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

كلية العلوم ظهر المهراز فاس
ⵜⴰⵖⵉⵍⵜ ⵏ ⵜⵣⵔⴼⵜ ⵎⵓⵔⴰⵣ ⴼⴰⵙ
FACULTE DES SCIENCES DHAR EL MAHRAZ FES

N° d'ordre : 86/2024

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

Formation Doctorale : ***STIC***

Discipline : ***Informatique***
Spécialité :***Informatique***
Laboratoire : ***LISAC***

THÈSE DE DOCTORAT

Système basé sur le profilage et la recommandation au service du tourisme Marocain

Présentée par :

Khalid AL FARARNI

Soutenue le 27/12/2024 devant le jury :

Nom et Prénom	Grade	Etablissement	Qualité
TAIRI Hamid	PES	Faculté des Sciences Dhar EL Mahraz, Fès USMBA	Président
EL BANNAY Omar	MCH	Ecole Nationale des Sciences Appliquées, Khouribga, USMS	Rapporteur
EL FAZAZY Khalid	PES	Faculté des Sciences Dhar EL Mahraz, Fès USMBA	Rapporteur
FARHAOUI Youssef	MCH	Faculté des Sciences et Techniques, Errachidia, UMI	Rapporteur
MAHRAZ Mohamed Adnane	MCH	Faculté des Sciences Dhar EL Mahraz, Fès USMBA	Examinateur
SABRI My Abdelouahed	PES	Faculté des Sciences Dhar EL Mahraz, Fès USMBA	Examinateur
AGHOUTANE Badraddine	PES	Faculté des Sciences, Meknès, UMI	Co-directeur de Thèse
YAHYAOUY Ali	PES	Faculté des Sciences Dhar EL Mahraz, Fès USMBA	Directeur de Thèse

Année Universitaire : 2023-2024

Année Universitaire : 2023-2024

Remerciements

Je remercie le bon DIEU, qui m'a donné la force, la volonté et le courage pour terminer ce modeste travail.

Ce travail n'a pu être possible que du fait de personnes qui y ont aimablement consacrés leur temps. Je tiens à exprimer d'abord toute ma gratitude au Pr. Ali YAHYAOUY, mon directeur de thèse, d'avoir accepté d'encadrer mon travail de thèse et d'avoir pris le temps de m'assister tout au long de cette période. Je tiens également à le remercier pour son soutien et sa disponibilité, ainsi que pour les nombreuses suggestions utiles qui ont contribué à la réussite de ce projet.

Je tiens également à remercier sincèrement le Pr. Badraddine AGHOUTANE, professeur assistant à l'université Moulay Ismail Meknès, pour son savoir, ses propositions, son entière disponibilité, ses conseils judicieux, son suivi par les réunions qui ont rythmé les différentes étapes de ce travail, sa patience et surtout sa gentillesse.

Mes remerciements vont également aux rapporteurs et membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner ce travail.

C'est un énorme remerciement que j'adresse à ma Mère et mon Père. Vous avez su à votre manière, par vos paroles et vos gestes, m'encourager et m'accompagner dans tous les moments de ce mémoire.

Je tiens à remercier tous mes amis et collègues pour leur soutien moral, et leurs grands efforts tout au long de cette préparation.

Mes remerciements les plus sincères sont adressés à toute personne que j'ai oubliée de citer ici et qui a contribué de près ou de loin à la réalisation de ce travail.

Résumé

Avec le développement d'Internet, des technologies et des moyens de communication, la production de données touristiques s'est multipliée à tous les niveaux (hôtels, restaurants, transports, patrimoine, événements touristiques, activités, etc. Cependant, la liste des possibilités offertes aux touristes par ces moteurs de recherche Web (ou même par des sites touristiques spécialisés) peut être accablante et les résultats pertinents sont généralement noyés dans le "bruit" informationnel, ce qui empêche, ou du moins ralentit le processus de sélection. Pour aider les touristes à planifier leur voyage et à trouver les informations qu'ils recherchent, de nombreux systèmes de recommandation ont été mis au point. Dans ce projet de recherche, nous présentons un aperçu des différentes approches de recommandation utilisées dans le domaine du tourisme. A partir de cette étude, une architecture et un cadre conceptuel pour le système de recommandation touristique sont proposés, basés sur une approche de recommandation hybride. Le système proposé va au-delà de la recommandation d'une liste d'attractions touristiques, adaptée aux préférences des touristes. Il peut être considéré comme un planificateur de voyage qui conçoit un programme détaillé, incluant des ressources touristiques hétérogènes, pour une durée de visite spécifique. Le but ultime est de développer un système de recommandation basé sur les technologies de données à grande échelle, l'intelligence artificielle et la recherche opérationnelle pour promouvoir le tourisme au Maroc, en particulier dans la région de Drâa-Tafilalet.

Mots-clés : Systèmes de recommandation, profilage des utilisateurs, filtrage basé sur le contenu, filtrage collaboratif, système de recommandation hybride, e-tourisme, planification de voyage

Abstract

With the development of the Internet, technology, and means of communication, the production of tourist data has multiplied at all levels (hotels, restaurants, transport, heritage, tourist events, activities, etc.), especially with the development of Online Travel Agency (OTA). However, the list of possibilities offered to tourists by these Web search engines (or even specialized tourist sites) can be overwhelming and relevant results are usually drowned in informational "noise", which prevents, or at least slows down the selection process. To assist tourists in trip planning and help them to find the information they are looking for, many recommender systems have been developed. In this research project, we present an overview of the various recommendation approaches used in the field of tourism. From this study, an architecture and a conceptual framework for tourism recommender system are proposed, based on a hybrid recommendation approach. The proposed system goes beyond the recommendation of a list of tourist attractions, tailored to tourist preferences. It can be seen as a trip planner that designs a detailed program, including heterogeneous tourism resources, for a specific visit duration. The ultimate goal is to develop a recommender system based on big data technologies, artificial intelligence, and operational research to promote tourism in Morocco, specifically in the Daraâ-Tafilalet region.

Keywords: Recommender Systems, User Profiling, Content-Based Filtering, Collaborative Filtering, Hybrid Recommender System, e-Tourism, Trip planning

Table de matières

Remerciements ii
Résumé... iii
Abstract iv
Tables des matières v
Liste des figures viii
Liste des tableaux ix
Liste des abréviations x
Introduction générale 11
1.Contexte de recherche 11
2. Objectifs de recherche 12
3. Problématique scientifique et motivation 13
4. Démarche méthodologique et organisation de la thèse 14
5. Contributions et productions scientifiques 16
Chapitre I : Systèmes de recommandation : Vue d'ensemble 18
I.1. Introduction 19
I.2. Les systèmes de recommandation 19
I.2.1. Définition 19
I.2.2. Historique 20
I.3. Approches de recommandation 22
I.3.1. Recommandation basée sur le contenu 22
I.3.2. Recommandation basée sur le filtrage collaboratif 23
I.3.2.1. Recommandation basée sur le voisinage (la mémoire) 25
I.3.2.2. Recommandation basée sur le modèle 25
I.3.3. Autres approches de recommandation 26
I.3.3.1. Recommandation démographique 26
I.3.3.2. Recommandation contextuelle 27
I.3.4. Recommandation basée sur le filtrage hybride 27
I.3.4.1. Techniques d'hybridation 28
I.4. Évaluation Les systèmes de recommandation 30
I.4.1. Techniques d'évaluation 31
I.4.1.1. Test en ligne 31
I.4.1.2. Études utilisateurs 31
I.4.1.3. Simulation (Évaluation hors-ligne) 32

I.5. Conclusion **34**
Chapitre II : Critique de l'existant et proposition d'une solution **35**
II.1. Introduction **36**
II.2. Travaux connexes : Les systèmes de recommandation touristiques **36**
II.2.1. Basés sur le contenu 36
II.2.2. Basés sur le filtrage collaboratif 38
II.2.3. Sensibles au contexte 39
II.2.4. Basés sur le filtrage hybride 39
II.3. Approche proposée **40**
II.3.1. Description du problème et solution proposée 40
II.3.2. Le cadre conceptuel de l'architecture proposée 42
II.3.3. Diagrammes de système 46
II.3.4. Méthodologie d'intégration de big data et IA dans l'architecture proposée. …… 48
II.4. Conclusion **50**
Chapitre III : L'IA au Service des Voyageurs : Révolution des Systèmes de Recommandation Intelligents **52**
III.1. Introduction **53**
III.2. Analyse du Tourisme Traditionnel **54**
III.2.1. Méthodes et Techniques Traditionnelles Employées dans le Tourisme 55
III.2.2. Défis et Limites du Tourisme Traditionnel 56
III.3. Transformation Numérique dans le Tourisme **58**
III.3.1. Importance de la Révolution Numérique pour le Secteur du Tourisme 59
III.3.2. Techniques Modernes Utilisées dans le Tourisme 60
III.3.3. Avantages et Opportunités Offertes par le Tourisme Numérique 61
III.4. L'Intégration des Systèmes de Recommandation dans le Tourisme **62**
III.4.1. Rôle des Systèmes de Recommandation dans la Personnalisation de l'Expérience Touristique 63
III.4.2. Cas d'Étude de Systèmes de Recommandation dans d'Autres Régions Touristiques 64
III.4.2.1. Cas d'Étude 1 : Recommandation dans la Région du Tyrol, Autriche 64
III.4.2.2. **Cas d'Étude 2 : Intégration de Systèmes de Recommandation à Séoul, Corée du Sud** 65
III.4.2.3. Cas d'Étude 3 : Systèmes de Recommandation dans les Îles Canaries, Espagne 65
III.5. Conclusion **66**
Chapitre IV : Implémentation de la solution : Un système de recommandation hybride pour la promotion du tourisme Marocain **69**

IV.1. Introduction 70
IV.2. L'architecture du système proposé 71
IV.3. Recueil des données (Contribution 1) 71
IV.3.1. Ensemble des données 71
IV.3.2. Profilage et formes de collecte d'informations 73
IV.3.2.1. Notion de Profil 73
IV.3.2.2. Formes de collecte d'informations 74
IV.4. Filtrage des données (Contribution 2) 76
IV.4.1. Construction du module basé sur le contenu 76
IV.4.1.1. Principe de fonctionnement 76
IV.4.1.2. Dataset et prétraitement 79
IV.4.1.3. Méthodes et modèles d'apprentissage 79
IV.4.1.4. Entraînement et résultats 83
IV.4.2. Construction du module collaboratif 85
IV.4.2.1. Principe de fonctionnement 85
IV.4.2.2. Dataset et prétraitement 90
IV.4.2.3. Méthodes et modèles d'apprentissage 90
IV.4.2.4. Entraînement et résultats 93
IV.4.3. Algorithme de filtrage hybride de l'approche proposée 94
IV.5. Interface du système proposé et visualisation des recommandations (Contribution 3) 95
IV.5.1. Interfaces du système proposé 96
IV.5.2. Visualisation des recommandations à base de profil utilisateur 99
IV.5.3. Dashboard administrateur 101
IV.6. Conclusion 102
Conclusion générale et perspectives 104
Bibliographie 107

Liste de figures

Fig. 1. Principe de recommandation basée sur le filtrage collaboratif 24
Fig. 2. Architecture de référence pour les systèmes de recommandation touristiques 41
Fig. 3. Organigramme du cadre conceptuel 43
Fig. 4. Profilage d'utilisateur 45
Fig. 5. Diagramme de cas d'utilisateur 47
Fig. 6. Diagramme de classe 48
Fig. 7. La version web et mobile de plateforme du patrimoine culturel et naturel de la région Daraâ-Tafilalet 49
Fig. 8. L'architecture du système proposé 71
Fig. 9. Les données Yelp 72
Fig. 10. Formulaire 1 de saisie de données utilisateurs 75
Fig. 11. Formulaire 2 de saisie de données utilisateurs 75
Fig. 12. Représentation vectorielle d'un corpus 77
Fig. 13. La fonction sigmoïde 80
Fig. 14. L'arbre de décision 80
Fig. 15. Principe de la méthode SVM 82
Fig. 16. Structure de ANN 82
Fig. 18. Principe de la méthode des k plus proches voisins 90
Fig. 19. Application de SVD 91
Fig. 20. Résultats de tests 94
Fig. 21. Page d'accueil 96
Fig. 23. Page blog 98
Fig. 24. Feed utilisateur 98
Fig. 25. Exemple du profil utilisateur 99
Fig. 26. Résultat de filtrage collaboratif à base de profile utilisateur 100
Fig. 27. Résultat de filtrage basé sur le contenu 100
Fig. 28. Planification de voyage 101
Fig. 28. Admin Dashboard 102

Liste des tableaux

Tableau 1. Matrice de confusion pour la classification des items 33
Tableau 2. Résultats de tests pour filtrage par contenue 84
Tableau 3. Résultats de tests pour filtrage collaboratif 93

Liste des abréviations

AI	Augmented Reality
FBC	Filtrage Basé Sur Le Contenu
FC	Filtrage Collaboratif
FI	Filtration de l'information
IA	Intelligence Artificielle
MAE	Mean Absolute Error
ML	Machine Learning
RI	Recherche d'information
SVM	Support Vector Machine
RMSE	Root Mean Square Error
SR	Système de Recommandation

Introduction générale

''Au Maroc, l'industrie touristique est un secteur économique important ; Le nombre d'arrivées de touristes au Royaume du Maroc est passé de 4,4 millions en 2001 à 9,3 millions en 2010, cependant, bon an mal an, les revenus touristiques représentent de 8% à 10 % du PIB du pays. Le tourisme représente une des forces majeures d'attractivité des territoires et, par conséquent, un puissant levier de leur développement économique. C'est également un secteur où les technologies de l'information et de la communication (TIC) jouent un rôle déterminant. ''

1. Contexte de recherche

Les TIC sont partout. Tout le monde, notamment les jeunes générations, est de plus en plus équipée d'ordinateurs, téléphones mobiles, smartphones, tablettes…, et avec les médias sociaux, la « vie en réseau » est devenue une réalité. Les usages « numériques » se multiplient, suscitent de nouvelles pratiques, de nouveaux besoins et de nouveaux modes de consommation, notamment des services touristiques et culturels. Consultation des informations touristiques en ligne, réservation et paiement de l'hébergement, places de concerts, festivals, spectacles via internet, échanges interactifs sur Facebook, forums ou blogs font désormais partie de la vie quotidienne. Pour une collectivité qui veut renforcer ou améliorer son offre touristique ou culturelle, le numérique devient un passage obligé !

Pour rester compétitifs, les acteurs du tourisme se doivent donc d'adopter une attitude proactive vis-à-vis du numérique. Ils doivent intéresser, satisfaire, pour donner l'envie aux touristes de promouvoir leur destination : bref créer du lien ! C'est l'un des avantages du numérique qui invite au voyage, rend la visite attractive, et simplifie la vie au touriste : un champ des possibles très large !

Aujourd'hui, ce sont près de huit touristes sur dix qui préparent leurs vacances sur Internet, et un sur trois qui reste derrière son écran pour acheter son séjour. Des fonctionnalités offertes par les sites web, les applications mobiles, permettent aussi d'accompagner le touriste durant

toutes les phases de son voyage, depuis la préparation jusqu'au partage en ligne, « avant/pendant/après ».

Ces pratiques génèrent d'immenses volumes de données exploitables pour le marketing des agences et des acteurs de tourisme pour séduire le consommateur et influencer ses décisions. Il s'agit d'un nouveau défi de taille pour toute destination touristique, qui doit être visible sur Internet et adaptée aux besoins et préférences des visiteurs.

D'où vient l'idée de mettre un système au profil du tourisme Marocain pour une gestion intelligente d'un territoire touristique, grâce à l'exploitation des données de masse obtenues depuis les sites web, les applications mobiles, la géolocalisation, les médias sociaux et les objets connectés. Dans ce cadre, nous proposons une solution basée sur l'analyse de données (TextMining), le profilage, les algorithmes de recommandations, les services Web et les ontologies, … pour promouvoir le tourisme Marocain, notamment celui de la destination Drâa-Tafilalet.

2. Objectifs de recherche :

Cette recherche a pour but de fournir un outil intelligent pour cibler et recommander l'offre touristique le plus adapté selon le profil des utilisateurs, de suivre et analyser leurs opinions afin d'améliorer l'expérience client et de visualiser les indicateurs clés sous forme de tableaux de bord pour prendre les décisions stratégiques adéquates.

Les objectifs de cette recherche peuvent alors être regroupés sous forme des briques fonctionnelles suivantes :

- Brique 1 : Collecter les données disponibles sur le Web et les réseaux sociaux en rapport avec le tourisme, les stocker dans un format compatible avec les technologies du Web sémantique. Ces données sont enrichies sémantiquement en utilisant les ontologies, qui fournissent des méthodes et des standards permettant un accès plus intelligent aux données.

- Brique 2 : Rechercher, analyser et sélectionner les données pertinentes pour mieux cibler des touristes potentiels. Ainsi, un système d'analyse et de recherche d'information sera élaboré permettant de récupérer juste les données comportant un certain mot-clé, concernant une personne (âge, nationalité, préférences personnelles, activités, …etc.). Ces

données seront utilisées pour déterminer le profil de l'utilisateur parmi les profils touristes que nous avons déjà définis (Aventuriers, jeunes, seniors, amateurs de conforts, touristes culturels, touristes d'affaires).

- Brique 3 : Proposer des parcours touristiques personnalisés et adaptées aux préférences des visiteurs en utilisant un système de recommandation sémantique et adaptatif. Une plateforme est en cours de développement, dans le cadre du projet permettant l'agrégation des données touristiques à partir de différentes sources. Cette plateforme regroupe des informations qui représentent les différents sites, objets touristiques et culturels de la région Drâa-Tafilalet (Musée, Bibliothèques, site touristiques et historiques, hôtels, …etc.).

- Brique 4 : Conserver et suivre les touristes, à travers leurs comptes Twitter ou Facebook, pour analyser leurs d'informations, notamment celles reliées à des objets touristiques (hôtels, restaurants, évènements…), données temporelles ou encore les informations sur leurs avis et sentiments exprimés. L'opinion publique est très importante dans le domaine du tourisme, puisqu'un individu peut influencer l'opinion d'un autre individu. Pour l'analyse et le traitement des données temporelles et d'opinions, nous pouvons utilisés et adaptés le concept d'opinion Mining et les normes SentiML et TimeML.

- Brique 5 : Elaborer périodiquement un tableau de bord touristique, pour les acteurs et les institutionnels du tourisme, permettant de visualiser un ensemble d'indicateurs et de statistiques par rapport à leur territoire dans l'objectif de prendre des décisions les plus efficaces.

3. Problématique scientifique et motivation

La problématique scientifique de ce projet, qui s'intéresse aux systèmes de recommandations pour les ressources touristiques, peut être exprimée de la manière suivante :

"Comment développer un système de recommandation hybride, intelligent et efficace qui facilite la découverte des ressources touristiques pertinentes et personnalisées pour améliorer l'expérience et la satisfaction des touristes dans la région de Drâa-Tafilalet ?

Cette problématique soulève plusieurs défis importants :

- Comment recueillir et analyser des données pertinentes (comme les préférences des touristes, leur historique de visites, leurs évaluations, les caractéristiques des activités) pour soutenir un système de recommandation ?
- Comment concevoir des algorithmes de recommandation qui peuvent exploiter efficacement ces données pour fournir des suggestions d'activités touristiques qui sont à la fois pertinentes et adaptées aux besoins individuels de chaque touriste ?
- Comment ces recommandations peuvent-elles être présentées de manière à encourager l'engagement des touristes et à optimiser leur expérience ?
- Comment évaluer l'efficacité de ces recommandations sur l'expérience touristique et la satisfaction des visiteurs ?
- Comment le système de recommandation peut-il s'adapter et évoluer avec le temps pour répondre aux besoins changeants des touristes et aux nouvelles offres touristiques ?

Ce sujet invite à une exploration en profondeur des possibilités offertes par l'exploration des données, et les technologies d'intelligence artificielle pour améliorer l'expérience de visiteurs en facilitant l'accès et l'utilisation des ressources touristiques.

4. Démarche méthodologique et organisation de la thèse

La méthodologie de ce projet s'articule autour de quatre étapes principales : premièrement, une revue de la littérature sur les systèmes de recommandation touristiques ; deuxièmement, l'élaboration d'un cadre conceptuel pour le développement des systèmes de recommandation touristiques hybrides et la description de l'architecture du système proposé ; troisièmement, la mise en œuvre de ce système en s'appuyant sur différentes techniques d'intelligence artificielle ; et enfin, l'évaluation du système développé, en donnant quelques propositions pour l'amélioration des résultats obtenus.

Dans un premier temps, une revue approfondie de la littérature situera ce travail dans le cadre des recherches actuelles sur les systèmes de

recommandation touristiques. Cela fournira les bases théoriques nécessaires et aidera à mettre en lumière les lacunes présentes dans ce domaine.

Ensuite, en s'appuyant sur une analyse des techniques d'intelligence artificielle utilisées dans les systèmes de recommandation touristiques, un cadre conceptuel pour le développement de ce type de système, en particulier ceux basés sur une approche hybride, sera présenté. L'architecture proposée sera conçue pour s'adapter aux besoins particuliers de chaque touriste.

En dernier lieu, une évaluation du système de recommandation sera effectuée pour en mesurer l'efficacité de l'approche proposée. Les indicateurs d'évaluation incluront la satisfaction des utilisateurs, la diversité et la qualité des recommandations proposées, ainsi que la capacité du système à s'adapter aux besoins changeants des touristes. Les résultats de cette évaluation serviront à identifier des pistes d'amélioration pour optimiser le système.

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

• **Introduction générale** - Etablissement du cadre pour notre recherche sur l'amélioration de l'accès et de la personnalisation des ressources touristiques à travers des outils intelligents, tels que les systèmes de recommandation.

• **Chapitre 1 : Systèmes de recommandation : Vue d'ensemble** - présente une vue générale sur les systèmes de recommandation. Nous définissons d'abord ce qu'est un système de recommandation. Ensuite nous détaillons ses différentes approches. Nous présentons enfin, les différentes techniques d'évaluation des systèmes de recommandation.

• **Chapitre 2 : Critique de l'existant et proposition d'une solution** - Consiste à présenter l'architecture générale des systèmes de recommandation touristiques, le cadre conceptuel de la solution proposé, tout en détaillant la démarche méthodologique d'intégration de big data et de techniques de l'intelligence artificielle dans l'architecture proposée.

• **Chapitre 3 : L'IA au Service des Voyageurs : Révolution des Systèmes de Recommandation Intelligents** -

• **Chapitre 4 : Implémentation de la solution : Un système de recommandation hybride pour la promotion du tourisme Marocain** - présente les détails d'implémentation et de mise en œuvre de notre système de recommandation touristique, ainsi qu'à la présentation des résultats obtenus. Ce travail comprendra le recueil des données, la conception et l'implémentation d'algorithmes de recommandation, et le développement d'une interface utilisateur conviviale.

Conclusion générale et perspectives – Fais un bilan des principales contributions, discute les aspects novateurs du système proposé et leur impact sur le secteur du tourisme marocain, puis examine les futures orientations de recherche dans ce domaine.

5. Contributions et productions scientifiques

Simultanément à notre recherche, nous avons accordé une grande importance à la diffusion de nos travaux à travers des communications et des articles publiés dans des revues renommées, tant au niveau national qu'international. Plusieurs documents ont été soumis, acceptés par des comités scientifiques et publiés dans de nombreuses revues internationales.

5.1. Publications dans des revues internationales (1er Auteur)

Intitulé de la publication	Date de la publication	Journal d'édition	Revue indexée	Impact facteur (SJR)
Hybrid recommender system for tourism based on big data and AI: A conceptual framework	12 Jan 2021	Big Data Mining and Analytics	SCOPUS et WOS	1.93
Tourism recommender systems: an overview	19 Jan 2022	International Journal of Cloud Computing	SCOPUS	0.17

5.2. Publications dans des revues internationales (Co-auteur)

Intitulé de la publication	Date de la publication	Journal d'édition	Revue indexée	Impact facteur (SJR)
An Approach based on Machine Learning Algorithms for the Recommendation of Scientific Cultural Heritage Objects	01-06-2020	International Journal of Advanced Computer Science and Applications (IJACSA)	SCOPUS et WOS	0.28

5.3. Publications sous forme de communication

Internationales Indexées		
Auteurs	**Nom et lieu de la manifestation**	**Titre de communication**
Khalid AL FARARNI, Badraddine AGHOUTANE, Jamal Riffi, Abdelouahed SABRI, Ali YAHYAOUY	The 1st international conference on Embedded Systems and Artificial Intelligence (ESAI'19), Fez, Morocco	Comparative Study on Approaches of Recommendation Systems
Khalid AL FARARNI, Badraddine AGHOUTANE, Loukmane MAADA, Fouad NAFIS, Jamal Riffi, Abdelouahed SABRI, Ali YAHYAOUY	The Sixth edition of the International Conference on Intelligent Systems and Computer Vision (ISCV 2024), Fez, Morocco.	Item Based Collaborative Filtering Tourism Recommender System Using Apache Mahout
Internationales Non indexés		
Khalid AL FARARNI, Badraddine AGHOUTANE, Ali YAHYAOUY	3ème édition de la Conférence Internationale Journées Scientifiques en Sciences Appliquées (JSSA'19), Larache, Morocco.	Un système de recommandation hybride et adaptative pour la promotion de la destination Maroc
Khalid AL FARARNI, Badraddine AGHOUTANE, Jamal Riffi, Abdelouahed SABRI, Ali YAHYAOUY	The First International Conference on scientific cultural heritage and advanced technologies: reality and horizons of the Daraa Tafilalet zone (SCHAT'19), Errachidia, Morocco.	Tourism Recommender Systems: A Survey

Chapitre I

Systèmes de recommandation : Vue d'ensemble

Chapitre I : Systèmes de recommandation : Vue d'ensemble 18
I.1. Introduction 19
I.2. Les systèmes de recommandation 19
I.2.1. Définition 19
I.2.2. Historique 20
I.3. Approches de recommandation 22
I.3.1. Recommandation basée sur le contenu 22
I.3.2. Recommandation basée sur le filtrage collaboratif 23
I.3.2.1. Recommandation basée sur le voisinage (la mémoire) 25
I.3.2.2. Recommandation basée sur le modèle 25
I.3.3. Autres approches de recommandation 26
I.3.3.1. Recommandation démographique 26
I.3.3.2. Recommandation contextuelle 27
I.3.4. Recommandation basée sur le filtrage hybride 27
I.3.4.1. Techniques d'hybridation 28
I.4. Évaluation Les systèmes de recommandation 30
I.4.1. Techniques d'évaluation 31
I.4.1.1. Test en ligne 31
I.4.1.2. Études utilisateurs 31
I.4.1.3. Simulation (Évaluation hors-ligne) 32
I.5. Conclusion 34

I.1. Introduction

À l'ère numérique, où les avancées technologiques transforment continuellement notre façon d'interagir avec l'information, les utilisateurs bénéficient d'un accès sans précédent à une multitude de données. Internet regorge de contenus diversifiés, allant des articles, et des vidéos, aux services et avis, ce qui devrait théoriquement faciliter la prise de décision. Cependant, dans ce contexte caractérisé par une surabondance d'informations, que l'on appelle infobésité ou déluge informationnel, il apparaît que les capacités humaines ne permettent pas l'analyse exhaustive de l'offre d'un corpus au sein d'une plateforme. Même dans le cadre de l'utilisation d'un moteur de recherche intégré, la recherche d'informations pertinentes peut s'avérer difficile pour les utilisateurs, car ils sont confrontés à un déluge d'informations. Pour aider l'esprit humain dans son processus de sélection, des systèmes de recommandation grand public ont vu le jour dans la dernière décennie du vingtième siècle.

Ce chapitre vise à examiner l'évolution et l'impact considérable de ces technologies avancées. Nous commencerons par analyser les systèmes de recommandation, en soulignant leur développement historique, les différentes approches et leur utilisation croissante dans divers domaines.

I.2. Les systèmes de recommandation

I.2.1. Définition

Selon une définition générale de Robin Burke [1,2], les systèmes de recommandations sont décrits comme : "Des systèmes capable de fournir des recommandations personnalisées permettant de guider l'utilisateur vers des ressources intéressantes et utiles au sein d'un espace de données important".

Au sens strict, ces systèmes sont définis comme des systèmes de filtrage de l'information visant à produire auprès d'un « client actif » (visiteur d'un site web, acheteur potentiel d'un produit, …) une liste personnalisée de suggestions (services, produits, liens à cliquer, …) en rapport avec ses préoccupations et attentes. Ces suggestions sont basées sur ce que l'utilisateur a acheté ou consulté précédemment mais aussi sur l'historique

d’activité d’autres acheteurs présentant un profil similaire.

C'est ainsi que l'on différencie les systèmes de recherche d'information, pour lesquels la demande de l'utilisateur de l'orienter et le guider vers des choix appropriés est explicite, et les systèmes de recommandation où la participation de l'utilisateur est non volontaire [3], induite notamment par des moteurs de traces [4].

I.2.2. Historique

L’historique des systèmes de recommandation remonte aux années 1990, bien qu’ils aient des racines plus anciennes dans le domaine de l’informatique et de l’intelligence artificielle. Voici un aperçu des principales étapes :

1. **Origines (années 1980) :** Les systèmes de recommandation sont reconnus assez tôt dans l’histoire de l’informatique ; « Information Lens System » [5] peut être considérée comme le premier système de recommandation. À l’époque, l’approche la plus commune pour le problème du partage d’informations dans l’environnement de messagerie électronique était la liste de distribution basée sur les groupes d’intérêt. La première définition pour le filtrage a été donnée aussi par Malone : " Même si le terme a une connotation littérale de laisser les choses dehors (filtrage négatif : enlèvement), nous l’utilisons ici dans un sens plus général qui consiste à sélectionner les choses à partir d’un ensemble plus large de possibilités (filtrage positif : sélection)".

2. **Début des systèmes de recommandation (années 1990) :**

    - **Collaborative Filtering :** Introduit par le système « Tapestry » [6], Il a été développé en 1992 par le centre de recherche de "Xerox" aux Etats Unis. Il s’agit d’un système de recommandation intégré à une application de mail électronique, permettant de recommander des listes de diffusion aux utilisateurs. Tapestry est fondé sur le Filtrage Collaboratif (FC) exploitant les annotations (les tags) des utilisateurs attribués aux listes de diffusion. L’analyse de ces annotations par le système permet de déterminer et de proposer les listes de diffusion qui sont pertinentes pour chaque utilisateur. Quelques années plus tard, un certain nombre de systèmes

académiques de recommandation ont vu le jour en 1994 et en 1995, tels que le système de recommandation d'articles d'actualités et de films développé par "GroupLens" [3] et le système de recommandation de musique "Ringo" proposé par [7].

- **Content-based Filtering :** Une approche qui recommandait des éléments similaires en se basant sur les caractéristiques des items eux-mêmes plutôt que sur les interactions des utilisateurs [8].

3. **Expansion (années 2000) :**
   - L'émergence de grandes plateformes comme Amazon et Netflix a popularisé les systèmes de recommandation. Netflix a commencé à utiliser des algorithmes avancés pour personnaliser les suggestions de films [9].
   - **Hybrid Models :** Les systèmes combinant filtrage collaboratif et filtrage basé sur le contenu sont devenus courants, permettant d'améliorer la précision des recommandations [1].
4. **Progrès technologiques (années 2010) :**
   - L'essor du Big Data et des techniques de machine learning, notamment les réseaux de neurones, a permis de développer des systèmes de recommandation plus sophistiqués.
   - Des entreprises comme Spotify et YouTube ont commencé à utiliser des algorithmes complexes pour analyser les comportements des utilisateurs et recommander des contenus adaptés [10].
5. **Développements récents (2020 et au-delà) :**
   - L'intégration de l'IA et du deep learning pour affiner encore plus les recommandations [11].
   - Une attention accrue à l'éthique et à la transparence dans les algorithmes de recommandation, notamment pour éviter les bulles de filtre et promouvoir la diversité des contenus.

Les systèmes de recommandation continuent d'évoluer, intégrant des aspects comme le traitement du langage naturel et l'analyse des sentiments pour affiner encore plus les suggestions. Ces systèmes sont devenus

indispensables dans divers secteurs pour aider à faire face à l'augmentation de la quantité d'informations et de produits disponibles et pour répondre à la demande croissante des utilisateurs. Dans le domaine touristique, les systèmes de recommandation pourraient être une aide précieuse lors de la préparation d'un voyage ou de la recherche d'un service parmi de nombreuses destinations, de nombreuses attractions et activités [12].

I.3. Approches de recommandation

Les systèmes de recommandation sont définis comme des systèmes de filtrage d'information permettant de recommander aux clients les offres les plus appropriées (produits, services, …) [13,14], par exemple des produits qui sont similaires à d'autres produits qu'ils ont déjà achetés et appréciés [15], ou des produits qu'ont déjà appréciés d'autres clients qui ont des goûts similaires [16].

Le principe est d'utiliser comme entrées les intérêts d'un utilisateur, colletées lors de sa navigation, pour prédire le degré d'intérêt que peut porter cet utilisateur pour un item donné [17]. Les approches employées pour estimer ces degrés d'appréciation sont nombreuses. Elles sont traditionnellement classées par la littérature en plusieurs catégories, en fonction de la manière dont ils collectent et analysent les informations des utilisateurs et filtrent la liste d'éléments [14,18].

I.3.1. Recommandation basée sur le contenu

Pour la proposition de recommandations aux utilisateurs, cette approche est fondée sur l'analyse des similarités de contenu entre les items précédemment consultés par les utilisateurs (ou examinent dans le présent) et ceux qui n'ont pas été encore consultés [1]. Le terme « contenu » fait référence aux attributs descriptifs (mots-clefs) des items qui sont utilisés pour formuler les recommandations.

Cette approche a ses racines dans la Recherche d'Information (IR) et l'Intelligence Artificielle [19]. En Recherche d'Information, il est considéré que les internautes voulant des recommandations sont engagés dans un processus de recherche d'information. L'utilisateur exprime un besoin ponctuel en donnant une requête (habituellement une liste de mots-clefs). Dans les systèmes de filtrage d'information (FI), le besoin est représenté

par le profil de l'utilisateur. En intelligence artificielle, la tâche de recommandation peut être exprimée comme un problème d'apprentissage qui exploite la connaissance passée des utilisateurs. De manière simple, les profils des utilisateurs sont sous la forme de vecteurs de mots-clefs et reflètent les intérêts à long terme de l'utilisateur. Souvent, il est préférable que le système apprenne le profil de l'utilisateur plutôt que d'imposer à celui-ci de le fournir. Cela implique généralement l'application de techniques de Machine Learning (ML). Leur but est d'apprendre à catégoriser de nouvelles informations en se basant sur les informations précédemment vues, et qui ont été libellées implicitement ou explicitement comme intéressant ou non par l'utilisateur. Avec ces libellés, les méthodes de Machine Learning sont capables de générer un modèle prédictif qui, étant donné un nouvel item, va aider à décider le degré d'intérêt que peut porter l'utilisateur pour l'item.

On distingue deux types de recommandation basée sur le contenu :

- **Recommandation basée sur les mots clefs :** Considère les caractéristiques des items que l'utilisateur a notés dans le passé afin de les corréler à son profil [2]. En effet, les profils des utilisateurs et des items sont représentés comme des vecteurs de mots-clefs pondérés. Ces mots-clés sont généralement soit extraits sur la base d'une indexation automatique, soit attribués manuellement [13]. Dans le but de recommander de nouveaux items intéressants, ces SR essayent de faire correspondre les attributs des items avec les préférences et les intérêts de l'utilisateur. Pour un nouvel item, le système effectue un calcul de similarité cosinus entre le vecteur de profil et le vecteur de l'item afin de prédire le score que pourrait porter l'utilisateur sur l'item.

- **Recommandation basée sur la sémantique :** La sémantique a été introduite par plusieurs méthodes dans le processus de recommandation [15]. Ces méthodes sont abordées en tenant compte de plusieurs critères :

  - Type de source de connaissance impliquée (lexique, ontologie…) ;

  - Techniques adoptées pour l'annotation ou la représentation d'items ;

  - Type de contenu inclus dans le profil utilisateur ;

  - Stratégie de correspondance entre items et profil.

I.3.2. Recommandation basée sur le filtrage collaboratif

Le filtrage collaboratif est une méthode de prédiction indépendante du domaine qui est utilisée pour des objets tels que la musique et les films qu'il est difficile, de décrire correctement à l'aide de métadonnées. Cette approche s'appuie sur les évaluations données par un ensemble d'utilisateurs sur un ensemble d'items, pour générer les prédictions ; Elle repose sur l'hypothèse que les utilisateurs qui ont des goûts antérieurs similaires sont susceptibles d'apprécier les mêmes items à l'avenir [9]. L'implémentation la plus simple et originale de cette approche est de recommander à un utilisateur donné les items hautement évalués dans le passé par d'autres utilisateurs qui présentent avec lui des similarités de préférences [20,21] ; On parle d'utilisateurs similaires. Pour identifier les utilisateurs similaires, le système de recommandation construit une base de données (matrice utilisateur x article) des préférences des utilisateurs pour les articles. Il compare ensuite les individus sur la base des similitudes dans leurs interactions (évaluations) passées afin de trouver des utilisateurs ayant des intérêts et des préférences similaires.

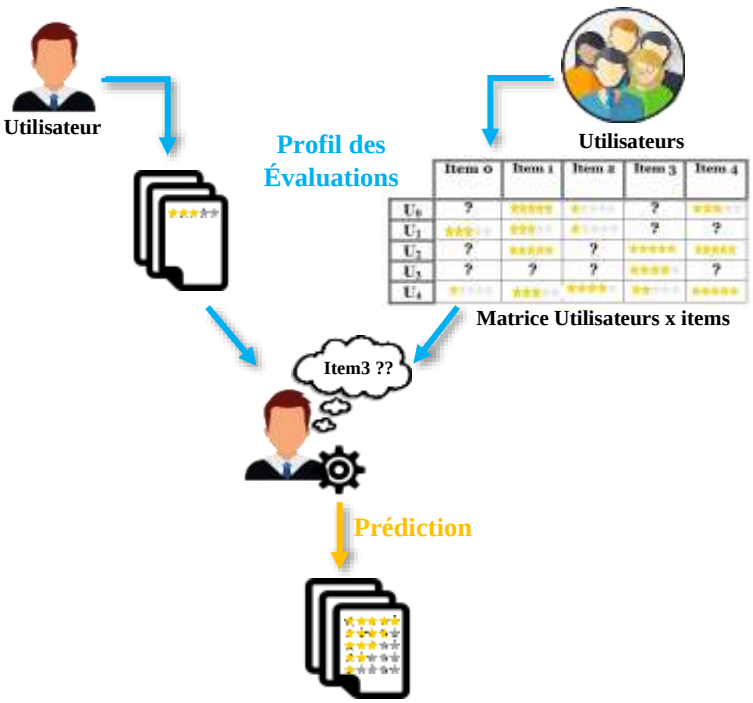


Fig. 1. Principe de recommandation basée sur le filtrage collaboratif

On distingue généralement deux sous-familles principales du filtrage collaboratif : les méthodes basées sur la mémoire (memory-based) et les

méthodes basées sur un modèle (model-based).

I.3.2.1. Recommandation basée sur le voisinage (la mémoire)

Les algorithmes de filtrage collaboratif basés sur la mémoire, appelés aussi basés sur les voisins [13] utilisent les notes des utilisateurs stockés en mémoire pour identifier des utilisateurs similaires et faire de la prédiction. L'idée clef est que la note d'un utilisateur u pour un nouvel item i est susceptible d'être similaire à celle donnée par un autre utilisateur v, si u et v ont noté d'autres items d'une manière similaire. De même, u est susceptible de noter deux items i et j de la même façon, si d'autres utilisateurs ont donné des notes similaires à ces deux items. On distingue alors deux grandes approches de filtrage collaboratif basé sur la mémoire :

- L'approche basée sur le voisinage des utilisateurs : utilise les notes d'autres utilisateurs pour évaluer l'intérêt d'un utilisateur pour un élément. Elle consiste à comparer les utilisateurs entre eux et à retrouver ceux ayant des goûts (notations) en commun, les notes d'un utilisateur u étant ensuite prédites selon ses voisins. Les plus proches voisins sont les utilisateurs les plus similaires à u dans leur notation.

- L'approche basée sur le voisinage des items : analyse les informations historiques afin de trouver des corrélations entre les articles fréquemment achetés et affiche des recommandations relatives aux articles corrélés. Il consiste à rapprocher les articles appréciés par les mêmes personnes et à prédire les notes des utilisateurs en fonction des articles les plus proches de ceux qu'ils ont déjà notés. Deux items sont considérés comme similaires si de nombreux utilisateurs du système les ont évalués de manière comparable.

I.3.2.2. Recommandation basée sur le modèle

Les méthodes basées sur un modèle ont été introduites dans les systèmes de recommandation pour résoudre certains des problèmes associés aux méthodes basées sur la mémoire. Ces problèmes comprennent notamment la mise à l'échelle des calculs et le manque de données [22,23]. Le passage à l'échelle pose un défi en termes de temps de calcul, en particulier lorsqu'il s'agit de manipuler de grandes quantités de données.

Le manque de données survient car il peut être difficile de trouver un nombre suffisant d'évaluations communes entre les utilisateurs, en particulier dans le cas de catalogues d'items volumineux. Sachant que le filtrage collaboratif utilise des mesures statistiques pour calculer la similarité, et que ces mesures demandent un nombre minimal de valeurs pour fonctionner, le FC est déstabilisé par le manque de données [24, 25].

Pour résoudre ces problèmes, les méthodes basées sur un modèle font appel à des techniques telles que la réduction de dimensionnalité et le clustering. Ces techniques permettent d'éliminer les utilisateurs ou les items non représentatifs, améliorant ainsi la précision des recommandations. Il est possible de ne conserver que les N voisins les plus similaires [3,26] ou ceux dont la similarité dépasse un certain seuil [7]. Cependant des études récentes montrent qu'engager les utilisateurs dissimilaires dans la prédiction peut aussi améliorer la performance [27,28]. Plusieurs méthodes ont été utilisées pour les algorithmes de recommandation basés sur un modèle. On peut citer, parmi les plus abouties, les méthodes de réduction de la dimension appelées SVD (décomposition en valeurs singulières) [13] les approches probabilistes [29], les approches basées sur le clustering [30] et les approches basées sur les règles d'association [31].

Les systèmes de recommandation basés sur des modèles offrent des capacités expressives hautes pour décrire les divers aspects des données. Ainsi, ils tendent à fournir des résultats plus précis que les systèmes basés sur le voisinage. Cependant, Le FC à base de modèle consomme normalement du temps pour construire le modèle et le mettre à jour. Il est aussi incapable de couvrir autant de diversité d'utilisateurs que la méthode à base de mémoire.

I.3.3. Autres approches de recommandation

I.3.3.1. Recommandation démographique

C'est une recommandation simple qui propose des items par rapport au profil démographique d'usager [32]. Elle comporte à catégoriser les usagers en plusieurs classes ou groupes par rapport aux informations

démographiques telles que le sexe, l'âge, la profession, la localisation, la langue, le pays, etc. Le principe de cette approche est que deux utilisateurs ayant évolué dans un environnement similaire partagent des gouts communs que deux utilisateurs ayant évolué dans des environnements différents et ne partageant donc pas les mêmes codes [33]. De nombreux sites utilisent cette solution pour proposer une offre de contenu "personnalisé". Par exemple, les utilisateurs sont redirigés vers un site Web particulier en fonction de leur langue ou de pays. Ces approches ont été très populaires dans la littérature de marketing, mais ont reçu peu d'attention dans le domaine des algorithmes de recommandation.

L'avantage principal d'une approche démographique est qu'il n'exige pas la connaissance de l'historique d'évaluations de l'utilisateur actif. Le problème principal de cette méthode est que le système recommande les mêmes articles aux gens avec des profils démographiques semblables, donc les recommandations sont trop générales (ou, au moins, pas très spécifiques pour un utilisateur donné [34]).

I.3.3.2. Recommandation contextuelle

Contrairement aux systèmes traditionnels qui se basent uniquement sur les préférences passées de l'utilisateur, ces systèmes tiennent compte de facteurs externes pour offrir des recommandations plus pertinentes et personnalisées. Les systèmes de recommandation contextuels sont conçus pour adapter leurs suggestions en fonction d'informations contextuelles spécifiques, comme l'emplacement, les conditions météorologiques et d'autres facteurs pertinents [35,36]. Selon [37], le contexte est défini comme « toute information pouvant être utilisée pour caractériser la situation d'une entité ». Cette entité peut être variée : elle peut représenter une personne, un lieu ou un objet jugé pertinent dans un contexte donné. L'inclusion de ces informations contextuelles permet d'affiner les recommandations, d'améliorer la personnalisation des suggestions et d'enrichir l'expérience des utilisateurs en tenant compte de leur environnement immédiat et de leurs besoins spécifiques.

I.3.4. Recommandation basée sur le filtrage hybride

La recommandation basée sur le filtrage hybride est une approche qui

combine plusieurs techniques de recommandation pour fournir des suggestions personnalisées aux utilisateurs. Cette méthode vise à surmonter les limites des approches de recommandation individuelles en exploitant les avantages de plusieurs méthodes simultanément. Elle vise également à améliorer la précision et la pertinence des suggestions en tirant parti des forces de chaque méthode tout en atténuant leurs faiblesses.

Selon [38], l'hybridation s'effectue en deux phases :

1. Effectuer de manière indépendante les filtrages des items via des méthodes collaboratives ou par le contenu (ou autres) pour générer des recommandations candidates (scores locaux) ;

2. Combiner ces ensembles de recommandations via des méthodes d'hybridation telles que la pondération, la mixtion, la cascade, la commutation, etc [1], afin de produire les recommandations finales pour les utilisateurs (scores finaux) [39].

I.3.4.1 Techniques d'hybridation

La différence entre les techniques d'hybridation réside dans la stratégie choisie pour combiner les scores locaux. L'hybridation peut être appliquée au niveau des notes, ou bien au niveau de la fonction de prédiction. Robin Burke résumait les techniques d'hybridation en sept techniques [2], dont :

- **Pondération (Weighted) :** Les notes calculées par les différentes techniques de recommandation sont combinées numériquement et pondérés afin de générer une seule recommandation. L'intérêt de cette méthode est que la combinaison est simple à réaliser et permet d'ajuster l'hybridation en fonction des performances. L'équation linéaire suivante peut être utilisée :

$$Score_{finale} = \alpha \times Score_1 + (1 - \alpha) \times Score_2 \ ; \quad 0 < \alpha < 1$$

Cette stratégie est flexible, il faut ajuster le paramètre pour quantifier la contribution de chaque approche. Un exemple pour la recommandation de films est disponible dans [40].

- **Commutation ou Technique d'alternance (Switching) :** C'est une technique qui permet de faire le choix d'un modèle de

recommandation parmi plusieurs, en se basant sur plusieurs critères. Le système se doit alors de définir les critères de commutation, ou les cas où l'utilisation d'une autre technique est recommandée. On peut considérer que le système tente de prédire la note par la première approche et utilise la deuxième en cas d'échec.

$$Score_{finale} = \begin{cases} Score_1, & si \quad Score_1 \neq null \\ Score_2, & si \quad Score_1 = null \end{cases}$$

La méthode "switching" introduit une complexité supplémentaire au processus de recommandation. En effet, le critère permettant le choix d'une technique doit être déterminé, ce qui requiert un autre niveau de paramétrage sur le système.

- **Technique mixte (Mixed) :** les recommandations issues de différentes techniques sont toutes présentées simultanément aux utilisateurs. Le problème qui peut ressortir quant à l'utilisation de cette méthode est la difficulté de calculer les scores pour ordonner une liste de recommandation, lorsque toutes les techniques recommandent les mêmes items mais avec des notes différentes. Pour pallier ce problème, on peut choisir par exemple une liste qui rassemble les cinq meilleurs items selon chaque technique, les agréger dans une seule liste, et présenter la liste à l'utilisateur [41], ou bien chaque approche calcule les scores de tous les items ; chaque item a le score le plus élevé, les items sont ensuite listés selon leurs scores, et les meilleurs sont recommandés.

$$Score_{finale} = \max(Score_1, Score_2)$$

- **Combinaison de caractéristiques (Features combination) :** Dans un hybride basé sur la combinaison de caractéristiques, les données provenant de techniques collaboratives sont traitées comme une caractéristique, et une approche basée sur le contenu est utilisé sur ces données.

- **Cascade :** La cascade implique un processus étape par étape. Dans ce cas, une technique de recommandation est appliquée en premier, produisant un ensemble de candidats potentiels. Puis, une deuxième technique raffine les résultats obtenus dans la première étape. Cette

méthode a pour avantage que si la première technique génère peu de recommandations, ou si ces recommandations sont ordonnées afin de permettre une sélection rapide, la deuxième technique ne sera plus utilisée.

Un exemple est donné dans [1]. Cette technique exige que l'item ait un score dans toutes les approches :

$$Score_{finale} = \begin{cases} Score_2, & si \ \ Score_1 > 0 \\ 0\,, & si\text{non} \end{cases}$$

- **Augmentation de caractéristiques (Features augmentation) :** est semblable à la cascade, mais dans ce cas-là les résultats obtenus (le classement ou la classification) de la première technique sont utilisées par le deuxième comme une caractéristique ajoutée.

Bien que les deux techniques, en cascade et par augmentation de caractéristiques, enchaînent deux recommandeurs, avec le premier possède une influence sur le second, elles sont fondamentalement très différentes. Dans un hybride par augmentation, les caractéristiques utilisées par le second recommandeur incluent la sortie du premier, telles que les évaluations par exemple. Dans un hybride en cascade, la deuxième technique n'utilise pas la sortie de la première technique dans la production de ses classements, mais les résultats des deux recommandeurs sont combinés de manière prioritaire.

- **Méta niveau (Meta-level) :** Dans un hybride basé sur méta niveau, une première **technique** est utilisée, mais différemment que la précédente méthode (augmentation de caractéristiques), non pas pour produire de nouvelles caractéristiques, mais pour produire un modèle. Et dans la deuxième étape, c'est le modèle entier qui servira d'entrée pour la deuxième technique [32].

I.4. Evaluation de systèmes de recommandation

Évaluer un système de recommandation permet de mesurer ses performances vis-à-vis de ses objectifs. De ce fait, les dimensions à évaluer diffèrent selon les objectifs fixés (prédiction des notes des utilisateurs pour les items, amélioration de qualité des recommandations, augmentation des

ventes, etc.).

La question de l'évaluation des systèmes de recommandation a été longuement étudiée dans le travail de [42]. Ce dernier répertorie les différents procédés permettant d'évaluer un système, ainsi que les dimensions pouvant être considérées durant l'évaluation.

I.4.1. Techniques d'évaluation

Trois techniques d'évaluation sont proposées par [43] : le test en ligne, études sur un échantillon d'utilisateurs et les simulations. Chacune présente un certain nombre d'avantages et d'inconvénients que nous allons étudier ci-dessous.

I.4.1.1. Test en ligne

Le test en ligne constitue une évaluation en temps réel de l'application sur de vrais utilisateurs du système [44]. Il peut se traduire par une simple comparaison des performances commerciales avant et après l'intégration du système de recommandation. Cette évaluation est menée sur un échantillon aléatoire d'utilisateurs, leurs réactions étant observées et comparées à celles du reste de la population.

L'avantage majeur de ce type d'évaluation réside dans l'obtention de résultats concrets, étant donné que les utilisateurs sont les utilisateurs finaux qui ne sont pas mis au courant de l'exécution du test. Néanmoins, ce type d'évaluation présente des inconvénients potentiels, notamment la possibilité de perdre des utilisateurs si le système recommande des items non pertinents. En conséquence, il est généralement recommandé de recourir à ce type d'évaluation après avoir effectué des tests hors ligne suffisants pour garantir la qualité des recommandations.

I.4.1.2. Études utilisateurs

Les études utilisateurs sont des tests qui consistent à recruter un groupe de volontaires, à leur demander d'utiliser un système de recommandation pour effectuer des tâches bien précises, et à observer leur comportement. Les participants peuvent également être interrogés sur leurs impressions sur l'expérience.

Ces études sont utiles pour comprendre comment les utilisateurs interagissent avec un système de recommandation et comment ce système influence leur comportement. Elles peuvent fournir des informations quantitatives et qualitatives sur l'efficacité et l'utilité du système.

Cependant, les études utilisateurs sont coûteuses et laborieuses. Il

peut être difficile de recruter un nombre suffisant de participants, et les tests doivent être répétés plusieurs fois pour assurer la fiabilité des résultats [42].

I.4.1.3. Simulation (Évaluation hors-ligne)

Simuler le comportement des utilisateurs présente une alternative rapide et peu coûteuse aux deux types d'évaluation cités précédemment. Cette méthode se base généralement sur une technique statistique appelée « validation croisée » (cross-validation, en anglais).

Essentiellement, la validation croisée requiert la disposition d'un jeu de données pouvant servir à reproduire le comportement des utilisateurs. Ce jeu de données contiendra par exemple les notes attribuées par des utilisateurs pour certains items. Ensuite, la validation croisée consiste à séparer les données disponibles en sous-ensembles. La première partie sert à faire la prédiction et la deuxième, à en valider l'algorithme.

La mesure d'évaluation doit s'accorder avec la stratégie suivie par le système. Deux classes de mesures d'évaluation sont utilisées : la classe des mesures de la qualité des prédictions, et la classe des mesures de qualité de recommandations [45].

a. Mesures de qualité des prédictions

Le but de ces mesures est de tester la ressemblance entre les valeurs prédites et les évaluations réelles faites par les utilisateurs. Nous citons dans cette section les deux mesures célèbres d'évaluation des prédictions.

- **Erreur Moyenne Absolue (MAE : Mean Absolute Error)**

La mesure MAE est une métrique couramment utilisée dans les systèmes de recommandation [7]. Elle évalue la précision des prédictions en calculant la moyenne des écarts absolus entre les valeurs prédites et les valeurs réelles. Formellement, elle se présente comme suit :

$$MAE = \frac{\sum_{(u,i)\in T} |r_{u,i} - \widehat{r_{u,i}}|}{|T|}$$

Où, $|T|$ est le nombre des évaluations utilisées dans le test, $r_{u,i}$ est l'évaluation réelle faite par l'utilisateur u à l'item i, et $\widehat{r_{u,i}}$ est la note prédite correspondante. Cette métrique permet d'évaluer la capacité globale du système de recommandation à prédire avec précision les notes attribuées par les utilisateurs.

- **La racine de l'erreur Quadratique Moyenne (RMSE : Root Mean Squared Error)**

La mesure RMSE est définie comme la racine carrée de la moyenne de l'erreur quadratique :

$$RMSE = \sqrt{\frac{\sum_{(u,i)\in T} (r_{u,i} - \widehat{r_{u,i}})^2}{|T|}}$$

Cette mesure est plus sensible à l'erreur parce que la différence est mise au carré avant qu'elle soit sommée.

b. Mesures de qualité des recommandations :

Ces mesures ne prennent pas en compte la valeur que le SR prédit, mais elles considèrent sa décision de recommander ou non un item. Le SR est récompensé pour les bonnes décisions (l'intégration des items pertinents dans sa liste de recommandations), et pénalisé pour les mauvaises (l'intégration des items non pertinents dans la liste, ou l'absence d'un item pertinent de la liste).

L'objectif est de mesurer la fréquence des bons et mauvais jugements portés par le système de recommandation à l'égard des items, et pour ce faire, la matrice de confusion (Tableau 1) permet de calculer diverses métriques d'évaluation, telles que :

		Prédiction	
		L'item est sélectionné	**L'item n'est pas sélectionné (raté)**
Réalité	**Pertinent**	Nb_{ps}	Nb_{pr}
	Impertinent	Nb_{imps}	Nb_{impr}

Tableau 1. Matrice de confusion pour la classification des items

- **<u>Précision (Precision) :</u>**

Elle est définie par le rapport entre le nombre d'items pertinents sélectionnés et le nombre total des items sélectionnés. La précision représente la possibilité qu'un item sélectionné soit pertinent. Formellement :

$$Précision = {Nb_{ps}}/{Nb_s}$$

Où $Nb_s = Nb_{ps} + Nb_{imps}$ est le nombre total des items sélectionnés.

- **<u>Rappel (Recall) :</u>**

C'est le rapport entre le nombre d'items pertinents sélectionnés et le nombre total des items pertinents disponibles. Ceci, représente la probabilité qu'un item pertinent soit sélectionné. Formellement :

$$Rappel = {Nb_{ps}}/{Nb_p}$$

Où $Nb_p = Nb_{ps} + Nb_{pr}$ est le nombre total des items pertinents.

Ces deux mesures (Précision et Rappel) nécessitent des compromis. En effet, tandis que la recommandation d'une longue liste d'items améliore le rappel, la précision peut se dégrader rapidement. C'est pourquoi une autre mesure, qui est la moyenne harmonique des deux dernières, est plus utile. Cette nouvelle mesure, appelée la F-Mesure [45], fait un équilibre entre la précision et le rappel.

- **F1-Score :**

La F-mesure est un compromis entre le rappel et la précision :

$$F1 = \frac{2 * Précision * Rappel}{Précision + Rappel}$$

- **La mesure d'évaluation de la Diversité :**

La diversité se rapporte à la diversité d'items dans une liste de recommandation. En d'autres termes, les items de la liste doivent être différents. Par exemple, il serait plus intéressant pour un site de voyage de proposer à l'utilisateur des destinations différentes plutôt qu'une même destination avec des commodités différentes.

La diversité d'une liste peut par exemple être mesurée à l'aide d'une mesure de similarité appliquée à chaque paire de ladite liste. La moyenne, le minimum ou le maximum de ces degrés de similarité pourront être utilisés pour définir la diversité.

I.5. Conclusion

Dans ce chapitre, nous avons présenté un état de l'art relatif aux techniques de recommandation, leur hybridation ainsi que les méthodes d'évaluation associées. Le filtrage collaboratif se propose de recommander aux utilisateurs certains articles qu'ils n'ont pas encore lus en se basant sur les opinions du groupe d'utilisateurs similaires. En revanche, le filtrage basé sur le contenu évalue si un texte est pertinent ou non pour un utilisateur, en fonction de son profil thématique. Néanmoins, ces deux aspects du profil semblent importants. Ainsi, une approche hybride combinant les goûts thématiques personnels et les opinions des utilisateurs du groupe paraissent la meilleure méthode pour fournir des recommandations. II est à noter que ces techniques de filtrage ne prennent pas en compte un bon nombre d'informations utiles pour expliquer le jugement qu'effectue l'utilisateur. Ceci requiert la prise en compte de différentes sources d'informations sur l'utilisateur : ses données démographiques (Filtrage démographique), son contexte actuel (Filtrage contextuel), ... et leurs corrélations avec le contenu des ressources qui lui sont proposés.

Chapitre II

Critique de l'existant et proposition d'une solution

Chapitre II : Critique de l'existant et proposition d'une solution 35
II.1. Introduction 36
II.2. Travaux connexes : Les systèmes de recommandation touristiques 36
II.2.1. Basés sur le contenu 36
II.2.2. Basés sur le filtrage collaboratif 38
II.2.3. Sensibles au contexte 39
II.2.4. Basés sur le filtrage hybride 39
II.3. Approche proposée 40
II.3.1. Description du problème et solution proposée 40
II.3.2. Le cadre conceptuel de l'architecture proposée 42
II.3.3. Diagrammes de système 46
II.3.4. Méthodologie d'intégration de big data et IA dans l'architecture proposée. …… 48
II.4. Conclusion 51

II.5. Introduction

Dans le domaine du tourisme, les systèmes de recommandation visent à faire correspondre les caractéristiques des ressources ou des attractions touristiques aux besoins des utilisateurs.

Comme il existe une multitude d'approches de recommandation, et due à la variété de leurs sources d'information, et la nature hétérogène des données touristiques, la mise en œuvre de ce type de systèmes, notamment ceux qui utilisent une approche hybride, rencontre plusieurs difficultés. L'objectif principal de ce chapitre est alors de contribuer à la conception des systèmes de recommandation touristiques en proposant un cadre qui clarifie le déroulement du processus de recommandation hybride et détaille chacune de ces étapes. La mise en œuvre du cadre proposé en plus de servir à rassembler les différentes approches de recommandation, va apporter une meilleure expérience au visiteur, en lui recommandant les items les plus pertinents et en l'aidant à personnaliser son parcours. Ensuite, nous donnons une brève description de notre système Big Data qui vise à intégrer notre système de recommandation et l'analyse des opinions en utilisant des techniques de Deep Learning [46].

Le présent chapitre s'organise de la façon suivante : Dans Sect. 2 nous présenterons brièvement une revue de la littérature des approches de recommandation les plus utilisées dans le domaine du tourisme. La section 3 présente la description de l'architecture proposée pour les systèmes de recommandation touristiques, ainsi que son cadre conceptuel et nous concluons dans la section 4, sur les perspectives de notre travail.

II.2. Travaux connexes : Les systèmes de recommandation touristiques

Les techniques de recommandation touristiques actuelles comportent plusieurs aspects novateurs et peuvent être classées de différentes manières, en fonction de la manière dont ils analysent les informations de l'utilisateur et filtrent la liste d'éléments [47] :

II.2.1. Basés sur le contenu

Pour la proposition de recommandations aux visiteurs potentiels, les

systèmes basés sur le contenu sont fondés sur l'analyse des similarités de contenu entre les items précédemment consultés par les utilisateurs (ou examinent dans le présent) et ceux qui n'ont pas été encore consultés [48,49].

Le filtrage basé sur le contenu est la technique la plus populaire et la plus répandue dans les systèmes de recommandation touristiques [50-56].

Bitonto et al., 2011 [57], par exemple, décrivent un système de recommandation basé sur le contenu pour le patrimoine culturel, qui est capable de suggérer à la fois des éléments matériels et immatériels qui ne se limitent pas à des événements mais incluent des légendes, des traditions culinaires, etc. En outre, pour chacun d'eux, les informations de service sont enrichies par des informations détaillées suggérées en fonction des intérêts de l'utilisateur cible. De plus, le système est en mesure de suggérer des éléments non seulement à un touriste général, mais également aux enseignants souhaitant organiser une visite éducative pour leurs élèves. Dans ce cas, les informations détaillées pourraient être utilisés dans des activités éducatives avant le voyage.

Une limite naturelle du filtrage basé sur le contenu est la nécessité de disposer d'une représentation générique et riche du contenu des items, ce qui n'est pas le cas des articles touristiques caractérisés par leur grande étendue et variété. On plus de ça, ce type de systèmes souffre généralement du problème de surspécialisation (Overspecialization) ; Par exemple, lorsqu'un touriste apprécie un événement ou un spectacle lors d'un voyage, cela ne veut pas dire qu'il voudra le revoir. Toutefois, le système lui suggérera, lorsqu'il utilisera une méthode basée sur le contenu, de revenir une seconde fois au même endroit avec le même type d'événement (Même s'il ne sera pas organisé !!), alors qu'il s'intéressera peut-être davantage aux objets qu'il n'a pas découverts lors du dernier voyage. Ainsi, ce type de systèmes souffre généralement du problème du démarrage à froid pour les nouveaux utilisateurs (pas d'historique de réservation ou de consultation). Cependant, quand de nouveaux items sont introduits dans le système, ils peuvent être recommandés directement, sans que cela ne nécessite un temps d'intégration comme c'est le cas pour les systèmes de recommandation basés sur une approche de filtrage collaboratif.

II.2.2. Basés sur le filtrage collaboratif

L'objectif de cette approche est de proposer aux visiteurs des destinations/activités dans lesquelles ils ne sont pas encore allés, mais qui pourraient leur plaire, en se basant sur les habitudes et les goûts de profils des utilisateurs similaires [58] ; La similarité de goût entre deux utilisateurs est calculée en se basant sur la similarité de leur historique de notation.

Une variété de systèmes collaboratifs existe dans le domaine de tourisme ; [59-63]

Le système VISIT [64], par exemple, applique des techniques d'analyse des sentiments (à l'aide de l'API Alchemy) pour analyser les actualités concernant une attraction donnée sur Twitter et Facebook et déterminer si les utilisateurs expriment des commentaires positifs ou négatifs à son sujet. Cette information est affichée en vert et en rouge par le système dans son interface, de sorte que l'utilisateur puisse facilement identifier les endroits que les visiteurs apprécient le plus aujourd'hui et qui ne le sont pas.

Les approches collaboratives dépassent certaines limitations des approches basées sur le contenu. Par exemple, il n'y a aucune analyse du sujet ou du contenu des items à recommander. Ce type de système est très efficace au cas où le contenu des items sont complexes ou difficilement définissable (ex. festivales, musées), il est compliqué ou impossible de l'analyser. L'utilisateur peut apercevoir divers domaines intéressants, car le principe du filtrage collaboratif ne se fonde absolument pas sur la dimension thématique des profils, et n'est pas soumis à l'effet « entonnoir ».

Néanmoins, ces systèmes présentent aussi quelques inconvénients :

- Recommandations limitées : Lorsque des items n'ont pas encore été notés ou libellés par des utilisateurs, les systèmes collaboratifs ne peuvent pas recommander ces items, car ils se basent sur ces informations pour les recommandations. Ainsi, ils ne peuvent pas trouver de voisins à un nouvel utilisateur qui n'a pas encore attribué de notes, ou a attribué peu de notes. Cela engendre des problèmes de démarrage à froid du côté des utilisateurs et du côté des items.

- Manque de données : Quand le catalogue d'items est grand, même un utilisateur hyper actif n'arrive pas à noter 1% du catalogue, et la matrice des notes sera quasi-vide. Il sera donc difficile de faire correspondre l'historique de voyages des utilisateurs et de trouver des notations communes entre les utilisateurs, il est presque impossible de rencontrer deux personnes qui vivent le même voyage, avec la même durée, les mêmes lieux d'intérêt et la même expérience.

II.2.3. Sensibles au contexte

Les systèmes de recommandation sont appelés sensibles au contexte lorsqu'ils utilisent le contexte dans son calcul pour prédire les éléments susceptibles d'intéresser l'utilisateur [37,65]. Parmi les éléments de contexte les plus utilisées dans les systèmes de recommandation touristiques, on trouve la géolocalisation, la météo, l'historique des visites, et le temps.

De nos jours, de nombreux appareils mobiles connectés à Internet, appelés aussi « Objets connectées », sont largement disponibles et utilisés afin de capturer et de fournir de nombreuses informations susceptibles d'enrichir le contexte actuel ainsi que ses variations.

Reference [66] propose un système de recommandation sensible au contexte et présentent une définition du concept de contexte à travers un méta-modèle. Une étude de cas appliquée a été menée à la ville de Tanger. Le système développé est composé de trois modules principaux: Le contexte, composé du profil de l'utilisateur, d'informations spatio-temporelles et environnementales (c'est-à-dire l'emplacement, les caractéristiques de l'appareil utilisée pour accéder à l'application, l'environnement physique externe, par exemple les conditions météorologiques) et des informations collectées sur la communauté (à partir de Facebook, Tweeter et d'autres applications de réseaux sociaux)); Le référentiel de contenu touristique, qui contient les données de services touristiques; et le système de recommandation.

II.2.4. Basés le filtrage hybride

L'hybridation de ces approches, afin de traiter les insuffisances de chaque technique utilisée seule et profiter de leurs points forts, a fait l'objet

de plusieurs travaux de recherche [67-71].

[72] propose une nouvelle architecture de système de recommandation pour les individus et les groupes, basée sur les caractéristiques des œuvres d'art, le contexte d'ensemble d'utilisateurs et l'affinité sociale entre ces utilisateurs. Le système est une hybridation de trois approches : l'approche basée sur le contenu, l'approche sociale et l'approche basée sur le contexte.

Même que l'hybridation rajoute encore plus de complexité au processus de recommandation [68]. En effet, elle requiert différentes sources de données et met en application plusieurs techniques à la fois. Elle nécessite ainsi des paramétrages supplémentaires liés à la combinaison de différentes méthodes. Les différents travaux qui s'intéressent à l'hybridation de techniques de recommandation ont démontré empiriquement que cette hybridation permet d'améliorer la précision des recommandations comparée par exemple au FC basé sur la mémoire ou à la technique basée sur le contenu [73-75]. Les systèmes hybrides permettent en outre de remédier à certains problèmes tels que le manque de données.

II.3. Approche proposée

II.3.1. Description du problème et solution proposée

Il est à noter que 90% des solutions actuelles sont généralement concentrés sur une seule catégorie d'articles (Hôtels, musées, sites touristiques…) [47], en fournissant juste des informations de services touristiques (insérées dans le système par l'administrateur ou par des experts) destinés à rendre le séjour plus agréable ; En addition, la plupart de ces travaux utilisent une approche unique, avec une nette prédominance pour les approches basées sur le contenu [47,76].

Pour toutes ces raisons, il est nécessaire de disposer d'un cadre conceptuel non seulement pour rassembler les approches de recommandation, mais aussi pour présenter les différentes ressources touristiques dans une seule architecture.

Rappelons que notre travail de recherche consiste à proposer une nouvelle architecture pour les systèmes de recommandation touristiques.

Cette architecture va se baser sur une approche de recommandation hybride, dont l'objectif est d'améliorer l'accès des utilisateurs aux ressources touristiques au niveau des systèmes de recherche d'information, tels que les portails touristiques et les Extranets documentaires des fournisseurs de services.

Autre aspect novateur de cette architecture, est que le système proposé va au-delà d'une liste d'attractions touristiques recommandées et peut être vu comme un planificateur qui a pour but la construction d'un programme complexe et détaillé d'une visite de plusieurs jours. Le client se verra ainsi proposer une liste diversifiée des ressources touristiques (monuments, activités, hôtels et spectacles…) qui répondent exactement à leur besoin spécifique et à leur préférence.

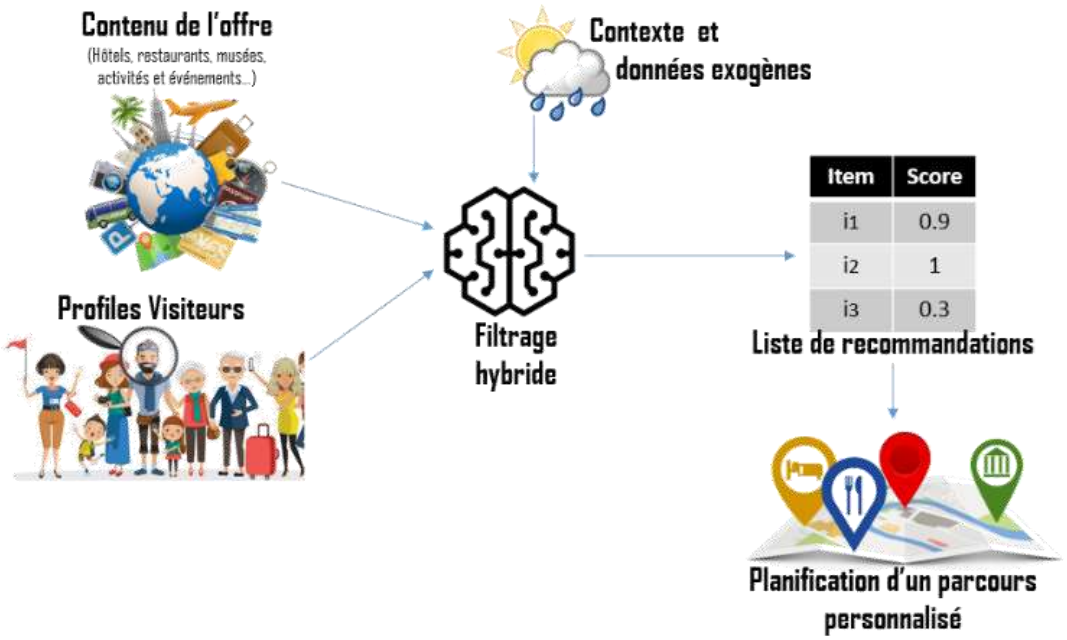


Fig. 2. Architecture de référence pour les systèmes de recommandation touristiques.

Nous proposons de décomposer l'architecture de système de recommandation touristique en cinq modules principaux (Figure 2) :

1. Les profils des visiteurs : contiendront notamment des informations pouvant être exploitées afin de déterminer les préférences des utilisateurs en termes d'items (données d'appréciations, informations sociales…).

2. Contenu de l'offre : contient les informations de services touristiques (tels que l'hébergement, la restauration, les sites

touristiques, les transports…), ainsi que le contenu multimédia associées.

3. Un méta-modèle contextuel : prend en compte de multiples facteurs impliqués dans la manipulation du contexte tels que le temps, l'espace, l'emplacement, la distance entre deux sites, les routes, l'historique des voyages touristiques, etc. afin de formuler une recommandation précise.

4. Le processus de filtrage hybride : renvoie une liste d'items avec les degrés d'appréciations que peut porter l'utilisateur cible à chaque item.

5. Un planificateur d'itinéraires : fait la sélection des items considérés pertinent pour l'utilisateur, et fait appel à des techniques de recherche opérationnelle pour corréler ces propositions sous forme d'un séjour.

II.3.2. Le cadre conceptuel de l'architecture proposée

Le cadre conceptuel de l'architecture proposée est constitué de trois sous-processus principaux, un processus dit de profilage des utilisateurs, un processus de sélection du contenu (filtrage) qui correspond le mieux aux profils des utilisateurs et un processus de planification d'itinéraire. Ces processus prennent place à l'intersection de différents domaines de recherche en informatique, notamment l'Intelligence Artificielle et la Recherche Opérationnelle.

Par exemple, en intelligence artificielle, le processus de profilage peut être exprimé comme un problème d'apprentissage qui exploite la connaissance passée des utilisateurs. Souvent, il est préférable que le système apprenne le profil de l'utilisateur plutôt que d'imposer à celui-ci de le fournir. Cela implique généralement l'application de techniques de Machine Learning (ML).

Le but du processus de filtrage est d'apprendre à catégoriser de nouvelles informations en se basant sur les informations précédemment vues, et qui ont été libellées implicitement ou explicitement comme intéressant ou non par l'utilisateur. Avec ces libellés, les méthodes de Machine Learning sont capables de générer un modèle prédictif qui, étant

donné un nouvel item, va aider à décider le degré d'intérêt que peut porter l'utilisateur pour l'item.

En recherche opérationnelle, le processus de planification d'itinéraire amène à la définition formelle d'un problème d'optimisation combinatoire qui est une variante d'un sous-problème du problème du voyageur de commerce. Pour résoudre ce genre de problème, nous pouvons compter sur les métaheuristiques afin de tendre vers une bonne solution en un temps raisonnable.

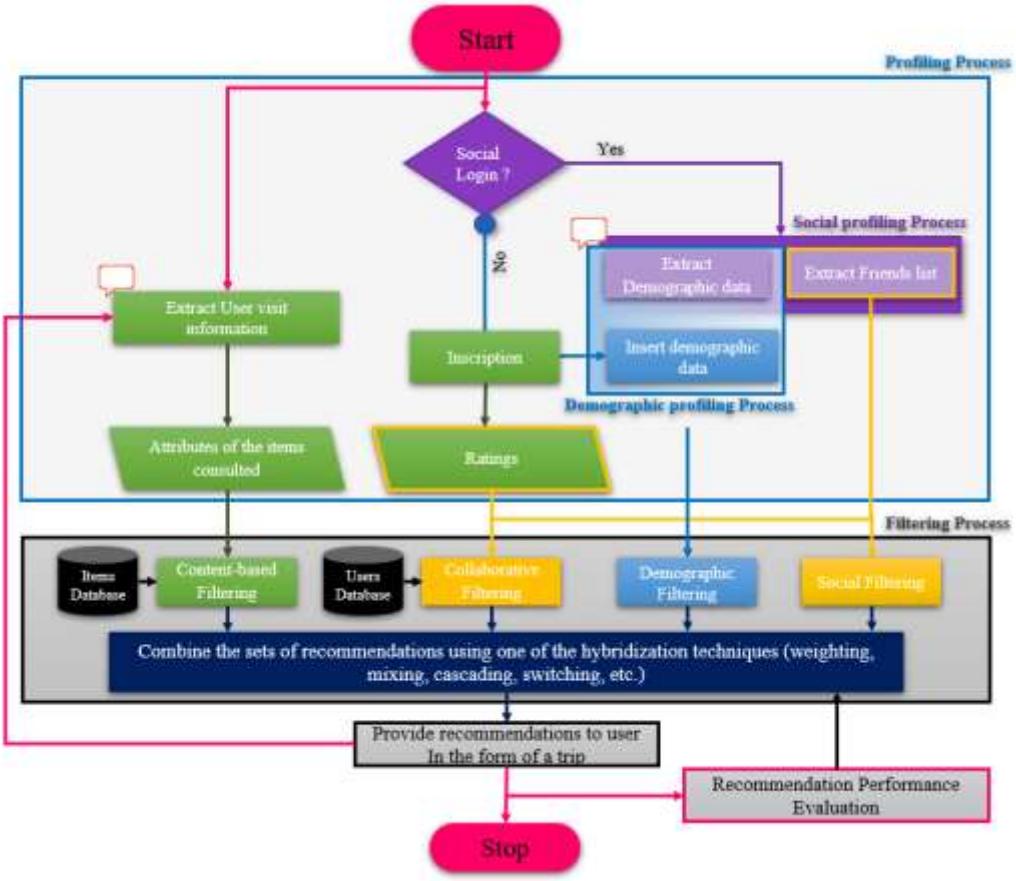


Fig. 3. Organigramme du cadre conceptuel

Le cadre conceptuel de l'architecture proposée (Figure 3) se compose alors de trois processus principaux à savoir un processus de profilage, un processus de filtrage, et le processus de planification d'itinéraire.

A. Processus de Profilage des utilisateurs :

La Collecte des données sur les utilisateurs ou le Profilage des utilisateurs est une étape très importante dans l'approche que nous présentons. Ce processus peut comporter quatre scénarios, à partir lesquelles les modules constituant le profil utilisateur peuvent être

extraits :

a. Inscription : l'utilisateur indique explicitement au système ses intérêts à travers un formulaire d'inscription ; en attribuant par exemple :

   • Des zones de commentaires, mots clés ou tags à sélectionner par l'utilisateur pour indiquer ces préférences : But de voyage, type d'items a consulté …

   • Des notes aux items consultés indiquant le degré d'appréciation d'un item par cet utilisateur.

   • Des attributs démographiques concernant l'utilisateur, tels que : l'âge, le sexe, la catégorie socio-professionnelle, la localité géographique, le statut personnel, etc. Certes, ces attributs ne fournissent pas d'informations sur les appréciations, mais ils permettent notamment d'affiner le profil utilisateur afin d'y adapter les recommandations. Par ailleurs, les données démographiques peuvent être utilisés pour calculer les recommandations lorsqu'il s'agit de nouveaux utilisateurs ; L'approche démographique est tout d'abord utilisée afin de résoudre le problème du démarrage à froid [77,78].

b. Le social login : Au lieu de créer un nouveau compte de connexion spécifiquement pour le système, l'utilisateur utilise ses informations de connexion existantes dans un réseau social tels que Facebook, Twitter ou Google+ pour se connecter [78-80]. Cette connexion nous permettre ensuite d'extraire les données démographiques ainsi que quelques informations concernant les relations de l'usager avec ces amis.

c. Consultation « même sans inscription » : Elle repose sur une observation et une analyse des comportements de l'utilisateur effectué de façon implicite dans l'application qui embarque le système de recommandation, le tout se fait en "arrière- plan" (en gros sans rien demander à l'utilisateur). Nous appellerons ces comportements "les traces d'usage". Ces traces peuvent inclure [81] :

   • Des indicateurs décrivant la manipulation tels que : des

"copier/coller" d'un texte à partir d'une page, la recherche d'un texte dans une page, l'ajout ou la suppression d'un item du panier ou la commande d'un item (dans le cadre des applications e-commerce), la sauvegarde ou l'impression d'une page, etc.

- Des indicateurs de navigation tels que : la fréquence et la durée de consultation, le nombre de clics et de survols de souris sur une page et sur des liens, le "scrolling", etc [76].

d. Le Contexte : repose sur l'intégration des informations contextuelles (localisation, temps, environnement physique…) pour la génération de parcours de visite dynamiques et personnalisés.

Les données récoltées sur l'utilisateur sont ensuite sélectionnées, analysées et sauvegardées sous forme des modules indépendants. Ces modules sont regroupés afin de construire le « user profil » (Figure 4). Un profil contiendra notamment des informations pouvant être exploitées afin de déterminer les préférences des utilisateurs en termes d'items (activités, sites touristiques…).

B. Processus de Filtrage :

En effet, l'adaptation de technique de recommandation repose entièrement sur le résultat de processus de profilage. Le processus prend en entrée l'ensemble des modules constituant le profil de l'utilisateur cible, à savoir :

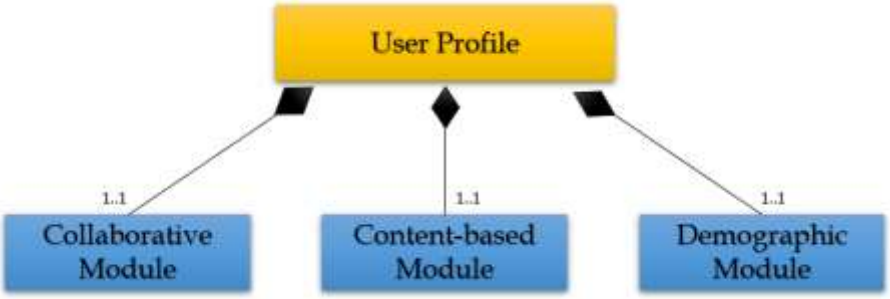


Fig. 4. Profilage d'utilisateur

- Le module basé sur le contenu : décrivant les caractéristiques des sites/activités touristiques que l'utilisateur a consultés dans le passé sous formes des vecteurs de mots-clefs obtenus après une phase d'indexation. Il est à signaler que ces mots-clefs sont

généralement soit extraits sur la base d'une indexation automatique lors de la consultation, soit attribués manuellement lors de l'inscription.

- Le module collaboratif : contenant les données d'évaluations des items consultés.
- Le module démographique : contient les attributs démographiques de l'utilisateur. Ces attributs peuvent être soit renseignés par l'utilisateur lui-même en remplissant le formulaire d'inscription, ou bien extraits par exemple à partir de son social login [77].

Une fois les modules de profil de l'utilisateur détectés, les approches de recommandation et la technique d'hybridation adéquate sont sélectionnées. Robin Burke a résumé les techniques d'hybridation en sept techniques [2] : pondération, mélange, cascade, commutation, combinaison de caractéristiques, augmentation de caractéristiques, et la méthode du méta-niveau.

Ce processus renvoie une liste d'items avec les degrés d'appréciations que peut porter l'utilisateur cible à chaque item (Figure 2).

C. Processus de Planification d'itinéraire :

Une fois que les degrés d'appréciations sont estimés, le système fait la sélection des items considérés pertinent pour l'utilisateur (dépassant un seuil donné par exemple), prenant en considération son contexte, et fait appel à des techniques de recherche opérationnelle pour corréler ces propositions sous forme d'un séjour.

II.3.3. Diagrammes du système

I.3.3.1. Diagramme de cas d'utilisateur

Le diagramme d'utilisateur se compose de deux acteurs clés : l'utilisateur (User) et l'administrateur (Admin). Les utilisateurs, représentant les visiteurs de la plateforme, bénéficient de plusieurs fonctionnalités. Ils peuvent consulter les articles touristiques disponibles, créer et gérer leur propre compte, créer des planifications de voyages personnalisées en sélectionnant des lieux et son nombre, ainsi que gérer leurs planifications existantes. D'autre part, l'administrateur est investi de

responsabilités plus étendues. Il a le pouvoir de gérer l'ensemble de la plateforme, y compris les articles, les utilisateurs, les notes attribuées, etc. De plus, l'administrateur analyse les performances du système, examine les tendances d'utilisation et évalue la qualité globale des recommandations et des contenus.

De plus, l'administrateur peut également jouer le rôle d'un utilisateur, ce qui signifie qu'il peut lire des articles touristiques, créer et gérer des plans de voyage.

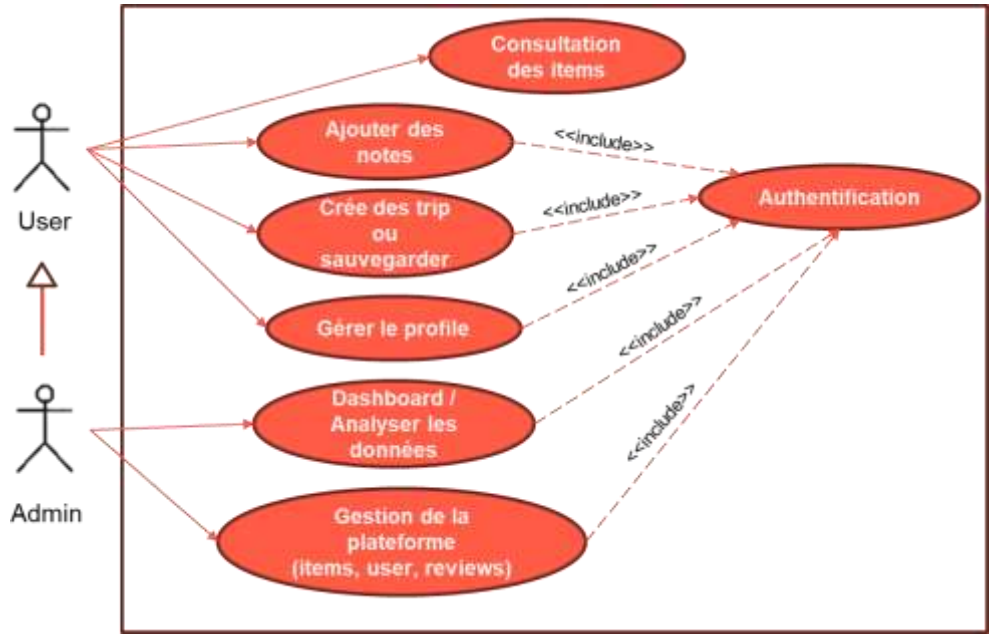


Fig. 5. Diagramme de cas d'utilisateur.

I.3.3.2. Diagramme de classe

Le diagramme de classe que nous avons élaboré pour notre plateforme touristique comprend cinq classes essentielles. Tout d'abord, la classe "User" représente les utilisateurs de la plateforme et contient des informations personnelles telles que le nom d'utilisateur, l'e-mail, la date de naissance et le mot de passe. Ensuite, la classe "User_Profile" est associée à chaque utilisateur et stocke les détails de ses activités (lecture d'articles et rédaction des notes). La classe "Item" concerne les sites touristiques et contient des données telles que des noms, des descriptions et des coordonnées de localisation (latitude et longitude).

Les "Reviews" sont consignées dans une classe distincte, contenant les avis des utilisateurs sur les sites touristiques, comprenant des notes et des

commentaires. Enfin, la classe "Trips" permet aux utilisateurs de planifier leurs voyages en assemblant différents sites touristiques pour créer des itinéraires personnalisés. Ces cinq classes sont interconnectées pour créer un système complet de recommandation touristique, où les profils des utilisateurs, les sites touristiques, les avis et les itinéraires de voyage sont tous pris en compte pour générer des recommandations adaptées à chaque utilisateur.

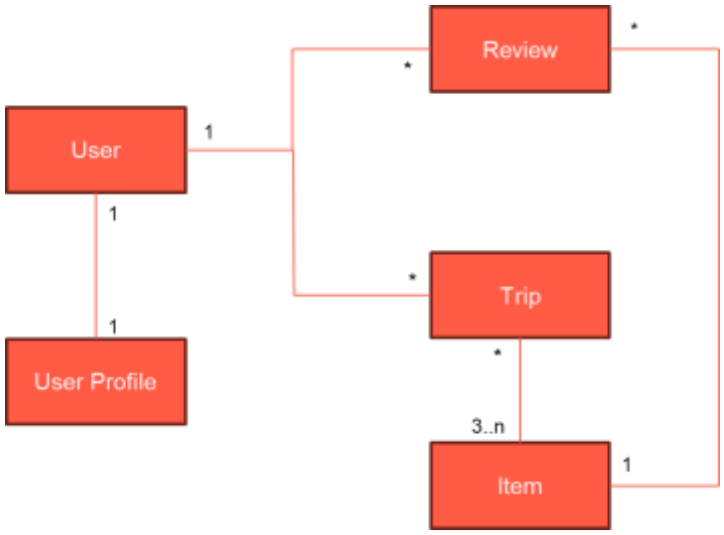


Fig. 6. Diagramme de classe

II.3.4. Méthodologie d'intégration de big data et IA dans l'architecture proposée

L'intégration du Big data et IA pour l'implémentation du système de recommandation proposé est l'une des pistes principales d'un projet qui vise à réaliser une solution Big data basée sur la recommandation hybride, l'analyse des sentiments et d'opinions en utilisant les techniques de Machine et Deep Learning [46] … Ce projet a pour objectif de fournir des outils intelligents pour cibler et recommander l'offre touristique le plus adapté selon le profil des utilisateurs, de suivre et analyser leurs opinions afin d'améliorer l'expérience client et prévoir la demande touristique au Maroc. Ce projet va aider les agences et les acteurs du tourisme Marocains, notamment ceux de la région Daraâ-Tafilalet, et leur permettre d'être plus disponibles sur Internet et d'offrir un meilleur service aux visiteurs. Pour ce faire, une méthodologie en quatre couches sera présentée qui décrit la démarche d'intégration de Big Data et de IA dans le système proposé :

Couche d'agrégation des données touristiques : Consiste à

mettre un très large éventail d'outils numériques pour accroître la visibilité, et l'attractivité des offres touristiques de la destination Daraâ-Tafilalet : Portails, Applications Mobiles/Web, Médias Sociaux, Réalité Augmentée, Reconstitutions de Monuments en 3D, Musées Virtuels, Bornes Interactives, Cartes et E-guides sur mobiles. Ces différents supports peuvent fournir une quantité considérable de contenus informatifs, mais aussi d'images et de vidéos. Ils proposent au visiteur une véritable « immersion » dans le territoire, et contribuent ainsi au rayonnement de la destination et à l'intensité de l'expérience touristique vécue. Dans ce sens, un ensemble de projets se sont déjà lancés pour conserver et valoriser le patrimoine culturel et naturel de la région [82] (Figure 7).

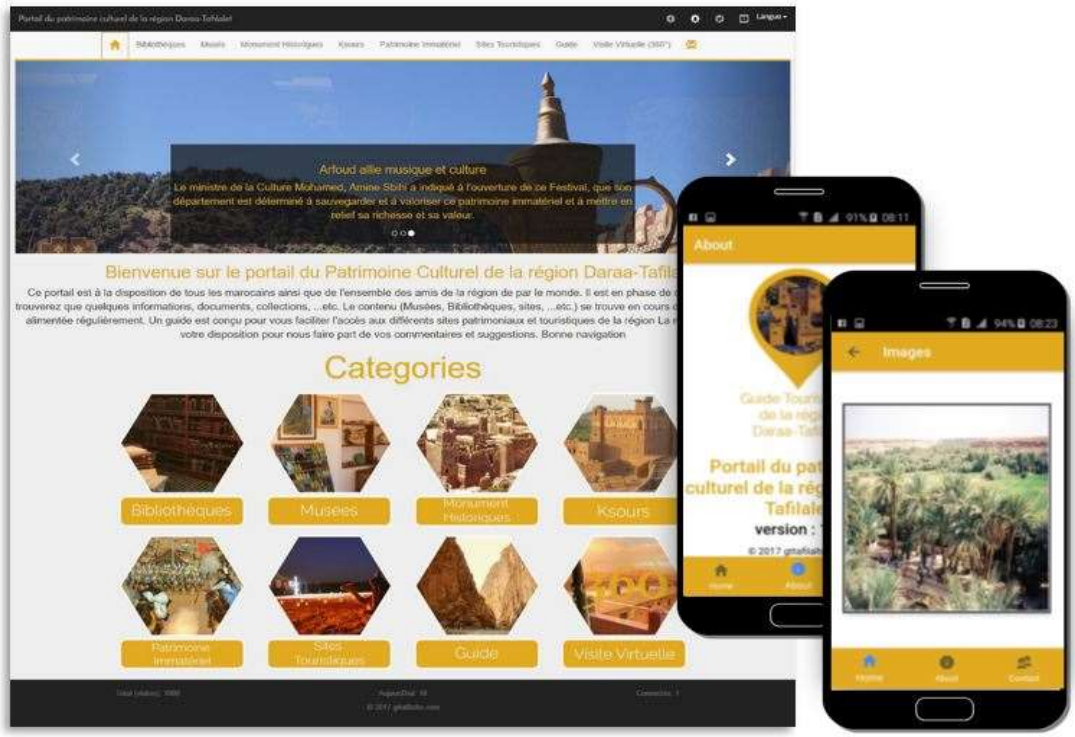


Fig. 7. La version web et mobile de plateforme du patrimoine culturel et naturel de la région Daraâ-Tafilalet

Les données touristiques sont souvent volumineuses et hétérogènes. Ce qui mène à l'utilisation du Big Data pour stocker ce volume important, et supporter et manipuler cette grande variété de données. Pour y parvenir, une large panoplie de solutions technologiques innovantes tels que des systèmes de gestion des bases de données NoSQL (Cassandra, MongoDB, …) et des systèmes de fichiers distribués (Hadoop HDFS), peuvent être adoptées. D'autres travaux de recherches [83,84] visent aussi à enrichir ces données sémantiquement en utilisant les technologies du Web sémantique, notamment les ontologies, pour une gestion et un accès plus intelligent.

Couche de Recommandation : vise à identifier le profile utilisateur en termes de préférences, en déterminant les modules (démographique, contextuelle, préférences ou ratings, …) qui le compose, choisir l'approche de recommandation appropriée à utiliser et développer l'algorithme adéquat à exécuter pour tirer profit des grands ensembles de données. Dans ce sens, les technologies big data offrent une implémentation à grande échelle de plusieurs techniques de Machine Learning et Deep Learning, y compris : Classification, Clustering, Association Rules, Regression, Collaborative Filtering, Recurrent Neural Network, … En se basant sur ces techniques, nous pouvons traiter et analyser en temps réel les différentes informations touristiques, exploiter les résultats obtenus afin de prédire les prochaines actions des touristes et leurs recommander des offres et des itinères appropriées. Par exemple, en analysant les trajets, mais aussi le temps passé devant les monuments, le système de recommandation peut mieux comprendre les attentes des visiteurs et y répondre.

Couche de visualisation des résultats : Cette couche va accompagner le touriste durant toutes les phases de son voyage, depuis la préparation jusqu'au partage en ligne, « avant/pendant/après ». En effet, maintenant quand on prépare un voyage, on commence par faire des recherches sur des sites, portails et applications mobiles pour s'informer et faire des choix, on prépare son parcours sur place avec des services et des contenus personnalisés, et on termine en se démultipliant sur les réseaux sociaux, blogs et forums. Cependant, pour personnaliser son parcours, le système va se basé sur des techniques de recherche opérationnelle et va présenter les résultats sur des cartes interactives.

Couche de validation de la solution proposée : Consiste à suivre et analyser les opinions et les sentiments des visiteurs présentés sur les blogs et les médias sociaux, par les compagnies de tourisme, pour comprendre leurs changements de désirs et partager les informations qui influencent leurs décisions. Les données de feedback sont généralement transformées, grâces aux technologies de visualisation et de BI, en des représentations graphiques (graphes, histogrammes, etc.) simples et faciles à comprendre. Ces représentations graphiques permettront aux

professionnels du domaine de tourisme d'avoir une vision claire, pour prendre les bonnes décisions et améliorer leurs stratégies. Ils pourront ainsi identifier les besoins des touristes, prédire leurs comportements futurs et proposer les ressources les mieux adaptées à chaque profil.

II.1. Conclusion

Afin d'aider les touristes dans le processus de sélection et de surmonter la surcharge d'informations, des systèmes de recommandation ont été développés au cours de la dernière décennie du vingtième siècle. Dans ce chapitre, nous avons effectué une revue de la littérature sur les systèmes de recommandation touristique actuels, puis nous avons présenté un cadre conceptuel pour la mise en œuvre des systèmes de recommandation touristique. L'architecture proposée basée sur le filtrage hybride vise à améliorer l'expérience du visiteur en lui recommandant les éléments les plus pertinents et en l'aidant à personnaliser son voyage.

Une fois que les ensembles d'éléments considérés comme pertinents pour le touriste sont sélectionnés, le système planifie un voyage approprié en combinant ces éléments à l'aide de techniques de recherche opérationnelle.

L'architecture de système de recommandation proposée sera alors l'objet d'une implémentation, par l'intermédiaire des nouvelles technologies Big Data et IA ; Nous avons choisi la région Drâa-Tafilalet comme destination pour concevoir et valider la solution proposée en procédant ensuite à l'extension de la plateforme à l'ensemble des régions du Royaume.

Chapitre III

L'IA au Service des Voyageurs : Révolution des Systèmes de Recommandation Intelligents

Chapitre III : L'IA au Service des Voyageurs : Révolution des Systèmes de Recommandation Intelligents 52
III.1. Introduction 53
III.2. Analyse du Tourisme Traditionnel 54
III.2.1. Méthodes et Techniques Traditionnelles Employées dans le Tourisme 55
III.2.2. Défis et Limites du Tourisme Traditionnel 56
III.3. Transformation Numérique dans le Tourisme 58
III.3.1. Importance de la Révolution Numérique pour le Secteur du Tourisme 59
III.3.2. Techniques Modernes Utilisées dans le Tourisme 60
III.3.3. Avantages et Opportunités Offertes par le Tourisme Numérique .. 61
III.4. L'Intégration des Systèmes de Recommandation dans le Tourisme ... 62
III.4.1. Rôle des Systèmes de Recommandation dans la Personnalisation de l'Expérience Touristique 63
III.4.2. Cas d'Étude de Systèmes de Recommandation dans d'Autres Régions Touristiques 64
III.4.2.1. Cas d'Étude 1 : Recommandation dans la Région du Tyrol, Autriche 64
III.4.2.2. **Cas d'Étude 2 : Intégration de Systèmes de Recommandation à Séoul, Corée du Sud** 65
III.4.2.3. Cas d'Étude 3 : Systèmes de Recommandation dans les Îles Canaries, Espagne 65
III.5. Conclusion 66

III.1. Introduction

Le tourisme a longtemps été une activité économique majeure, contribuant de manière significative au développement économique, à la création d'emplois, et à la promotion culturelle des régions [85]. La région de Drâa-Tafilalet, riche en patrimoine culturel, en paysages naturels variés, et en traditions ancestrales, est un exemple typique de région où le tourisme joue un rôle crucial dans le développement régional.

Cependant, comme dans de nombreuses autres régions, le tourisme dans Drâa-Tafilalet a été principalement dominé par des méthodes traditionnelles de marketing et de gestion, qui, bien qu'efficaces à une certaine époque, montrent aujourd'hui leurs limites [86] face à la montée en puissance des technologies numériques.

Historiquement, le tourisme traditionnel reposait sur des stratégies de marketing de masse telles que les brochures, les publicités dans les médias imprimés, et les foires touristiques. Ces méthodes avaient pour objectif d'attirer un large public, sans distinction particulière des besoins ou préférences individuelles des voyageurs. Les agences de voyage jouaient également un rôle central dans la planification et la réservation des voyages, offrant aux touristes des packages standardisés et peu personnalisés, avec l'évolution des attentes des consommateurs et l'avènement des nouvelles technologies, ce modèle de tourisme a commencé à montrer des signes de faiblesse. Les voyageurs modernes recherchent désormais des expériences plus personnalisées, authentiques, et instantanément accessibles, ce qui a rendu les approches traditionnelles de plus en plus obsolètes.

L'ère numérique a apporté une transformation radicale au secteur du tourisme, changeant la façon dont les destinations sont découvertes, réservées, et consommées [87]. L'intégration des technologies numériques dans le tourisme, souvent désignée sous le terme de "révolution numérique", a introduit de nouvelles dynamiques qui ont bouleversé les paradigmes existants. Cette révolution est caractérisée par l'utilisation généralisée de l'Internet, des smartphones, des réseaux sociaux, des plateformes de réservation en ligne, et de la réalité augmentée pour créer des expériences touristiques plus interactives et immersives.

En plus des plateformes de réseaux sociaux ont joué un rôle

déterminant dans la transformation numérique du tourisme [88]. Les plateformes telles qu'Instagram, Facebook, et Twitter sont devenues des outils puissants pour la promotion touristique, permettant aux destinations de se connecter directement avec les voyageurs potentiels. Les voyageurs eux- mêmes contribuent à cette dynamique en partageant leurs expériences en temps réel, ce qui influence les décisions d'autres utilisateurs et crée des tendances touristiques. Cette interaction constante a transformé le tourisme en une activité plus participative et communautaire, où les avis et les recommandations jouent un rôle central dans la prise de décision des voyageurs.

La région de Drâa-Tafilalet, avec son riche patrimoine et sel touristique encore sous- exploité, se trouve à un carrefour entre le tourisme traditionnel et la modernisation numérique. L'intégration des technologies numériques dans cette région pourrait non seulement améliorer l'attractivité de la région, mais aussi permettre une meilleure gestion des flux touristiques [89], une diversification de l'offre touristique, et une conservation plus efficace des ressources culturelles et naturelles.

Ainsi, ce chapitre se propose d'examiner en profondeur cette transformation du tourisme, en commençant par une analyse des méthodes traditionnelles de promotion et de gestion touristique, suivie d'une exploration des innovations numériques qui ont révolutionné le secteur. Nous discuterons également de l'impact de ces technologies sur la région de Drâa-Tafilalet, en soulignant les défis et les opportunités qu'elles représentent pour le futur du tourisme dans cette région.

III.2. Analyse du Tourisme Traditionnel

Le tourisme traditionnel a longtemps été le pilier central du développement économique et culturel de nombreuses régions à travers le monde. Ce modèle, bien que solide à son époque, reposait sur des pratiques et des infrastructures qui, aujourd'hui, semblent dépassées par l'avènement des technologies numériques. Toutefois, pour comprendre la transformation du tourisme moderne, il est crucial d'examiner en détail les méthodes et les techniques qui ont façonné le tourisme traditionnel, ainsi que les défis inhérents à ce modèle.

Le tourisme traditionnel était profondément ancré dans les stratégies de marketing de masse, où l'objectif principal était d'attirer le plus grand nombre de touristes possible sans distinction majeure de leurs préférences ou de leurs besoins individuels. Cette approche, bien qu'efficace pour générer un volume élevé de visiteurs, présentait des limitations significatives en termes de personnalisation de l'expérience touristique. Les agences de voyage et les offices du tourisme jouaient un rôle central dans la diffusion de l'information, s'appuyant principalement sur des supports physiques tels que les brochures, les magazines, les affiches, et les foires touristiques. Ces canaux de communication, bien qu'essentiels pour l'époque, étaient limités par leur capacité à adapter les messages aux besoins spécifiques des différents segments de marché.

III.2.1. Méthodes et Techniques Traditionnelles Employées dans le Tourisme

Les méthodes de promotion dans le tourisme traditionnel reposaient sur une combinaison de publicité imprimée, de participation à des événements et de collaborations avec des agences de voyage. Les brochures, par exemple, étaient des outils essentiels pour la communication des destinations touristiques. Elles étaient souvent distribuées dans des lieux stratégiques tels que les aéroports, les agences de voyage, et les foires touristiques. Ces brochures fournissaient des informations clés sur les attractions locales, les hébergements, et les activités disponibles, mais leur contenu était généralement statique et ne pouvait pas être mis à jour facilement pour refléter les changements ou les nouvelles offres.

Les magazines spécialisés dans le tourisme étaient également des outils de promotion importants. Ces publications ciblaient souvent un public plus aisé et étaient utilisées pour promouvoir des destinations prestigieuses ou des expériences haut de gamme. Les articles publiés dans ces magazines mettaient en avant des aspects spécifiques des destinations, mais la portée de ces publications était souvent limitée à un segment spécifique de la population. Les affiches et les panneaux publicitaires dans les grandes villes ou les aéroports visaient à capter l'attention des passants et à susciter l'intérêt pour une destination particulière. Cependant, comme pour les

brochures, ces supports étaient limités par leur incapacité à offrir une interaction directe avec les touristes potentiels.

Les foires et salons touristiques, où les offices du tourisme et les prestataires de services exposaient leurs offres, étaient également des événements clés dans le calendrier du tourisme traditionnel. Ces événements permettaient aux acteurs du secteur de rencontrer directement les clients potentiels, de présenter leurs produits et services, et de nouer des relations avec des partenaires commerciaux. Toutefois, la participation à ces événements nécessitait des ressources considérables et n'était pas toujours accessible aux petites entreprises ou aux destinations moins connues.

Le rôle des agences de voyage ne peut pas être sous-estimé dans le contexte du tourisme traditionnel. Ces agences servaient de point de contact principal pour les voyageurs, offrant des services de conseil, de réservation, et d'organisation de voyages. Elles proposaient souvent des packages touristiques standardisés qui incluaient le transport, l'hébergement, et parfois des visites guidées. Cependant, cette standardisation limitait la possibilité pour les voyageurs de personnaliser leurs expériences en fonction de leurs préférences individuelles. Les agences de voyage dépendaient fortement des commissions versées par les hôtels et les compagnies aériennes, ce qui influençait parfois les recommandations qu'elles faisaient à leurs clients.

III.2.2. Défis et Limites du Tourisme Traditionnel

Malgré ses succès, le tourisme traditionnel était confronté à plusieurs défis qui ont progressivement révélé ses limites, surtout à mesure que les attentes des consommateurs évoluaient et que les nouvelles technologies faisaient leur apparition.

L'un des principaux défis du tourisme traditionnel résidait dans la difficulté à atteindre un public large sans l'aide des moyens numériques. Les stratégies de marketing étaient souvent coûteuses et nécessitaient des investissements importants en publicité imprimée, en participation à des foires, et en collaborations avec des agences de voyage. Cette approche excluait de facto de nombreuses petites destinations et entreprises

touristiques, qui n'avaient pas les ressources nécessaires pour se promouvoir efficacement à grande échelle.

De plus, la dépendance à des informations statiques constituait une autre limitation majeure. Les brochures, magazines, et autres supports imprimés ne pouvaient être mis à jour qu'à intervalles réguliers, ce qui signifiait que les informations qu'ils contenaient pouvaient rapidement devenir obsolètes. Par exemple, les horaires des attractions touristiques, les prix des billets, et les disponibilités des hébergements pouvaient changer, mais les brochures ne pouvaient pas refléter ces changements en temps réel. Cela entraînait souvent de la confusion et des frustrations chez les voyageurs, qui se retrouvaient parfois avec des informations incorrectes ou dépassées.

L'incapacité à personnaliser les offres touristiques en fonction des préférences individuelles des voyageurs était également une faiblesse notable du tourisme traditionnel. Les packages touristiques standardisés offerts par les agences de voyage ne prenaient pas en compte les préférences spécifiques des voyageurs, telles que leurs intérêts personnels, leurs habitudes de voyage, ou leurs besoins particuliers. Cette uniformité dans l'offre réduisait la satisfaction des clients, qui avaient souvent l'impression de vivre une expérience touristique impersonnelle et générique.

Un autre défi majeur était le manque d'interaction directe entre les prestataires de services et les clients. Dans le modèle traditionnel, les agences de voyage servaient d'intermédiaires entre les clients et les fournisseurs de services touristiques, ce qui limitait la possibilité pour les prestataires de recueillir des retours directs de la part des voyageurs. Cette absence de communication directe empêchait les prestataires de services d'adapter rapidement leurs offres en fonction des besoins et des attentes changeants des clients. De plus, les clients avaient souvent peu de moyens pour donner leur avis ou partager leurs expériences, ce qui limitait l'amélioration continue des services touristiques.

En outre, le tourisme traditionnel n'était pas toujours en mesure de s'adapter rapidement aux changements du marché ou aux nouvelles tendances. Par exemple, les changements dans les préférences des

voyageurs vers des expériences plus authentiques, immersives, et personnalisées n'ont pas été facilement intégrés dans les offres touristiques traditionnelles, qui étaient souvent rigides et standardisées. Cette incapacité à évoluer avec les attentes des consommateurs a conduit à un déclin progressif de l'attractivité des destinations qui s'appuyaient uniquement sur des méthodes traditionnelles de promotion.

Enfin, le tourisme traditionnel présentait des défis logistiques importants, notamment en ce qui concerne la gestion des flux touristiques et la distribution des services. Les infrastructures touristiques étaient souvent conçues pour accueillir un grand nombre de visiteurs, mais elles n'étaient pas toujours flexibles ou efficaces pour répondre aux variations saisonnières ou à la demande imprévisible. De plus, la concentration des touristes dans certaines zones géographiques spécifiques entraînait des problèmes de sur fréquentation, avec des impacts négatifs sur les ressources naturelles, la culture locale, et la qualité de vie des résidents.

Ainsi, bien que le tourisme traditionnel ait permis à de nombreuses régions de développer leur offre touristique et d'attirer des visiteurs du monde entier, ses limitations ont progressivement révélé la nécessité d'un changement vers des modèles plus flexibles, personnalisés, et adaptés aux nouvelles réalités du marché mondial. Cette transition a été facilitée par l'émergence des technologies numériques, qui ont ouvert de nouvelles perspectives pour le secteur touristique, rendant possible une interaction plus directe, une personnalisation accrue, et une gestion plus efficace des ressources.

III.3. Transformation Numérique dans le Tourisme

Le tourisme, secteur historiquement dépendant de méthodes traditionnelles pour attirer et servir les visiteurs, a subi une transformation radicale au cours des dernières décennies grâce à l'émergence des technologies numériques. Ce bouleversement, souvent qualifié de "révolution numérique", a profondément modifié la manière dont les destinations touristiques sont perçues, commercialisées et consommées par les voyageurs. En effet, le passage au numérique a non seulement permis une amélioration significative des services offerts aux touristes, mais a également réinventé la chaîne de valeur de l'industrie touristique

dans son ensemble.

Le développement de nouvelles technologies telles que l'internet, les smartphones, et les plateformes de réservation en ligne a donné naissance à des formes innovantes de tourisme, notamment l'e-tourisme, qui place le numérique au cœur de l'expérience touristique. Cette transformation numérique a également donné lieu à des stratégies marketing plus ciblées, permettant aux prestataires de services de mieux répondre aux attentes de voyageurs de plus en plus exigeants. Ce chapitre explore en détail l'importance de cette révolution numérique dans le secteur du tourisme, les techniques modernes qui y sont associées, ainsi que les avantages et opportunités qu'elle a engendrés.

III.3.1. Importance de la Révolution Numérique pour le Secteur du Tourisme

L'importance de la révolution numérique dans le secteur du tourisme ne peut être sous- estimée. À une époque où les voyageurs sont de plus en plus connectés et informés, la capacité à accéder à des informations précises et actualisées en temps réel est devenue essentielle pour attirer et satisfaire les clients. La transformation numérique a permis au secteur du tourisme de s'adapter aux attentes croissantes des voyageurs modernes, qui recherchent des expériences personnalisées, interactives, et instantanément accessibles.

Les innovations technologiques ont redéfini la manière dont les touristes planifient, réservent, et vivent leurs voyages. Par exemple, l'accès à des plateformes numériques comme les sites web et les applications mobiles a rendu possible la recherche d'informations sur les destinations, la comparaison des prix, et la réservation de services en quelques clics. Les outils de réservation en ligne, tels que Booking.com ou Airbnb, ont révolutionné le marché de l'hébergement en offrant aux utilisateurs la possibilité de choisir parmi une vaste gamme d'options adaptées à leurs préférences et à leur budget. De plus, les systèmes de paiement en ligne sécurisés ont facilité les transactions, réduisant ainsi les barrières à la réservation.

La révolution numérique a également amélioré la communication entre

les prestataires de services et les clients. Les plateformes sociales, comme Facebook, Instagram, et Twitter, permettent aux destinations touristiques et aux entreprises du secteur de se connecter directement avec les clients, de partager du contenu en temps réel, et de recevoir des retours immédiats sur leurs services. Cette interaction directe a non seulement renforcé l'engagement des clients, mais a également permis aux entreprises de mieux comprendre les attentes et les préférences des voyageurs, et d'ajuster leurs offres en conséquence.

Enfin, la transformation numérique a favorisé le développement de nouvelles formes de tourisme. L'e-tourisme, qui intègre les technologies numériques à toutes les étapes de l'expérience touristique, a ouvert la voie à une personnalisation accrue des voyages, permettant aux touristes de créer des itinéraires sur mesure qui répondent à leurs besoins spécifiques. Cette personnalisation est renforcée par l'utilisation de technologies telles que le Big Data et l'intelligence artificielle, qui permettent de collecter et d'analyser d'énormes quantités de données sur les comportements des utilisateurs, afin de proposer des recommandations personnalisées en temps réel.

III.3.2. Techniques Modernes Utilisées dans le Tourisme

La transformation numérique du secteur du tourisme s'appuie sur une variété de techniques modernes qui ont redéfini la manière dont les services touristiques sont commercialisés, consommés, et gérés. Parmi ces techniques, l'e-tourisme, les plateformes numériques, et la réalité augmentée jouent un rôle central.

L'e-tourisme représente l'intégration des technologies numériques dans tous les aspects du voyage, de la planification à la réalisation, en passant par la réservation et l'évaluation post- voyage. Cette approche englobe l'utilisation de sites web, d'applications mobiles, et de plateformes en ligne pour rechercher des destinations, comparer des offres, réserver des services, et partager des expériences. L'e-tourisme permet aux voyageurs d'accéder à des informations actualisées, de personnaliser leurs itinéraires, et de bénéficier de recommandations basées sur leurs préférences et comportements antérieurs.

Les plateformes numériques, telles que les sites web, les applications mobiles, et les réseaux sociaux, jouent un rôle crucial dans la promotion des destinations touristiques et la gestion des interactions avec les clients. Les sites web des offices de tourisme et des entreprises du secteur offrent une vitrine numérique pour les destinations, permettant aux visiteurs de découvrir des attractions, de consulter des avis, et de réserver des services. Les applications mobiles, quant à elles, offrent une accessibilité accrue, permettant aux utilisateurs de planifier leurs voyages et d'accéder à des informations en temps réel, où qu'ils se trouvent.

Les réseaux sociaux ont également transformé la manière dont les destinations sont commercialisées. Les plateformes comme Instagram et Facebook sont devenues des outils de promotion puissants, où les images et les vidéos partagées par les utilisateurs et les influenceurs jouent un rôle clé dans l'inspiration des voyages. Ces plateformes permettent également aux destinations d'engager une conversation directe avec les voyageurs, de partager des mises à jour en temps réel, et de gérer leur réputation en ligne.

La réalité augmentée (RA) est une autre technologie innovante qui a été intégrée dans l'expérience touristique pour enrichir la visite des sites et des attractions. La RA permet de superposer des informations numériques, telles que des descriptions, des images, ou des vidéos, sur des éléments du monde réel, offrant ainsi une expérience immersive et interactive aux visiteurs. Par exemple, les applications de réalité augmentée peuvent fournir des informations historiques sur un monument, ou permettre aux utilisateurs de visualiser des reconstitutions de sites archéologiques en temps réel. Cette technologie améliore non seulement la compréhension et l'appréciation des sites par les visiteurs, mais elle contribue également à rendre l'expérience touristique plus engageante et mémorable.

III.3.3. Avantages et Opportunités Offertes par le Tourisme Numérique

Le passage au numérique a apporté une multitude d'avantages au secteur du tourisme, créant de nouvelles opportunités pour les entreprises, les destinations, et les voyageurs.

L'un des principaux avantages du tourisme numérique est la capacité à offrir des expériences hautement personnalisées. Grâce aux technologies

comme le Big Data et l'intelligence artificielle, les entreprises peuvent analyser les comportements des utilisateurs et proposer des recommandations sur mesure, adaptées aux préférences individuelles des voyageurs. Cette personnalisation augmente la satisfaction des clients, qui se sentent mieux compris et mieux servis, et contribue à fidéliser les clients.

Le tourisme numérique a également amélioré la gestion des flux touristiques, en permettant une meilleure répartition des visiteurs dans l'espace et dans le temps. Par exemple, les applications mobiles et les systèmes de réservation en ligne peuvent être utilisés pour encourager les visiteurs à explorer des sites moins fréquentés, ou pour réserver des créneaux horaires spécifiques pour visiter des attractions populaires, réduisant ainsi la sur fréquentation et préservant les ressources naturelles et culturelles.

De plus, le numérique a transformé la manière dont les destinations touristiques gèrent leur image et interagissent avec les clients. Les réseaux sociaux et les plateformes d'avis en ligne permettent aux destinations de surveiller leur réputation en temps réel, de répondre rapidement aux critiques, et de mettre en avant leurs atouts auprès d'un large public. Cette gestion proactive de la réputation aide les destinations à attirer plus de visiteurs et à se différencier dans un marché de plus en plus compétitif.

Le tourisme numérique a également ouvert de nouvelles perspectives pour le développement durable du secteur. En facilitant l'accès à des informations précises et actualisées, les technologies numériques permettent aux visiteurs de faire des choix plus éclairés et plus respectueux de l'environnement. Par exemple, les applications mobiles peuvent encourager les visiteurs à utiliser des moyens de transport écologiques, à choisir des hébergements écoresponsables, ou à respecter les règles de conduite sur les sites naturels sensibles. De plus, la gestion en temps réel des flux touristiques contribue à réduire l'impact environnemental du tourisme en limitant la sur fréquentation des sites et en protégeant les écosystèmes locaux.

Enfin, le tourisme numérique a permis de réduire les barrières à l'entrée pour de nombreuses petites entreprises et destinations, en leur offrant des outils abordables pour se promouvoir et se connecter directement avec les

clients. Les plateformes de réservation en ligne, les réseaux sociaux, et les sites d'avis ont démocratisé l'accès au marché touristique mondial, permettant à des destinations autrefois méconnues de gagner en visibilité et d'attirer un public international.

III.4. L'Intégration des Systèmes de Recommandation dans le Tourisme

À mesure que l'industrie touristique continue de se numériser, les systèmes de recommandation sont devenus des outils essentiels pour offrir des expériences personnalisées aux voyageurs. Ces systèmes, qui analysent les préférences, les comportements, et les données des utilisateurs, jouent un rôle crucial dans la création d'itinéraires sur mesure, la suggestion de sites et d'activités, et l'amélioration générale de l'expérience touristique. Grâce à eux, les destinations peuvent non seulement répondre aux attentes des visiteurs, mais aussi anticiper leurs besoins, renforçant ainsi la satisfaction et la fidélité des clients. Cette section explore en détail le rôle de ces systèmes dans la personnalisation de l'expérience touristique et présente des études de cas illustrant leur efficacité dans différentes régions.

III.4.1. Rôle des Systèmes de Recommandation dans la Personnalisation de l'Expérience Touristique

Les systèmes de recommandation, initialement popularisés par des plateformes de commerce électronique comme Amazon et des services de streaming comme Netflix, ont trouvé des applications significatives dans le secteur du tourisme. Ces systèmes utilisent diverses techniques, allant de la recommandation basée sur le contenu à la recommandation collaborative, pour analyser les données des utilisateurs et proposer des suggestions qui correspondent à leurs préférences individuelles.

La personnalisation est devenue une priorité dans le tourisme moderne, où les voyageurs cherchent des expériences uniques adaptées à leurs goûts spécifiques. Les systèmes de recommandation permettent aux prestataires de services touristiques de répondre à cette demande en proposant des activités, des sites, et des services qui correspondent étroitement aux intérêts des utilisateurs. Par exemple, un système de recommandation pourrait suggérer des visites de musées pour un amateur d'art, des

excursions en plein air pour un passionné de nature, ou des expériences gastronomiques pour un gourmet, en fonction de leurs historiques de recherche, de leurs comportements d'achat, ou même de leurs interactions sur les réseaux sociaux.

L'un des avantages majeurs des systèmes de recommandation est leur capacité à traiter de grandes quantités de données pour identifier des tendances et des préférences cachées, souvent inaccessibles par des méthodes traditionnelles de marketing. Ces systèmes peuvent intégrer diverses sources de données, telles que les réservations passées, les avis en ligne, les historiques de navigation, et les données démographiques, pour construire un profil détaillé de chaque utilisateur. À partir de ces profils, ils sont capables de générer des recommandations personnalisées en temps réel, offrant ainsi une expérience utilisateur plus engageante et pertinente.

En outre, les systèmes de recommandation jouent un rôle clé dans la gestion des flux touristiques, en orientant les visiteurs vers des sites moins fréquentés ou en dehors des heures de pointe, ce qui contribue à une meilleure gestion des ressources et à la réduction de la sur fréquentation. Par exemple, un système pourrait recommander la visite d'un site touristique populaire tôt le matin ou en fin de journée, lorsque l'affluence est moindre, ou suggérer des attractions alternatives offrant des expériences similaires.

Les systèmes de recommandation ne se contentent pas de répondre aux attentes des voyageurs ; ils contribuent également à accroître leur satisfaction en leur offrant des expériences sur mesure. En suggérant des activités et des sites qui correspondent aux intérêts spécifiques des utilisateurs, ces systèmes permettent aux voyageurs de découvrir des aspects de la destination qu'ils n'auraient peut-être pas envisagés, enrichissant ainsi leur expérience globale.

III.4.2. Cas d'Étude de Systèmes de Recommandation dans d'Autres Régions Touristiques

Plusieurs destinations touristiques à travers le monde ont intégré avec succès des systèmes de recommandation pour améliorer leur offre et attirer davantage de visiteurs. Ces cas d'étude illustrent comment les technologies

de recommandation peuvent être utilisées pour personnaliser l'expérience touristique, augmenter la satisfaction des clients, et optimiser la gestion des flux de visiteurs.

III.4.2.1. Cas d'Étude 1 : Recommandation dans la Région du Tyrol, Autriche

La région du Tyrol, en Autriche, a mis en place plusieurs systèmes de recommandation innovants pour améliorer l'expérience des touristes visitant ses nombreuses attractions naturelles [90,91]. En utilisant une plateforme basée sur l'intelligence artificielle, le Tyrol offre aux visiteurs des suggestions personnalisées pour des randonnées, des excursions en montagne, et des activités de plein air. Ce système analyse les préférences des utilisateurs, telles que la difficulté des sentiers, la durée des randonnées, et les centres d'intérêt (par exemple, la faune, la flore, les vues panoramiques), pour proposer des itinéraires sur mesure qui répondent aux attentes individuelles des visiteurs. Ce projet a non seulement amélioré la satisfaction des touristes, mais a également aidé à mieux répartir les visiteurs sur les différents sites, réduisant ainsi la pression sur les zones les plus populaires.

III.4.2.2. Cas d'Étude 2 : Intégration de Systèmes de Recommandation à Séoul, Corée du Sud

Séoul, la capitale dynamique de la Corée du Sud, a adopté quelques systèmes de recommandation pour améliorer l'expérience des touristes en ville [92,93]. Grâce à une application mobile alimentée par un algorithme de recommandation, les visiteurs peuvent recevoir des suggestions personnalisées pour des restaurants, des boutiques, des événements culturels, et des attractions touristiques. Le système utilise des données en temps réel pour adapter les recommandations en fonction de la localisation actuelle de l'utilisateur, de ses préférences exprimées, et des tendances touristiques actuelles dans la ville. Par exemple, si un utilisateur montre un intérêt pour l'architecture moderne, l'application peut suggérer des visites de bâtiments emblématiques, tels que la Tour Lotte ou le Dongdaemun Design Plaza. Ce système a contribué à augmenter la durée de séjour des touristes à Séoul, tout en leur offrant une expérience plus enrichissante et personnalisée.

III.4.2.3. Cas d'Étude 3 : Systèmes de Recommandation dans les

Îles Canaries, Espagne

Les Îles Canaries, une destination touristique populaire en Espagne, ont intégré plusieurs systèmes de recommandation pour promouvoir un tourisme durable et responsable [94,95]. En utilisant une plateforme de recommandation, les autorités locales orientent les visiteurs vers des zones moins connues et encouragent des activités respectueuses de l'environnement, telles que la randonnée sur des sentiers écologiques, la visite de réserves naturelles, et la participation à des programmes de conservation. Le système propose également des recommandations basées sur la météo et les conditions environnementales, aidant les touristes à choisir les activités les plus appropriées en fonction du climat local. Ce système a permis de diversifier l'offre touristique des Îles Canaries et de répartir les flux touristiques de manière plus équilibrée sur l'archipel, tout en promouvant un tourisme plus durable.

Ces cas d'étude montrent clairement que les systèmes de recommandation peuvent jouer un rôle crucial dans l'amélioration de l'expérience touristique. Ils permettent non seulement de personnaliser les offres en fonction des préférences individuelles, mais aussi d'optimiser la gestion des flux touristiques, de promouvoir des pratiques durables, et d'augmenter la satisfaction globale des visiteurs. En adoptant ces technologies, les destinations touristiques peuvent non seulement attirer plus de visiteurs, mais aussi offrir des expériences mémorables qui favorisent la fidélité des clients.

III.5. Conclusion

La transformation numérique a profondément modifié le secteur du tourisme, bouleversant les méthodes traditionnelles et ouvrant de nouvelles perspectives pour l'amélioration de l'expérience des visiteurs. Cette évolution est marquée par une transition significative du tourisme traditionnel, caractérisé par des pratiques telles que la publicité imprimée, les brochures et les agences de voyage physiques, vers un tourisme numérique sophistiqué qui exploite les technologies modernes pour offrir des services plus personnalisés, interactifs et accessibles en temps réel. L'impact de cette transformation est palpable à travers divers aspects du secteur, notamment la personnalisation accrue des offres, l'amélioration de

l'accessibilité à l'information, et l'optimisation des services touristiques grâce à des outils numériques avancés.

L'intégration des systèmes de recommandation joue un rôle crucial dans cette révolution numérique, en permettant aux destinations de proposer des expériences sur mesure adaptées aux préférences et aux comportements individuels des voyageurs. Ces systèmes exploitent des algorithmes sophistiqués pour analyser les données des utilisateurs et offrir des recommandations personnalisées qui répondent précisément aux attentes des visiteurs. Par exemple, en utilisant des données historiques, des préférences exprimées, et des tendances émergentes, les systèmes de recommandation peuvent suggérer des activités, des attractions et des services qui enrichissent l'expérience touristique, rendant chaque voyage plus mémorable et pertinent.

Les avantages de cette personnalisation sont multiples. En permettant aux visiteurs de découvrir des aspects de la destination qui correspondent à leurs intérêts spécifiques, les systèmes de recommandation non seulement augmentent la satisfaction des clients, mais ils contribuent également à une gestion plus efficace des ressources touristiques. En dirigeant les flux de visiteurs vers des sites moins fréquentés ou en offrant des recommandations basées sur des périodes de moindre affluence, ces systèmes aident à réduire la surfréquentation des attractions populaires, contribuant ainsi à une expérience plus agréable pour tous et à une meilleure gestion des infrastructures.

Dans le contexte spécifique de la région de Drâa-Tafilalet, riche en patrimoine culturel et naturel, les systèmes de recommandation peuvent jouer un rôle déterminant dans la valorisation et la préservation de ces ressources uniques. En proposant des recommandations adaptées aux intérêts des visiteurs, ces systèmes peuvent aider à mettre en avant les trésors cachés de la région, stimuler l'exploration de sites moins connus, et promouvoir des activités qui mettent en valeur le patrimoine local. De plus, en intégrant des aspects tels que les conditions météorologiques locales, les événements culturels en cours, et les particularités géographiques, les recommandations peuvent être encore plus ciblées et pertinentes, offrant une expérience plus immersive et authentique aux voyageurs.

Les exemples de mise en œuvre réussie de systèmes de recommandation dans diverses régions touristiques à travers le monde montrent l'efficacité de ces technologies dans l'amélioration de l'expérience touristique. Les cas d'étude des Îles Canaries, du Tyrol, et de Séoul illustrent comment ces systèmes peuvent être utilisés pour promouvoir des pratiques de tourisme durable, gérer les flux touristiques, et offrir des recommandations pertinentes qui répondent aux besoins des visiteurs. En appliquant des approches similaires à Drâa-Tafilalet, la région peut tirer parti des systèmes de recommandation pour attirer un public plus large, offrir des expériences enrichissantes, et soutenir le développement durable du tourisme local.

Toutefois, il est essentiel de reconnaître que la mise en œuvre de systèmes de recommandation dans le tourisme nécessite une approche soigneusement planifiée. Les défis liés à la collecte et à la gestion des données, à la protection de la vie privée des utilisateurs, et à l'intégration des technologies dans les infrastructures existantes doivent être pris en compte. Une attention particulière doit également être accordée à l'adaptation des recommandations aux spécificités culturelles et locales de la région, afin d'assurer une pertinence et une sensibilité appropriées.

En conclusion, la transformation numérique et l'intégration des systèmes de recommandation représentent des opportunités considérables pour réinventer le secteur du tourisme. En exploitant ces technologies avancées, les destinations comme Drâa-Tafilalet peuvent non seulement améliorer l'expérience des visiteurs, mais aussi promouvoir une gestion plus équilibrée et durable des ressources touristiques. En combinant innovation technologique avec sensibilité culturelle et respect des spécificités locales, il est possible de créer des expériences touristiques qui sont à la fois enrichissantes pour les visiteurs et bénéfiques pour la communauté locale. La clé du succès réside dans l'engagement à utiliser ces outils de manière réfléchie et stratégique, en mettant l'accent sur la personnalisation, la durabilité, et la satisfaction des besoins des voyageurs tout en préservant et en valorisant les richesses culturelles et naturelles de la région.

Chapitre IV

Implémentation de la solution : Un système de recommandation hybride pour la promotion du tourisme Marocain

Chapitre IV : Implémentation de la solution : Un système de recommandation hybride pour la promotion du tourisme Marocain 69

IV.1. Introduction 70
IV.2. L'architecture du système proposé 71
IV.3. Recueil des données (Contribution 1) 71
IV.3.1. Ensemble des données 71
IV.3.2. Profilage et formes de collecte d'informations 73
IV.3.2.1. Notion de Profil 73
IV.3.2.2. Formes de collecte d'informations 74
IV.4. Filtrage des données (Contribution 2) 76
IV.4.1. Construction du module basé sur le contenu 76
IV.4.1.1. Principe de fonctionnement 76
IV.4.1.2. Dataset et prétraitement 79
IV.4.1.3. Méthodes et modèles d'apprentissage 79
IV.4.1.4. Entraînement et résultats 83
IV.4.2. Construction du module collaboratif 85
IV.4.2.1. Principe de fonctionnement 85
IV.4.2.2. Dataset et prétraitement 90
IV.4.2.3. Méthodes et modèles d'apprentissage 90
IV.4.2.4. Entraînement et résultats 93
IV.4.3. Algorithme de filtrage hybride de l'approche proposée 94
IV.5. Interface du système proposé et visualisation des recommandations (Contribution 3) 95
IV.5.1. Interfaces du système proposé 96
IV.5.2. Visualisation des recommandations à base de profil utilisateur 99
IV.5.3. Dashboard administrateur 101
IV.6. Conclusion 102

III.1. Introduction

L'avènement du développement technologique et l'explosion des données ont considérablement transformé le domaine du tourisme et l'expérience du voyageur moderne. Ces avancées ont ouvert de nouvelles perspectives passionnantes pour améliorer la façon dont les gens planifient, vivent et se souviennent de leurs voyages, avec un rôle central joué par les systèmes de recommandation touristique. Dans ce contexte en évolution rapide, le voyageur d'aujourd'hui est confronté à une multitude d'options pour chaque aspect de son voyage, que ce soit la recherche de destinations, d'hébergements, de restaurants ou d'activités. Les données, alimentées par la technologie, permettent désormais aux voyageurs de prendre des décisions plus éclairées et de personnaliser leur expérience en fonction de leurs préférences individuelles.

Conformément aux conclusions de cette étude, notre objectif est de développer un système visant à alléger la charge de travail des agences et des intervenants du secteur touristique au Maroc, en particulier dans la région de Drâa-Tafilalet. Ce système a pour ambition d'accroître leur disponibilité tout en améliorant la qualité des services proposés aux visiteurs. De plus, il offre aux touristes et aux voyageurs une opportunité exceptionnelle de vivre une expérience de voyage à la fois agréable et enrichissante. Cette solution s'appuiera sur l'utilisation des évaluations et des profils des clients pour recommander de manière personnalisée les offres touristiques les plus pertinentes. Ainsi, les clients se verront présenter des listes d'activités, d'hébergements et d'offres qui correspondent parfaitement à leurs besoins et à leurs préférences spécifiques.

L'objectif principal de ce chapitre alors était de concrétiser l'approche théorique à travers un système fonctionnel, de décomposer les composantes essentielles de notre architecture, démontrant ainsi sa viabilité et mettant en lumière nos concepts novateurs. Il présente en détail les techniques de l'IA utilisées pour développer la solution, les données exploitées, ainsi que les jeux de données employés pour les tests et la validation des recommandations. Nous y aborderons également les résultats obtenus à la suite des différentes phases d'implémentation, avant

de conclure avec une présentation de l'interface utilisateur (UI), élément essentiel pour garantir une expérience fluide et intuitive pour les utilisateurs finaux.

IV.2. L'architecture du système proposé

L'architecture du notre système de recommandation est composé de trois parties, comme elle est illustrée sur la figure 8 : (i) profilage et recueil des données, (ii) la sélection du contenu (filtrage) qui correspond le mieux aux profiles des utilisateurs (iii) visualisation des données et validation de résultats [96].

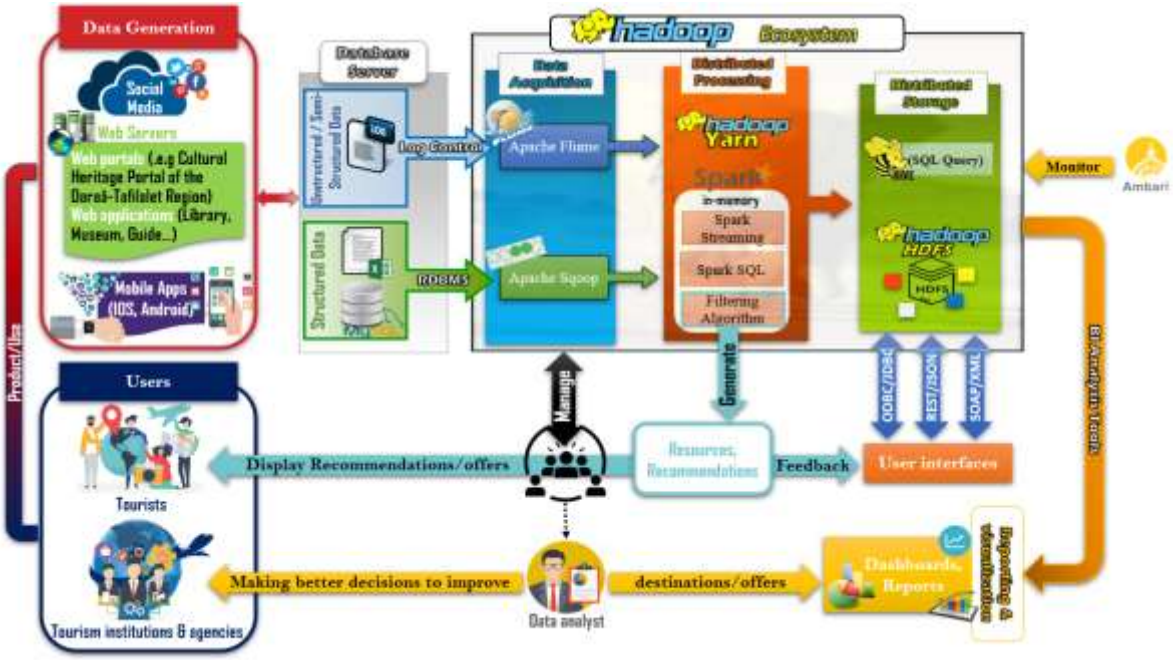


Fig. 8. L'architecture du système proposé

IV.3. Recueil des données (Contribution 1)

IV.3.1. Ensemble des données

Deux ensembles de données seront utilisés pour former et tester les différents composant du système proposé. Le premier est le célèbre ensemble de données Yelp (https://www.yelp.com/dataset/) et le second est l'ensemble de données Drâa-Tafilalet [96].

Les données Yelp constitue une précieuse ressource pour nombreuses entreprises et chercheurs, offrant un aperçu détaillé des avis et des notations des utilisateurs concernant une vaste gamme de services et

d'entreprises, notamment les restaurants, les hôtels, les commerces de détail, et bien d'autres. Ces données comprennent des informations telles que les commentaires des utilisateurs, les évaluations numériques, les attributs des entreprises (comme le type de cuisine pour les restaurants), les heures d'ouverture et les coordonnées géographiques.

Ce jeu de données a été utilisé pour tester les performances des différents modèles de prédictions des activités touristes.

L'ensemble de données contient des informations sur 150 346 entreprises dans 11 zones métropolitaines, avec 6 990 280 avis d'utilisateurs. L'ensemble de données est fourni dans six fichiers JSON : entreprise, utilisateur, avis, conseil, check-in et photo. Chaque fichier comprend un objet JSON par ligne, donnant des informations sur un seul type d'objet. Chaque fichier comprend un seul objet JSON par ligne, fournissant des informations sur un type d'objet spécifique. Le fichier « business.json » contient des données sur les catégories, le nom de l'entreprise, la localisation, le type d'entreprise, ainsi qu'un identifiant et un score uniques. Le fichier « user.json » contient les métadonnées pertinentes de l'utilisateur et la matrice de son réseau social. Le fichier « review.json » comprend un commentaire, une note, une date, un identifiant unique de l'utilisateur et un identifiant unique de l'entreprise. Le fichier « checkin.json » contient les données d'enregistrement pour une entreprise donnée.



Fig. 9. Les données Yelp

Le système proposé est également testé sur la base de données provenant de la zone de Drâa-Tafilalet. Cette zone est l'une des plus célèbres destinations touristiques du Maroc et se distingue par un grand nombre d'attractions touristiques. Le jeu de données Drâa-Tafilalet contient 1737 avis notés par 876 utilisateurs sur 70 objets touristiques. À l'aide des outils Big Data, les chercheurs ont collecté à partir de diverses sources d'information des informations concernant les sites touristiques,

les clients et même les avis et notations. Une fois les données collectées, les résultats sont automatiquement enregistrés dans des fichiers JSON, en utilisant la même structure que l'ensemble de données Yelp.

IV.3.2. Profilage et formes de collecte d'informations

IV.3.2.1. Notion de Profil

De façon générale, le profil d'un objet est un ensemble de caractéristiques permettant de l'identifier ou de le représenter. Deux types de profils peuvent être exploités dans les systèmes de recommandation, correspondant aux deux entités utilisées dans ces systèmes : l'utilisateur et l'item ;

Le profil de document/item : il correspond à la description d'un document qui est souvent réduite, en RI ou FI, à une liste de mots-clés pondérés décrivant le contenu sémantique du document. Les mots-clés et leurs poids sont obtenus en général par une opération d'indexation [97]. Plusieurs travaux permettent actuellement de décrire les documents en utilisant également d'autres critères que ceux liés à leur contenu effectif. On peut citer par exemple les métadonnées du Dublin Core13 pour la description de ressources : Titre, Auteur, Sujet, etc. Nous pouvons également citer les travaux de Lainé-Cruzel [98] qui permettent de définir des propriétés liées à l'ensemble d'un document (profession de l'auteur, type de document, etc.) ainsi que celles relatives à des parties de documents (type d'unité documentaire, forme discursive, style, etc.) afin de restreindre les documents pertinents (du point de vue du sujet dont ils traitent) aux seuls documents exploitables et réellement utilisables.

Le profil utilisateur : il s'agit de la description des caractéristiques d'un utilisateur (données démographiques, centres d'intérêt, préférences, etc.). Plusieurs approches d'acquisition des éléments d'un profil utilisateur existent et peuvent être regroupées en approches manuelles et en approches automatiques ou semi-automatiques [1]. Parmi les approches automatiques ou semi-automatiques, nous pouvons distinguer le 'profiling' [30] et les approches par stéréotypes [99]. Le profiling consiste à scruter, enregistrer et analyser les actions et successions d'actions d'un utilisateur lors de différentes sessions de recherche pour déterminer son profil. En

revanche, l'approche par stéréotype est fondée sur l'identification de groupes d'utilisateurs et la détermination des caractéristiques clés de chaque groupe ;

IV.3.2.2. Formes de collecte d'informations

A. Collecte des données explicites

Dans le cadre de la collecte des données, nous avons aussi mis en place un autre processus de collecte de données tant implicites qu'explicites pour enrichir et affiner nos modèles de prédiction. Les données explicites sont directement fournies par les utilisateurs, telles que leurs avis, évaluations des différents sites touristiques, mais aussi des préférences déclarées lors de l'inscription des utilisateurs. Ces informations sont cruciales car elles reflètent directement les expériences et les opinions des utilisateurs sur les destinations touristiques. Les utilisateurs expriment leurs goûts et leurs préférences à travers des notes et des commentaires, ce qui nous permet d'obtenir une vue claire et détaillée de ce qu'ils recherchent en matière de voyage. Ces informations sont collectées entre autres lors de l'inscription des utilisateurs à travers les formulaires d'inscription suivantes :

Ce formulaire permet de récupérer des informations pertinentes telles que l'âge de l'utilisateur, sa localisation, son pays et sa ville de naissance.

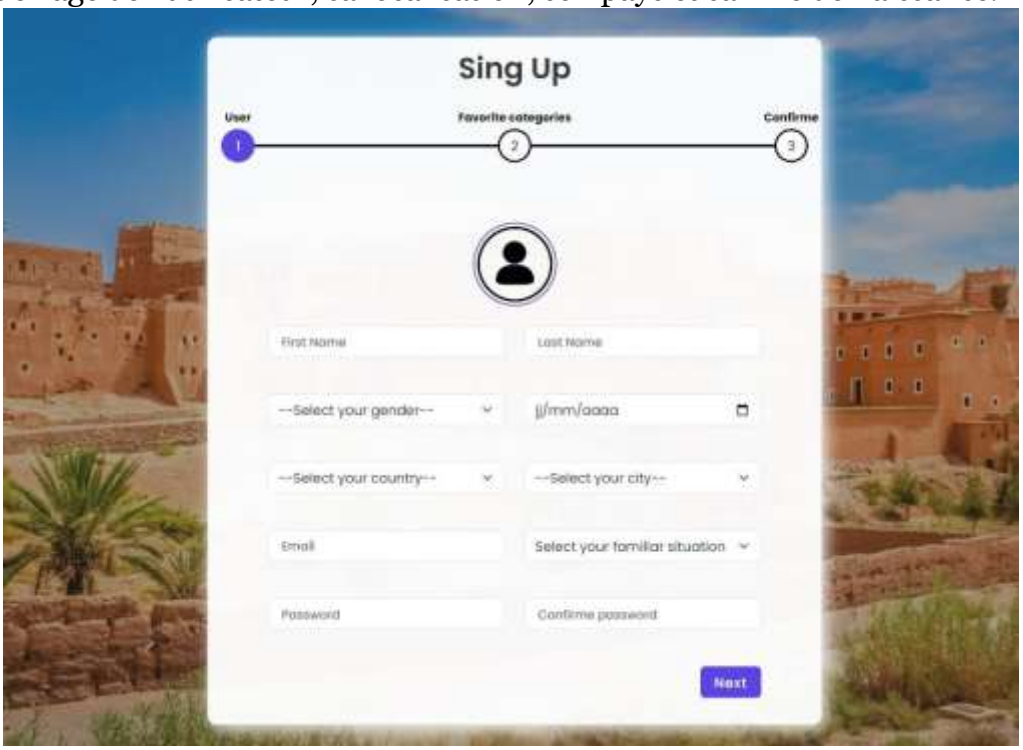


Fig. 10. Formulaire 1 de saisie de données utilisateurs

Nous avons aussi la deuxième page d'inscription. Elle permet à l'utilisateur de choisir les catégories qu'il aime ou préfère. Cette étape nous permet d'obtenir une connaissance préalable des préférences de l'utilisateur malgré l'absence d'activité à l'intérieur de la plateforme, ce qui permet de fournir des recommandations appropriées et d'éviter les problèmes de démarrage à froid.

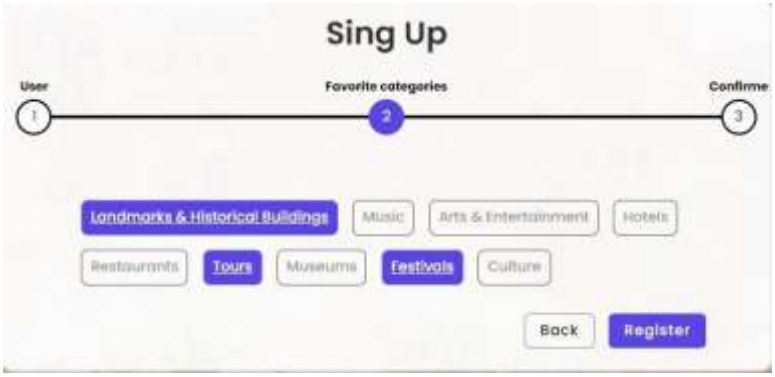


Fig. 11. Formulaire 2 de saisie de données utilisateurs

B. Collecte des données implicites

En parallèle, nous avons également collecté des données implicites, qui sont générées par les comportements et les interactions des utilisateurs sur notre plateforme. Cela inclut des informations telles que les pages visitées, le temps passé sur chaque page, les recherches effectuées et les articles consultés. Ces données, bien que moins directes, offrent des aperçus précieux sur les préférences sous-jacentes des utilisateurs, souvent révélées par leurs actions et comportements en ligne.

Ces données ont été pu récolter à l'aide de google Analytics. Car elle fournit un ensemble de fonctionnalités propre à la collecte des données utilisateurs sur les applications web. Et présente aussi, une interface utilisateur user-friendly qui permet d'avoir des interactions en temps réel au niveau de la plateforme.

En combinant ces types de données, nous sommes en mesure de développer des modèles de recommandations plus sophistiqués et personnalisés. Les données explicites nous fournissent des indicateurs clairs des préférences des utilisateurs, tandis que les données implicites nous aident à comprendre les tendances comportementales et les intérêts émergents. En intégrant ces informations dans nos modèles de prédiction, nous pouvons mieux personnaliser les recommandations touristiques en

fonction des goûts individuels et des tendances observées au niveau global. Cette approche holistique nous permet de fournir des recommandations plus précises et engageantes, offrant ainsi une expérience utilisateur améliorée et répondant aux besoins diversifiés de notre audience.

IV.4. Filtrage des données (Contribution 2)

IV.4.1. Construction du module basé sur le contenu

IV.4.1.1. Principe de fonctionnement

La recommandation basée sur le contenu est une technique qui utilise les informations descriptives des items (articles, services, produits, etc.) pour générer des recommandations personnalisées à un utilisateur. En se basant sur les caractéristiques spécifiques des items et le profil de l'utilisateur, cette méthode trouve des items similaires à ceux que l'utilisateur a déjà appréciés.

Le système compare le profil utilisateur aux métadonnées (informations descriptives) des nouveaux items disponibles pour trouver ceux qui sont les plus similaires à ce que l'utilisateur a déjà aimé.

Il est à signaler que ces mots-clés sont généralement soit extraits sur la base d'une indexation automatique, soit attribués manuellement. Pour ce qui est des systèmes de recommandation touristiques, le contenu est plutôt structuré et représenté par des métadonnées définies au préalable et valables pour tous les items [8].

Ainsi, les approches de ce type considèrent les mots comme de simples séquences non ordonnées. Le principe général est de représenter un corpus, contenant un ensemble de documents formés de descripteurs (souvent des mots) par une matrice de type documents/descripteurs. Cette matrice va être constituée de m × n cellules telle que représentée dans la figure 12.

Formellement, tout document est représenté par un vecteur poids sur des termes, où chaque poids indique le degré d'association entre le document et le terme. Soit $D = \{d_1, d_2, ..., d_m\}$ dénotant un ensemble de documents ou corpus, et $T = \{t_1, t_2, ..., t_n\}$ le dictionnaire, ou l'ensemble des mots du corpus. T est obtenu en appliquant des opérations de traitement

du langage naturel, comme l'atomisation (tokenization), l'élimination des mots vides de sens, et la troncature (stemming) [19]. Chaque document dj est représenté par un vecteur dans un espace vectoriel à n dimensions, tel que dj = {w_{1j}, w_{2j}, ..., w_{nj}}, où w_{kj} est le poids du terme t_k dans le document d_j.

$$d_j \rightarrow \begin{pmatrix} W_{1,1} & \cdots & W_{1,n} \\ \vdots & \ddots & \vdots \\ W_{m,1} & \cdots & W_{m,n} \end{pmatrix} \quad t_i^T \downarrow$$

Fig. 12. Représentation vectorielle d'un corpus

Cette représentation est fréquemment nommée la représentation ou matrice de Salton et fut pour la première fois utilisée dans le cadre du système de recherche d'information SMART [100].

La représentation de documents dans le modèle d'espace vectoriel fait apparaître deux difficultés : la pondération des termes et la mesure de similarité des vecteurs caractéristiques.

Le schéma de pondération de terme le plus communément utilisé est la pondération TF-IDF ("Term Frequency-Inverse Document Frequency") [101]. Il s'agit d'une mesure statistique qui permet d'évaluer l'importance d'un mot dans un document ou dans un item faisant partie d'une collection ou d'un corpus [8].

Le principe de cette mesure est que les mots-clés paraissant dans beaucoup d'items ne permettent pas de distinguer un item pertinent d'un autre qui ne l'est pas. Or, les mots-clés qui sont rares et communs à quelques items définissent plus la similarité de contenu ainsi que la pertinence d'un item par rapport à un autre.

- **TF signifie 'Term Frequency'**, et désigne une mesure proportionnelle à la fréquence du terme dans le document (pondération locale), l'idée sous-jacente est que plus un terme est fréquent dans un document, plus il est important dans la description de ce document [102].

- **IDF signifie 'Inverted Document Frequency' :** ce facteur mesure l'importance d'un terme dans toute la collection. Un terme qui apparaît souvent dans la base documentaire ne doit pas avoir le même impact qu'un

terme moins fréquent [103].

Ces conjectures sont bien illustrées par la fonction TF-IDF :

$$\mathrm{TFIDF}(t_k, d_j) = \mathrm{TF}(t_k, d_j).\log\frac{|D|}{|d_k|}$$

Où $|D|$ dénote le nombre de documents dans le corpus, et $|d_k|$ représente le nombre de documents de la collection dans lesquels le terme t_k apparait au moins une fois, avec :

$$\mathrm{TF}(t_k, d_j) = \frac{n_{k,j}}{\sum_K n_{k,j}}$$

$n_{k,j}$: Le nombre d'apparition de t_k dans le document d_j.

K : Le nombre total de termes dans d_j.

$$w_{tk,\mathrm{dj}} = \frac{\mathrm{TFIDF}(t_k,d_j)}{\sqrt{\sum_{k=1}^{K}(\mathrm{TFIDF}(t_k,d_j))^2}}$$

Après la pondération des termes, il faut définir une mesure de similarité des vecteurs caractéristiques. Cette mesure de similarité est requise pour déterminer la proximité entre le profil utilisateur et les items potentiels. Il existe de nombreuses mesures de similarité, mais la plus largement utilisée est la similarité cosinus :

$$\boldsymbol{User_Profil}\,(u) = \{(t_j^u, w_j^u)\},\ j = 1 \dots k$$

$$\boldsymbol{Item_Contenu}\,(d) = \{(t_i^d, w_i^d)\},\ i = 1 \dots n$$

$$Sim\,(u,d) = \frac{\sum_{k=1}^{|K|} w_{k,u} \times w_{k,\mathrm{dj}}}{\sqrt{\sum_{k=1}^{|K|} {w_{k,u}}^2 \times \sum_{k=1}^{|K|} {w_{k,\mathrm{dj}}}^2}}$$

Avec :

|K| : est le nombre de termes présentés à la fois dans les deux profiles utilisateur et item.

$w_{k,u}$: Le poids du terme t_k dans le vecteur du profile utilisateur u.

w_{k,d_j} : Le poids du terme t_k dans le vecteur d'item d_j.

Le profil utilisateur est dynamique et s'améliore au fil du temps grâce aux retours (feedbacks) fournis par l'utilisateur. Par exemple, si l'utilisateur indique qu'il a aimé ou n'a pas aimé une recommandation, cette information est utilisée pour ajuster le profil et améliorer la qualité des recommandations futures.

IV.4.1.2. Dataset et prétraitement

Tout d'abord, les catégories ou mots-clés de chaque item ont été utilisé pour les représenter sous forme de vecteur numérique à l'aide de TF-IDF.

La construction du profil utilisateur basée sur les vecteur TF-IDF des items, Dans ce processus, chaque utilisateur se voit attribuer un profil unique qui reflète ses préférences et ses intérêts. Pour cela, on commence par analyser toutes les interactions passées de l'utilisateur avec des items, en utilisant des vecteurs TF-IDF pour mesurer l'importance des mots clés dans chaque interaction. L'utilisateur attribue ensuite des notes à ces éléments, qui sont utilisées comme facteurs de pondération. Le profil utilisateur final est obtenu en calculant une moyenne pondérée des vecteurs TF-IDF des éléments notés, ce qui reflète les intérêts spécifiques de l'utilisateur.

IV.4.1.3. Méthodes et modèles d'apprentissage

A. La régression logistique

La régression logistique est un modèle d'apprentissage automatique supervisé qui est Utilisé pour prédire une variable dépendante binaire ou catégorielle en fonction d'un ensemble de variables indépendantes, Basée sur la fonction logistique, également appelée fonction sigmoïde, qui permet de modéliser la probabilité d'appartenance à une classe.

Estime les coefficients qui maximisent la probabilité de la variable dépendante donnée les variables indépendantes.

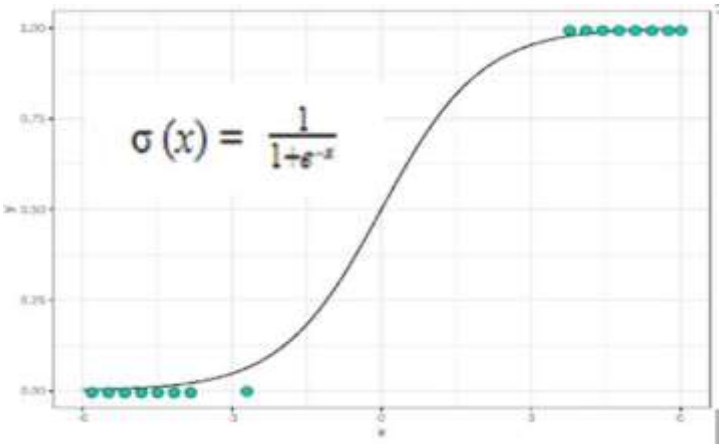


Fig. 13. La fonction sigmoïde

B. L'arbre de décision

L'arbre de décision est une technique d'apprentissage automatique et d'analyse de données utilisée pour prendre des décisions en formulant des règles à partir des données. Il est largement utilisé dans la classification et la régression, ainsi que dans la prise de décision. Structure de modèle qui représente des décisions et des actions sous forme d'un arbre inversé, où chaque nœud représente une caractéristique (ou attribut) et chaque branche représente une règle de décision.

Construit en sélectionnant de manière itérative les caractéristiques les plus importantes pour diviser les données en sous-ensembles homogènes.

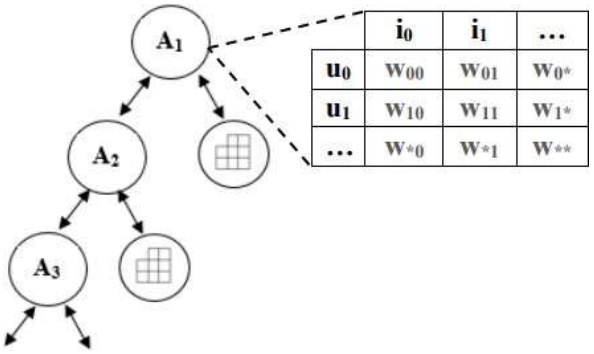


Fig. 14. L'arbre de décision

C. Multinomial Naïve Bayes

Le modèle de classification Multinomial Naïve Bayes est un algorithme d'apprentissage automatique couramment utilisé pour résoudre des problèmes de classification de textes, notamment dans le domaine du traitement du langage naturel (NLP) et de la fouille de textes. Il est basé sur le théorème de Bayes et constitue une extension du modèle Naïve Bayes, adapté aux données catégoriques ou discrètes.

$$\mathbf{P(B/A)} = \frac{\mathbf{P(A \cap B)}}{\mathbf{P(A)}}$$

Le modèle Multinomial Naïve Bayes est particulièrement approprié pour les tâches de classification de textes où les caractéristiques sont des comptages ou des fréquences de mots, de phrases ou de termes. Il est souvent utilisé dans des applications telles que la classification de documents, la détection de spam, la catégorisation de textes, la

classification de sentiments, etc.

D. Support Vector Machine

La machine à vecteurs de support (SVM) est un algorithme d'apprentissage automatique utilisé pour la classification et la régression. Elle est largement utilisée dans divers domaines, y compris la classification d'images, la classification de textes, la bio-informatique et bien d'autres. L'objectif principal des SVM est de trouver un hyperplan optimal dans un espace multidimensionnel qui sépare les exemples de différentes classes avec la plus grande marge possible.

L'hyperplan est déterminé en utilisant un sous-ensemble d'exemples d'entraînement appelés vecteurs de support, qui sont les points les plus proches de la frontière de décision.

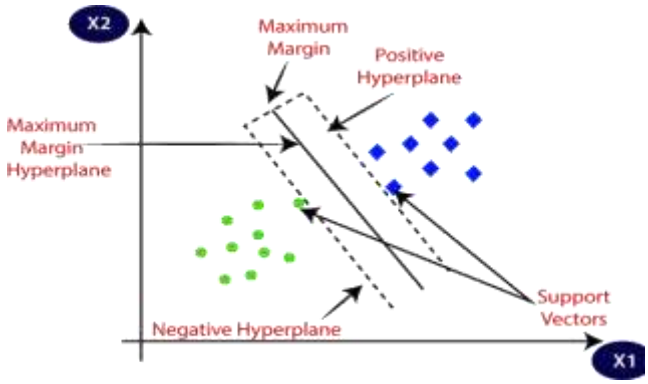


Fig. 15. Principe de la méthode SVM

E. Réseaux de neurones artificiels

Les réseaux de neurones, également appelés réseaux de neurones artificiels (ANN) sont au cœur des algorithmes d'apprentissage en profondeur. Leur nom et leur structure sont inspirés du cerveau humain, imitant la façon dont les neurones biologiques se signalent les uns aux autres.

Les réseaux de neurones artificiels (ANN) sont composés d'une couche de nœuds, contenant une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Chaque nœud, ou neurone artificiel, se connecte à un autre et a un poids et un seuil associés. Si la sortie d'un nœud individuel est supérieure à la valeur de seuil spécifiée, ce nœud est activé, envoyant des données à la couche suivante du réseau. Sinon, aucune donnée n'est transmise à la couche

suivante du réseau.

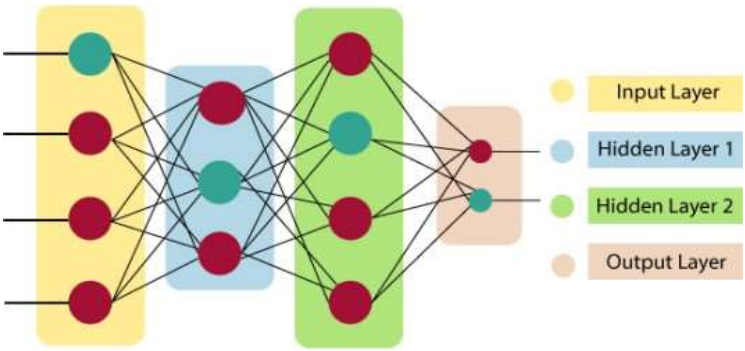


Fig. 16. Structure de ANN

IV.4.1.4. Entraînement et résultats

Dans cette partie, Cinque modèles ont étés conçus pour prédire les avis des utilisateurs : régression logistique, L'arbre de décision, Multinomial naive bayes, SVM, et ANN, pour implémenter ces modèles, les packages tels que Scikit-Learn ver 1.1.1, Tensorflow ver.2.3.0 et Keras ver. 2.4.3 et le langage de programmation Python ont été utilisé.

Après de nombreuses expériences et tests, nous sommes arrivés aux hyperparamètres optimaux de chaque modèle à l'idée de l'algorithme GridSearchCV avec une validation croisée (en divisant les données en 5 ensembles K = 5) , pour SVM, nous avons utilisé le noyau de la fonction de base radiale (rbf), gamma = 'scale' et paramètre de régularisation C = 20 et pour régression logistique nous avons utilisé inverse de la force de régularisation = 1000, penalty = l2 (ridge régularisation) et pour L'arbre de décision nous avons utilisé ccp_alpha = 0, criterion = 'entropy' et pour Multinomial naive bayes nous avons utilisé alpha = 10, fit_prior = True. Et pour l'architecture du modèle ANN utilisée dans cette étude est illustrée à la figure 17.

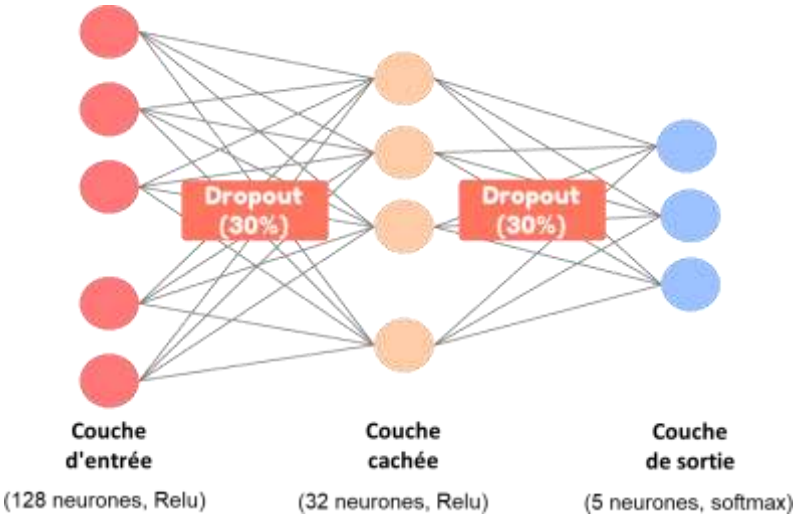


Fig. 17. Architecture de ANN

Ce modèle suit une structure séquentielle, où les couches sont empilées les unes sur les autres. Tout d'abord, il y a une couche d'entrée spécifiant la forme de l'entrée en fonction du nombre de caractéristiques dans vos données d'entraînement. Ensuite, une couche dense avec 128 neurones et une fonction d'activation 'relu' est utilisée pour apprendre des représentations complexes

des données. Après chaque couche dense, une couche de dropout avec un taux de 0,3 est ajoutée pour régulariser le modèle en désactivant aléatoirement certains neurones pendant l'entraînement. Une autre couche dense avec 32 neurones suit, encore une couche de dropout, puis la couche de sortie avec 5 neurones, utilisant 'softmax' comme fonction d'activation prédire les avis des utilisateurs {1, 2, 3, 4, 5}.

Après apprentissage et tests, et à travers les matrices de confusion des résultats des tests pour chaque modèle, nous concluons les résultats représentés dans le tableau 2 suivant :

	Precision	Recall	F1 Score
LogisticRegression	65.66%	68.78%	64.31%
Decision Tree	92.79%	92.83%	92.80%
MultinomialNB	35.14%	59.28%	44.13%
SVM	93.57%	93.53%	93.39%
Model_ANN	93.79%	93.83%	92.80%

Tableau 2. Résultats de tests pour filtrage par contenue

L'analyse des performances de nos modèles révèle des résultats variables. Parmi les modèles évalués, le modèle de Deep Learning se distingue avec une précision de 93,79%, un rappel de 93,83% et un score F1 de 93,80%. Cela en fait le choix le plus solide pour cette tâche de classification. Les modèles SVM (Support Vector Machine) et l'arbre de décision affichent également d'excellentes performances, avec des scores de précision, de rappel et de score F1 proches de 93%. En revanche, la régression logistique présente des performances plus modestes, avec une précision de 65,66% et un score F1 de 64,31%. Le modèle MultinomialNB (Naïve Bayes) montre également des résultats mitigés, avec une précision de seulement 35,14% et un score F1 de 44,13%. En conclusion, pour cette tâche de classification multiclasse, les modèles d'arbre de décision, SVM et de Deep Learning se démarquent comme les choix les plus performants, tandis que la régression logistique et Naïve

Bayes semblent moins adaptés.

IV.4.2. Construction du module collaboratif

IV.4.2.1. Principe de fonctionnement

Il fonctionne sur le principe que les personnes qui étaient d'accord dans le passé le seront également dans le futur. Les concepts mathématiques clés derrière le filtrage collaboratif sont les mesures de similarité (FC basé sur la mémoire) et la factorisation matricielle (FC basé sur le modèle).

A. FC basé sur le voisinage des utilisateurs

Nous pouvons distinguer deux phases essentielles en FC basé sur le voisinage utilisateurs : la phase d'identification du voisinage et la phase de calcul des prédictions.

En se basant sur le profil d'un utilisateur u, le système recherche les utilisateurs v qui lui sont les plus similaires. Plusieurs mesures ont été exploitées dans le cadre du FC basé sur la mémoire dans le but d'évaluer les similarités d'appréciations entre utilisateurs et identifier les utilisateurs voisins. Parmi ces mesures nous pouvons citer : le coefficient de corrélation de Pearson [104], la mesure basée sur le cosinus [105], la corrélation de Spearman [3], "Mean squared difference" (qui représente une mesure de dissimilarité) [7], etc.

Les mesures les plus populaires sont la mesure basée sur le cosinus et le coefficient de corrélation de Pearson.

- Cosinus des vecteurs :

Dans cette méthode les utilisateurs sont considérés comme deux vecteurs de même origine dans un espace de m dimensions, m est égal au nombre d'items évalués par les deux utilisateurs. Plus deux utilisateurs sont similaires, plus l'angle entre leurs vecteurs respectifs est petit. Empiriquement, la similarité entre deux utilisateurs u et v est calculée par la formule du Cosinus suivante :

$$Sim_{Cosinus}\,(u,v) = \frac{\sum_{i=1}^{m} r_{u,i} \times r_{v,i}}{\sqrt{\sum_{i=1}^{m} {r_{u,i}}^2 \times \sum_{i=1}^{m} {r_{v,i}}^2}}$$

m : est le nombre d'items ayant été voté à la fois par l'utilisateur u et v.

$r_{u,i}$ $(resp.\ r_{v,i})$: est la note de l'utilisateur u (resp. v) pour l'item i.

- Corrélation de Pearson (Cosinus Centré)

La corrélation de Pearson est une méthode issue des statistiques. La valeur de cette mesure varie entre -1 (les deux utilisateurs sont complètement opposés) et +1 (les deux utilisateurs sont complètement corrélés). Sa formule générale est la suivante :

$$Sim_{Pearson}\ (u,v) = \frac{\sum_{i=1}^{m}(r_{u,i}-\overline{r_u})\times(r_{v,i}-\overline{r_v})}{\sqrt{\sum_{i=1}^{m}\ (r_{u,i}-\overline{r_u})^2 \times \sum_{i=1}^{m}\ (r_{v,i}-\overline{r_v})^{\ 2}}}$$

m : est le nombre d'items ayant été voté à la fois par l'utilisateur u et v.

$r_{u,i}$ $(resp.\ r_{v,i})$: est la note de l'utilisateur u (resp. v) pour l'item i.

$\overline{r_u}$ $(resp.\ \overline{r_v})$: est la moyenne des votes de l'utilisateur u (resp. v).

L'inconvénient des deux mesures Pearson et cosinus, est que le calcul des similarités devient non fiable voire impossible, lorsque le système dispose de peu d'items co-notés entre utilisateurs. Afin de pallier ce problème, certaines extensions ont été proposées notamment par [29], telle que "La note par défaut " consistant à attribuer une valeur par défaut à une note manquante. Mais l'enjeu à ce niveau est de savoir quelle valeur par défaut choisir (appréciation positive, négative ou bien neutre) et d'évaluer son impact sur le calcul des similarités.

Une fois que toutes les similarités de l'utilisateur cible u par rapport aux autres utilisateurs sont calculées et que les N utilisateurs les plus similaires qui constituent le voisinage de cet utilisateur cible sont définis, La note $r_{u,i}$ de l'utilisateur u sur l'item i peut être prédite par la moyenne des notes de ces voisins:

$$\widehat{r_{u,i}} = \frac{1}{|N(u)|} \sum_{v\in N} r_{v,i}$$

N(u) : est l'ensemble de voisins de l'utilisateur u.

Un problème avec cette méthode par moyenne est qu'elle ne prend pas en compte le fait que les voisins peuvent avoir des niveaux différents de similarité. En effet, on peut prédire que la note de l'utilisateur u est plus à

même de se rapprocher de la note des voisins avec une plus grande similarité. Ainsi, une solution à ce problème est de pondérer la contribution de chaque voisin par leur similarité à u. Cependant, si la somme de ces poids ne fait pas 1, les notes prédites peuvent être en dehors des valeurs autorisées. De ce fait, il est courant de normaliser ces poids de telle sorte que la note prédite devienne :

$$\widehat{r_{u,i}} = \frac{\sum_{v \in N(u)} S_{u,v} \times r_{v,i}}{\sum_{v \in N(u)} S_{u,v}} \quad \textit{Avec } S_{u,v} = \textit{Sim } (u,v)$$

Néanmoins, un problème majeur du FC est la notation des utilisateurs. En effet, si un utilisateur considère que la perfection n'existe pas, il n'affectera jamais la note maximale à un item et donc repartir ses notes de 1 à 4 (si les notes possibles sont de 1 à 5). À l'inverse, un utilisateur différent peut, s'il n'aime pas noter trop sévèrement, repartir les notes qu'il attribue de 2 à 5. Pour pallier ce problème, la moyenne des notes de l'utilisateur u est introduite à la formule de la note prédite :

$$\widehat{r_{u,i}} = \overline{r_u} + \frac{\sum_{v \in N(u)} S_{u,v} \times (r_{v,i} - \overline{r_u})}{\sum_{v \in N(u)} S_{u,v}} \quad \textit{Avec } S_{u,v} = \textit{Sim } (u,v)$$

B. FC basé sur le voisinage des items

Alors que les méthodes basées sur le voisinage utilisateur s'appuient sur l'avis d'utilisateurs partageant les mêmes idées pour prédire une note, les approches basées sur les items [9] [106] prédisent la note d'un utilisateur u pour un item i en se basant sur les notes de u pour des items similaires à i. Dans de telles approches, deux items sont similaires si plusieurs utilisateurs du système les ont notés d'une manière similaire.

En effet, au lieu de mesurer la corrélation entre des utilisateurs, les ratings sont utilisés pour mesurer la corrélation entre les items, en s'aidant toujours de coefficients de corrélation de Pearson, mais cette fois-ci appliqué aux items :

$$Sim_{Pearson}\ (i,j) = \frac{\sum_{u=1}^{n} (r_{u,i} - \overline{r_i}) \times (r_{u,j} - \overline{r_j})}{\sqrt{\sum_{u=1}^{n} (r_{u,i} - \overline{r_i})^2 \times \sum_{u=1}^{n} (r_{u,j} - \overline{r_j})^2}}$$

n : est nombre d'utilisateurs qui ont votés pour les deux items.

$r_{u,i}$ $(resp.\ r_{u,j})$: est la note de l'utilisateur u pour l'item i (resp. j).

$\overline{r_i}$ $(resp.\ \overline{r_j})$: est la moyenne des votes attribués à l'item i (resp. j).

Une fois que la similarité parmi les items a été calculée, la prochaine étape est de prévoir pour l'utilisateur cible u, une valeur pour l'item i. Une manière commune est de capturer comment l'utilisateur a évalué les items similaires [32]. La valeur prévue est basée sur la somme pondérée des estimations de l'utilisateur et peut être calculée à l'aide de la formule suivante :

$$\widehat{r_{u,i}} = \overline{r_i} + \frac{\sum_{j \in N(i)} S_{i,j} \times (r_{u,j} - \overline{r_i})}{\sum_{j \in N(i)} S_{i,j}} \quad Avec\ S_{i,j} = Sim(i,j)$$

N(i) : est l'ensemble de voisins de l'item i.

Une fois les prédictions calculées, le système de FC recommande à l'utilisateur actif les items ayant les valeurs de prédiction les plus élevées.

C. Modèle de factorisation matricielle

La factorisation matricielle est largement utilisée dans les systèmes de recommandation pour prédire les préférences des utilisateurs et recommander des éléments (films, produits, livres, etc.) qu'ils pourraient aimer. L'idée est de décomposer une matrice d'utilisateurs-items en matrices plus petites représentant les caractéristiques des utilisateurs et des items. Voici comment cela fonctionne dans ce contexte :

On commence avec une matrice R de taille m × n, où m est le nombre d'utilisateurs et n le nombre d'items. Chaque élément $r_{i,j}$ de la matrice représente l'interaction (comme une note ou une évaluation) de l'utilisateur i avec l'item j. La matrice est généralement sparse (sparse matrix), c'est-à-dire qu'elle contient de nombreuses valeurs manquantes, correspondant aux items non évalués par les utilisateurs.

L'objectif est de prédire les valeurs manquantes dans la matrice R, c'est-à-dire estimer les notes des items que les utilisateurs n'ont pas encore évalués. Pour ce faire, R est approximée par le produit de deux matrices de rang inférieur :

$$\boldsymbol{R \approx PQ^T}$$

Où :

P est une matrice m×k représentant les **caractéristiques latentes des utilisateurs**, où k est le nombre de facteurs latents.

Q est une matrice n×k représentant les **caractéristiques latentes des items**.

Les lignes de P et Q sont des vecteurs de caractéristiques latentes, et le produit scalaire entre les vecteurs des utilisateurs et des items donne une estimation de la note.

Pour estimer les matrices P et Q, on minimise la fonction de coût qui mesure l'erreur entre les notes réelles et les notes prédites :

$$\mathrm{J} = \sum_{(\mathrm{i,j})\in \text{observé}} (r_{ij} - {p_i}^T q_j)^2 + \lambda(\|\mathrm{P}\|^2 + \|\mathrm{Q}\|^2)$$

Où :

(i, j) ∈ observé Signifie que la somme est effectuée sur l'ensemble des notes connues.

Le terme $\lambda(\|\mathrm{P}\|^2 + \|\mathrm{Q}\|^2)$ est une régularisation pour éviter le surapprentissage.

La minimisation peut être effectuée à l'aide de la descente de gradient stochastique (SGD) ou de l'alternating least squares (ALS).

Une fois les matrices P et Q apprises, les valeurs manquantes dans la matrice R sont estimées par le produit scalaire $\widehat{r_{u,\imath}} = {p_i}^T q_j$, où p_i est le vecteur des caractéristiques latentes de l'utilisateur i, et q_j est celui de l'item j.

Les recommandations sont faites en suggérant à chaque utilisateur les items qui ont les plus fortes prédictions de notes.

Plusieurs méthodes de réduction de la dimension ont été utilisées pour les algorithmes de recommandation basés sur un modèle. On peut citer, parmi les plus abouties, la décomposition en valeurs singulières (SVD), SVD++, et la factorisation en matrices non négatives (NMF).

IV.4.2.2. Dataset et prétraitement

Pour ce type de recommandation, seul le tableau des notes (Reviews) peut être utilisé, avec exactement trois colonnes : la colonne «id_user », qui contient le numéro d'identification de l'utilisateur qui a créé l'avis, la colonne « id_business », qui contient le numéro d'identification du site touristique ou item et la colonne « stars », qui contient la note.

Après cela, nous convertissons les colonnes « id_user » et « id_business » d'une chaîne de caractère en chiffres en utilisant la fonction LabelEncoder de scikit-learn. Et Nous ajustons également les données au format de Surprise, puis divise les données en ensembles d'entraînement (80% = 12802) et de test (20% = 3200).

IV.4.2.3. Méthodes et modèles d'apprentissage

A. K plus proches voisins (KNN)

L'algorithme KNN (ou K-Nearest Neighbors) est un algorithme de Machine Learning qui appartient à la classe des algorithmes d'apprentissage supervisé, Il fonctionne en trouvant les K voisins les plus proches d'un point dans un espace de caractéristiques, puis en utilisant ces voisins pour la classification (dans le cas de problèmes de classification) ou pour la prédiction de valeurs (dans le cas de problèmes de régression). Le choix de K est crucial et dépend du problème spécifique. KNN est simple à comprendre mais sensible à la taille de l'ensemble de données et à la normalisation des caractéristiques.

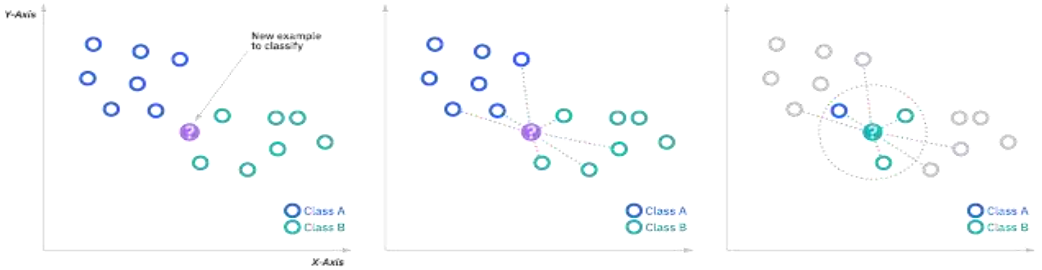


Fig. 18. Principe de la méthode des k plus proches voisins

B. Décomposition en Valeurs Singulières (SVD)

La SVD (Singular Value Decomposition en anglais), est une technique fondamentale en algèbre linéaire et en analyse des données. Elle est largement

utilisée dans de nombreux domaines, y compris l'apprentissage automatique, la recommandation, la compression d'images, le traitement du signal et bien d'autres. La SVD permet de décomposer une matrice en trois autres matrices, ce qui permet de représenter et de comprendre les données d'une manière plus simple et plus informative.

La SVD décompose une matrice M de dimensions m x n en trois matrices distinctes :

- Matrice U : Une matrice orthogonale de dimension m x m, où les colonnes sont appelées vecteurs singuliers gauche.
- Matrice Σ (Sigma) : Une matrice diagonale de dimension m x n, où les éléments sur la diagonale sont les valeurs singulières de A, classées par ordre décroissant.
- Matrice V^T (V transposé) : Une matrice orthogonale de dimension n x n, où les lignes sont appelées vecteurs singuliers droits.

$$\mathbf{M} = \boldsymbol{U} * \boldsymbol{\Sigma} * \boldsymbol{V}^{\mathrm{T}}$$

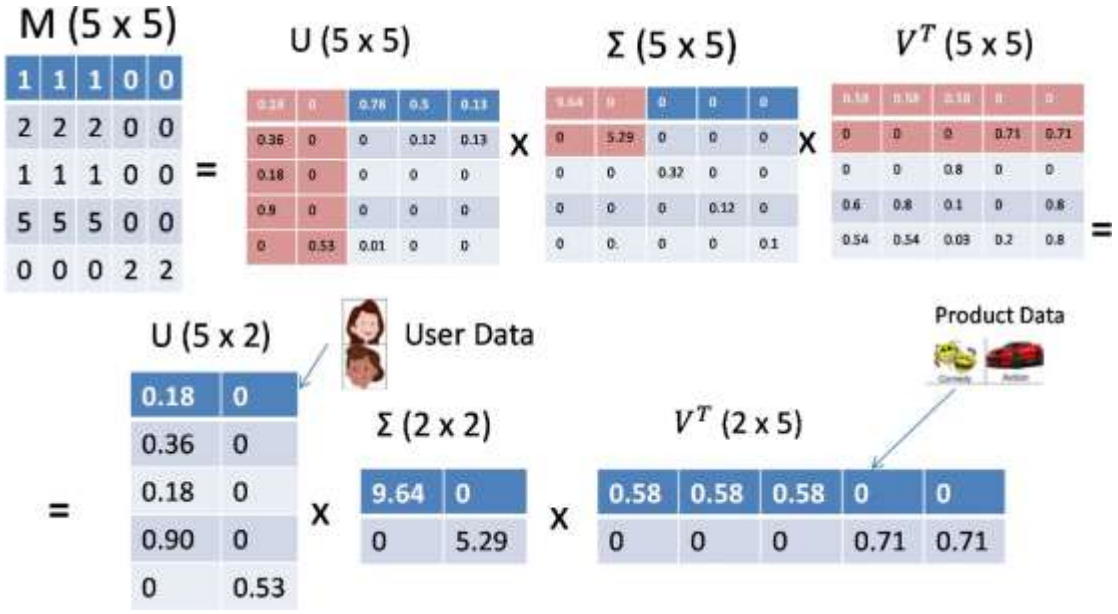


Fig. 19. Application de SVD

C. Décomposition en Valeurs Singulières plus (SVD++)

La SVD++ (Singular Value Decomposition Plus) est une amélioration de la technique de Décomposition en Valeurs Singulières (SVD) qui a été spécifiquement développée pour les systèmes de recommandation

collaboratifs. Elle vise à résoudre un problème commun dans les systèmes de recommandation, à savoir la "sparsité" des données, c'est-à-dire le fait que la plupart des utilisateurs notent et interagissent uniquement avec un petit sous-ensemble des éléments disponibles.

La SVD++ étend la SVD traditionnelle en prenant en compte l'ensemble des données d'interaction utilisateur-élément, y compris les données implicites (comme le simple fait de consulter un élément sans laisser de note explicite). Elle repose sur l'idée que les utilisateurs qui ont interagi avec un élément de manière explicite ou implicite ont potentiellement une affinité plus forte avec cet élément que ceux qui n'ont eu aucune interaction.

D. La factorisation en matrices non négatives (NMF)

La factorisation en matrices non négatives est une méthode générale de décomposition matricielle, introduite par Lee et Seung (1999). Elle permet d'approximer toute matrice V de taille (n × m) et dont les éléments sont tous positifs, grâce à une décomposition de la forme $V \approx W H$, où W et H sont des matrices (n × k) et (k × m). La matrice V contient les vecteurs réels de dimension m, la matrice W contient les vecteurs correspondants dans un espace de dimension $k < m$, et la matrice de passage H contient les vecteurs de base. L'avantage de la factorisation matricielle non négative par rapport aux autres techniques de factorisation matricielle est que l'interprétation est plus intuitive car les valeurs non négatives facilitent l'analyse des relations entre utilisateurs et articles. Et convient bien aux données contenant uniquement des valeurs non négatives (comme les systèmes de notation ou d'évaluation). En comparaison, la SVD ne permet d'interpréter ni les vecteurs de base ni les nouvelles coordonnées réduites, à cause de la présence de coefficients négatifs.

E. CoClustering

Également connu sous le nom de biclustering, est une technique d'analyse de données et de clustering qui vise à regrouper simultanément à la fois les lignes et les colonnes d'une matrice de données. Contrairement aux méthodes de clustering traditionnelles qui se concentrent uniquement sur les

observations (lignes) ou les variables (colonnes), le CoClustering cherche à identifier des groupes de lignes similaires et des groupes de colonnes similaires dans une matrice de données.

L'objectif principal du CoClustering est de découvrir des structures cachées dans les données, ce qui peut être utile dans divers domaines tels que la biologie, la recommandation de contenu, la classification de documents, l'analyse des médias sociaux, etc. Cette technique permet d'identifier des sous-populations de lignes et de colonnes qui ont des comportements similaires, ce qui peut faciliter la compréhension des relations entre les observations et les variables.

IV.4.2.4. Entraînement et résultats

Pour l'apprentissage, nous avons utilisé l'algorithme GridSearchCV avec une validation croisée (en divisant les données en 5 ensembles K = 5) afin d'identifier les hyperparamètres optimaux. Après cette étape, nous testons tous nos modèles (KNN, SVD, SVD++, NMF et CoClustering) en utilisant les données de test, qui contiennent 20 % de dataset, ce qui équivaut à 3200 line. Les résultats sont présentés dans la figure 20 et le tableau 3 :

Métrique	KNN (User -based)	KNN (Item -based)	SVD	SVD++	NMF	CoClustering
MAE	0.683684	0.688457	0.699150	0.698941	0.746975	0.761836
RMSE	0.935086	0.939016	0.938548	0.938440	1.017767	1.017992

Tableau 3. Résultats de tests pour filtrage collaboratif

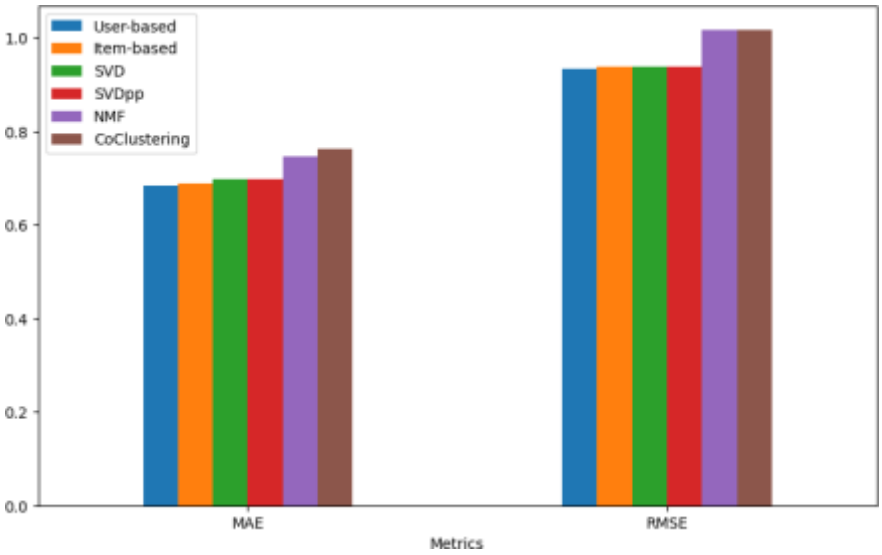


Fig. 20. Résultats de tests

Dans cette étude comparative des performances des algorithmes de recommandation, six approches différentes ont été évalué : User-based KNN, Item-based KNN, SVD, SVDpp, NMF et CoClustering. Nous avons utilisé les métriques MAE (Mean Absolute Error) et RMSE (Root Mean Square Error) pour mesurer leur précision. Les résultats montrent que les approches KNN(User-based) et KNN(Item-based) ont des performances similaires, avec des valeurs de MAE de 0,683684 et 0,688457 respectivement, ainsi que des RMSE de 0,935086 et 0,939016. Et les méthodes de factorisation matricielle SVD et SVDpp présentent des performances également proches, toutes se situant dans une fourchette de précision acceptable. NMF obtient des résultats légèrement moins précis, tandis que l'algorithme CoClustering a les performances les moins bonnes parmi les six approches évaluées, avec un MAE de 0,761836 et un RMSE de 1,017992.

IV.4.3. Algorithme de filtrage hybride de l'approche proposée

Les résultats précédents mettent en évidence que le système de recommandation basé sur le filtrage collaboratif n'a pas atteint les performances souhaitées, comme en témoignent les taux d'erreur élevés. Par conséquent, une approche d'hybridation a été adopté en combinant les modèles du système de recommandation basé sur le filtrage par contenu (le modèle SVM et le modèle de deep learning), qui a montré des performances plus satisfaisantes. Cette décision découle de la volonté d'améliorer la qualité des recommandations proposées aux utilisateurs.

Les résultats de l'hybridation entre le modèle SVM et le modèle de Deep Learning en utilisant la méthode weighted montrent que la précision du modèle combiné varie en fonction des poids attribués à chaque modèle. Lorsque le modèle de Deep Learning est fortement pondéré (0.95) par rapport au modèle SVM (0.05), la précision est de 91.41%. Cette précision reste relativement constante même lorsque les poids sont ajustés. Cependant, lorsque les poids sont plus équilibrés, par exemple, avec un poids de 0.5 pour chaque modèle, la précision augmente considérablement à 93.47%. Lorsque le modèle SVM est fortement pondéré (0.95) par rapport au modèle de Deep Learning (0.05), la précision reste similaire à celle du modèle de Deep Learning seul, ce qui suggère que le modèle de Deep Learning contribue de manière significative aux performances globales.

En général, ces résultats indiquent que l'hybridation de modèles de SVM et de Deep Learning peut améliorer la précision, en particulier lorsque les deux modèles sont équilibrés en termes de poids. Cependant, il est important de noter que l'ajustement des poids doit être effectué avec précaution en fonction des objectifs spécifiques de l'application, car des poids inappropriés peuvent avoir un impact négatif sur les performances.

IV.5. Interfaces du système proposé et visualisation des recommandations (Contribution 3)

Une plateforme user-friendly a été mise en place pour faciliter l'exploration et la découverte d'activités touristiques adaptées aux préférences uniques de chaque utilisateur. Notre site web de recommandation des activités touristiques a été conçu avec une interface intuitive et un design épuré, offrant une navigation fluide et agréable. Grâce à une technologie avancée de filtrage et de recommandation, les utilisateurs peuvent facilement trouver des activités qui correspondent à leurs intérêts personnels et à leurs besoins spécifiques. En utilisant une combinaison de données explicites et implicites, nous fournissons des recommandations personnalisées basées sur les préférences de chaque individu, telles que les avis, les évaluations et les comportements de navigation.

En outre, notre plateforme intègre des outils de recherche puissants et des options de filtrage détaillées, permettant aux utilisateurs de découvrir des

activités touristiques en fonction de divers critères, tels que la localisation, le type d'activité, et les préférences culturelles. Les utilisateurs peuvent également consulter des profils détaillés pour chaque activité, comprenant des photos, des descriptions, et des évaluations par d'autres voyageurs. Nous nous engageons à offrir une expérience utilisateur exceptionnelle en continuant à améliorer notre plateforme, en intégrant les retours des utilisateurs et en adaptant les recommandations en fonction des tendances émergentes. Notre objectif est de rendre la planification des voyages aussi simple et agréable que possible, en offrant des recommandations pertinentes et enrichissantes.

IV.5.1. Interfaces du système proposé

A. La page d'accueil

L'interface de la page d'accueil, illustrée par la figure 21 suivante, est conçue pour fournir une expérience sur mesure aux utilisateurs, avec une section de recommandations qui s'adapte aux préférences par pays soit des sites les plus populaires du moment dans le pays de l'utilisateur, mais aussi d'un filtrage collaboratif.



Fig. 21. Page d'accueil

B. Page d'authentification

L'internaute va sur le portail, clique sur le bouton de connexion ''Login'',

pour accéder à la page d'authentification.

Le processus de connexion dans notre plateforme utilisant JSON Web Tokens (JWT) commence par l'utilisateur saisissant ses identifiants (email et mot de passe). Le serveur vérifie ces informations, puis génère un jeton (tokens) contenant des données sur l'utilisateur, signé pour garantir son intégrité. Ce jeton est renvoyé au client, qui l'utilise pour accéder aux ressources sécurisées ultérieurement. Le serveur vérifie la validité du jeton pour autoriser l'accès aux ressources. Une fois le jetons expiré, l'utilisateur doit se reconnecter. Ce processus offre une sécurité accrue et simplifie la gestion des sessions, améliorant ainsi l'expérience utilisateur.

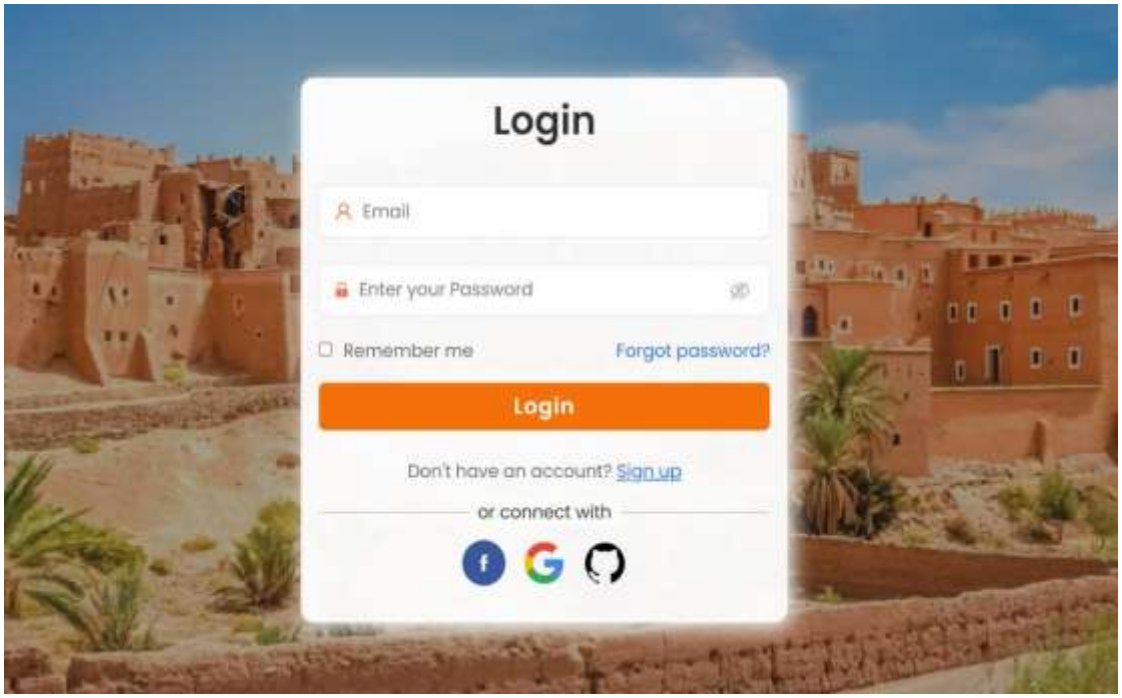


Fig. 22. Page d'authentification

C. Le Blog

La page blog, fournie un aperçu des différents lieu touristiques et permet à l'utilisateur d'effectuer un filtrage beaucoup plus spécifique en fonction de ses préférences. Il inclut aussi une barre de recherche permettant de trouver du contenu adapté.



Fig. 23. Page blog

D. Le feed utilisateur

Le feed utilisateur, est conçue pour fournir une expérience beaucoup plus immersive 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 et d'un ensemble de commandes applicables pour chaque vidéo.



Fig. 24. Feed utilisateur

IV.5.2. Visualisation des recommandations à base de profil utilisateur

A. Profil de l'utilisateur

Le profil de l'utilisateur comprend à la fois des informations personnelles telles que le nom, le prénom, le sexe, le pays, la ville, etc., ainsi que des mots-clés qui décrivent les préférences et les caractéristiques distinctives de l'utilisateur. De plus, il comprend des classifications qui indiquent les centres d'intérêt de l'utilisateur et son niveau d'affinité et de préférence pour chacune d'entre elles. Ce profil permet à la plateforme de personnaliser les recommandations et les expériences en fonction des préférences et des caractéristiques de l'utilisateur.



Fig. 25. Exemple du profil utilisateur

B. Résultat de filtrage collaboratif à base de profile utilisateur

Sur la page d'accueil, la section Destinations populaires affiche une collection de touristiques populaires auprès des autres utilisateurs.



Fig. 26. Résultat de filtrage collaboratif à base de profile utilisateur

C. Résultat de filtrage basé sur le contenu

Sur la page d'un article, si l'utilisateur n'est pas connecté, le système recommande des articles en utilisant le filtrage basé sur le contenu sans tenir en compte des préférences de l'utilisateur (absence de profil utilisateur). La situation est différente lorsque l'utilisateur est connecté car le système prend en compte de ses préférences, c'est-à-dire en utilisant le profil utilisateur.

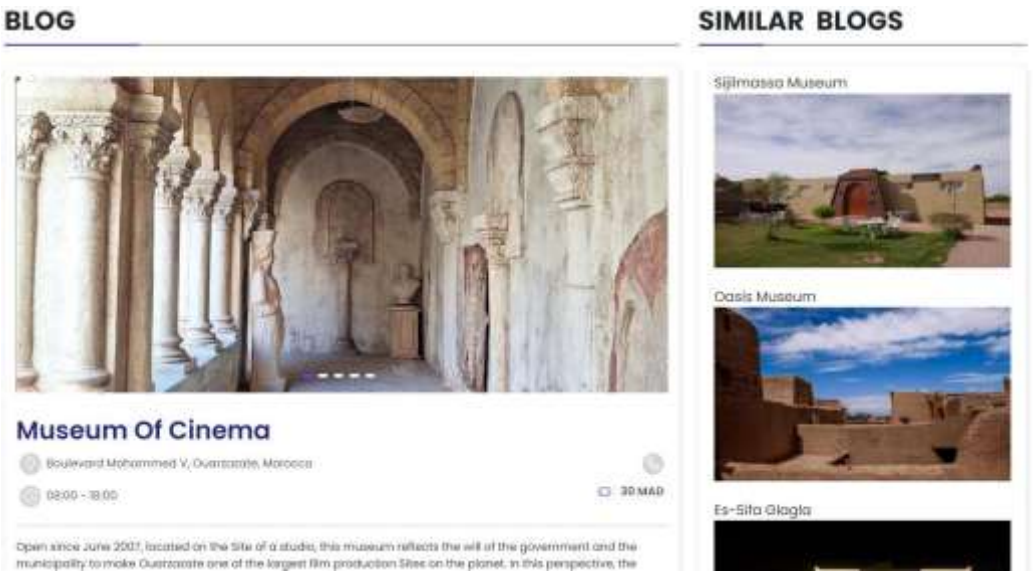


Fig. 27. Résultat de filtrage basé sur le contenu

D. La page de planification des voyages

L'utilisateur peut planifier et gérer ses voyages via la page "Trip Planning". Pour créer un plan de voyage, l'utilisateur choisit le lieu de début et de fin du voyage, ainsi que le nombre de lieux qu'il souhaite visiter, puis appuie sur le bouton " Generate itinerary".

Le système recommande le nombre de lieux souhaités par l'utilisateur, qui sont en alternance avec ses préférences, puis organise ces lieux pour créer un plan d'itinéraire raccourci qui passe par tous ces lieux. L'utilisateur peut également ajouter d'autres lieux. Ou supprimez-les du plan.

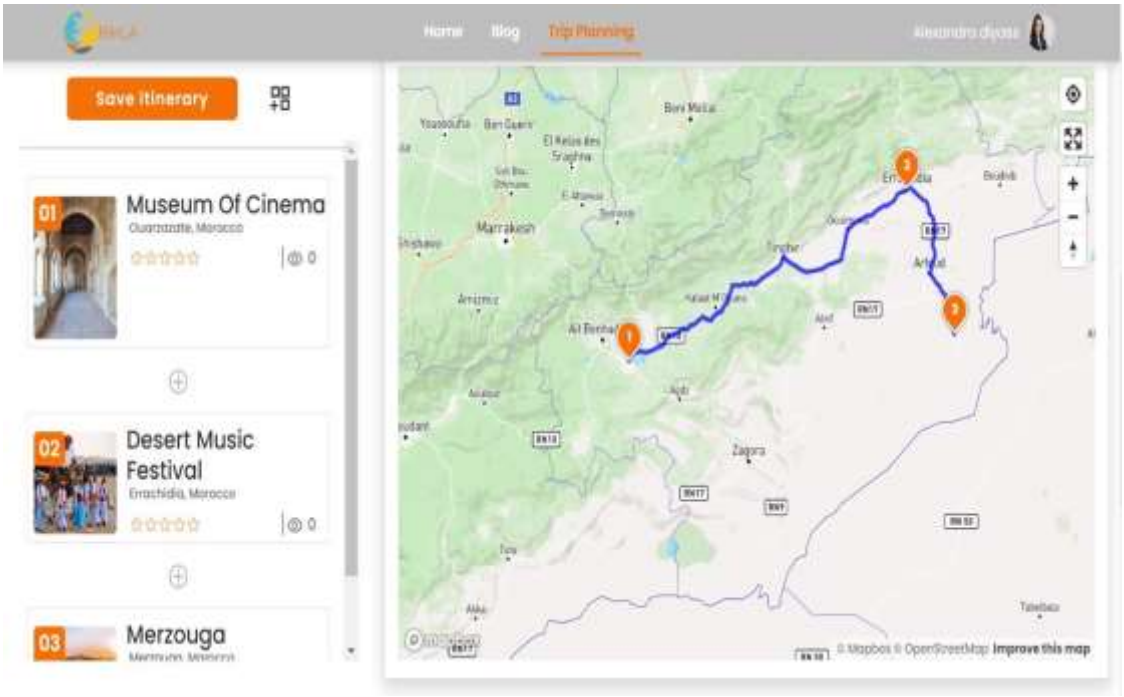


Fig. 28. Planification de voyage

IV.5.3. Dashboard administrateur

L'administrateur est considéré comme l'un des principaux acteurs de la plateforme, car il a un rôle important à jouer dans sa gestion et son développement. En tant que gestionnaire, l'administrateur est responsable de la création et de la maintenance des lieux touristiques, des utilisateurs et des évaluations. Il peut ajouter, supprimer ou modifier ces informations, ainsi que les attribuer aux bonnes catégories.

L'administrateur joue également un rôle essentiel dans le suivi et le test du système de recommandations. Il peut analyser les répartitions des évaluations des utilisateurs et les taux d'acceptation ou de rejet de l'itinéraire de voyage.

Ces informations peuvent être utilisées pour améliorer le système de recommandations et fournir des recommandations plus pertinentes aux utilisateurs.

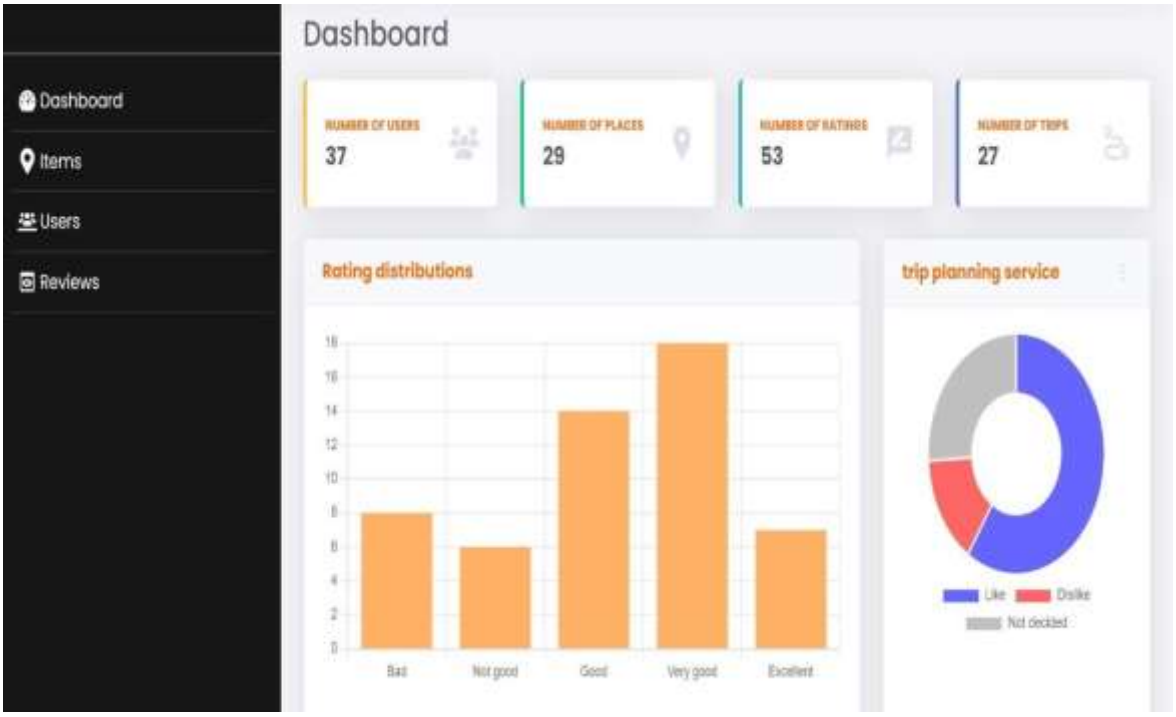


Fig. 29. Admin Dashboard

IV.6. Conclusion

L'implémentation de la solution a permis de concrétiser l'ensemble des concepts théoriques et méthodologiques explorés au cours de ce projet. En s'appuyant sur des outils et technologies modernes, nous avons développé un système de recommandation hybride performant, capable d'offrir des suggestions personnalisées et pertinentes aux voyageurs de la région du Drâa-Tafilalet.

Les résultats obtenus ont démontré l'efficacité de l'algorithme de recommandation, avec des performances satisfaisantes en termes de précision et de pertinence.

L'interface utilisateur développée dans ce cadre offre une expérience intuitive et fluide, facilitant la navigation et l'accès aux différentes ressources touristiques. Ce chapitre a ainsi montré comment les différentes composantes du projet se sont articulées pour donner naissance à une solution fonctionnelle, prête à être déployée et éventuellement améliorée pour de futurs besoins.

Cette implémentation constitue une base solide pour des itérations futures, permettant d'améliorer davantage les capacités du système en fonction des retours utilisateurs et de l'évolution des technologies.

Conclusion Générale & Perspectives

Cette thèse vise à améliorer le tourisme au Maroc, en particulier dans la région du Drâa-Tafilalet, en créant un système de recommandation hybride qui utilise des techniques de Deep et de Machine Learning (ML) pour informer et guider les touristes vers leurs points d'intérêt [46,96]. Les visiteurs pourront ainsi découvrir un large éventail d'attractions matérielles (hôtels, restaurants, Ksours (châteaux), monuments historiques, musées, églises, etc.) et immatériels (festivals, musiques, artisans traditions, défilés religieux, etc.), correspondant aux préférences et aux attentes spécifiques de chaque utilisateur.

En analysant les implémentations potentielles des systèmes de recommandation intelligents dans le secteur du tourisme, les auteurs introduisent une architecture hybride qui combine diverses techniques de filtrage, telles que les algorithmes d'arbre de décision, les k plus proches voisins (kNN), et les machines à vecteurs de support (SVM). Deux ensembles de données, Yelp et Drâa-Tafilalet, sont utilisés pour évaluer la performance du système. Les résultats expérimentaux, basés sur les mesures RMSE et MAE, indiquent que l'approche hybride, utilisant une pondération adaptée, offre une précision de prédiction supérieure à celle des approches traditionnelles séparées.

Le système développé dépasse la simple génération de recommandations individuelles et peut être considéré comme un assistant intelligent, capable d'élaborer un itinéraire personnalisé et optimisé pour une expérience complète et immersive sur plusieurs jours.

Par exemple, si un touriste potentiel souhaite passer quelques jours à la région avec sa famille (adultes) du 2 au 15 avril (14 jours) et qu'il est intéressé à découvrir les sites touristiques et peut voyager dans un rayon de 300 km. Le système fait alors ses prédictions suivant les profiles (contextuel, évaluations) renseignés explicitement par l'utilisateur ou constitués implicitement (réseaux sociaux, articles consultées), et recommande tout d'abord les zones d'intérêt sur une carte interactive sous forme d'un trajet. Une fois que l'utilisateur se concentre (clic) sur un article particulier, ses informations de

service touristique sont également affichées, enrichies d'un contenu multimédia illustrant l'item.

Notre système de recommandation touristique se révèle être un outil puissant pour les voyageurs de la région, offrant une expérience de planification de voyage personnalisée et simplifiée. Les résultats expérimentaux confirment son efficacité et sa convivialité, en faisant un atout précieux pour les amateurs de voyages à la recherche d'aventures uniques et sur mesure. Cependant, plusieurs perspectives peuvent être envisagées pour enrichir et étendre les fonctionnalités du système :

1. Amélioration continue de la personnalisation

Adaptation en temps réel : Intégrer des données en temps réel, telles que les conditions météorologiques, les événements locaux, ou les changements de disponibilité, pour ajuster les recommandations.

2. Extension des sources de données

Intégration de l'Internet des Objets (IoT) : Collecter des données à partir de dispositifs IoT, tels que les capteurs dans les musées, les hôtels, ou les sites touristiques, pour fournir des recommandations plus précises basées sur la fréquentation ou les préférences des visiteurs.

3. Amélioration de l'expérience utilisateur

Utilisation de la réalité augmentée (AR) et de la réalité virtuelle (VR) : Offrir aux utilisateurs la possibilité de visualiser des attractions ou des itinéraires avant leur visite, améliorant ainsi la planification et l'engagement.

Assistant vocal intelligent : Développer des interfaces vocales qui permettent aux utilisateurs d'interagir de manière plus naturelle avec le système, facilitant l'accès aux informations et aux recommandations.

4. Internationalisation et adaptation culturelle

Support multilingue : Étendre les fonctionnalités du système pour offrir des recommandations dans plusieurs langues, en tenant compte des spécificités culturelles et des préférences des touristes internationaux.

En conclusion, ce projet apporte une contribution importante à notre compréhension des systèmes de recommandation dans le secteur touristique. Il établit une base solide pour le développement de solutions technologiques

plus personnalisées et pertinentes, tout en mettant en lumière l'impact essentiel de l'IA dans l'enrichissement de l'expérience touristique à l'ère numérique. Les perspectives et les défis abordés ouvrent la voie à de nouvelles innovations et recherches dans ce domaine en constante évolution.

Bibliographie

[1] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4):331–370.

[2] Burke R. (2007) Hybrid Web Recommender Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg.

[3] Resnick P., Neophytos I., Bergstrom P., Mitesh S., & Riedl J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In CSCW94 - Conference on Computer Supported Cooperative Work, pages 175–186. Chapel Hill, Addison-Wesley, 1994.

[4] Resnick P. & Varian H. R. (1997). Recommender systems. In Communications of the ACM, volume 40, pages 56-58.

[5] Malone T., Brobst S., Cohen S., Grant K. & Turbak F. (1987). Intelligent information des systèmes de partage. In Communications of the ACM, volume 30, pages 390–402.

[6] David Goldberg, David Nichols, Brian M. Oki, & Douglas Terry. Using collaborative filtering to weave an information tapestry. Communications of the Association of Computing Machinery, 35(12):61–70, 1992.

[7] Maes P. and Shardanand U. (1995). Social information filtering: algorithms for Automating “Word of Mouth”. In the SIGCHI conference on Human factors in computing systems, Denver, Colorado, United States. ACM Press/Addison-Wesley Publishing Co.

[8] Michael J., Pazzani & Daniel Billsus. Content-based recommendation systems. In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, The Adaptive Web, number 4321 in Lecture Notes in Computer Science, pages 325–341. Springer Berlin Heidelberg, January 2007.

[9] Linden G., Smith B. and York J. (2003). Amazon.com Recommendations: Item-to-item Collaborative Filtering. IEEE Internet Computing 7, 76-80.

[10] Koren, Y., Bell, R., & Volinsky, C. (2009). "Matrix Factorization Techniques for Recommender Systems." Computer Science Department, Stanford University.

[11] Zhang, S., et al. (2019). "Deep Learning Based Recommender System: A Survey and New Perspectives." ACM Computing Surveys.

[12] Sebastia, Laura & García, Inmaculada & Onaindia, Eva & Guzmán Alvarez, César. (2009). E-Tourism: A tourist recommendation and planning application. International Journal of Artificial Intelligence Tools. 18. 717-738. 10.1142/S0218213009000378.

[13] F. Ricci, L. Rokach and B. Shapira, Recommender Systems Handbook, Springer US, Boston, MA, 2011, Ch. Introduction to Recommender Systems Handbook, pp. 1–35. doi: 10.1007/978-0-387-85820-3_1.

[14] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, June 2005, doi: 10.1109/TKDE.2005.99.

[15] de Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). Semantics- aware content-based recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), Recommender systems handbook (pp. 119–159). Springer. doi:10.1007/ 978-1-4899-7637-6_4.

[16] Loh, Stanley & Lorenzi, Fabiana & Lichtnow, Daniel. (2003). A Tourism Recommender System Based on Collaboration and Text Analysis.. J. of IT & Tourism. 6. 157-165. 10.3727/1098305031436980.

[17] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, “Mobile Recommender Systems in Tourism,” J. Network and Com- puter Applications, vol. 39, pp. 319–333, Mar. 2014, doi: 10.1016/ j.jnca.201304.006

[18] Rao, K & Talwar, V. (2008). Application Domain and Functional Classification of Recommender Systems—A Survey. DESIDOC Journal of Library & Information Technology. 28. 10.14429/djlit.28.3.174.

[19] Baeza-Yates & Ribeiro-Neto (1999). Modern Information Retrieval. Addison-Wesley.

[20] Schafer J., Frankowski D., Herlocker J., & Sen S. (2007). Collaborative filtering recommender systems. The adaptive web, page 291–324, 2007.

[21] Su, X., & Khoshgoftaar, T.M. (2009). A Survey of Collaborative Filtering Techniques. Adv. Artif. Intell., 2009, 421425:1-421425:19.

[22] Gui-Rong Xue, Chenxi Lin, Qiang Yang, WenSi Xi, Hua-Jun Zeng, Yong Yu, and Zheng Chen. Scalable collaborative filtering using cluster-based smoothing. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR ’05, page 114–121, New York, NY, USA, 2005. ACM.

[23] Su X., Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. Advances in Artificial Intelligence, Janvier 2009.

[24] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’08, page 426–434, New York, NY, USA, 2008. ACM.

[25] Gábor Takács, István Pilászy, Bottyán Németh, & Domonkos Tikk (2007). Major components of the gravity recommendation system. SIGKDD Explor. Newsl. 9(2):80–83, 2007.

[26] Herlocker J., Konstan J., Borchers A., & Riedl J. (1999). An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval.

[27] Wei Zeng, Ming-Sheng Shang, Qian-Ming Zhang, Linyuan Lü, & Tao Zhou. Can dissimilar users contribute to accuracy and diversity of personalized recommendation?. International Journal of Modern Physics C, 21(10):1217–1227, October 2010.

[28] Wei Zeng, Yu-Xiao Zhu, Linyuan Lü, & Tao Zhou. Negative ratings play a positive role in information filtering. Statistical Mechanics and its Applications, 390(23–24):4486–4493, November 2011.

[29] Breese J., Heckerman D., & Kadie C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In 14th Annual Conference on Uncertainty in Artificial Intelligence, pages 43–52.

[30] Cho Y. H., Kim J. K., Kim S. H. , A personalized recommender system based on web usage mining and decision tree induction, Expert System with Applications, vol. 23, n° 3, pages 329-342, 2002.

[31] Heckerman, D., Chickering, D., Meek, C., Rounthwaite, R., and Kadie, C. (2000). Dependency networks for inference, collaborative Filtering, and data visualization. Technical Report MSR-TR-00-16, Microsoft Research, Red- mond, WA.

[32] Arnautu O. R. (2012). Mures: Un système de recommandation de musique. Master's thesis, La Faculté des arts et des sciences Université de Montréal.

[33] Bouchindhomme & Rochlitz (1992). Dans le flou artistique. Éléments d'une théorie de la "rationalité esthétique". Pages 203–238.

[34] Montaner M., López B., De La Rosa J. L., (2003). A Taxonomy of Recommender Agents on the Internet, Artificial Intelligence Review, Vol. 19, pp.285-330.

[35] Fenza, G., Fischetti, E., Furno, D., & Loia, V. (2011). A hybrid context aware system for tourist guidance based on collaborative filtering. In Proceedings of IEEE int. conference on fuzzy systems (pp. 131–138).

[36] Meehan, K., Lunney, T., Curran, K., & McCaughey, A. (2013). Context-aware intelligent recommendation system for tourism. In Pervasive computing and communications workshops (percom workshops), 2013 ieee international conference on (pp. 328– 331). doi:10.1109/PerComW.2013.6529508.

[37] Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M., Steggles, P.: Towards a better understanding of context and context-awareness. In: Handheld and Ubiquitous Computing, pp. 304–307. Springer, Berlin, Heidelberg (1999)

[38] Schein A., Popescul A., Ungar L. H. & Pennock D. M. (2002). Methods and metrics for cold-start recommendations. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'02), pages 253–260, New York, USA. ACM.

[39] Nguyen A. T. (2006). COCoFil2: Un nouveau système de filtrage collaboratif basé sur le modèle des espaces de communautés. PhD thesis, university Joseph Fourier-Grenoble I.

[40] Bamshad Mobasher, Xin Jin, & Yanzan Zhou (2004). Semantically enhanced collaborative filtering on the web. In Bettina Berendt, Andreas Hotho, Dunja Mladenic, Maarten van Someren, Myra Spiliopoulou, and Gerd Stumme, editors, Web Mining: From Web to Semantic Web, number 3209 in Lecture Notes in Computer Science, pages 57–76. Springer Berlin Heidelberg, January 2004.

[41] Smyth B. & Cotter P. (2000). A personalized TV listings service for the digital TV age. Knowledge-Based Systems, 13(2):53–59, 2000.

[42] Guy Shani & Asela Gunawardana. Evaluating recommendation systems. In Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor, editors, Recommender Systems Handbook, pages 257–297. Springer US, January 2011.

[43] Guy Shani (2010). Tutorial on evaluating recommender systems. In Proceedings of the fourth ACM conference on Recommender systems, RecSys '10, pages 1-1. ACM, 2010.

[44] Ron Kohavi, Roger Longbotham, Dan Sommerfield, & Randal M. Henne. Controlled experiments on the web: survey and practical guide. Data Mining and Knowledge Discovery, 18(1):140–181, February 2009.

[45] Jon Herlocker, Joseph A. Konstan, & John Riedl (2002). An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information Retrieval, 5(4):287–310, October 2002.

[46] Boulaalam, O., Aghoutane, B., El Ouadghiri, D., Moumen, A., Malinine, M. (2018). Proposal of a Big data System Based on the Recommendation and Profiling Techniques for an Intelligent Management of Moroccan Tourism. Procedia Computer Science. 134. 346-351. 10.1016/j.procs.2018.07.200.

[47] Borrás, J., Moreno, A., & Valls, A. (2014). Intelligent tourism recommender systems: A survey. Expert Systems with Applications, 41(16), 7370–7389. doi:10.1016/ j.eswa.2014.06.007.

[48] Lim, K. H., Chan, J., Leckie, C., & Karunasekera, S. (2015). Personalized tour recommendation based on user interests and points of interest visit durations. In Proceedings of the 24th international conference on artificial intelligence. In IJCAI'15 (pp. 1778–1784). AAAI Press.

[49] Brilhante, I. R., Macedo, J. A., Nardini, F. M., Perego, R., & Renso, C. (2015). On planning sightseeing tours with TripBuilder. Information Processing & Management, 51(2), 1–15. doi:10.1016/j.ipm.2014.10.003.

[50] Badriyah, Tessy & Azvy, Sefryan & Yuwono, Wiratmoko & Syarif, Iwan. (2018). Recommendation system for property search using content based filtering method. 25-29. 10.1109/ICOIACT.2018.8350801.

[51] Santos, Filipe & Almeida, Ana & Martins, C. & Moura Oliveira, Paulo & Gonçalves, Ramiro. (2016). Tourism Recommendation System based in user's profile and functionality levels. 93-97. 10.1145/2948992.2948995.

[52] Herzog, D. & Wörndl, W.. (2014). A travel recommender system for combining multiple travel regions to a composite trip. CEUR Workshop Proceedings. 1245. 42-47.

[53] Gyorodi, Robert & Győrödi, Cornelia & Dersidan, Mihai. (2013). An Extended Recommendation System using Data Mining Implemented for Smart Phones. INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY 11 (3), ISSN 2277-3061, Publisher IJCT Foundation, United States, Impact factor 1.341, pag. 2360-2372, http://ijctonline.com/ojs/index.php/ijct/index. 11. 2360-2372. 10.24297/ijct.v11i3.1169.

[54] Ruotsalo, Tuukka & Haav, Krister & Stoyanov, Antony & Roche, Sylvain & Fani, Elena & Deliai, Romina & Mäkelä, Eetu & Kauppinen, Tomi & Hyvönen, Eero. (2013). SMARTMUSEUM: a mobile recommender system for the web of data. Web Semantics: Science, Services and Agents on the World Wide Web. 20. 50–67. 10.1016/j.websem.2013.03.001.

[55] Yu, Chien-Chih & Chang, Hsiao-ping. (2009). Personalized Location-Based Recommendation Services for Tour Planning in Mobile Tourism Applications. 38-49. 10.1007/978-3-642-03964-5_5.

[56] Wang, Zehao & Liu, Bing. (2019). Tourism recommendation system based on data mining. Journal of Physics: Conference Series. 1345. 022027. 10.1088/1742-6596/1345/2/022027.

[57] Di Bitonto, Pierpaolo & Laterza, Maria & Roselli, Teresa & Rossano, Veronica. (2011). A recommendation system to promote local cultural heritage. JE-LKS. JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY. 7. 97-107. 10.20368/1971-8829/556.

[58] R., Logesh & V, Subramaniyaswamy. (2016). A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users. Computational Intelligence and Neuroscience. 2016. 1-28. 10.1155/2016/1291358.

[59] Ravi, L., & Vairavasundaram, S. (2016). A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users. Computational Intelligence and Neuroscience, 2016.

[60] Bahramian, Zahra & Abbaspour, Rahim & Claramunt, Christophe. (2018). Toward Geospatial Collaborative Tourism Recommender Systems. 10.4018/978-1-5225-5088-4.ch010.

[61] Alvarado-Uribe, Joanna & Gomez, Andrea & Molina, Germán & Gonzalez-Mendoza, Miguel & Parra Meroño, María Concepción & Jara, Antonio J.. (2018). Towards the Development of a Smart Tourism Application Based on Smart POI and

Recommendation Algorithms: Ceutí as a Study Case. 904-916. 10.1007/978-3-319-61542-4_92.

[62] Ravi L, Vairavasundaram S. A Collaborative Location Based Travel Recommendation System through Enhanced Rating Prediction for the Group of Users. Comput Intell Neurosci. 2016;2016:1291358. doi: 10.1155/2016/1291358. Epub 2016 Mar 16. PMID: 27069468; PMCID: PMC4812910.

[63] Nafs, Aliakbar & Arash, Nik & Nafs, & Javidi, M.M.. (2008). A preference list based recommender system for tourism industry, design and implementation. MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings e-Commerce 2008.

[64] K Meehan, T Lunney, K Curran, and A McCaughey, "VISIT: VirtualIntelligent System for Informing Tourists", PGNET, Liverpool, 2012.

[65] Bahramian, Zahra & Abbaspour, Rahim & Claramunt, Christophe. (2017). A CONTEXT-AWARE TOURISM RECOMMENDER SYSTEM BASED ON A SPREADING ACTIVATION METHOD. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLII-4/W4. 333-339. 10.5194/isprs-archives-XLII-4-W4-333-2017.

[66] Aarab, Zineb & Elghazi, Asmae & Saidi, Rajaa & Rahmani, Moulay Driss. (2018). Toward a Smart Tourism Recommender System: Applied to Tangier City. https://doi.org/10.1007/978-3-319-74500- 8_59.

[67] Lucas, J. P., Luz, N., Moreno, M. N., Anacleto, R., Figueiredo, A. A., & Martins, C. (2013). A hybrid recommendation approach for a tourism system. Expert Systems with Applications, 40(9), 3532–3550. doi:10.1016/j.eswa.2012.12.061.

[68] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System," J. Phys. Conf. Ser., vol. 1000, no. 1, 2018, doi: 10.1088/1742-6596/1000/1/012101.

[69] Soares, J.D., Suyoto, & Santoso, A.J. (2017). M-Guide: Hybrid Recommender System Tourism In East-Timor. 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), 303-309.

[70] Kbaier, Mohamed & Masri, Hela & Krichen, Saoussen. (2017). A Personalized Hybrid Tourism Recommender System. 244-250. 10.1109/AICCSA.2017.12.

[71] S. Renjith and C. Anjali, "A personalized mobile travel recommender system using hybrid algorithm," 2014 1st Int. Conf. Comput. Syst. Commun. ICCSC 2014, no. December, pp. 12–17, 2014, doi: 10.1109/COMPSC.2014.7032612.

[72] M. Hong, J. J. Jung, F. Piccialli, and A. Chianese, Social recommendation service for cultural heritage, Personal and Ubiquitous Computing, vol. 21, no. 2, pp. 191–201, 2017.

[73] Balabanovic M., Shoham Y., Fab: Content-Based, Collaborative Recommendations. Communications of the ACM, vol. 40, n°. 3, pages 66-72, 1997.

[74] Michael J., & Pazzani (1999). A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Rev., 13(5-6):393–408, 1999.

[75] Melville P., Mooney R. & Nagarajan R. (2002). Content boosted collaborative filtering for improved recommendations. In Proceedings of the Eighteenth national conference on Artificial intelligence, pages 187–192, Menlo Park, CA, USA. American Association for Artificial Intelligence.

[76] Fararni, K. A., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. (2021). Tourism recommender systems: An overview. International Journal of Cloud Computing, 10(5/6), 603-612. https://doi.org/10.1504/IJCC.2021.120397

[77] Nguyen, A. ; Denos, N. et Berrut, C. (2006). Exploitation des données disponibles à froid pour améliorer le démarrage à froid dans les systèmes de filtrage d'information. In Actes du XXIV Congrès d'INFORSID, pages 81–95.

[78] Vozalis M., & Margaritis K. (2006). On the enhancement of collaborative filtering by demographic data. Web Intelligence and Agent Systems : An International Journal (WIAS), 4(2):117–138.

[79] Kavitha, S. & Jobi, Vijay & Sridhar, Rajeswari. (2017). Tourism Recommendation Using Social Media Profiles. 10.1007/978-981-10-3174-8_22.

[80] Menk, A. ; Sebastia, L. et Ferreira, R. (2019) «Recommendation Systems for Tourism Based on Social Networks: A Survey», ArXiv190312099 Cs, mars 2019, Consulté le: déc. 08, 2019. [En ligne]. Disponible sur: http://arxiv.org/abs/1903.12099.

[81] Mark Claypool, Phong Le, Makoto Wased, and David Brown. (2001). Implicit interest indicators. In Proceedings of the 6th international conference on Intelligent user interfaces (IUI '01). Association for Computing Machinery, New York, NY, USA, 33–40. DOI: https://doi.org/10.1145/359784.359836

[82] Badraddine AGHOUTANE et all. "Développement de plateformes de catalogage, de gestion et de diffusion en ligne du patrimoine culturel de la région d'Errachidia", dans le cadre du programme (2016) Stratégique de Recherche de l'UMI.

[83] NAFIS, F., YAHYAOUY, A., AGHOUTANE, B. (2018) "Towards a generic ontology for describing heritage data". Conférence Afro-Méditerranéenne Sur La Recherche Multidisciplinaire & Applications 07-08 Juillet 2018, IAV - Rabat, Maroc.

[84] NAFIS, F., YAHYAOUY, A., AGHOUTANE, B. (2019) "Ontologies for classification of cultural heritage data". The 5 th International conference on Wireless Technologies, Embedded and Intelligent Systems Wits'2019, April 03-05, Fez-Morocco.

[85] United Nations World Tourism Organization (UNWTO). (2020). Tourism and Sustainable Development Goals – Journey to 2030. UNWTO Report.

[86] Middleton, V.T.C., & Clarke, J. (2001). Marketing in Travel and Tourism. Butterworth-Heinemann

[87] Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism Management, 29(4), 609-623.

[88] Sigala, M., Christou, E., & Gretzel, U. (2012). Social Media in Travel, Tourism and Hospitality: Theory, Practice and Cases. Ashgate Publishing, Ltd.

[89] Rifai, T. (2017). Tourism and the Digital Transformation. International Tourism Forum. UNWTO.

[90] Egger, R., & Buhalis, D. (2019). Smart tourism: Implications for the Tyrol region. Journal of Tourism Futures, 5(1), 17-28. https://doi.org/10.1108/JTF-01-2019-0004

[91] Kossmann, M., & Cottam, H. (2020). Data-driven recommendations for tourism in Tyrol: Enhancing visitor experiences through analytics. International Journal of Information Management, 54, 102184. https://doi.org/10.1016/j.ijinfomgt.2020.102184

[92] Lee, S. M., & Choi, S. (2020). Mobile technologies in tourism: How recommendation systems influence tourist behavior in Seoul. Journal of Hospitality and Tourism Technology, 11(4), 523-537. https://doi.org/10.1108/JHTT-04-2020-0065

[93] Park, S., & Ryu, K. (2019). Context-aware recommendation systems in tourism: The case of Seoul. Computers in Human Behavior, 100, 1-10. https://doi.org/10.1016/j.chb.2019.06.022

[94] Rodríguez, A., & González, A. (2019). Trends in tourist preferences: The role of recommender systems in the Canary Islands. Tourism Management, 70, 46-58. https://doi.org/10.1016/j.tourman.2018.07.003

[95] Sánchez-Cañizares, S. M., & Pino, J. (2021). Ecotourism and sustainable practices in the Canary Islands: The role of personalized recommendations. Tourism Management Perspectives, 38, 133-142. https://doi.org/10.1016/j.tmp.2020.12.006

[96] K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," in Big Data Mining and Analytics, vol. 4, no. 1, pp. 47-55, March 2021, doi: 10.26599/BDMA.2020.9020015.

[97] Rijsbergen C. J. V, Information Retrieval. Second edition, Butterworths, 1979. Rucker J., Polanco M. J., Siteseer: Personalized Navigation for the Web, Communications of the ACM, vol. 40, n° 3, pages 73-75, 1997.

[98] Lainé-Cruzel S., ProfilDoc : Filtrer une information exploitable, Bulletin des Bibliothèques de France, n° 5, pages 60-65, http://www.enssib.fr/bbf/bbf-99-5/10_lainecruzel.pdf, 1999.

[99] Shapira B., Shoval P., Hanani U., Stereotypes In Information Filtering Systems, Information Processing & Management, vol. 33, n° 3, pages 273-287, 1997.

[100] Salton G. & Lesk M. E. (1965). The SMART automatic document retrieval systems and illustration. Communications of the ACM 8, 391-398.

[101] Salton G. (1989). Automatic text processing: the transformation, analysis, and retrieval of information by computer. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[102] Wahiba Nesrine Zemirli, thèse Doctorat: «Modèle d'accès personnalisé à l'information basé sur les Diagrammes d'Influence intégrant un profil utilisateur évolutif»,Université Paul Sabatier de Toulouse III, 25/1/ 2009.

[103] Hicham Chebili, Mémoire magister: «Agrégation des résultats dans la recherche d'information Semi-structurée », Ecole supérieure d'informatique (E .S .I) Oued-Smar Algérie, 03/05/2011.

[104] Jonathan L., Herlocker J., Joseph A. Konstan, Al Borchers, & John Riedl (1999). An algorithmic framework for performing collaborative filtering. In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '99, page 230–237, New York, NY, USA, 1999. ACM.

[105] Sarwar B., Karypis G., Konstan J. & Riedl J. (2000). Application of dimensionality reduction in recommender system - a case study. In ACM WebKDD 2000 Web Mining for E-commerce Workshop.

[106] Deshpande M., & Karypis G. (2004). Item-based top-N recommendation algorithms. ACM Transaction on Information Systems, 143-177.