont ils sont informés des nouveautés ou des changements sur la plateforme.

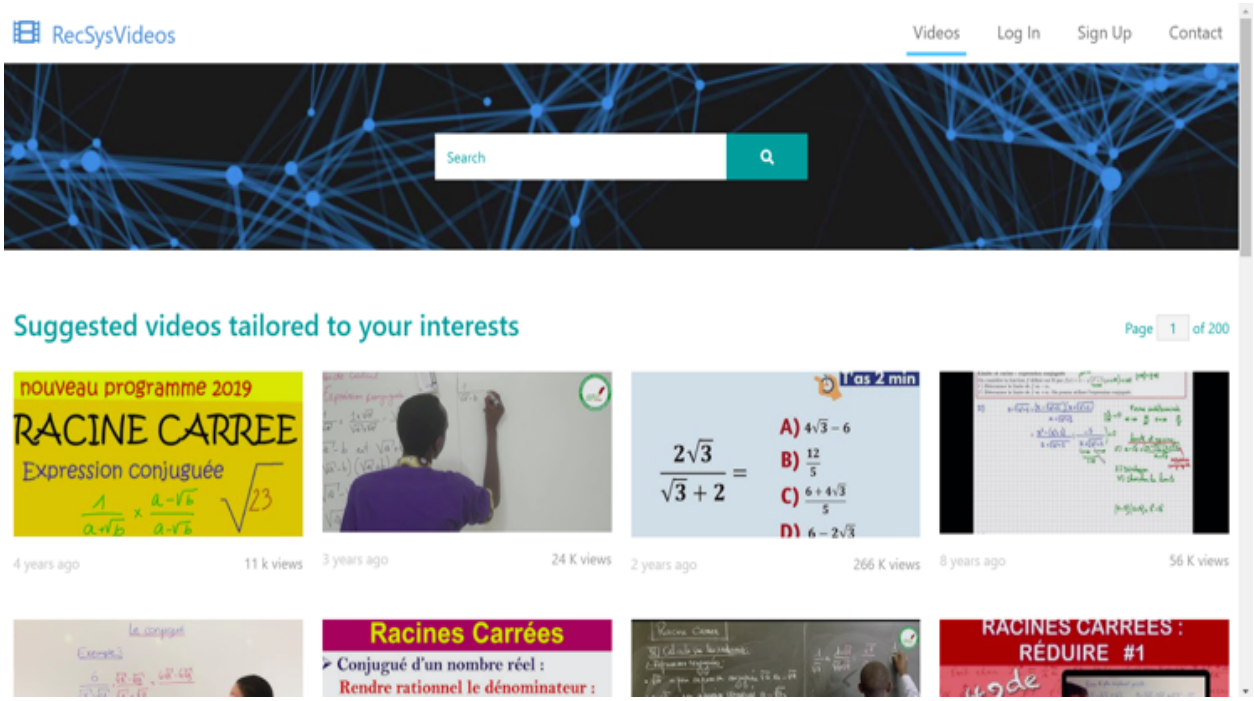


*Figure 18: Page d'accueil de l'Interface du système proposé*

## 4.4. Résultat et discussion

### 4.4.1. Évaluation

#### 4.4.1.1. Fréquence d'utilisation de YouTube

Avant de commencer le développement du système proposé, nous avons diffusé un questionnaire en ligne à 75 participants. Ce groupe était composé de stagiaires du Centre Régional des Métiers de l'Éducation et de la Formation (CRMEF) à Meknès et d'enseignants affiliés à la direction provinciale de l'éducation nationale de Taounate, au Maroc, où les tests ont eu lieu. L'objectif de ce sondage était de déterminer les habitudes d'utilisation de la plateforme YouTube et ses fonctionnalités associées.

Selon les données illustrées à la figure 19, une majorité significative des participants, représentant 54,7%, accède à YouTube tous les jours, ce qui suggère une intégration profonde de cette plateforme dans leur quotidien. Un autre segment notable de 21,3% se connecte à YouTube chaque semaine, montrant une régularité dans l'utilisation mais avec une fréquence moindre. Ensuite, un groupe plus restreint de 16% visite YouTube quelques fois par mois, indiquant une utilisation plus occasionnelle ou spécifique à certains besoins. La catégorie "quelques fois par an" ne semble pas être représentée dans le graphique, ce qui laisse penser que cette fréquence d'utilisation est soit très rare, soit complètement absente parmi les répondants. Finalement, un petit pourcentage, non spécifié dans votre résumé mais visible comme la portion non colorée sur le graphique, ne visite pas du tout YouTube, soulignant l'existence d'une minorité qui n'utilise pas cette plateforme.

Cette répartition indique que la majorité des participants sont des utilisateurs réguliers de YouTube, ce qui peut informer les décideurs sur l'importance d'intégrer YouTube comme un outil pédagogique ou de formation dans les programmes éducatifs au Maroc.

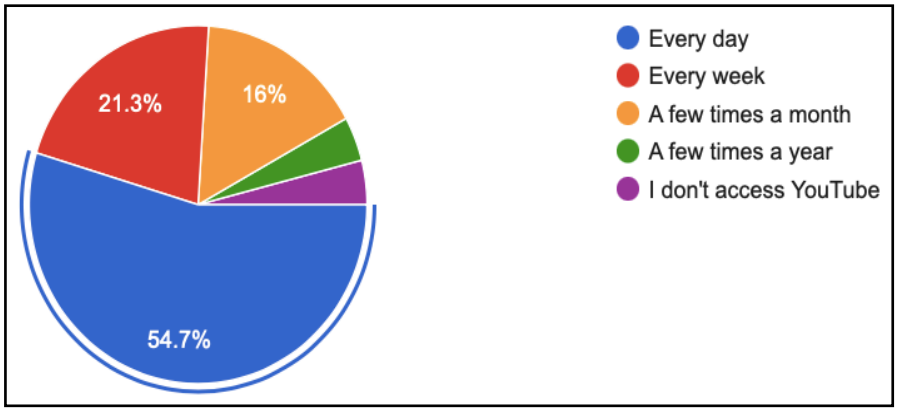


*Figure 19: Fréquence d'utilisation de YouTube.*

#### 4.4.1.2. Usage éducatif de YouTube

La Figure 20 met en lumière l'utilisation de YouTube à des fins éducatives parmi les participants. Elle révèle que 25,7% d'entre eux regardent des cours vidéo sur YouTube tous les jours, ce qui dénote une utilisation intensive de la plateforme pour l'éducation. La majorité des sondés, soit 41,9%, utilisent YouTube pour l'éducation de manière hebdomadaire, indiquant que, bien que l'utilisation ne soit pas quotidienne, elle reste une composante régulière de leur apprentissage. Un autre 17,6% se connectent à YouTube pour apprendre quelques fois par mois,

ce qui suggère qu'ils se servent de la plateforme pour des besoins éducatifs spécifiques ou ponctuels. Un plus petit pourcentage, 8,1%, se tourne vers YouTube pour l'éducation quelques fois par an, ce qui pourrait correspondre à une utilisation occasionnelle pour des besoins très ciblés, comme les révisions ou l'acquisition de nouvelles compétences. Finalement, il y a une part non spécifiée de participants qui n'utilisent pas du tout YouTube pour des cours vidéo, soulignant l'existence d'une minorité qui se passe de cette ressource pour leur formation ou leur enseignement. Cela donne un aperçu de la variété des fréquences d'utilisation de YouTube comme outil éducatif au sein de la population étudiée.

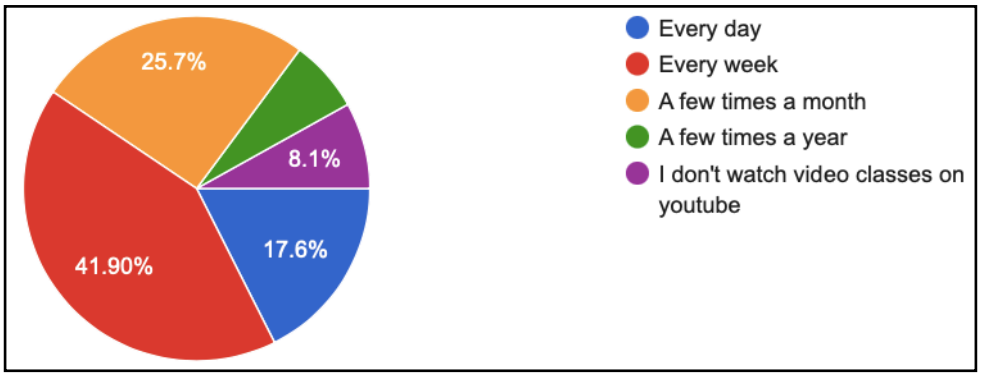


*Figure 20: Utilisation de YouTube à des fins éducatives.*

Ces données mettent en exergue l'importance cruciale de YouTube dans le paysage éducatif contemporain. Au-delà de sa fonction traditionnelle de divertissement, YouTube s'affirme comme un vecteur d'enseignement incontournable. Grâce à sa facilité d'accès et à la diversité de son offre de contenus, la plateforme se révèle être un atout considérable non seulement pour les apprenants et les formateurs mais également pour ceux qui poursuivent l'apprentissage en autonomie.

Les critères déterminants pour les utilisateurs lors de la sélection d'une vidéo sont avant tout la qualité visuelle et sonore, puis viennent le titre et la miniature de la vidéo. Pour tester le système recommandé, des vidéos ont été recherchées sur YouTube, reconnu pour son large éventail de contenus et son efficace moteur de recherche. Il est crucial de mentionner que, dans une première approche, aucune identification de l'utilisateur n'est prise en compte, pour éviter d'influencer les résultats. Par exemple, une recherche sur le mot "conjugué", généralement associé aux maths, a proposé des contenus variés : une vidéo sur la conjugaison chimique, une leçon sur la conjugaison des verbes en français et, en fin de liste, une sur le conjugué en

mathématiques. Cet éventail illustre la complexité de dénicher des ressources éducatives pertinentes parmi tant d'autres. Tout contenu non pertinent est considéré comme du "bruit".

En reproduisant cette recherche avec identification de l'utilisateur, les résultats ont été plus ciblés sur le concept de "conjugué en mathématiques", plaçant la vidéo précédemment en bas de liste en haut des recommandations. Cela suggère que YouTube s'appuie sur des éléments comme l'historique de visionnage pour affiner ses suggestions. Néanmoins, des bruits peuvent persister. Les utilisateurs sont donc invités à peaufiner leurs requêtes pour dénicher le contenu adéquat, ce qui peut être fastidieux. Une solution serait d'avoir un système de recommandation spécifiquement axé sur le contenu éducatif, excluant les vidéos musicales, humoristiques ou d'autres distractions. De cette manière, seuls les contenus les plus adaptés, basés sur le thème et les retours des utilisateurs, seraient mis en avant.

L'API YouTube, lorsqu'elle est utilisée pour la recherche de vidéos, ne donne pas toujours des résultats satisfaisants, même avec l'ajout de mots synonymes à la recherche. Les vidéos renvoyées peuvent être dans d'autres langues, aborder des sujets totalement différents ou simplement être de mauvaise qualité. Cela renforce l'argument selon lequel enrichir une recherche avec des synonymes peut être bénéfique. Toutefois, il ne suffit pas seulement d'ajouter de bons termes de recherche ; il est aussi crucial de filtrer les résultats inappropriés pour réduire le "bruit" envoyé à l'utilisateur. Par exemple, en utilisant l'enrichissement des requêtes, l'API YouTube a suggéré, en troisième place, une vidéo sur le "système conjugué" en chimie, et en quatrième, une vidéo sur la conjugaison en français. Ces deux vidéos, bien que pertinentes pour certains, étaient considérées comme du bruit dans notre contexte et donc non pertinentes. Cette expérience souligne le fait que, bien que puissant, l'algorithme de recommandation de YouTube peut être optimisé pour des besoins éducatifs spécifiques.

À partir de cette prise de conscience, un système dédié a été conçu, intégrant des techniques d'intelligence artificielle pour recommander des vidéos pertinentes selon le domaine d'étude recherché. Comme évoqué précédemment, notre solution proposée combine le "filtrage basé sur le contenu" et le "filtrage basé sur les évaluations". Mais avant de plonger dans les résultats de cette combinaison, il est intéressant de jeter un œil aux recommandations générées uniquement par la méthode de filtrage basé sur le contenu.

Le système a priorisé la précision des termes de recherche par rapport aux évaluations des utilisateurs lors de la formulation des recommandations. Cela a conduit à des recommandations comme la première vidéo, qui, bien qu'étant pertinente en termes de titre, peut ne pas l'être en termes de qualité de contenu. Sa faible popularité après une année sur YouTube suggère qu'elle n'a peut-être pas été bien reçue par les utilisateurs. Cette observation souligne l'importance d'un équilibre entre la pertinence des termes et la qualité perçue, comme le suggèrent les évaluations des utilisateurs.

L'inclusion du terme "conjugué" dans les titres des neuf premières recommandations montre l'efficacité de l'algorithme à identifier les vidéos basées sur les termes clés recherchés. Cependant, cela ne garantit pas que le contenu de la vidéo est de qualité ou pertinent pour l'utilisateur. Les recommandations 3 et 4, en revanche, semblent être de bonnes suggestions basées à la fois sur la qualité visuelle et sonore et sur leur pertinence par rapport au sujet recherché. Cela suggère que, bien que le filtrage basé sur le contenu soit utile pour la pertinence du sujet, d'autres facteurs, tels que les évaluations des utilisateurs et la qualité du contenu, doivent également être pris en compte pour améliorer la précision des recommandations.

Il est intéressant de constater que la vidéo, initialement en quatrième position lors du filtrage basé sur le contenu, a grimpé à la première place lors de l'utilisation de la méthode de fusion. Ceci est dû à l'ajout du score FBC de 4 au score FC de 22, donnant à la vidéo un score total de 26. Ce score était le plus bas dans le calcul réalisé par l'algorithme de fusion. Par ailleurs, d'autres vidéos de la même chaîne figuraient parmi les sept recommandations suivantes. Leur qualité de production supérieure, tant en termes d'audio que de visuel, pourrait expliquer les évaluations positives qu'elles ont accumulées. Bien que les vidéos en troisième et quatrième positions ne contenaient pas l'expression "conjugué" dans leurs titres ou tags, elles étaient néanmoins pertinentes, abordant le thème complémentaire de la "racine carrée conjuguée". Elles étaient d'ailleurs les seules dans le top 10 à ne pas mentionner "conjugué". Sans le seuil fixé pour le score technique FBC, davantage de telles recommandations auraient pu apparaître. Les scores inférieurs à 0,30 ont été exclus, les rendant inapplicables dans la méthode de fusion, ce qui a permis de privilégier des recommandations plus pertinentes et mieux classées.

Effectivement, une autre suggestion concernait le "Système conjugué de la chimie", qui était classée en 10e position. Bien qu'elle partage le terme "conjugué" avec les termes de recherche, elle n'était pas pertinente au contexte recherché. Sa présence dans la liste est attribuable aux

bonnes évaluations qu'elle a reçues et à son score élevé de 0,72 en technique CBF. Néanmoins, comme il s'agit d'un terme homonyme – ayant la même orthographe et la même prononciation mais relevant d'un domaine différent – cette recommandation est jugée inappropriée. Même si elle est placée relativement loin dans le classement, l'utilisation de techniques pour éliminer les homonymes peut être envisagée pour améliorer la précision des recommandations. Les retours et évaluations des recommandations peuvent être une piste à explorer. De plus, l'exploitation des ontologies peut s'avérer bénéfique pour pallier ce genre de problème.

Les retours d'évaluation indiqués ci-après démontrent que le système élaboré a recueilli un niveau d'approbation encourageant de la part de ses utilisateurs. Dans cette optique, des sélections préalablement filtrées ont été présentées à un ensemble de 21 participants au Centre Régional de l'Éducation et de la Formation des Professions à FES-MEKNES, Maroc. Ce groupe était principalement constitué d'enseignants stagiaires en informatique, au nombre de 17, ainsi que de 4 formateurs spécialisés en informatique. Ils ont été sollicités pour évaluer la pertinence des vidéos suggérées, en se focalisant sur le thème "conjugué maths" et en prenant en compte les potentiels éléments perturbateurs (telles que les vidéos non éducatives, en d'autres langues, hors-sujet ou de faible qualité).

Quatre listes distinctes ont été présentées, chacune ayant ses caractéristiques spécifiques, comme illustré dans le tableau 6:

**Liste A :** vidéos suggérées par YouTube suite à une requête sur "conjugué" sans se connecter.

**Liste B :** vidéos suggérées par YouTube après recherche du terme ciblé avec connexion, supposant que l'utilisateur ait précédemment visionné des vidéos relatives au thème.

**Liste C :** vidéos retournées par l'API YouTube, dont les termes de recherche ont été optimisés grâce à l'enrichissement de la requête.

**Liste D :** vidéos identifiées par le système développé, qui, outre l'amélioration de la requête, intègre une stratégie de recommandation mixte ainsi que le classement YouTube.

*Tableau 6: Analyse du Taux de rejet et des opinions des utilisateurs sur différentes listes de recommandations vidéo.*

| Liste | Taux de rejet | Opinions des utilisateurs |
|---|---|---|
| A | 80% | 80% de rejet, les utilisateurs ont rapporté nécessiter beaucoup d'effort pour trouver une vidéo intéressante sur cette liste. Cela indique que les recommandations générées sans connexion de l'utilisateur peuvent ne pas être suffisamment ciblées. |
| B | 75% | Avec un taux de rejet de 75%, les utilisateurs ont également mentionné devoir fournir plus d'effort pour dénicher une vidéo intéressante. Cette liste bénéficie de la connexion de l'utilisateur, ce qui suggère que les vidéos visionnées précédemment ont un impact modéré sur la pertinence des suggestions. |
| C | 60% | Le taux de rejet baisse à 60% ici, où les utilisateurs ont trouvé que des efforts étaient nécessaires mais légèrement réduits par rapport aux listes A et B. Cette liste utilise des termes de recherche optimisés par l'API YouTube, ce qui semble améliorer légèrement la pertinence des recommandations. |
| D | 15% | Cette liste montre un net progrès avec un taux de rejet de seulement 15%. Les utilisateurs ont indiqué avoir atteint leur objectif avec les trois premières recommandations, ce qui souligne l'efficacité du système développé qui combine une stratégie de recommandation mixte et le classement YouTube pour affiner les résultats. |

Outre l'amélioration des requêtes dans la dernière sélection, diverses méthodes ont été mises en œuvre, notamment l'intelligence artificielle. L'algorithme de fusion, qui a été largement discuté dans la stratégie de filtrage combiné des systèmes de recommandation, en est un exemple. Ces éléments ont sans doute influencé les retours positifs obtenus. Parmi les autres points saillants, les commentaires élogieux concernant la distinction entre les vidéos "éducatives" et "non éducatives" sont notables. Ces commentaires ont mis en avant la capacité à éliminer la majorité des vidéos hors sujet. L'exploitation d'autres données relatives à la vidéo, telles que la langue, la durée et les avis des utilisateurs de YouTube, a été utile pour trier les vidéos. Cela a notamment permis d'exclure les vidéos qui n'étaient pas dans la langue de l'utilisateur.

Les vidéos sélectionnées ont été soumises à deux techniques de recommandation. D'abord, le filtrage basé sur le contenu a été utilisé, lequel évalue la ressemblance entre les termes du sujet cherché et ceux du titre et de la description de la vidéo. Puis, il classe les vidéos selon leur pertinence vis-à-vis du sujet abordé. La seconde technique est le filtrage collaboratif. Elle réalise des analyses basées sur les notations des utilisateurs de YouTube et produit une liste de vidéos classées selon leur pertinence par rapport à ces évaluations. Dans la liste définitive, la

priorité est donnée aux vidéos présentes dans les deux listes. Les autres sont ajoutées alternativement selon leur classement dans les listes d'origine.

### 4.4.2. Discussion

L'analyse comparative des systèmes de recommandation met en évidence des éléments cruciaux influençant leur performance et leur justesse dans le cadre des recommandations de vidéos éducatives. À cet égard : La décision de privilégier la reconnaissance textuelle à partir des images, comme c'est le cas pour VCSR, ou de s'appuyer sur les sous-titres aisément disponibles sur des plateformes actuelles telles que YouTube, est fondamentale. La reconnaissance textuelle dans les images pourrait engendrer des complications et des erreurs, alors que YouTube propose un accès direct et intuitif aux sous-titres, facilitant grandement la consultation du contenu pour des utilisateurs avec des besoins linguistiques ou auditifs variés. Le choix d'une vidéothèque est primordial. La popularité internationale de YouTube, sa riche collection de contenus éducatifs, son interface intuitive, ses fonctionnalités de reconnaissance vocale et son large éventail de contenus en font une option privilégiée pour des objectifs éducatifs, comme le démontre cette étude. L'introduction d'algorithmes d'apprentissage pour évaluer les vidéos est une démarche prospective. Cela permet au système de constamment adapter et perfectionner ses suggestions. En contraste, les systèmes de recommandation conventionnels se basant exclusivement sur des standards préétablis pourraient peiner à s'adapter aux changements des attentes des utilisateurs et à la dynamique du contenu. L'adoption d'un algorithme de recommandation mixte, mêlant des techniques basées sur le contenu et sur les avis, renforce efficacement la pertinence des suggestions. L'évaluation basée sur le contenu garantit la concordance thématique, tandis que les recommandations axées sur les avis reflètent les goûts des utilisateurs, favorisant ainsi leur contentement tout en couvrant un éventail d'intérêts et de besoins. En somme, ce débat souligne la pertinence de se tourner vers des plateformes contemporaines telles que YouTube et des technologies de pointe, qui simplifient des tâches comme la récupération de sous-titres et proposent une expérience utilisateur plus immersive. De surcroît, l'introduction d'algorithmes d'apprentissage et de techniques de recommandation combinées manifeste une réaction adaptative face à l'évolution du contenu éducatif et des attentes des utilisateurs, réitérant la volonté d'optimiser l'accessibilité et la justesse des recommandations vidéo pour une audience internationale.

## 4.5. Conclusion

Pour conclure, cette contribution met l'accent sur l'élaboration et le déploiement d'un système de recommandation destiné aux vidéos éducatives présentes sur YouTube. Le système en question tire parti des méthodes de filtrage basées sur le contenu et collaboratif afin de proposer des suggestions sur-mesure, en fonction des préférences thématiques de l'utilisateur et des évaluations des vidéos. Les constatations révèlent que la prise en compte des retours des utilisateurs de YouTube et de la langue d'origine des vidéos améliore sensiblement la précision des recommandations. Les apports majeurs de cette étude résident dans la création d'un système de recommandation polyvalent, adaptable à une variété de contextes d'application. La conception et les spécificités techniques de ce système peuvent servir de modèle pour d'autres projets similaires, tandis que ses API REST assurent une intégration et une utilisation fluides. Les retours des utilisateurs témoignent d'un taux d'adoption positif, confirmant ainsi la pertinence du système pour guider les individus dans leur quête de contenus éducatifs de qualité sur YouTube.

Néanmoins, bien que le système dévoilé affiche des performances prometteuses, il laisse entrevoir des marges d'amélioration et de recherche pour l'avenir. Parmi ces axes d'amélioration, nous pouvons citer l'analyse sentimentale des commentaires laissés par les utilisateurs, la reconnaissance des chaînes vidéo fiables, la gestion des vidéos hors sujet, ou encore l'exploitation des sous-titres pour optimiser le filtrage basé sur le contenu. Une étude plus poussée des métadonnées associées aux vidéos et le recours à des algorithmes d'apprentissage automatique pour l'évaluation de la qualité figurent également parmi les pistes envisageables pour les étapes suivantes.

En somme, ce chapitre met en lumière le succès de la mise en place d'un système de recommandation spécifique aux vidéos éducatives sur YouTube. Ses fonctionnalités s'avèrent être un atout de taille pour guider les utilisateurs vers des contenus pédagogiques pertinents et de premier choix. Ces résultats enrichissent le champ des systèmes de recommandation et suggèrent des voies d'amélioration dans le cadre de l'éducation non formelle. Les futures initiatives pourront s'inspirer de ces découvertes pour perfectionner et diversifier le système, au bénéfice des apprenants et des formateurs dans leur exploration des ressources pédagogiques sur le web.

# CONCLUSION GÉNÉRALE ET PERSPECTIVES

# Conclusion générale et perspectives

À travers cette thèse, nous avons exploré le potentiel transformateur des systèmes de recommandation intelligents dans le domaine de l'éducation, spécifiquement pour la recommandation de ressources pédagogiques vidéo. Cette recherche s'est appuyée sur une confluence de domaines tels que l'apprentissage intelligent, l'analyse de données éducatives, et les technologies de recommandation, pour répondre à des défis uniques posés par le paysage numérique en constante évolution de l'éducation.

Nos investigations ont démontré que, lorsqu'ils sont bien conçus et mis en œuvre, les systèmes de recommandation intelligents peuvent significativement améliorer l'accès à des ressources éducatives pertinentes et personnalisées. Les méthodes que nous avons développées ont prouvé leur efficacité en tenant compte des préférences et des besoins des utilisateurs, intégrant des aspects tels que la confiance et les interactions sociales pour affiner les recommandations. De plus, l'adoption d'une approche hybride combinant le filtrage basé sur le contenu et le filtrage collaboratif a également apporté des avantages notables.

L'application de ces solutions à des plateformes réelles comme YouTube a révélé des améliorations significatives dans la manière dont les apprenants accèdent et interagissent avec les contenus pédagogiques. En permettant une personnalisation plus profonde et en guidant les utilisateurs vers des ressources plus pertinentes, ces systèmes ont le potentiel de rendre l'apprentissage en ligne plus efficace et engageant.

Toutefois, notre travail soulève également de nouvelles questions et ouvre des pistes pour des recherches futures. Parmi celles-ci, l'incorporation de l'analyse sentimentale des retours utilisateurs et l'exploitation des sous-titres pour améliorer le filtrage basé sur le contenu représentent des domaines prometteurs. De plus, la gestion des problématiques de confidentialité et d'éthique liées à l'utilisation des données utilisateurs dans les systèmes de recommandation nécessite une attention continue.

Une exploration plus approfondie des mécanismes permettant de reconnaître les sources fiables et de filtrer les contenus hors sujet pourrait également contribuer à l'élaboration de systèmes de recommandation plus robustes et précis. Par ailleurs, l'application de ces systèmes dans

différents contextes éducatifs, notamment dans l'éducation formelle et l'apprentissage tout au long de la vie, représente une autre voie intéressante pour des recherches ultérieures.

En conclusion, cette thèse contribue de manière significative à notre compréhension de l'apprentissage intelligent et des systèmes de recommandation dans l'éducation. Elle offre un cadre pour le développement futur de technologies éducatives plus adaptatives et personnalisées, et met en lumière le rôle crucial que la technologie peut jouer dans l'amélioration de l'expérience d'apprentissage à l'ère numérique. Les perspectives et les défis identifiés dans ce travail ne manqueront pas d'inspirer de futures innovations et recherches dans ce domaine passionnant.

# Publications

## - Revues Internationales

**TIMMI Mohamed**, LAAOUINA Loubna, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. Educational Video Recommender System. *International Journal of Information and Education Technology*, 2024, vol. 14, no 3. Indexes in Scopus.

**TIMMI Mohamed**, LAAOUINA Loubna, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. Recommendation System based on User Trust and Ratings. *International Journal of Advanced Computer Science and Applications*, 2022, vol. 13, no 7. Indexes in Scopus.

## - Conférences Internationales

**TIMMI Mohamed**, LAAOUINA Loubna, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. *"*Comparative Evaluation of YouTube, Facebook, TikTok, and Instagram as Online Education Tools". In : *International Conference on Intelligent Systems and Computer Vision (ISCV 2024)*. (IEEE).

**TIMMI Mohamed**, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. **«** Advancing Towards Intelligent E-Learning Systems ». In : The 7th *International Conference on Big Data end Internet of Things (BDIoT'24)*. (Springer International Publishing).

**TIMMI Mohamed**, LAAOUINA Loubna, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. *"*Learning analytics: a review of the literatur"e. In : *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE, 2022. p. 1-4.

**TIMMI Mohamed**, JEGHAL Adil, EL GAROUANI Said, and YAHYAOUY Ali. *"*The review of objectives, methods, tools, and algorithms for educational data mining". In : *International Conference on Digital Technologies and Applications*. Cham : Springer International Publishing, 2021. p. 177-188.

# Bibliographie

[1] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Past, present, and future of smart learning: a topic-based bibliometric analysis," *International Journal of Educational Technology in Higher Education,* vol. 18, pp. 1-29, 2021.

[2] A. Bhutoria, "Personalized education and artificial intelligence in the United States, China, and India: A systematic review using a human-in-the-loop model," *Computers and Education: Artificial Intelligence,* vol. 3, p. 100068, 2022.

[3] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data,* vol. 9, no. 1, p. 59, 2022.

[4] Y. Lin *et al.*, "A survey on reinforcement learning for recommender systems," *IEEE Transactions on Neural Networks and Learning Systems,* 2023.

[5] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, "Recommender systems challenges and solutions survey," in *2019 international conference on innovative trends in computer engineering (ITCE)*, 2019: IEEE, pp. 149-155.

[6] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simulation Modelling Practice and Theory,* vol. 113, p. 102375, 2021.

[7] S. A. Thorat, G. Ashwini, and M. Seema, "Survey on Collaborative and Content-based Recommendation Systems," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2023: IEEE, pp. 1541-1548.

[8] D. P. D. Rajendran and R. P. Sundarraj, "Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings," *International Journal of Information Management Data Insights,* vol. 1, no. 2, p. 100027, 2021.

[9] D. Jannach, P. Pu, F. Ricci, and M. Zanker, "Recommender systems: Past, present, future," *Ai Magazine,* vol. 42, no. 3, pp. 3-6, 2021.

[10] R. S. Sharma, A. A. Shaikh, and E. Li, "Designing Recommendation or Suggestion Systems: looking to the future," *Electronic Markets,* vol. 31, pp. 243-252, 2021.

[11] E. Calvano, G. Calzolari, V. Denicolò, and S. Pastorello, "Artificial Intelligence, Algorithmic Recommendations and Competition," *Algorithmic Recommendations and Competition (May 14, 2023),* 2023.

[12] T. Balaji, C. S. R. Annavarapu, and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Computer Science Review,* vol. 40, p. 100395, 2021.

[13] W. Wongvilaisakul, P. Netinant, and M. Rukhiran, "Dynamic Multi-Criteria Decision Making of Graduate Admission Recommender System: AHP and Fuzzy AHP Approaches," *Sustainability,* vol. 15, no. 12, p. 9758, 2023.

[14] G. Fan, C. Zhang, J. Chen, P. Li, Y. Li, and V. C. Leung, "Improving Rating Prediction in Multi-Criteria Recommender Systems Via a Collective Factor Model," *IEEE Transactions on Network Science and Engineering,* 2023.

[15] T. B. Ogunseyi, C. B. Avoussoukpo, and Y. Jiang, "A systematic review of privacy techniques in recommendation systems," *International Journal of Information Security,* pp. 1-14, 2023.

[16] Z. Halim, S. Hussain, and R. H. Ali, "Identifying content unaware features influencing popularity of videos on youtube: A study based on seven regions," *Expert Systems with Applications,* vol. 206, p. 117836, 2022.

[17] B. Alhijawi, M. A. Alrub, and M. Al-Fayoumi, "Generalized Ethereum Blockchain-based recommender system framework," *Information Systems,* vol. 111, p. 102113, 2023.

[18] M. T. Ramakrishna *et al.*, "HCoF: Hybrid Collaborative Filtering Using Social and Semantic Suggestions for Friend Recommendation," *Electronics,* vol. 12, no. 6, p. 1365, 2023.

[19] K. Ludwig *et al.*, "Divided by the algorithm? The (limited) effects of content-and sentiment-based news recommendation on affective, ideological, and perceived polarization," *Social Science Computer Review,* p. 08944393221149290, 2023.

[20] G. Onofrei, R. Filieri, and L. Kennedy, "Social media interactions, purchase intention, and behavioural engagement: The mediating role of source and content factors," *Journal of Business Research,* vol. 142, pp. 100-112, 2022.

[21] A. M. A. Ausat, "The Role of Social Media in Shaping Public Opinion and Its Influence on Economic Decisions," *Technology and Society Perspectives (TACIT),* vol. 1, no. 1, pp. 35-44, 2023.

[22] B. Yu and Y. W. Leung, "Establishing ties or strengthening friendships? Students' use of online social networks in intercultural friendship development," *Information Technology & People,* 2023.

[23] S. Jhaver, A. Q. Zhang, Q. Z. Chen, N. Natarajan, R. Wang, and A. X. Zhang, "Personalizing content moderation on social media: User perspectives on moderation choices, interface design, and labor," *Proceedings of the ACM on Human-Computer Interaction,* vol. 7, no. CSCW2, pp. 1-33, 2023.

[24] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, and M. Wang, "A neural influence diffusion model for social recommendation," in *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, 2019, pp. 235-244.

[25] M. TIMMI, L. LAAOUINA, A. JEGHAL, S. EL GAROUANI, and A. YAHYAOUY, "Recommendation System based on User Trust and Ratings," *International Journal of Advanced Computer Science and Applications,* vol. 13, no. 7, 2022.

[26] V. Mohammadi, A. M. Rahmani, A. M. Darwesh, and A. Sahafi, "Trust-based recommendation systems in Internet of Things: a systematic literature review," *Human-centric Computing and Information Sciences,* vol. 9, no. 1, pp. 1-61, 2019.

[27] H. Yin, Q. Wang, K. Zheng, Z. Li, J. Yang, and X. Zhou, "Social influence-based group representation learning for group recommendation," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019: IEEE, pp. 566-577.

[28] C. Song, B. Wang, Q. Jiang, Y. Zhang, R. He, and Y. Hou, "Social recommendation with implicit social influence," in *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2021, pp. 1788-1792.

[29] R. Abolghasemi, P. Engelstad, E. Herrera-Viedma, and A. Yazidi, "A personality-aware group recommendation system based on pairwise preferences," *Information Sciences,* vol. 595, pp. 1-17, 2022.

[30] S. Reyssier and S. Simonian, "Effet d'un EIAH sur la motivation des élèves à faire des mathématiques: caractéristiques des élèves et style de l'enseignant," *Canadian Journal of Education/Revue canadienne de l'éducation,* 2021.

[31] C. De Medio, C. Limongelli, F. Sciarrone, and M. Temperini, "MoodleREC: A recommendation system for creating courses using the moodle e-learning platform," *Computers in Human Behavior,* vol. 104, p. 106168, 2020.

[32] G. George and A. M. Lal, "Review of ontology-based recommender systems in e-learning," *Computers & Education,* vol. 142, p. 103642, 2019.

[33] M. Markowski, H. Bower, R. Essex, and C. Yearley, "Peer learning and collaborative placement models in health care: a systematic review and qualitative synthesis of the literature," *Journal of Clinical Nursing,* vol. 30, no. 11-12, pp. 1519-1541, 2021.

[34] D. Shi, T. Wang, H. Xing, and H. Xu, "A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning," *Knowledge-Based Systems,* vol. 195, p. 105618, 2020.

[35] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems,* vol. 30, 2017.

[36] C. Conati, O. Barral, V. Putnam, and L. Rieger, "Toward personalized XAI: A case study in intelligent tutoring systems," *Artificial intelligence,* vol. 298, p. 103503, 2021.

[37] J. Monsalve-Pulido, J. Aguilar, E. Montoya, and C. Salazar, "Autonomous recommender system architecture for virtual learning environments," *Applied Computing and Informatics,* vol. 20, no. 1/2, pp. 69-88, 2024.

[38] B. Li *et al.*, "A personalized recommendation framework based on MOOC system integrating deep learning and big data," *Computers and Electrical Engineering,* vol. 106, p. 108571, 2023.

[39] P. Rishad, N. Saurav, L. Laiju, J. Jayaraj, G. P. Kumar, and C. Sheela, "Application of fuzzy logic in video recommendation system for syllabus driven E-learning platform," in *AIP Conference Proceedings*, 2021, vol. 2336, no. 1: AIP Publishing.

[40] C.-L. Tang, J. Liao, H.-C. Wang, C.-Y. Sung, Y.-R. Cao, and W.-C. Lin, "Supporting online video learning with concept map-based recommendation of learning path," in

*Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1-8.

[41] N. Belarbi, N. Chafiq, M. Talbi, A. Namir, and H. Benlahmar, "A recommender system for videos suggestion in a SPOC: A proposed personalized learning method," in *Big Data and Smart Digital Environment*, 2019: Springer, pp. 92-101.

[42] T. Silveira, M. Zhang, X. Lin, Y. Liu, and S. Ma, "How good your recommender system is? A survey on evaluations in recommendation," *International Journal of Machine Learning and Cybernetics,* vol. 10, pp. 813-831, 2019.

[43] J. Zhu, B. G. Patra, H. Wu, and A. Yaseen, "A novel NIH research grant recommender using BERT," *PloS one,* vol. 18, no. 1, p. e0278636, 2023.

[44] X. Cai, Z. Hu, P. Zhao, W. Zhang, and J. Chen, "A hybrid recommendation system with many-objective evolutionary algorithm," *Expert Systems with Applications,* vol. 159, p. 113648, 2020.

[45] F. Wang, H. Zhu, G. Srivastava, S. Li, M. R. Khosravi, and L. Qi, "Robust collaborative filtering recommendation with user-item-trust records," *IEEE Transactions on Computational Social Systems,* 2021.

[46] X. Wang, R. Zhang, Y. Sun, and J. Qi, "Doubly robust joint learning for recommendation on data missing not at random," in *International Conference on Machine Learning*, 2019: PMLR, pp. 6638-6647.

[47] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *applied sciences,* vol. 10, no. 21, p. 7748, 2020.

[48] K. Balog, F. Radlinski, and S. Arakelyan, "Transparent, scrutable and explainable user models for personalized recommendation," in *Proceedings of the 42nd international acm sigir conference on research and development in information retrieval*, 2019, pp. 265-274.

[49] N. Balasubramaniam, M. Kauppinen, A. Rannisto, K. Hiekkanen, and S. Kujala, "Transparency and explainability of AI systems: From ethical guidelines to requirements," *Information and Software Technology,* vol. 159, p. 107197, 2023.

[50] F. L. da Silva, B. K. Slodkowski, K. K. A. da Silva, and S. C. Cazella, "A systematic literature review on educational recommender systems for teaching and learning: research trends, limitations and opportunities," *Education and Information Technologies,* vol. 28, no. 3, pp. 3289-3328, 2023.

[51] R. Deffayet, T. Thonet, J.-M. Renders, and M. de Rijke, "Offline evaluation for reinforcement learning-based recommendation: a critical issue and some alternatives," in *ACM SIGIR Forum*, 2023, vol. 56, no. 2: ACM New York, NY, USA, pp. 1-14.

[52] A. Sun, "Take a Fresh Look at Recommender Systems from an Evaluation Standpoint," in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 2629-2638.

[53] C. Shi, X. Wang, S. Luo, H. Zhu, J. Ye, and R. Song, "Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework," *Journal of the American Statistical Association,* vol. 118, no. 543, pp. 2059-2071, 2023.

[54] Y. Gambo and M. Z. Shakir, "Evaluating students' experiences in self-regulated smart learning environment," *Education and Information Technologies,* vol. 28, no. 1, pp. 547-580, 2023.

[55] J. Zhang, Q. Jing, Y. Liang, H. Jiang, and N. Li, "Smart learning environments in school: design principles and case studies," in *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*: Springer, 2023, pp. 3659-3686.

[56] M. A. Haque *et al.*, "Sustainable and efficient E-learning internet of things system through blockchain technology," *E-Learning and Digital Media,* p. 20427530231156711, 2023.

[57] W. Villegas-Ch and J. García-Ortiz, "Enhancing Learning Personalization in Educational Environments through Ontology-Based Knowledge Representation," *Computers,* vol. 12, no. 10, p. 199, 2023.

[58] G. Yadav, Y.-J. Tseng, and X. Ni, "Contextualizing Problems to Student Interests at Scale in Intelligent Tutoring System Using Large Language Models," *arXiv preprint arXiv:2306.00190,* 2023.

[59] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining and Knowledge Discovery,* vol. 10, no. 3, p. e1355, 2020, doi: 10.1002/widm.1355.

[60] W. Chango, J. A. Lara, R. Cerezo, and C. Romero, "A review on data fusion in multimodal learning analytics and educational data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 12, no. 4, p. e1458, 2022.

[61] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining and Knowledge Discovery,* 2020, doi: 10.1002/widm.1355.

[62] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Applied Sciences,* vol. 10, no. 11, p. 3894, 2020.

[63] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, "A systematic review of deep learning approaches to educational data mining," *Complexity,* vol. 2019, 2019.

[64] C. Jalota and R. Agrawal, "Analysis of educational data mining using classification," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019: IEEE, pp. 243-247.

[65] S. Hussain, R. Atallah, A. Kamsin, and J. Hazarika, "Classification, clustering and association rule mining in educational datasets using data mining tools: A case study," in *Cybernetics and Algorithms in Intelligent Systems: Proceedings of 7th Computer Science On-line Conference 2018, Volume 3 7*, 2019: Springer, pp. 196-211.

[66] H.-C. Hung, I.-F. Liu, C.-T. Liang, and Y.-S. Su, "Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education," *Symmetry,* vol. 12, no. 2, p. 213, 2020.

[67] D. N. Sulisawati, L. Lutfiyah, F. Murtinasari, and L. Sukma, "Differences of Visual, Auditorial, Kinesthetic Students in Understanding Mathematics Problems," *Malikussaleh Journal of Mathematics Learning (MJML),* vol. 2, no. 2, pp. 45-51, 2019.

[68] C. Fischer *et al.*, "Mining big data in education: Affordances and challenges," *Review of Research in Education,* vol. 44, no. 1, pp. 130-160, 2020.

[69] H. Li, Y. Xiong, C. V. Hunter, X. Guo, and R. Tywoniw, "Does peer assessment promote student learning? A meta-analysis," *Assessment & Evaluation in Higher Education,* vol. 45, no. 2, pp. 193-211, 2020.

[70] X. Xia, "Learning behavior mining and decision recommendation based on association rules in interactive learning environment," *Interactive Learning Environments,* vol. 31, no. 2, pp. 593-608, 2023.

[71] C. Baek and T. Doleck, "Educational data mining versus learning analytics: A review of publications from 2015 to 2019," *Interactive Learning Environments,* vol. 31, no. 6, pp. 3828-3850, 2023.

[72] R. Saadatdoost, A. T. H. Sim, and E. Yadegaridehkordi, "The Coursera Community Framework: exploring the MOOC as a Community of Practice," *Journal of e-Learning and Knowledge Society,* vol. 19, no. 1, pp. 53-66, 2023.

[73] N. Z. Dina and N. Juniarta, "Analysis of the Educational Application User's Satisfaction Levels by Using the Integrated Entropy Weight-VIKOR Method," *DESIDOC Journal of Library & Information Technology,* vol. 43, no. 2, 2023.

[74] G. K. Renu and K. Sharma, "A NOVEL APPROACH TO MAKING EFFICIENT E-LEARNING PROCESS IN THE USING BIG DATA."

[75] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics,* vol. 37, pp. 13-49, 2019/04/01/ 2019, doi: https://doi.org/10.1016/j.tele.2019.01.007.

[76] C. Guzmán-Valenzuela, C. Gómez-González, A. Rojas-Murphy Tagle, and A. Lorca-Vyhmeister, "Learning analytics in higher education: a preponderance of analytics but very little learning?," *International Journal of Educational Technology in Higher Education,* vol. 18, pp. 1-19, 2021.

[77] K. Y. Koay and W. C. Poon, "Understanding students' cyberslacking behaviour in e-learning environments: Is student engagement the key?," *International Journal of Human–Computer Interaction,* vol. 39, no. 13, pp. 2573-2588, 2023.

[78] R. F. Kizilcec and D. Davis, "Learning Analytics Education: A case study, review of current programs, and recommendations for instructors," in *Practicable Learning Analytics*: Springer, 2023, pp. 133-154.

[79] A. Kukkar, R. Mohana, A. Sharma, and A. Nayyar, "Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms," *Education and Information Technologies,* pp. 1-30, 2023.

[80] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, "Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course," *International Journal of Educational Technology in Higher Education,* vol. 20, no. 1, pp. 1-23, 2023.

[81] H. Waheed, S.-U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: A neural network approach," *Expert Systems with Applications,* vol. 213, p. 118868, 2023.

[82] J. Sahni, "Is Learning Analytics the Future of Online Education?: Assessing Student Engagement and Academic Performance in the Online Learning Environment," *International Journal of Emerging Technologies in Learning (Online),* vol. 18, no. 2, p. 33, 2023.

[83] A. Tlili *et al.*, "Automatic modeling learner's personality using learning analytics approach in an intelligent Moodle learning platform," *Interactive Learning Environments,* vol. 31, no. 5, pp. 2529-2543, 2023.

[84] R. Heiser, M. E. D. Stritto, A. Brown, and B. Croft, "Amplifying Student and Administrator Perspectives on Equity and Bias in Learning Analytics: Alone Together in Higher Education," *Journal of Learning Analytics,* vol. 10, no. 1, pp. 8-23, 2023.

[85] R. Yousefi Nooraie, J. E. Sale, A. Marin, and L. E. Ross, "Social network analysis: An example of fusion between quantitative and qualitative methods," *Journal of Mixed Methods Research,* vol. 14, no. 1, pp. 110-124, 2020.

[86] A. Alam and A. Mohanty, "Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics," in *International Conference on Communication, Networks and Computing*, 2022: Springer, pp. 166-177.

[87] R. Kasperski and I. Blau, "Social capital in high-schools: teacher-student relationships within an online social network and their association with in-class interactions and learning," *Interactive Learning Environments,* vol. 31, no. 2, pp. 955-971, 2023.

[88] A. Sayal *et al.*, "Visual Sentiment Analysis Using Machine Learning for Entertainment Applications," in *2023 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 2023: IEEE, pp. 1-8.

[89] E. H. Korkut and E. Surer, "Visualization in virtual reality: a systematic review," *Virtual Reality,* pp. 1-34, 2023.

[90] A. M. F. Yousef and A. R. Khatiry, "Cognitive versus behavioral learning analytics dashboards for supporting learner's awareness, reflection, and learning process," *Interactive Learning Environments,* vol. 31, no. 9, pp. 5460-5476, 2023.

[91] M. A. Almulla and W. M. Al-Rahmi, "Integrated social cognitive theory with learning input factors: the effects of problem-solving skills and critical thinking skills on learning performance sustainability," *Sustainability,* vol. 15, no. 5, p. 3978, 2023.

[92] S. Buckingham Shum, "Embedding learning analytics in a university: Boardroom, staff room, server room, classroom," in *Practicable Learning Analytics*: Springer, 2023, pp. 17-33.

[93] N. S. Raj and V. Renumol, "A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020," *Journal of Computers in Education,* vol. 9, no. 1, pp. 113-148, 2022.

[94] S. Venkatesan, "A Recommender System Based on Matrix Factorization Techniques Using Collaborative Filtering Algorithm," *neuroquantology,* vol. 21, no. 5, p. 864, 2023.

[95] L. Huang, H. Ma, X. He, and L. Chang, "Leveraging Multisource Information in Matrix Factorization for Social Collaborative Filtering," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020: IEEE, pp. 1-8.

[96] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex & Intelligent Systems,* vol. 7, pp. 439-457, 2021.

[97] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data," *Expert Systems with Applications,* vol. 149, p. 113248, 2020.

[98] P. Ma, L. Wang, and J. Qin, "A Low-Rank Tensor Factorization Using Implicit Similarity in Trust Relationships," *Symmetry,* vol. 12, no. 3, p. 439, 2020.

[99] C. Park, D. Kim, J. Oh, and H. Yu, "Improving top-K recommendation with truster and trustee relationship in user trust network," *Information Sciences,* vol. 374, pp. 100-114, 2016.

[100] M. Al-Ghamdi, H. Elazhary, and A. Mojahed, "Evaluation of Collaborative Filtering for Recommender Systems," *International Journal of Advanced Computer Science and Applications,* vol. 12, no. 3, 2021.

[101] B. Yang, Y. Lei, J. Liu, and W. Li, "Social collaborative filtering by trust," *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 8, pp. 1633-1647, 2016.

[102] H. Chen, S. Wang, N. Jiang, Z. Li, N. Yan, and L. Shi, "Trust-aware generative adversarial network with recurrent neural network for recommender systems," *International Journal of Intelligent Systems,* vol. 36, no. 2, pp. 778-795, 2021.

[103] Y. Ruan and A. Durresi, "A survey of trust management systems for online social communities–trust modeling, trust inference and attacks," *Knowledge-Based Systems,* vol. 106, pp. 150-163, 2016.

[104] F.-S. Hsieh, "Trust-based recommendation for shared mobility systems based on a discrete self-adaptive neighborhood search differential evolution algorithm," *Electronics,* vol. 11, no. 5, p. 776, 2022.

[105] C. Xu, A. S. Ding, and K. Zhao, "A novel POI recommendation method based on trust relationship and spatial–temporal factors," *Electronic Commerce Research and Applications,* vol. 48, p. 101060, 2021.

[106] J. Shokeen, C. Rana, and P. Rani, "A trust-based approach to extract social relationships for recommendation," in *Data Analytics and Management*: Springer, 2021, pp. 51-58.

[107] K. Zhang, X. Liu, W. Wang, and J. Li, "Multi-criteria recommender system based on social relationships and criteria preferences," *Expert Systems with Applications,* vol. 176, p. 114868, 2021.

[108] S. Raza and C. Ding, "A regularized model to trade-off between accuracy and diversity in a news recommender system," in *2020 IEEE International Conference on Big Data (Big Data)*, 2020: IEEE, pp. 551-560.

[109] H. Zhang, I. Ganchev, N. S. Nikolov, and M. Stevenson, "UserReg: A simple but strong model for rating prediction," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021: IEEE, pp. 3595-3599.

[110] J. Shokeen and C. Rana, "Social recommender systems: techniques, domains, metrics, datasets and future scope," *Journal of Intelligent Information Systems,* vol. 54, no. 3, pp. 633-667, 2020.

[111] A. Rahim, M. Y. Durrani, S. Gillani, Z. Ali, N. U. Hasan, and M. Kim, "An efficient recommender system algorithm using trust data," *The Journal of Supercomputing,* vol. 78, no. 3, pp. 3184-3204, 2022.

[112] W. Zhang, X. Zhang, H. Wang, and D. Chen, "A deep variational matrix factorization method for recommendation on large scale sparse dataset," *Neurocomputing,* vol. 334, pp. 206-218, 2019.

[113] R. Chen, Y.-S. Chang, Q. Hua, Q. Gao, X. Ji, and B. Wang, "An enhanced social matrix factorization model for recommendation based on social networks using social interaction factors," *Multimedia Tools and Applications,* pp. 1-31, 2020.

[114] Y. Pan, F. He, H. Yu, and H. Li, "Learning adaptive trust strength with user roles of truster and trustee for trust-aware recommender systems," *Applied Intelligence,* vol. 50, no. 2, pp. 314-327, 2020.

[115] S. Xu, H. Zhuang, F. Sun, S. Wang, T. Wu, and J. Dong, "Recommendation algorithm of probabilistic matrix factorization based on directed trust," *Computers & Electrical Engineering,* vol. 93, p. 107206, 2021.

[116] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in neural information processing systems*, 2008, pp. 1257-1264.

[117] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM conference on Information and knowledge management*, 2008, pp. 931-940.

[118] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009, pp. 203-210.

[119] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the fourth ACM conference on Recommender systems*, 2010, pp. 135-142.

[120] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 287-296.

[121] Z. Shen, M. Pritchard, S. Tan, and C. Noteboom, "Educative Sensemaking on Social Media: An Empirical Investigation of Informal Learning on YouTube," in *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.

[122] Y. Deldjoo, M. Schedl, P. Cremonesi, and G. Pasi, "Recommender systems leveraging multimedia content," *ACM Computing Surveys (CSUR),* vol. 53, no. 5, pp. 1-38, 2020.

[123] B. Hui, L. Zhang, X. Zhou, X. Wen, and Y. Nian, "Personalized recommendation system based on knowledge embedding and historical behavior," *Applied Intelligence,* pp. 1-13, 2022.

[124] R. Mervin, D. Mavalaru, and T. Thomas, "Knowledge-Driven and Intelligent Computing in Healthcare," *Handbook of Intelligent Healthcare Analytics: Knowledge Engineering with Big Data Analytics,* pp. 167-188, 2022.

[125] M. Abbas, M. U. Riaz, A. Rauf, M. T. Khan, and S. Khalid, "Context-aware Youtube recommender system," in *2017 international conference on information and communication technologies (ICICT)*, 2017: IEEE, pp. 161-164, doi: https://ieeexplore.ieee.org/document/8320183.

[126] M. Tavakoli, S. Hakimov, R. Ewerth, and G. Kismihok, "A recommender system for open educational videos based on skill requirements," in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, 2020: IEEE, pp. 1-5.

[127] S. Qin, R. Menezes, and M. Silaghi, "A recommender system for youtube based on its network of reviewers," in *2010 IEEE Second International Conference on Social Computing*, 2010: IEEE, pp. 323-328.

[128] J. Sultana, "A survey on YouTube recommendation system," 2020.

[129] J. C. Yang, Y. T. Huang, C. C. Tsai, C. I. Chung, and Y. C. Wu, "An automatic multimedia content summarization system for video recommendation," *Journal of Educational Technology & Society,* vol. 12, no. 1, pp. 49-61, 2009.

[130] A. A. Vera and C. M. González, "Educational Resource Recommender Systems Using Python and Moodle," in *International Conference on Computational Science and Its Applications*, 2022: Springer, pp. 15-30.

[131] R. Ilieva, Y. Yonkov, and A. Tonchev, "Building Independent Models of YouTube Content Networks as a Tool for Monitoring Health Misinformation," in *2022 XXXI International Scientific Conference Electronics (ET)*, 2022: IEEE, pp. 1-6.

[132] T. Alrashed, J. Almahmoud, A. X. Zhang, and D. R. Karger, "Scrapir: making web data apis accessible to end users," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1-12.

[133] S. Garg, Saurabh, and M. Breja, "Social network analysis of YouTube: a case study on content diversity and genre recommendation," in *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1*, 2021: Springer, pp. 25-37.

[134] M. Kurdi, N. Albadi, and S. Mishra, "“Think before you upload”: an in-depth analysis of unavailable videos on YouTube," *Social Network Analysis and Mining,* vol. 11, no. 1, p. 48, 2021.

[135] A. C. Munaro, R. Hübner Barcelos, E. C. Francisco Maffezzolli, J. P. Santos Rodrigues, and E. Cabrera Paraiso, "To engage or not engage? The features of video content on YouTube affecting digital consumer engagement," *Journal of consumer behaviour,* vol. 20, no. 5, pp. 1336-1352, 2021.

[136] R. F. Alhujaili and W. M. Yafooz, "Sentiment analysis for youtube videos with user comments," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021: IEEE, pp. 814-820.

[137] A. R. Shaikh, H. Alhoori, and M. Sun, "YouTube and science: models for research impact," *Scientometrics,* vol. 128, no. 2, pp. 933-955, 2023.

[138] H. Bouazza, B. Said, and F. Z. Laallam, "A hybrid IoT services recommender system using social IoT," *Journal of King Saud University-Computer and Information Sciences,* vol. 34, no. 8, pp. 5633-5645, 2022.