

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université A. Mira de Béjaïa  
Faculté des Sciences Exactes  
Département d'Informatique



## Mémoire de fin de cycle

En vue de l'obtention du diplôme de master recherche en Informatique  
Option : Systèmes d'informations avancés (SIA)

### THÈME

Système de recommandation basé sur les techniques de  
l'intelligence artificielle

### Réalisé par :

M<sup>lle</sup> IDIR Ludmilia

M<sup>lle</sup> AYAD Yasmine

Soutenu le : 01/07/2025

Devant le jury composé de :

<b>Président</b>	M <sup>r</sup>	ACHROUFENE Achour	U.A/Mira Béjaïa.
<b>Examinatrice</b>	M <sup>me</sup>	BOULAHROUZ Djamilia	U.A/Mira Béjaïa.
<b>Examineur</b>	M <sup>r</sup>	BOUCHEBBAH Fatah	U.A/Mira Béjaïa.
<b>Examinatrice</b>	M <sup>me</sup>	SAAD Narimane	U.A/Mira Béjaïa.
<b>Encadrante</b>	M <sup>me</sup>	AIT HACENE Souhila	U.A/Mira Béjaïa.

Promotion : 2024/2025



## Remerciements

Avant tout, nous tenons à exprimer notre profonde gratitude envers notre Dieu, le miséricordieux et tout-puissant, qui nous a octroyé le courage nécessaire pour mener à bien notre parcours universitaire. Que sa guidance continue de nous éclairer tout au long de notre vie future.

Nous souhaitons également remercier particulièrement Madame AIT HACENE Souhila. Sans son aide et son encadrement, ce travail n'aurait pas pu être réalisé et ne serait pas aussi riche. Nous la remercions sincèrement de nous avoir soutenus dans cette démarche, d'avoir consacré le temps nécessaire à sa réalisation, et de nous avoir prodigué des remarques et des retours pertinents, ainsi que de son souci du détail, contribuant ainsi de manière significative à l'aboutissement de ce mémoire.

Nous remercions Monsieur ACHEROUFENE Achour pour l'honneur qu'il nous a fait en acceptant de présider le jury. Nos remerciements s'adressent également à Madame BOULAH-ROUZ Djamila, à BOUCHEBBAH Fatah et à SAAD Narimane, membres du jury, pour l'intérêt manifesté envers notre travail en acceptant de l'examiner et de l'enrichir par leurs précieuses suggestions.

Nous exprimons notre reconnaissance à nos familles pour leur soutien moral inconditionnel et leur présence réconfortante lors des moments difficiles.

Nous tenons à remercier chaleureusement l'ensemble des enseignants qui ont assuré notre formation. Leur encadrement pédagogique de qualité et leur présence d'esprit remarquable ont été déterminants dans notre réussite. Nous leur adressons notre plus sincère gratitude.

# Dédicaces

## À mes parents, piliers silencieux de mon monde, Mohamed et Fouzia

Il n'existe pas de mots assez vastes pour contenir ce que vous avez semé en moi. Vous m'avez portée bien avant mes premiers pas, dans vos rêves, vos peurs et vos espoirs. Vous m'avez élevée sans jamais me façonner à votre image, mais en m'apprenant à devenir la mienne et à croire en mes capacités. C'est dans vos mains fatiguées que j'ai appris la valeur du travail. C'est avec un immense honneur, mais aussi beaucoup de modestie, que je vous dédie le fruit de mon travail, un travail qui n'aurait jamais vu le jour sans votre soutien et votre présence à mes côtés.

## À mon grand-père Nour Dine et à ma grand-mère Ayou

Je vous envoie, à travers ce modeste travail, ma profonde gratitude et ma reconnaissance ainsi que mes sincères remerciements pour votre soutien et vos encouragements qui m'ont suivi tout au long de ma vie. Que Dieu vous garde pour moi. Vous êtes les meilleurs grands-parents du monde.

## À la mémoire de mon grand-père Hachemi et à ma grand-mère Tassadit

Je n'ai pas pu réaliser ce travail en ta présence, cher grand-père, et voilà que je le dédie à ta mémoire. Tu me manques énormément et je prie Dieu chaque jour pour qu'il t'accueille dans son vaste paradis. Merci à toi aussi, grand-mère, pour ta gentillesse et tes prières.

## À mes grandes sœurs Lylia et Leticia et ma petite sœur Lina

Merci d'être mes repères, mes confidentes, et mes piliers. Votre force, votre amour et vos conseils m'inspirent chaque jour. Vous avez toujours été là pour moi, dans les bons comme dans les mauvais moments, et je vous en serai toujours reconnaissante. **Lylia**, malgré la distance, tu continues à veiller sur moi et à me faire sourire, comme si tu n'étais jamais vraiment loin. **Léticia**, tu ne m'as jamais quittée, tu as toujours été là, présente et aimante... Et rien ne pourra jamais remplacer cela. Et toi, **Lina**, ma petite sœur... Malgré ton jeune âge, tu as été d'un soutien immense. Ma binôme de bêtises, toujours là pour me faire rire et me remonter le moral. Merci d'être toi.

## À mes petits frères Yahia et Rayane

En témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous, je vous dédie ce modeste travail. Que Dieu vous bénisse et vous guide vers le meilleur.

## À mon encadrante, Madame AIT HACENE Souhila

Merci d'avoir été un repère et un soutien tout au long de mon parcours. Vous m'avez transmis bien plus que des savoirs, une rigueur et une éthique. Dans le silence de vos conseils, j'ai appris à me dépasser et à penser autrement. Vous avez ma profonde reconnaissance.

## À mes amis Chaimaa, Dihia et Sonia

Merci pour votre soutien et vos encouragements, qui m'ont aidé à surmonter je ne sais combien de difficultés, et surtout, merci de m'avoir fait rire dans mes moments difficiles.

### **À ma famille**

Merci pour votre présence bienveillante, vos encouragements et votre chaleur familiale.

### **À ma binôme, ma partenaire de route, Yasmine**

Merci pour ta présence constante, ton énergie, ton humour dans les moments de tension et ta détermination dans ceux de doute. À deux, on a porté bien des charges, relevé des défis, parfois en silence, parfois en éclat de rire, mais toujours avec cette solidarité sans faille qui fait toute la différence. Je suis fière d'avoir cheminé avec toi.

# Dédicaces

## **À mes parents, Kamel et Farida.**

Vous êtes les fondations de ma force, les battements discrets derrière chacun de mes pas. Grâce à vos sacrifices silencieux, à votre amour inconditionnel et à votre présence rassurante, j'ai pu avancer, même quand le chemin semblait flou. Aujourd'hui, c'est avec une émotion immense que je vous dédie cette étape : je suis fier d'être votre premier enfant à soutenir un mémoire, et j'espère, de tout mon cœur, que vos regards brillent de la même fierté que le mien.

## **À mes sœurs, Ikram et Anfal, et à mon frère Zaki.**

Merci pour votre amour simple, vos blagues qui réconfortent, et pour être toujours là, silencieusement mais puissamment.

## **À mon grand-père.**

Ton absence me serre le cœur. Tu voulais tant être là, vivre ce moment à mes côtés... Je sais que, d'où tu es, tu veilles sur moi avec fierté. Ce mémoire t'est dédié, avec toute mon affection et mon respect.

## **À mes amies précieuses : Sarah, Hayat, Kahina, Cerine, Lydia, Zahra, Ryna, Zara, Dyhia, Elissa et Rania.**

Vous êtes plus que des amies : vous êtes mes âmes sœurs d'amitié, celles qui ont semé de la lumière sur mon chemin, apaisé mes angoisses et partagé mes rires les plus sincères.

## **À ma famille et aux voisins (C. Habta).**

Merci pour votre présence bienveillante, vos encouragements et votre chaleur familiale.

## **À ma binôme, Ludmilia.**

J'ai eu la chance de croiser ton chemin. Travailler avec toi a été un plaisir rare, une vraie collaboration faite de respect, d'écoute et d'efficacité. Merci d'avoir été la meilleure binôme que l'on puisse espérer.

## **À Mayas, mon ami de toujours, et à Hichem, une rencontre précieuse.**

Merci pour votre présence constante, votre soutien indéfectible, les moments partagés, et surtout pour avoir cru en moi lorsque j'en doutais. Votre bienveillance a été une force tout au long de ce parcours.

## **Et enfin... à moi-même.**

Pour avoir cru en mes rêves, même lorsqu'ils semblaient lointains. Pour chaque pas fait dans la douleur, pour chaque larme transformée en force. Pour ne pas avoir abandonné. Ce mémoire est la preuve que je suis capable d'aller au bout de ce que j'entreprends. Et je suis fier de la femme que je deviens.

*Yasmine Ayad*

# Table des matières

<b>Table des matières</b>	<b>I</b>
<b>Liste des figures</b>	<b>V</b>
<b>Liste des tableaux</b>	<b>VII</b>
<b>Liste des acronymes</b>	<b>VIII</b>
<b>Introduction générale</b>	<b>1</b>
<b>Chapitre 1: Systèmes de recommandations (SR) et intelligence artificielle (IA)</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Systèmes de recommandations (SR) . . . . .	3
1.2.1 Définition des SR . . . . .	3
1.2.2 Types de systèmes de recommandation . . . . .	4
A. SR basé sur le filtrage collaboratif (collaborative filtering, CF) . . . . .	4
B. SR basée sur le contenu (content-based filtering, CBF) . . . . .	5
C. SR basé sur les connaissances (Knowledge-Based Recommender Systems, KBRS) . . . . .	6
D. SR hybrides (Hybrid Recommender Systems, HRS) . . . . .	6
E. SR basé sur le contexte (Context-Aware Recommender System, CARS) . . . . .	6
1.2.3 Limites et défis des systèmes de recommandation . . . . .	7
A. Le problème du démarrage à froid (Cold Start) . . . . .	7
B. Le problème de scalabilité (scalability problem) . . . . .	7
C. Le problème de la surspécialisation (overspecialization) . . . . .	7
D. Le problème de sparsité des données (Data Sparsity Problem) . . . . .	8

## TABLE DES MATIÈRES

---

E. Le problème de confidentialité et de sécurité des données (Privacy and Data Security Problem) . . . . .	8
1.2.4 Les domaines d'application des systèmes de recommandation . . . . .	8
1.3 L'IA et les systèmes de recommandation . . . . .	8
1.4 Intelligence artificielle . . . . .	10
1.4.1 Définition et évolution de l'intelligence artificielle . . . . .	10
1.4.2 Apprentissage automatique (machine learning, ML) . . . . .	10
A. Apprentissage supervisé . . . . .	11
B. Apprentissage non supervisé . . . . .	11
C. Apprentissage semi-supervisé . . . . .	11
D. Apprentissage par renforcement . . . . .	12
1.4.3 Apprentissage profond (deep learning, DL) . . . . .	12
A. Couches des réseaux de neurones . . . . .	12
B. Architecture des réseaux de neurones profonds . . . . .	13
1.4.4 Limites des approches mono-tâche de l'IA . . . . .	13
1.4.5 L'apprentissage multi-tâches (Multi-Task Learning, MTL) . . . . .	14
A. Principe de base du MTL . . . . .	14
B. Fonctionnement du MTL . . . . .	14
C. Types d'apprentissage multi-tâches . . . . .	15
D. Avantages et défis du MTL . . . . .	15
1.5 Conclusion . . . . .	15
<b>Chapitre 2: État de l'art sur les travaux dédiés aux SR</b> . . . . .	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Revue de la littérature . . . . .	17
2.2.1 Articles de synthèse sur les SR . . . . .	17
2.2.2 Articles de synthèse sur les approches hybrides en SR . . . . .	18
2.2.3 Articles de synthèse sur les systèmes de recommandation basés sur l'intelligence artificielle . . . . .	19

## TABLE DES MATIÈRES

---

2.2.4	Articles de synthèse sur l'intelligence artificielle appliquée au filtrage collaboratif et aux approches par graphes . . . . .	20
2.2.5	Articles de synthèse sur les systèmes de recommandation basés exclusivement sur l'intelligence artificielle . . . . .	21
2.3	Conclusion . . . . .	30
<b>Chapitre 3: Approche proposée : système de recommandation hybride avec apprentissage multi-tâches (HMTL-RS)</b>		<b>31</b>
3.1	Introduction . . . . .	31
3.2	Approche proposée : système de recommandation hybride avec apprentissage multi-tâche intégré (HMTL-RS) . . . . .	31
3.2.1	Première étape : Collecte et prétraitement des données . . . . .	32
3.3.1.1	Collecte de données . . . . .	32
3.3.1.2	Prétraitement des données . . . . .	33
3.2.2	Deuxième étape : Système de recommandation basé sur le contenu . . . . .	38
3.2.3	Troisième étape : Système de recommandation hybride . . . . .	39
3.2.4	Quatrième étape : HMTL-RS . . . . .	40
3.3.4.1	Première version : MTL à trois tâches . . . . .	40
3.3.4.2	Évolution vers HMTL-RS : un MTL à cinq tâches intégré à un système hybride . . . . .	41
3.3	Conclusion . . . . .	42
<b>Chapitre 4: Évaluation des performances de HMTL-RS</b>		<b>43</b>
4.1	Introduction . . . . .	43
4.2	Plateformes et outils de développement . . . . .	43
4.2.1	Environnement de développement . . . . .	43
4.2.2	Langage de programmation . . . . .	44
4.2.3	Bibliothèques de Python . . . . .	44
4.3	Métriques d'évaluation de l'approche HMTL-RS . . . . .	45
4.4	Évaluation 1 : Calcul de métriques pour HTML-RS . . . . .	46
4.4.1	Analyse des courbes d'apprentissage . . . . .	46

## TABLE DES MATIÈRES

---

4.5	Évaluation 2 : Comparaison de HMTL-RS avec un modèle monotâche . . . . .	48
4.6	Conclusion . . . . .	50
	<b>Conclusion générale</b>	<b>51</b>
	<b>References</b>	<b>52</b>

## Table des figures

1.1	Modèle classique d'un SR basé sur CF [1] . . . . .	4
1.2	Modèle classique de CBF [1] . . . . .	5
1.3	Modèle classique de KBRS [1] . . . . .	6
1.4	Modèle classique de HRS [1] . . . . .	7
1.5	Architecture hiérarchique des domaines de l'intelligence artificielle [2] . . . . .	11
1.6	Diagramme schématique d'un réseau de neurones [3] . . . . .	13
3.1	Schéma récapitulatif de l'architecture du HMTL-RS . . . . .	32
3.2	Capture d'écran du dataset initial Airline Dataset.csv . . . . .	33
3.3	Capture d'écran du dataset Airline _Dataset _Modifie.csv avec les variables syn- thétiques . . . . .	34
3.4	Capture d'écran du second dataset train.csv . . . . .	34
3.5	Capture d'écran du dataset après la fusion Airline _Dataset _Final.csv . . . . .	35
3.6	Capture d'écran du dataset Airline _Dataset _Final.csv avec les nouvelles va- riables aléatoires . . . . .	36
3.7	Capture d'écran de la suite du dataset Airline _Dataset _Final.csv contenant les variables aléatoires . . . . .	36
3.8	Capture d'écran de la première partie du dataset d'entraînement . . . . .	37
3.9	Capture d'écran de la première partie du dataset de validation . . . . .	38
3.10	Capture d'écran de la première partie du dataset de test . . . . .	38
3.11	Capture d'écran de l'interface utilisateur . . . . .	42

## TABLE DES FIGURES

---

4.1	Présentation des courbes moyennes d'entraînement et de validation obtenues lors de l'évaluation du modèle HMTL-RS . . . . .	47
4.2	Histogramme comparatif des performances du HMTL-RS et du STL pour les différentes tâches . . . . .	49

## Liste des tableaux

1.1	Comparaison entre les différents types de systèmes de recommandation . . . . .	9
2.1	Synthèse complète des travaux sur les systèmes de recommandation . . . . .	29
4.1	Résumé des principales métriques d'évaluation pour HMTL-RS . . . . .	46
4.2	Résultats obtenus sur les deux premiers folds . . . . .	48

## Liste des acronymes

<b>A</b>	ACC	Accuracy
	AUC	Aire sous la courbe ROC
<b>C</b>	CF	Collaborative Filtering
	CBF	Content-Based Filtering
	CARS	Context-Aware Recommender System
	CNN	Convolutional Neural Networks
	$R^2$	Coefficient of determination
<b>D</b>	DL	Deep Learning
<b>G</b>	GAN	Generative Adversarial Networks
<b>H</b>	HRS	Hybrid Recommender Systems
	HMTL-RS	Hybrid and Multi-Task Learning-based Recommender System
	HSTL-RS	Hybrid and Single-Task Learning-based Recommender System
<b>I</b>	IA	Intelligence Artificielle
<b>K</b>	KBRS	Knowledge-Based Recommender Systems
<b>M</b>	ML	Machine Learning,
	MTL	Multi-Task Learning
	MSE	Mean Squared Error
	MAE	Mean Absolute Error
<b>R</b>	RNN	Recurrent Neural Networks,
	RMSE	Root Mean Squared Error
<b>S</b>	SR	Systèmes de Recommandations
	STL	Single-Task Learning

# Introduction générale

Dans un monde en constante évolution et qui est de plus en plus connecté, la prise de décision devient un processus complexe face à une multitude d'options disponibles. C'est dans ce contexte que sont apparus les systèmes de recommandation (SR), développés pour orienter les choix des utilisateurs en se basant sur leurs préférences et besoins, en particulier dans les domaines où les quantités de données sont importantes.

Les systèmes de recommandation ont pour objectif de suggérer des contenus ou des produits pertinents à un utilisateur en se basant sur ses préférences passées. Ces systèmes ont longtemps reposé sur des méthodes traditionnelles telles que le filtrage collaboratif, les approches basées sur le contenu ou les modèles hybrides. Cependant, ces techniques montrent aujourd'hui leurs limites en termes d'efficacité et de précision. En réponse à cela, de plus en plus de SR exploitent les avancées de l'intelligence artificielle (IA), notamment le machine learning et le deep learning. Ces approches permettent de générer des recommandations plus personnalisées, plus diversifiées et mieux adaptées aux besoins spécifiques des utilisateurs.

Cependant, malgré leur évolutivité, les systèmes de recommandation continuent de faire face à des défis majeurs tels que le démarrage à froid, la surspécialisation ou encore la sparsité des données qui limitent la qualité des recommandations proposées. Le démarrage à froid fait référence à la capacité du système à générer des recommandations pour de nouveaux utilisateurs ou de nouveaux éléments. La surspécialisation correspond au fait que le système propose des éléments souvent trop similaires à ceux déjà consommés, ce qui nuit à la diversité. Quant à la sparsité de données, elle est due au manque d'interactions enregistré entre les utilisateurs et les éléments.

Dans ce travail, nous proposons un système de recommandation hybride, intitulé Hybrid and Multi-Task Learning-based Recommender System (HMTL-RS), combinant le filtrage collaboratif, le filtrage basé sur le contenu et l'apprentissage multitâche (MTL). Notre modèle, MTL, repose sur le traitement de cinq tâches, à savoir la prédiction de la note utilisateur, l'estimation de la nouveauté, l'évaluation de la diversité, la mesure de similarité et la prise en compte du contexte utilisateur. L'intégration du MTL dans notre système a permis l'apprentissage conjoint de tâches complémentaires afin de générer des recommandations à la fois plus pertinentes, diversifiées et généralisables.

Le système HMTL-RS est basé sur la combinaison de plusieurs contributions clés, notamment :

- La collecte et le prétraitement des données en rassemblant et préparant des données provenant de plusieurs datasets, assurant ainsi leur qualité et leur pertinence.
- Mise en œuvre d'un SR avec filtrage basé sur le contenu, exploitant les caractéristiques explicites des vols et les préférences des utilisateurs pour générer des recommandations personnalisées.
- Développement d'un SR hybride combinant filtrage collaboratif et filtrage basé sur le

contenu, afin de pallier les limites observées en matière de diversité et de capacité de généralisation des recommandations.

- Renforcement du SR hybride avec une architecture d'apprentissage multitâche à cinq objectifs, permettant de surmonter les problèmes de démarrage à froid et de surspécialisation, tout en améliorant la personnalisation des recommandations et l'interaction utilisateur via une interface dédiée.

Pour l'évaluation de notre proposition, nous avons choisi de l'appliquer au domaine de la recommandation de voyage en raison de la disponibilité et de la richesse des données dans ce secteur.

Pour démontrer l'efficacité de notre approche, nous avons effectué une première évaluation de HMTL-RS basée sur le calcul des valeurs de différentes métriques standards telles que RMSE, MAE,  $R^2$ , précision, rappel et F1-score. Par la suite, nous avons implémenté un second SR nommé Hybrid Single-Task Learning-based Recommender System HSTL-RS dans lequel le MTL a été remplacé par un apprentissage STL monotâche (Single Task Learning). Contrairement au STL, qui se concentre sur l'apprentissage d'une seule tâche à la fois, le MTL permet une meilleure évaluation des performances en termes de qualité prédictive et d'efficacité du système sur des tâches multiples. Les résultats expérimentaux obtenus démontrent une amélioration significative confirmant la pertinence de l'intégration conjointe du MTL et de l'architecture hybride.

Ce mémoire est structuré en quatre chapitres :

- Le chapitre 1, intitulé "Systèmes de recommandations (SR) et intelligence artificielle (IA)" explore les concepts fondamentaux des systèmes de recommandation et leur évolution avec l'intelligence artificielle.
- Le chapitre 2, appelé "État de l'art sur les travaux dédiés aux SR" présente un état de l'art détaillé des travaux liés aux systèmes de recommandations.
- Le chapitre 3 intitulé "Approche proposée : système de recommandation hybride avec apprentissage multi-tâches intégré ((HMTL-RS))" décrit notre approche HMTL-RS, ses modules et les différentes contributions.
- Le chapitre 4 nommé "Évaluation des performances de HMTL-RS" consacré à l'évaluation expérimentale de notre système et à l'analyse des résultats obtenus.

# Chapitre 1

## Systemes de recommandations (SR) et intelligence artificielle (IA)

### 1.1 Introduction

Dans un monde où l'information et les options se multiplient, la prise de décision devient de plus en plus complexe. Pour faciliter ce processus, il est nécessaire de sélectionner une option parmi plusieurs alternatives en fonction d'objectifs, de contraintes et d'informations. C'est dans ce contexte que les systèmes de recommandation interviennent, en analysant les préférences et les comportements des utilisateurs pour fournir des suggestions personnalisées. Ces dernières années, nous remarquons l'intégration de l'intelligence artificielle dans les SR dans le but d'améliorer leurs performances. Ainsi, grâce à l'intelligence artificielle, ces systèmes ont été révolutionnés par l'utilisation d'algorithmes avancés comme l'apprentissage automatique et profond, ce qui a permis de retourner des recommandations plus adaptées aux besoins individuels des utilisateurs.

Ce chapitre présente les fondements des systèmes de recommandation et leur évolution avec l'intelligence artificielle, ainsi que les défis auxquels les SR sont confrontés.

### 1.2 Systemes de recommandations (SR)

Cette section explore les bases théoriques des SR et leur évolution récente.

#### 1.2.1 Définition des SR

Les systèmes de recommandations sont des outils basés sur des algorithmes et des techniques d'intelligence artificielle, élaborés pour assister la prise de décision, en particulier lorsque la quantité de données est importante. Ils fournissent des suggestions basées sur des informations personnalisées, afin de prédire et de proposer des services ou des informations pertinentes dans divers contextes [4, 5, 6, 7, 8, 9].

## 1.2.2 Types de systèmes de recommandation

Les systèmes de recommandation sont classés en plusieurs types, chacun adapté à des besoins spécifiques et exploitant les données de manière distincte. Pour répondre aux divers besoins des utilisateurs et aux contraintes des domaines d'application, plusieurs approches ont été développées. Nous citons dans ce qui suit quelques types de SR.

### A. SR basé sur le filtrage collaboratif (collaborative filtering, CF)

Le filtrage collaboratif est l'une des techniques les plus utilisées dans les systèmes de recommandation. Il consiste à faire des recommandations en se basant sur les similarités des sélections effectuées par différents utilisateurs, car il repose sur l'hypothèse selon laquelle des utilisateurs similaires ont les mêmes préférences [10, 11, 8, 12]. La figure 1.1 illustre un modèle classique de CF.

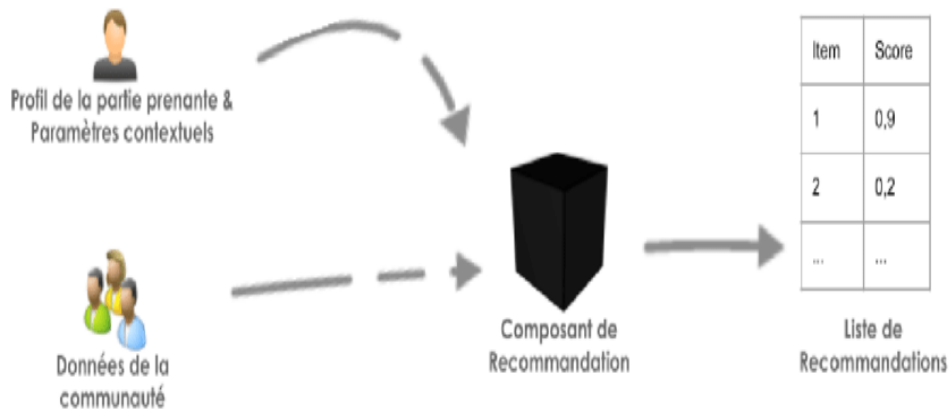


Fig. 1.1 – Modèle classique d'un SR basé sur CF [1]

Voici un exemple d'utilisation de SR basé sur le filtrage collaboratif pour les sites de commerce électronique. Quand deux utilisateurs achètent des couches pour bébés, ils se verront recommander des produits similaires à ceux achetés par d'autres utilisateurs ayant effectué des achats similaires. Cette approche se base également sur les notes attribuées aux produits et les compare entre les différents utilisateurs. Il existe deux classes principales de recommandation collaborative, basée sur la mémoire et basée sur le modèle.

**a) SR collaboratif basé sur la mémoire** : Deux catégories sont distinguées dans la littérature : les CF basés sur l'utilisateur et les CF basés sur les éléments.

#### Filtrage collaboratif basé sur l'utilisateur

Ce modèle repose sur l'idée que chaque utilisateur appartient à un groupe d'utilisateurs ayant des comportements et des préférences similaires. En se basant sur cette similitude, le système génère des recommandations et des suggestions pour l'utilisateur cible [10, 11, 8].

## Filtrage collaboratif basé sur les éléments

Ce modèle se base sur l'identification du contenu apprécié par l'utilisateur. Il calcule ensuite le degré de similarité entre ces différentes propriétés pour trouver de nouveaux produits similaires et les recommander à l'utilisateur. Ce système repose donc principalement sur les caractéristiques des objets et non sur les préférences des autres utilisateurs [10, 11, 8].

## b) SR collaboratif basé sur le modèle

Cette technique repose sur un modèle statistique pour évaluer les utilisateurs. Elle se base sur le calcul de la probabilité qu'un utilisateur apprécie un nouvel objet en s'appuyant sur les évaluations qu'il a déjà effectuées. Le modèle estime ainsi la probabilité d'intérêt pour un objet à partir des comportements passés de l'utilisateur et, parfois, de ceux d'utilisateurs similaires [10, 11, 8].

## B. SR basée sur le contenu (content-based filtering, CBF)

Cette technique de recommandation repose sur deux éléments : le profil de l'utilisateur et les caractéristiques intrinsèques des objets. Le système analyse les objets qu'un utilisateur a appréciés pour créer un profil personnalisé. En s'appuyant sur ce profil, qui contient les évaluations positives de l'utilisateur, le système compare les objets à recommander avec ceux déjà présents dans le profil, puis propose des recommandations pertinentes [8, 4, 9]. La figure 1.2 illustre un modèle classique de CBF.

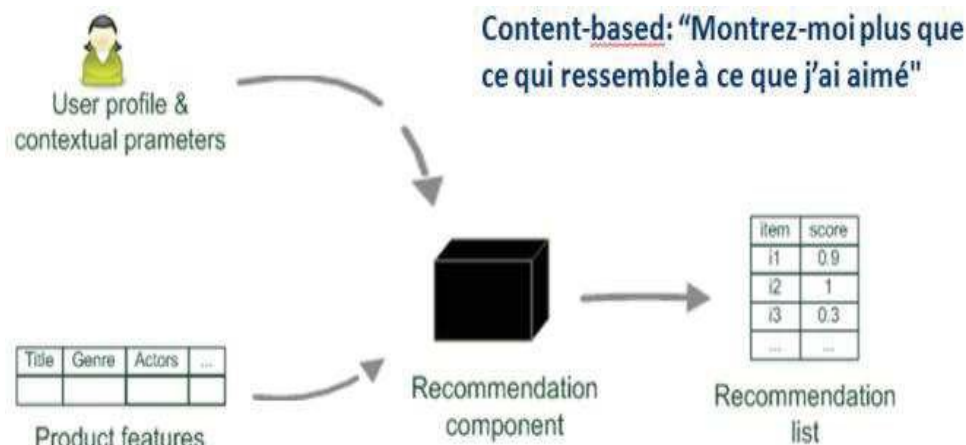


Fig. 1.2 – Modèle classique de CBF [1]

Nous distinguons deux catégories :

## a) CBF basées sur les attributs

Cette approche utilise les caractéristiques intrinsèques des éléments pour trouver une correspondance avec le profil de l'utilisateur. Ensuite, elle se concentre principalement sur chaque attribut en analysant ses différents aspects afin de fournir des recommandations pertinentes [8, 4, 9].

## b) CBF basé sur le raisonnement par cas

Contrairement aux approches basées sur les attributs, cette méthode n'analyse pas les attributs en détail, mais effectue une comparaison globale des nouveaux objets avec ceux appréciés dans le profil de l'utilisateur [8, 4, 9].

## C. SR basé sur les connaissances (Knowledge-Based Recommender Systems, KBRS)

Ce type de système de recommandation est fondé sur une base de connaissances établie à partir des préférences des utilisateurs, déclarées explicitement via des questions ou des formulaires. Cette méthode se concentre sur les besoins spécifiques des utilisateurs et recommande, en fonction de ces derniers, les objets les plus adaptés à leurs attentes [9, 8, 4, 12]. La figure 1.3 illustre un modèle classique de KBRS.

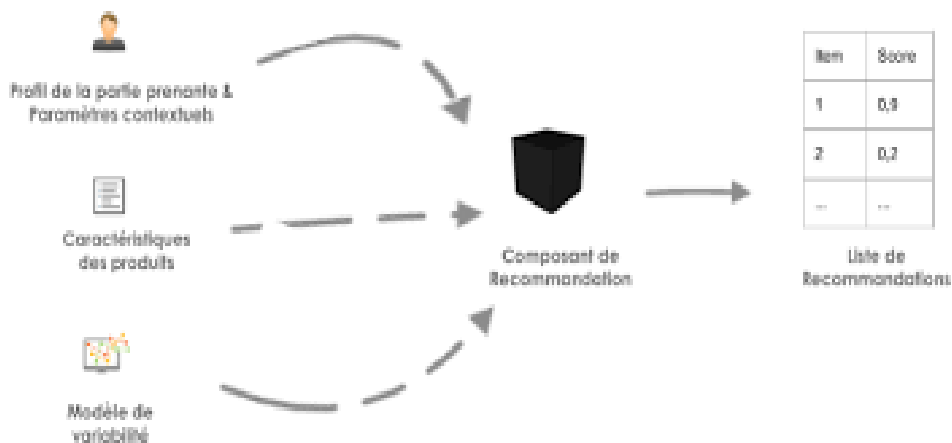


Fig. 1.3 – Modèle classique de KBRS [1]

## D. SR hybrides (Hybrid Recommender Systems, HRS)

Les systèmes hybrides jouent un rôle essentiel dans l'amélioration des performances des systèmes de recommandation. Ce type de système combine plusieurs méthodes afin d'exploiter les points forts de chaque méthode et de minimiser l'impact de leurs points faibles [4, 9, 5]. La figure 1.4 illustre un modèle classique de HRS.

## E. SR basé sur le contexte (Context-Aware Recommender System, CARS)

Les systèmes de recommandation basés sur le contexte, contrairement aux autres SR, ne se limitent pas aux préférences des utilisateurs ou aux caractéristiques des objets. Ils prennent également en compte des informations contextuelles pour générer des recommandations. Ces dispositifs modifient leurs propositions en fonction de divers paramètres environnementaux et individuels de l'utilisateur lors de l'échange, ce qui favorise l'intégration d'un plus grand nombre de critères dans le mécanisme de recommandation et renforce sa pertinence [4, 9, 5].

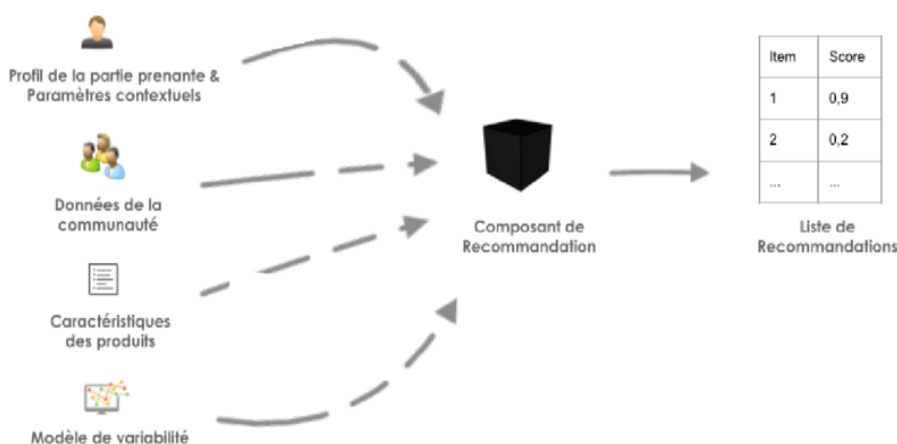


Fig. 1.4 – Modèle classique de HRS [1]

### 1.2.3 Limites et défis des systèmes de recommandation

Les systèmes de recommandation rencontrent plusieurs défis majeurs qui influencent leur performance et leur efficacité, parmi lesquels nous citons :

#### A. Le problème du démarrage à froid (Cold Start)

Le démarrage à froid est un problème récurrent dans les systèmes de recommandation lorsque le système manque de données [13, 4]. Ce problème se manifeste dans deux cas :

**Nouvel utilisateur :** Lorsqu'un nouvel utilisateur rejoint une plateforme, aucune information sur ses préférences n'est disponible tant qu'il n'a pas évalué ou aimé un article. Par conséquent, le système ne peut pas lui proposer des recommandations pertinentes [13, 4].

**Nouvel élément ou produit :** Lorsqu'un nouvel élément est ajouté sur le marché, mais qu'il n'a pas encore été évalué, il devient difficile de l'intégrer dans les recommandations du système [4, 13].

#### B. Le problème de scalabilité (scalability problem)

Le problème de scalabilité survient lorsque le nombre d'utilisateurs et d'éléments augmente. À mesure que ces volumes augmentent, la quantité de données à traiter croît, ce qui rend les calculs plus complexes à effectuer et exigeant en termes de stockage. Cela complique également la génération de recommandations, en particulier celles en temps réel [13, 14].

#### C. Le problème de la surspécialisation (overspecialization)

Ce problème réside dans le fait que les systèmes de recommandation proposent des suggestions trop similaires aux choix précédents des utilisateurs, ce qui limite la diversité des recommandations et la découverte de nouveaux produits. Cela rend les choix proposés aux utilisateurs vraiment restreints en termes d'options, ce qui peut mener à leur désintérêt en raison de la monotonie des recommandations [13, 15, 8].

### D. Le problème de sparsité des données (Data Sparsity Problem)

Ce problème survient généralement lorsque plusieurs utilisateurs n'évaluent pas la majorité des éléments ou ne les consultent même pas. Cette situation entraîne une précision réduite et une difficulté à prédire ce que les utilisateurs pourraient aimer en raison du manque d'évaluations et d'interactions entre les utilisateurs et les éléments. Cela conduit à un manque de personnalisation et à des suggestions moins pertinentes, obligeant souvent les systèmes à recommander uniquement des éléments populaires ou déjà bien évalués [13, 14, 8].

### E. Le problème de confidentialité et de sécurité des données (Privacy and Data Security Problem)

Les systèmes de recommandation collectent d'énormes quantités de données sur les utilisateurs, tant au niveau personnel que comportemental, afin d'effectuer des recommandations pertinentes. Cependant, cela expose ces données à des risques de piratage, mettant ainsi en danger la vie privée des utilisateurs. En conséquence, les utilisateurs peuvent éprouver des réticences à partager des volumes importants de données personnelles [13, 8, 16].

Plusieurs études ont proposé des comparaisons entre les différents SR existants [17, 18, 19, 20]. Ainsi, la table 1.1 offre une comparaison détaillée des divers SR discutés dans la section précédente.

### 1.2.4 Les domaines d'application des systèmes de recommandation

Les systèmes de recommandation sont utilisés dans divers domaines pour fournir des recommandations personnalisées et faciliter la prise de décision. Parmi ces domaines, nous trouvons :

- **Commerce électronique** Les SR augmentent les ventes en fournissant des recommandations personnalisées de produits basées sur l'historique d'achat et les préférences des utilisateurs.
- **Santé** Les systèmes de recommandation améliorent la prise en charge des patients et la recherche médicale en suggérant des plans de traitement ou de prévention personnalisés.
- **Tourisme** Les SR offrent des recommandations personnalisées sur les destinations, les hôtels et les activités selon les préférences des utilisateurs, leur historique et plusieurs autres critères tels que le budget et autre.
- **Agriculture** Les systèmes de recommandations sont utilisés dans ce domaine pour optimiser les récoltes ou encore pour suggérer des cultures adaptées au sol.
- **Films et livres** Recommandations de films, de séries et de livres selon les préférences des utilisateurs et l'historique de visionnage/lecture.

## 1.3 L'IA et les systèmes de recommandation

L'intégration de l'intelligence artificielle a transformé et optimisé les systèmes de recommandation. En s'appuyant sur des techniques sophistiquées comme l'apprentissage automatique

## CHAPITRE 1. SYSTÈMES DE RECOMMANDATIONS (SR) ET INTELLIGENCE ARTIFICIELLE (IA)

Type de SR	Avantages	Inconvénients
SR basée sur le contenu	<ul style="list-style-type: none"> <li>— Pas de dépendance aux données d'autres utilisateurs.</li> <li>— Recommandations explicites et transparentes.</li> <li>— Personnalisation forte.</li> <li>— Résistance au problème de démarrage à froid pour les utilisateurs.</li> <li>— Contrôle sur la diversité des recommandations.</li> </ul>	<ul style="list-style-type: none"> <li>— Manque de diversité dans les recommandations.</li> <li>— Dépendance aux informations de contenu.</li> <li>— Problème de surspécialisation.</li> <li>— Difficulté à saisir les préférences implicites.</li> <li>— Stagnation du contenu recommandé.</li> </ul>
Filtrage collaboratif	<ul style="list-style-type: none"> <li>— Basé sur les comportements réels des utilisateurs.</li> <li>— Facilité d'implémentation.</li> <li>— Aucune connaissance préalable des éléments nécessaire.</li> <li>— Adaptation dynamique aux préférences des utilisateurs.</li> <li>— Découverte de nouvelles préférences.</li> </ul>	<ul style="list-style-type: none"> <li>— Démarrage à froid.</li> <li>— Problème de scalabilité.</li> <li>— Sparsité des données.</li> <li>— Manque de diversité et risque de sur-ajustement.</li> </ul>
SR basée sur les connaissances	<ul style="list-style-type: none"> <li>— Précision sans dépendance aux données historiques.</li> <li>— Transparence et explicabilité.</li> <li>— Recommandations adaptées aux besoins spécifiques.</li> <li>— Pas de dépendance aux comportements passés.</li> <li>— Gestion des préférences complexes.</li> </ul>	<ul style="list-style-type: none"> <li>— Nécessité d'une base de connaissances riche et précise.</li> <li>— Difficulté à gérer les préférences implicites.</li> <li>— Limitation dans la découverte de nouveaux éléments.</li> <li>— Faible flexibilité face aux évolutions des préférences.</li> </ul>
Systèmes hybrides	<ul style="list-style-type: none"> <li>— Amélioration de la précision des recommandations.</li> <li>— Réduction des problèmes de démarrage à froid.</li> <li>— Meilleure gestion de la diversité des recommandations.</li> <li>— Plus de flexibilité.</li> <li>— Amélioration de la robustesse et de la fiabilité.</li> </ul>	<ul style="list-style-type: none"> <li>— Complexité accrue.</li> <li>— Coût de calcul élevé.</li> <li>— Difficulté d'équilibrage entre différentes méthodes.</li> <li>— Problème de personnalisation excessive.</li> <li>— Dépendance à la qualité des données.</li> </ul>
SR basée sur le contexte	<ul style="list-style-type: none"> <li>— Pertinence accrue des recommandations.</li> <li>— Adaptabilité au comportement et aux préférences évolutives.</li> <li>— Amélioration de l'expérience utilisateur.</li> <li>— Optimisation des performances en temps réel.</li> <li>— Réduction de la saturation des recommandations.</li> </ul>	<ul style="list-style-type: none"> <li>— Complexité de la collecte des données contextuelles.</li> <li>— Problèmes de confidentialité et de sécurité des données.</li> <li>— Besoins accrus en calcul et en stockage.</li> <li>— Complexité de la modélisation du contexte.</li> <li>— Dépendance à des contextes spécifiques.</li> </ul>

Tab. 1.1 – Comparaison entre les différents types de systèmes de recommandation

et l'apprentissage profond, l'IA permet de modéliser avec précision les préférences des utilisateurs, d'analyser divers types de données et d'adapter les recommandations. Ces techniques surpassent les limitations des approches classiques en offrant des prédictions plus exactes et des suggestions fortement personnalisées, tout en tenant compte du contexte et de l'évolution des comportements des utilisateurs. De ce fait, l'IA est devenue essentielle pour développer des systèmes de recommandation intelligents, performants et capables de s'ajuster dynamiquement aux attentes des utilisateurs [21, 22].

### 1.4 Intelligence artificielle

Cette section a pour objectif d'examiner les fondements conceptuels de l'intelligence artificielle ainsi que ses évolutions récentes.

#### 1.4.1 Définition et évolution de l'intelligence artificielle

L'intelligence artificielle a émergé durant les années cinquante avec des concepts théoriques posés par des pionniers tels qu'Alan Turing, qui s'est interrogé sur la capacité des machines à « penser ». Depuis, ce domaine n'a cessé d'évoluer, s'étendant progressivement à de nombreux secteurs. Cette évolution s'explique principalement par les avancées en puissance de calcul et par l'augmentation considérable des données numériques. Son impact croissant soulève à la fois des opportunités prometteuses et des défis éthiques majeurs.

L'IA englobe un large éventail de techniques visant à doter les machines de capacités cognitives comparables à celles de l'être humain, telles que l'apprentissage, le raisonnement, la perception ou encore la prise de décision. Ces techniques de l'IA reposent sur des approches variées, allant des approches classiques telles que les règles logiques, les systèmes experts, la planification automatique à l'apprentissage automatique et aux réseaux de neurones [21, 23, 24]. La figure 1.5 illustre l'architecture hiérarchique des domaines de l'intelligence artificielle.

#### 1.4.2 Apprentissage automatique (machine learning, ML)

L'apprentissage automatique, ou machine learning, est un domaine dynamique de l'intelligence artificielle. Elle repose sur la création d'algorithmes sophistiqués qui permettent aux ordinateurs d'acquérir des connaissances et de s'améliorer à partir de données brutes, sans nécessiter une programmation explicite pour chaque tâche. En exploitant des techniques statistiques et des algorithmes spécialisés, le machine learning donne aux machines la capacité d'identifier des motifs complexes et des structures significatives au sein de vastes ensembles de données. Ce champ de recherche et d'application a connu une expansion remarquable ces dernières années, transformant de nombreux aspects de notre quotidien [25, 26, 22]. L'apprentissage automatique repose sur quatre types principaux d'apprentissage, chacun ayant ses spécificités et ses applications pratiques :

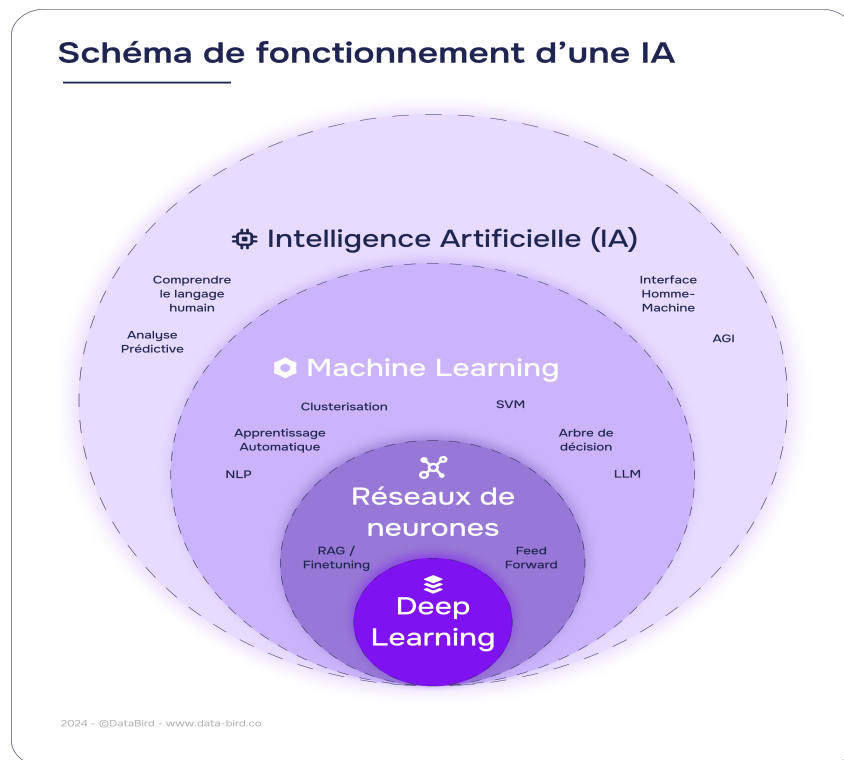


Fig. 1.5 – Architecture hiérarchique des domaines de l'intelligence artificielle [2]

## A. Apprentissage supervisé

Ce type d'apprentissage repose sur des données étiquetées, où chaque entrée est associée à une sortie connue. Le modèle apprend à partir de ces exemples pour prédire les sorties de nouvelles données non étiquetées. Il est principalement utilisé pour des tâches comme la classification, qui vise à attribuer à chaque observation une étiquette discrète parmi un nombre fini de classes, et la régression, qui vise à prédire une valeur numérique continue à partir de ces caractéristiques. Cette approche est particulièrement efficace lorsque les relations entre les données sont bien définies et exploitables ; on la retrouve notamment dans des algorithmes tels que les machines à vecteurs (SVM) ou les arbres de décisions (DT) [27, 28, 26].

## B. Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, qui utilise des données étiquetées, l'apprentissage non supervisé analyse des données brutes non étiquetées pour découvrir des structures cachées, des corrélations ou des regroupements. Il repose sur des techniques comme le clustering et la réduction de dimensionnalité pour extraire des informations significatives et révéler les caractéristiques essentielles des données [27, 28, 26].

## C. Apprentissage semi-supervisé

L'apprentissage semi-supervisé combine des données étiquetées (en petite quantité) et non étiquetées pour améliorer les performances du modèle. Il exploite les structures cachées dans

les données non étiquetées, offrant une solution efficace lorsque l'étiquetage est coûteux ou complexe, comme en médecine. Des techniques comme la propagation d'étiquettes permettent de tirer parti de ces données [27, 28, 26].

### D. Apprentissage par renforcement

L'apprentissage par renforcement consiste à ce qu'un agent apprenne à prendre des décisions optimales dans un environnement pour maximiser une récompense. Après chaque action, il reçoit un retour (récompense ou punition) qui ajuste son comportement afin d'améliorer ses performances progressivement. Particulièrement adapté aux environnements complexes et dynamiques, comme les jeux, la robotique ou la conduite autonome, il excelle dans la prise de décision séquentielle et l'adaptation en temps réel [27, 28, 26].

#### 1.4.3 Apprentissage profond (deep learning, DL)

Le deep learning constitue une branche avancée de l'apprentissage automatique qui se distingue par l'utilisation de réseaux de neurones artificiels complexes constitués de nombreuses couches interconnectées. Ces architectures, inspirées d'une représentation simplifiée du cerveau humain et utilisant des réseaux neuronaux profonds avec de multiples couches cachées, apprennent des représentations hiérarchiques des données, traitant efficacement des informations complexes comme les images, le texte et l'audio [29, 30, 31].

### A. Couches des réseaux de neurones

Un réseau de neurones standard est généralement organisé en trois types de couches fondamentales :

- **Couche d'entrée** Elle reçoit les données initiales, généralement sous forme de vecteurs numériques, comme des pixels pour une image ou des mots pour un texte [30, 29, 28].
- **Couches cachées** Ces couches internes effectuent des transformations non linéaires sur les données, permettant au modèle d'apprendre des caractéristiques complexes à divers niveaux d'abstraction. Un réseau de neurones profond peut inclure des dizaines, voire des centaines, de ces couches pour extraire des informations de plus en plus riches [29, 22, 28].
- **Couche de sortie** Elle produit le résultat final du modèle, qu'il s'agisse d'une prédiction, d'une classification ou d'une autre forme de réponse basée sur les données traitées [30, 29, 28]. La figure 1.6 illustre une représentation d'un réseau de neurones.

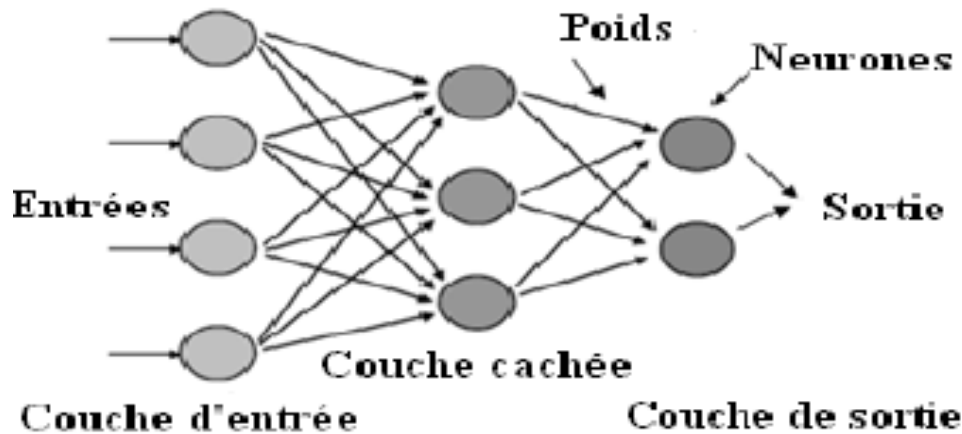


Fig. 1.6 – Diagramme schématisique d'un réseau de neurones [3]

## B. Architecture des réseaux de neurones profonds

Les réseaux de neurones profonds utilisent diverses architectures pour répondre à des besoins spécifiques. Parmi les architectures les plus couramment utilisées, nous citons :

**Les réseaux convolutifs (Convolutional Neural Networks, CNN) :** Les CNN sont des architectures de choix pour l'analyse d'images et la reconnaissance visuelle. Leur force réside dans l'utilisation de filtres convolutifs qui extraient des caractéristiques locales en examinant les pixels adjacents au sein d'une image. Ces réseaux neuronaux sont devenus omniprésents dans le domaine de la vision par ordinateur [22, 28, 31].

**Les réseaux récurrents (Recurrent Neural Networks, RNN) :** Les RNN sont conçus pour le traitement de données séquentielles telles que les séries temporelles, le texte ou l'audio. Grâce à leur mémoire interne, ils sont capables de capturer les relations et les dépendances qui existent entre les éléments qui se suivent dans une séquence [29, 28, 31].

**Les réseaux de neurones génératifs (Generative Adversarial Networks, GAN) :** Les GAN fonctionnent grâce à deux composants principaux : un générateur qui crée des données et un discriminateur qui tente de distinguer les données réelles de celles produites par le générateur. Les GAN sont fréquemment utilisés pour la création d'images réalistes, de vidéos et de morceaux de musique [22, 28, 31].

**Les transformateurs (Transformers)** Initialement conçus pour le traitement du langage naturel, les transformateurs sont idéaux pour l'identification des relations complexes entre les éléments d'une séquence, comme les mots dans une phrase, sans égard à leur ordre. Cette architecture a permis la création de modèles novateurs tels que GPT et BERT [29, 28, 31].

### 1.4.4 Limites des approches mono-tâche de l'IA

Bien que les approches existantes de l'intelligence artificielle, basées sur l'apprentissage mono-tâche, aient permis des avancées importantes, elles présentent des limitations notables. Premièrement, elles exigent souvent de grands volumes de données étiquetées pour chaque tâche, ce qui pose problème lorsque ces données sont rares ou coûteuses à obtenir. Deuxièmement,

mement, ces modèles sont peu flexibles et ont du mal à s'adapter à des contextes nouveaux ou évolutifs, limitant leur capacité de généralisation. De plus, ils ne peuvent généralement pas réutiliser les connaissances acquises d'une tâche à une autre, ce qui restreint leur polyvalence et leur efficacité dans divers environnements. Enfin, malgré leurs performances croissantes, ces systèmes restent inférieurs aux capacités cognitives humaines et sont souvent mis en difficulté face à des situations imprévues ou ambiguës [21, 29]. Ainsi, pour combler ces lacunes, le multi-task learning se présente comme une approche prometteuse.

### 1.4.5 L'apprentissage multi-tâches (Multi-Task Learning, MTL)

L'apprentissage multi-tâches est une approche IA novatrice qui consiste à entraîner un seul modèle pour réaliser plusieurs tâches simultanément. Contrairement aux méthodes classiques où chaque tâche est isolée, le MTL met à profit les liens entre les tâches pour partager les connaissances et optimiser les performances [32, 33, 34]. En partageant ces connaissances, le MTL améliore la généralisation, diminue le surapprentissage et optimise l'usage des données, surtout quand certaines tâches ont peu d'exemples étiquetés. Il réduit aussi les coûts de développement et renforce la robustesse face à divers environnements. Aujourd'hui, le MTL est largement appliqué dans le traitement du langage naturel, la vision par ordinateur et la bioinformatique, contribuant à des systèmes intelligents plus adaptables et efficaces [35, 34, 36].

#### A. Principe de base du MTL

Dans le MTL, un modèle unique est entraîné pour exécuter plusieurs tâches simultanément. Pour ce faire, les premières couches du réseau, souvent appelées couches partagées (shared layers), apprennent des représentations générales qui sont communes à l'ensemble des tâches. Ces représentations servent de base commune pour toutes les tâches. Par la suite, des couches spécifiques à chaque tâche (task-specific layers) sont utilisées pour capturer les caractéristiques propres à chacune d'entre elles [35, 36]. L'architecture MTL favorise le transfert de connaissances entre les différentes tâches, permettant d'améliorer la généralisation du modèle, notamment pour les tâches où les données disponibles sont limitées. En mutualisant les informations apprises dans les couches partagées tout en adaptant les couches spécifiques, le MTL renforce non seulement l'efficacité globale du modèle, mais aussi sa robustesse face à des contextes variés [34, 33].

#### B. Fonctionnement du MTL

Le Multi-Task Learning repose sur l'optimisation simultanée de plusieurs fonctions de perte, chacune associée à une tâche spécifique. La fonction de perte globale du modèle est généralement une combinaison pondérée de ces pertes individuelles, permettant de réguler l'importance accordée à chaque tâche durant l'apprentissage. Cette pondération peut être fixe (définie manuellement) ou dynamique, avec des méthodes ajustant automatiquement les poids en fonction de facteurs comme la difficulté ou la vitesse de convergence de chaque tâche [32, 34].

En pratique, le MTL peut être implémenté à travers différentes architectures, parmi lesquelles on distingue principalement :

**Le partage strict des paramètres (hard parameter sharing)** Les couches initiales du réseau sont entièrement communes à toutes les tâches, seules les couches finales dédiées à chaque tâche sont distinctes [35, 36].

**Le partage souple des paramètres (soft parameter sharing)** Chaque tâche possède son propre modèle, mais des techniques de régularisation sont employées pour encourager la similarité entre les paramètres de ces modèles [35, 36].

**Les architectures modulaires ou dynamiques (modular and dynamic architectures)** qui modifient la structure du modèle en fonction des interdépendances entre les tâches [35, 36].

En résumé, le MTL est une approche d'apprentissage automatique qui peut utiliser les modèles de deep learning, notamment les réseaux de neurones, pour entraîner un seul modèle qui permettra de résoudre plusieurs tâches (régression et classification) simultanément afin d'améliorer les performances globales.

### C. Types d'apprentissage multi-tâches

L'apprentissage multi-tâches se présente sous plusieurs formes (supervisé, non supervisé, par renforcement), selon les tâches et l'architecture du modèle. Structurellement, il peut impliquer un partage complet ou partiel des couches, ou utiliser des mécanismes d'attention pour gérer les interactions complexes entre les tâches. Le MTL s'applique à des tâches similaires (classification d'images) ou différentes (classification + régression), offrant une grande adaptabilité aux besoins concrets [33, 35].

### D. Avantages et défis du MTL

Le multi-task learning présente de nombreux bénéfices, notamment une amélioration de la capacité de généralisation des modèles, une réduction du risque de surapprentissage, ainsi qu'une meilleure exploitation des données, particulièrement dans les situations où certaines tâches disposent de ressources limitées. En partageant des représentations communes, le modèle peut capitaliser sur des informations complémentaires issues des différentes tâches, renforçant ainsi ses performances globales. Cependant, cette approche implique plusieurs défis, tels que la sélection judicieuse des tâches à regrouper, la gestion équilibrée de leurs contributions respectives ainsi que la résolution des conflits qui peuvent survenir entre des tâches faiblement corrélées. Par conséquent, la réussite du MTL dépend d'une conception rigoureuse de l'architecture du modèle et d'une définition cohérente des objectifs d'apprentissage [34, 35].

## 1.5 Conclusion

L'intelligence artificielle constitue le pilier central des systèmes de recommandation modernes, permettant d'automatiser et d'optimiser le processus de suggestion afin d'offrir une

## **CHAPITRE 1. SYSTÈMES DE RECOMMANDATIONS (SR) ET INTELLIGENCE ARTIFICIELLE (IA)**

---

expérience utilisateur plus personnalisée et efficace. En exploitant des techniques avancées, ces SR parviennent à s'adapter dynamiquement aux préférences et aux besoins des utilisateurs, améliorant ainsi la pertinence et la diversité des recommandations. Cependant, leur déploiement soulève des défis majeurs, tant sur le plan technique qu'éthique, notamment en ce qui concerne la transparence des algorithmes, la gestion des biais et la protection des données personnelles. Afin de mieux comprendre ces enjeux et d'analyser les approches existantes, le chapitre suivant sera consacré à un état de l'art détaillé des systèmes de recommandation, mettant en lumière leurs évolutions, leurs limites et les solutions proposées dans la littérature.

## Chapitre 2

# État de l'art sur les travaux dédiés aux SR

### 2.1 Introduction

Ce chapitre constitue une partie fondamentale du mémoire, car il permet de situer notre travail par rapport aux approches et méthodologies développées dans le domaine des systèmes de recommandation. Son objectif est de fournir une vision et une critique des approches utilisées dans les systèmes de recommandation. En examinant les travaux antérieurs, nous cherchons à identifier les forces et les limites des solutions existantes, ainsi qu'à mettre en lumière les opportunités d'innovation. Cette analyse nous permettra de définir les bases théoriques et techniques sur lesquelles s'appuiera notre propre système de recommandation.

### 2.2 Revue de la littérature

Afin de mieux cerner les avancées récentes et les différentes approches employées dans le domaine des systèmes de recommandation, nous présentons certains travaux de la littérature.

#### 2.2.1 Articles de synthèse sur les SR

Nous commençons cette section par la présentation de quelques états de l'art sur les SR. Athira et Selvam (2018) [17] analysent les principales approches utilisées dans les systèmes de recommandation pour l'e-learning. Ils examinent comment ces méthodes – filtrage collaboratif, filtrage basé sur le contenu, systèmes hybrides et techniques d'intelligence artificielle (réseaux de neurones, algorithmes évolutionnaires) – améliorent la personnalisation de l'apprentissage. L'étude compare ces approches selon leur performance, précision, complexité et adaptabilité, en soulignant l'efficacité des systèmes hybrides. L'article propose enfin une synthèse comparative, utile aux chercheurs et concepteurs de plateformes e-learning.

Nawara et Kashef (2021) [37] proposent une synthèse sur les systèmes de recommandation sensibles au contexte dans l'IoT, appelés IoT-CARS. Ils montrent comment l'intégration d'informations contextuelles (localisation, temps, préférences) améliore la pertinence des recommandations. Trois approches sont présentées : le filtrage précontexte, le filtrage postcontexte

et la modélisation du contexte. Les méthodes de modélisation (graphiques, objets, ontologies) sont comparées selon la complexité et l'évolutivité. L'étude souligne les avantages des IoT-CARS dans des domaines comme la santé ou l'éducation, tout en pointant des défis tels que l'hétérogénéité des données, la confidentialité et le démarrage à froid.

Dans leur étude, Rathod et al. (2024) [38] s'intéressent aux systèmes de recommandation e-learning pour les personnes autistes. Ils analysent les approches existantes et soulignent des défis majeurs, tels que l'absence de principes de conception adaptés et les limites techniques. Les approches sont classées en trois catégories : filtrage basé sur le contenu, filtrage collaboratif et méthodes hybrides. Les auteurs examinent aussi l'usage d'ontologies et de techniques contextuelles pour mieux répondre aux besoins spécifiques des apprenants autistes. L'article conclut sur l'importance de ces systèmes pour une personnalisation efficace, tout en rappelant les défis liés à l'évolutivité, la confidentialité et le démarrage à froid.

Nous passons maintenant à l'étude de l'article publié par Jaiswal et al. (2020) [39], qui présente CRSAS, un système de recommandation collaboratif destiné au secteur agricole indien. Ce système vise à faciliter l'accès des agriculteurs aux aides, subventions et informations agricoles adaptées à leurs besoins. Fonctionnant via une application web, CRSAS exploite deux approches de filtrage collaboratif : l'une centrée sur l'utilisateur, et l'autre sur les éléments recommandés. Les étapes principales incluent la création de profils, l'analyse des requêtes et la génération de recommandations. Le système se distingue par sa capacité à atténuer le problème de démarrage à froid, tout en améliorant l'accès aux ressources gouvernementales.

### 2.2.2 Articles de synthèse sur les approches hybrides en SR

Certains travaux proposent des approches hybrides, tels que [40, 41, 12]. Tout d'abord, nous présentons un état de l'art sur les SR hybrides pour le commerce électronique, publié par Bodduri et al. en 2024 [41]. Cette étude vise à identifier les tendances, méthodes et défis des systèmes hybrides, notamment face à la rareté des données, au démarrage à froid et à la précision. Les auteurs suivent une méthodologie en quatre étapes : formulation des questions, sélection des études, synthèse des données et évaluation critique. Trois axes guident l'analyse : l'évolution des approches hybrides, la diversité des combinaisons d'algorithmes et les problèmes traités. L'étude met en avant l'essor de l'apprentissage profond et la combinaison fréquente du filtrage collaboratif avec le filtrage basé sur le contenu, souvent renforcés par des approches basées sur la connaissance ou le contexte. Les auteurs concluent que les SR hybrides sont prometteurs pour le commerce électronique, tout en soulignant les défis liés à la complexité algorithmique et à l'optimisation à grande échelle.

Yin et al. présentent dans [40] (2022) une approche pour les systèmes de recommandation séquentielle, intégrant les préférences des utilisateurs à court et long terme avec des informations spatiotemporelles (localisation et moment). Ils proposent le modèle UTSR, basé sur des réseaux GRU et un mécanisme d'attention, pour capturer ces dépendances. UTSR prend en compte les besoins immédiats et passés tout en adaptant les recommandations au contexte spatiotemporel. Les expérimentations sur des jeux de données réels montrent une nette amélioration des perfor-

mances (HR@K et NDCG@K) par rapport aux méthodes existantes. Les auteurs concluent que leur approche favorise une personnalisation plus fine, utile notamment pour les réseaux sociaux et le commerce électronique.

### 2.2.3 Articles de synthèse sur les systèmes de recommandation basés sur l'intelligence artificielle

Dans la dernière partie de cet état de l'art, nous présentons des travaux [42, 43, 44, 12, 45] qui exploitent les outils de l'intelligence artificielle dans les SR, une tendance croissante ces dernières années en raison de leurs capacités à améliorer les performances de ces systèmes. Nous détaillons dans ce qui suit ces travaux, en commençant par le travail de Gündüz [42] publié en 2023, qui analyse l'impact croissant de l'intelligence artificielle sur l'aviation et l'hôtellerie. L'étude montre comment ces secteurs utilisent l'IA pour personnaliser les services, optimiser les prix, automatiser la gestion et améliorer l'interaction client via des chatbots et des recommandations intelligentes. Grâce à ses capacités prédictives, l'IA permet d'adapter les offres en temps réel et d'anticiper les besoins des voyageurs. L'auteur conclut que l'IA est un levier stratégique clé pour la transformation numérique, tout en soulevant des défis liés à l'éthique, à la protection des données et à l'adoption technologique.

Priyanka et al. (2023) proposent dans [43] une approche novatrice pour l'agriculture intelligente, combinant l'Internet des objets (IoT) et le machine learning afin de développer un système de recommandation de cultures adapté aux conditions environnementales en temps réel. Leur objectif est d'aider les agriculteurs à prendre des décisions éclairées en fonction de paramètres tels que l'humidité, la température et le pH du sol. L'étude montre comment les capteurs IoT, associés à des modèles prédictifs, permettent de recommander les cultures les plus adaptées à chaque situation locale. Les auteurs analysent également les systèmes existants, en mettant en avant les bénéfices en termes de précision, de durabilité et de rendement, tout en signalant des défis liés à la connectivité, aux coûts et à l'adoption technologique. Ils concluent que l'intégration de l'IoT et du machine learning constitue une avancée majeure pour l'agriculture de précision, contribuant à améliorer la productivité tout en répondant aux enjeux de durabilité et de sécurité alimentaire.

Dans cet article [44], Li et al. (2023) proposent une revue systématique des systèmes de recommandation basés sur l'apprentissage profond, avec une classification des différentes approches et techniques utilisées. Les auteurs explorent comment les modèles d'apprentissage profond, tels que les réseaux de neurones convolutifs, les réseaux de neurones récurrents et d'autres architectures avancées, sont appliqués pour améliorer les performances des systèmes de recommandation. L'objectif est d'identifier les avantages et les limites des techniques actuelles tout en mettant en lumière les défis à relever dans ce domaine. Les auteurs proposent une classification des modèles en trois catégories principales : les modèles de filtrage collaboratif, les modèles basés sur le contenu et les modèles hybrides. Ils abordent également l'utilisation spécifique des CNN pour l'extraction de caractéristiques à partir de données visuelles et des RNN pour modéliser les séquences temporelles dans les préférences des utilisateurs. L'étude

conclut que l'utilisation des modèles d'apprentissage profond permet d'améliorer la précision et la personnalisation des recommandations par rapport aux méthodes traditionnelles. Cependant, les auteurs soulignent que des défis demeurent, notamment concernant l'évolutivité, le coût computationnel élevé et la gestion de la protection des données personnelles des utilisateurs.

### 2.2.4 Articles de synthèse sur l'intelligence artificielle appliquée au filtrage collaboratif et aux approches par graphes

Certains papiers combinent le filtrage collaboratif avec des techniques de l'IA : dans [12] les auteurs Peng et al. (2024) proposent un système de recommandation de films combinant le filtrage collaboratif et l'apprentissage par renforcement profond (deep reinforcement learning, DRL) afin d'améliorer la précision et la personnalisation, notamment pour les nouveaux utilisateurs. Leur modèle intègre trois composantes clés : la factorisation matricielle via la décomposition en valeurs singulières (SVD) pour extraire des représentations latentes des utilisateurs et des films ; la méthode acteur-critique du DRL pour optimiser les recommandations séquentielles ; et l'algorithme Deep Deterministic Policy Gradient (DDPG) pour un apprentissage adaptatif et continu, particulièrement efficace pour les recommandations Top-N. L'évaluation sur le dataset MovieLens, selon la précision, le rappel et le F1 score, montre que le modèle surpasse les approches traditionnelles, en offrant des recommandations ciblées même avec peu de données. Les auteurs soulignent que le DRL permet de générer des recommandations initiales pertinentes et de les ajuster au fil des interactions. Toutefois, des défis persistent, notamment en matière de personnalisation avancée, de scalabilité et de gestion du démarrage à froid.

Tandis que dans [45] (2021), Dadoun et al. explorent l'application des systèmes de recommandation pour personnaliser les offres aériennes et améliorer l'expérience utilisateur. Ils analysent comment l'exploitation des préférences clients, du contexte de voyage et de l'historique d'interaction permet aux compagnies aériennes de proposer des offres plus ciblées. Leur approche repose sur un modèle hybride combinant filtrage collaboratif, filtrage basé sur le contenu et algorithmes de machine learning, afin de générer des recommandations dynamiques, sensibles aux saisons, aux comportements individuels et aux stratégies commerciales. Contrairement aux méthodes classiques de tarification dynamique, cette approche met l'accent sur une personnalisation fine des services. L'étude conclut que ces systèmes renforcent la conversion des offres, la satisfaction client et l'efficacité commerciale, tout en soulignant la nécessité d'une adaptation continue aux évolutions technologiques et aux spécificités du secteur aérien.

Au lieu de combiner l'IA avec les SR basiques, Dadoun et al. (2021) présentent dans [46] une méthode pour les systèmes de recommandation de destinations touristiques, combinant l'apprentissage multi-tâches et les graphes de connaissances. Leur approche vise à prédire plusieurs cibles liées au choix de voyage, en exploitant les relations complexes entre les entités touristiques. L'architecture repose sur une modélisation par graphe de connaissances intégrée à un cadre d'apprentissage multi-tâches. Elle fusionne les interactions utilisateurs avec des informations sémantiques sur les destinations, grâce à des embeddings partagés qui capturent les liens

entre utilisateurs, lieux, activités et préférences. Les résultats montrent que cette combinaison améliore à la fois la précision des recommandations et leur personnalisation. Cette approche innovante ouvre la voie à des systèmes de recommandation plus intelligents, capables de mieux comprendre la complexité des préférences des voyageurs et de s'adapter aux profils variés.

### 2.2.5 Articles de synthèse sur les systèmes de recommandation basés exclusivement sur l'intelligence artificielle

D'autres travaux proposent des approches basées exclusivement sur l'IA, comme dans [47] Kan et al. (2024) proposent un système de recommandation de vols personnalisé combinant prédiction de liens et apprentissage profond, afin d'optimiser la pertinence des suggestions tout en tenant compte de la fiabilité des vols (retards, annulations). Leur modèle hybride s'appuie sur plusieurs composantes : la prédiction de liens pour identifier des profils utilisateurs similaires et filtrer les vols, un algorithme K-Means pour structurer les données aériennes, et un réseau neuronal convolutif pour estimer les risques de perturbation en fonction de données contextuelles (météo, historiques, caractéristiques opérationnelles). Les résultats de ces prédictions alimentent un mécanisme de pondération ajustant les recommandations selon la fiabilité estimée. Les expériences montrent une précision de 95,13 % pour la prédiction des perturbations, surpassant les méthodes existantes, et une précision de 72,31 % pour les recommandations, soit une amélioration notable de 15,6%. L'étude conclut que cette approche allie efficacement personnalisation et fiabilité, tout en soulignant l'importance d'une adaptation continue face aux spécificités du secteur aérien.

Tejas et al. (2023) dans [48] explorent l'usage de l'IA pour automatiser la création d'itinéraires et optimiser la réservation de vols et d'hébergements. En s'appuyant sur des modèles de traitement du langage naturel (comme GPT-3) et des algorithmes de machine learning, ils génèrent des suggestions personnalisées en temps réel selon la destination, le budget et les préférences de l'utilisateur. Implémentée dans une application web MERN (MongoDB, Express.js, React, Node.js), leur solution agrège des données provenant d'API de réservation, de tendances du marché et d'avis clients pour affiner ses recommandations. Les résultats montrent que cette approche permet de créer rapidement des itinéraires personnalisés, plus efficaces que les méthodes traditionnelles. L'étude conclut que l'intégration de l'IA simplifie la planification de voyages tout en améliorant la précision des recommandations, tout en rappelant les défis liés à la qualité des données et à l'évolution constante des préférences des utilisateurs.

J. Schoonderbeek (2020) [49] propose une approche de prévision du trafic aérien basée sur l'apprentissage profond multi-tâches (Deep Multi-Task Learning, DMTL). Cette méthode permet de prédire simultanément plusieurs variables liées aux passagers, comme le nombre de voyageurs par vol ou itinéraire, en exploitant les corrélations entre ces tâches.

L'étude met en avant l'efficacité du DMTL pour améliorer la précision des prévisions et optimiser l'apprentissage de modèles complexes, en capturant mieux les tendances spatio-temporelles des déplacements aériens. Les résultats montrent des performances nettement supérieures aux modèles classiques, une meilleure utilisation des données et une forte adaptabilité aux variations

des itinéraires et des saisons. Ce travail souligne ainsi le potentiel du DMTL pour la prévision en aviation, en fournissant des outils robustes pour optimiser la gestion du transport aérien et appuyer la prise de décision stratégique.

Dans [50], Ding et al. (2024) introduisent une approche innovante pour les systèmes de recommandation, combinant apprentissage profond, apprentissage fédéré et apprentissage multi-tâches. Leur modèle, nommé PF-MSMTrec, vise à prédire plusieurs objectifs simultanément, tels que le taux de clics (CTR) et le taux de conversion après clic (CTCVR), tout en garantissant la confidentialité des données réparties entre différents clients. Reposant sur l'architecture multi-gate mixture-of-experts (MMoE), enrichie par des techniques de séparation des paramètres et d'agrégation personnalisée, leur approche réduit les conflits d'optimisation entre les tâches et scénarios, ce qui améliore la précision globale. L'étude met en évidence l'efficacité de cette méthode en matière d'optimisation multi-tâches, de protection des données via l'apprentissage fédéré et de gestion équilibrée des objectifs grâce à des mécanismes tels que la normalisation par lots fédérée. Cette approche ouvre la voie à des systèmes de recommandation plus robustes, respectueux de la vie privée et adaptés à des environnements complexes et multi-scénarios.

Gu et al. (2022) proposent dans [51] une méthode innovante pour estimer la localisation géographique d'images touristiques sans métadonnées, en combinant apprentissage profond et apprentissage multi-tâches. Le modèle GLML (Geolocation via Multi-task Learning) intègre simultanément deux tâches : la classification et la recherche d'images similaires, afin d'estimer la position géographique d'une image. Il exploite des techniques de multi-global pooling pour extraire des caractéristiques visuelles robustes et représentatives, améliorant ainsi la précision de la localisation. Les expérimentations menées sur sept ensembles de données publics démontrent l'efficacité du modèle GLML, surpassant les méthodes existantes en termes de précision de localisation. Cette approche ouvre de nouvelles perspectives pour les systèmes de recommandation de voyages, en permettant de suggérer des destinations basées sur des images partagées sans informations géographiques explicites.

Wu et al. (2025) [52] introduisent un modèle de recommandation pour les vidéos courtes, associant apprentissage multi-tâches et analyse des comportements utilisateurs. Leur approche vise à prédire simultanément plusieurs objectifs, comme l'engagement et la satisfaction, en intégrant les interactions passées et les préférences contextuelles. L'architecture repose sur un cadre multi-tâches, combiné à des mécanismes d'analyse comportementale, qui ajustent dynamiquement les recommandations selon les habitudes des utilisateurs. Le modèle exploite des embeddings pour représenter les préférences, tout en prenant en compte des facteurs contextuels comme l'heure ou les tendances sociales. Les résultats montrent que cette approche améliore nettement la personnalisation, en proposant des recommandations mieux adaptées aux profils et aux comportements individuels. Ce modèle constitue une avancée importante pour les systèmes de recommandation de vidéos, en renforçant leur capacité à anticiper et répondre aux attentes des utilisateurs.

Bouni et al. (2022) proposent dans [53] un système de recommandation agricole fondé sur

l'apprentissage par renforcement profond (Deep Reinforcement Learning, DRL). Ce système vise à aider les agriculteurs à sélectionner les cultures les mieux adaptées aux caractéristiques du sol et aux conditions environnementales. L'étude met en avant l'efficacité du DRL pour optimiser la prise de décision en tenant compte de multiples facteurs, tels que le climat et les propriétés du sol. Les résultats montrent une nette amélioration de la précision des recommandations par rapport aux méthodes traditionnelles, avec des suggestions plus personnalisées et adaptées aux environnements dynamiques, où les conditions changent fréquemment. Grâce à sa capacité d'apprentissage continu, le modèle s'adapte efficacement aux variations, résolvant des défis classiques comme la sélection optimale des cultures et la variabilité des rendements. Ce système représente ainsi une avancée majeure pour l'agriculture intelligente, favorisant des pratiques plus durables et des rendements optimisés.

Nous clôturons cette section avec un tableau comparatif 2.1 des articles étudiés, mettant en évidence les différences et les similitudes entre les approches proposées.

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Rathod et al., 2024	Éducation spécialisée	Systèmes adaptés aux troubles autistiques avec approche hybride	Ontologies, modélisation contextuelle, Adaptation personnalisée	Besoins spécifiques des personnes autistes (TSA) Problème du démarrage à froid, confidentialité	Personnalisation complexe, manque de données.
Priyanka et al., 2023	Agriculture intelligente	Recommandation intelligente pour les cultures par ML	Machine Learning, IoT, Analyse de données environnementales	Conditions météorologiques imprévisibles, variabilité des sols	Qualité des données, disponibilité et précision des données IoT, efficacité énergétique.
Kan et al., 2024	Aviation	Recommandation personnalisée de vols avec prédiction de liens, CNN	Réseaux neuronaux convolutifs (CNN), clustering K-Means, Graphes de données	Complexité des données aériennes, critères multiples	Scalabilité pour grands ensembles de données, précision des prédictions, gestion des perturbations.

*Suite à la page suivante.*

Tab. 2.1 – Suite

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Jaiswal et al., 2020	Agriculture	Filtrage collaboratif	Profils utilisateurs, requêtes agricoles, modèles de similarité utilisateur-produit	Manque d'accès aux ressources, sensibilisation aux programmes	Problème du démarrage à froid, accessibilité des programmes et manque de diversité dans les recommandations.
Dadoun et al., 2021	Aviation	Personnalisation des offres aériennes grâce à l'approche hybride	Modèles d'apprentissage automatique, NDC	Contrôle des offres par les GDS externes, personnalisation limitée	Adoption technologique, flexibilité des modèles et intégration aux systèmes existants.
Tejas et al. 2023	Aviation et planification de voyages	Automatisation des recommandations à l'aide de ML et de GPT-3	Technologies MERN, API de réservation	Complexité de la planification, préférences évolutives	Qualité des données, complexité des préférences et adaptation aux préférences dynamiques.

*Suite à la page suivante.*

Tab. 2.1 – Suite

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Gündüz, 2023	Aviation et Hôtellerie	Recommandation contextuelle et analyse thématique via IA	IA, NLP, chatbots, optimisation des prix	Manque de personnalisation, gestion manuelle des services	Adoption limitée, défis éthiques.
Athira et Selvam, 2018	E-learning	Revue systématique et comparative sur les SR en e-learning	Comparaison de modèles de recommandation	Diverse user profiles, data protections	Efficacité des recommandations et absence de tests empiriques approfondis.
Yin et al., 2022	IA et personnalisation	Recommandation séquentielle	Réseaux de neurones récurrents (GRU), mécanismes d'attention	Précision des recommandations séquentielles, adaptation aux préférences	Complexité du modèle.
Nawara et Kashef, 2021	IoT	Approche contextuelle	Systèmes adaptatifs et modélisation IoT	Intégration des recommandations en fonction du contexte IoT	Problèmes de démarrage à froid, protection de la vie privée.
Li et al., 2023	IA et Systèmes de recommandation	Revue systématique des approches basées sur le deep learning	Réseaux neuronaux, architectures CNN et RNN	Classification des SR utilisant le Deep Learning	Difficulté d'implémentation à grande échelle.

*Suite à la page suivante.*

Tab. 2.1 – Suite

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Bodduluri et al.	E-commerce	Revue systématique des systèmes hybrides	Combinaison de filtrage collaboratif et basé sur le contenu	Problème du démarrage à froid, rareté des données	Scalabilité et mise à jour des modèles.
Bouni et al., 2022	Agriculture intelligente	Recommandation optimisée novatrice	Deep Reinforcement Learning (DRL)	Amélioration de la précision des recommandations agricoles	Contraintes de données et d'apprentissage.
Peng et al., 2024	Films /Divertissement	Système hybride innovant	Deep RL et filtrage collaboratif	Optimisation, amélioration de la précision des recommandations	Temps de calcul élevé, complexité des modèles.
J.Schoonderbeek, 2020	Aviation	Apprentissage multi-tâches pour la prévision de passagers	Réseaux neuronaux profonds, Multi-Task Learning (MTL)	Problème du démarrage à froid (cold-start), prévision de la demande passagers, saisonnalité, erreurs de prévision	Manque de données historiques précises, gestion des données peu représentatives.

*Suite à la page suivante.*

Tab. 2.1 – Suite

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Ding et al., 2024	Recommandation hybride	Apprentissage fédéré et multi-tâches pour recommandations personnalisées	Apprentissage profond, multitâches, apprentissage fédéré (FedAvg)	Protection des données privées, optimisation multi-tâches, confidentialité des utilisateurs, démarrage à froid, difficulté d'agrégation de données	Adaptation aux évolutions rapides du marché, efficacité de la fédération.
Dadoun et al., 2021	Tourisme	Recommandation de destinations touristiques avec Multi-task Learning	Graphes de connaissances, Multi-Task Learning (MTL), Embeddings partagés	Prédiction de destinations touristiques, personnalisation des recommandations, démarrage à froid	Complexité des graphes de connaissances, besoin de données complètes.
Gu et al., 2022	Recommandation touristique	Géolocalisation d'images et recommandation touristique avec MTL	Réseaux neuronaux, Multi-Task Learning (MTL), Embeddings	Prédiction de la géolocalisation des images, démarrage à froid, données manquantes	Besoin de grandes quantités de données étiquetées, imprécision des géolocalisations sans métadonnées.

*Suite à la page suivante.*

Tab. 2.1 – *Suite*

Référence	Domaine	Approche	Méthodes utilisées	Problèmes abordés	Limites
Wu et al., 2025	Médias sociaux	Recommandation de vidéos avec prise en compte du comportement utilisateur	Multi-Task Learning, Embeddings, Réseaux neuronaux profonds (DNN)	Personnalisation des recommandations vidéo, gestion de comportements complexes, diversité des préférences	Problème de démarrage à froid, gestion de la diversité des préférences.

Tab. 2.1 – Synthèse complète des travaux sur les systèmes de recommandation

Dans la comparaison fournie dans la table 2.1, nous avons examiné plusieurs genres de SR en mettant en avant leurs points forts et points faibles dans différents champs d'application. Bien que ces systèmes soient couramment exploités dans des secteurs tels que le commerce électronique, l'IoT, l'agriculture et l'éducation, notre étude s'est orientée vers leur application dans l'aviation et, plus précisément, la recommandation de vols. Ces orientations répondent aux défis spécifiques du secteur, tels que l'ajustement aux préférences des voyageurs.

### 2.3 Conclusion

Dans ce chapitre, nous avons mené une analyse approfondie des différents systèmes de recommandation (SR) présents dans la littérature, en mettant en évidence leurs avantages et leurs limites selon les domaines d'application. Les systèmes fondés sur l'apprentissage profond se sont révélés particulièrement performants en matière de personnalisation et d'adaptation au contexte utilisateur, tout en soulevant des défis liés à la complexité computationnelle et à la scalabilité. Au vu des approches examinées, nous avons choisi de concentrer notre attention sur les systèmes de recommandation appliqués au domaine de l'aviation, et plus spécifiquement à la recommandation de vols. Ce choix oriente la suite de notre travail, qui portera dans le prochain chapitre sur la méthodologie adoptée pour concevoir et développer notre solution.

## **Chapitre 3**

# **Approche proposée : système de recommandation hybride avec apprentissage multi-tâches (HMTL-RS)**

### **3.1 Introduction**

Les voyageurs sont souvent confrontés à une multitude d’options de vols, ce qui rend la phase de sélection complexe et peu adaptée à leurs préférences et besoins spécifiques. Afin de résoudre cette problématique, nous proposons une approche intitulée Hybrid and Multi-Task Learning-based Recommender System (HMTL-RS). Dans ce chapitre, nous présenterons en détail notre approche qui combine un système de recommandation hybride (filtrage collaboratif et basé sur le contenu) avec une architecture d’apprentissage multi-tâche. Cette combinaison permet de tirer parti des avantages de chaque méthode tout en améliorant la capacité de généralisation du modèle grâce à l’apprentissage simultané de plusieurs fonctions pertinentes. Le système est appliqué au domaine de la recommandation de vols afin de fournir aux voyageurs des suggestions plus précises, personnalisées et adaptées à leurs attentes.

### **3.2 Approche proposée : système de recommandation hybride avec apprentissage multi-tâche intégré (HMTL-RS)**

L’approche que nous présentons, nommée HMTL-RS (System Recommendation – Hybride Multi-Tâche pour les vols), repose sur une architecture hybride intégrant plusieurs paradigmes de recommandation. Elle combine les apports du filtrage collaboratif, du filtrage basé sur le contenu ainsi que de l’apprentissage multi-tâche dans le but de générer des recommandations à la fois personnalisées, diversifiées et robustes en prenant en compte les contraintes pratiques et les préférences des utilisateurs. Cette architecture a été développée de manière progressive au fil de nos travaux suivant une logique incrémentale visant à améliorer continuellement la qua-

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

lité et les performances du système. Chaque étape a permis d'introduire de nouveaux modules ou d'optimiser les composants existants, allant d'un simple système basé sur le contenu à un système multi-tâche intégrant des représentations vectorielles (embedding) apprises automatiquement. La figure 3.1 représente les différentes phases de cette construction.

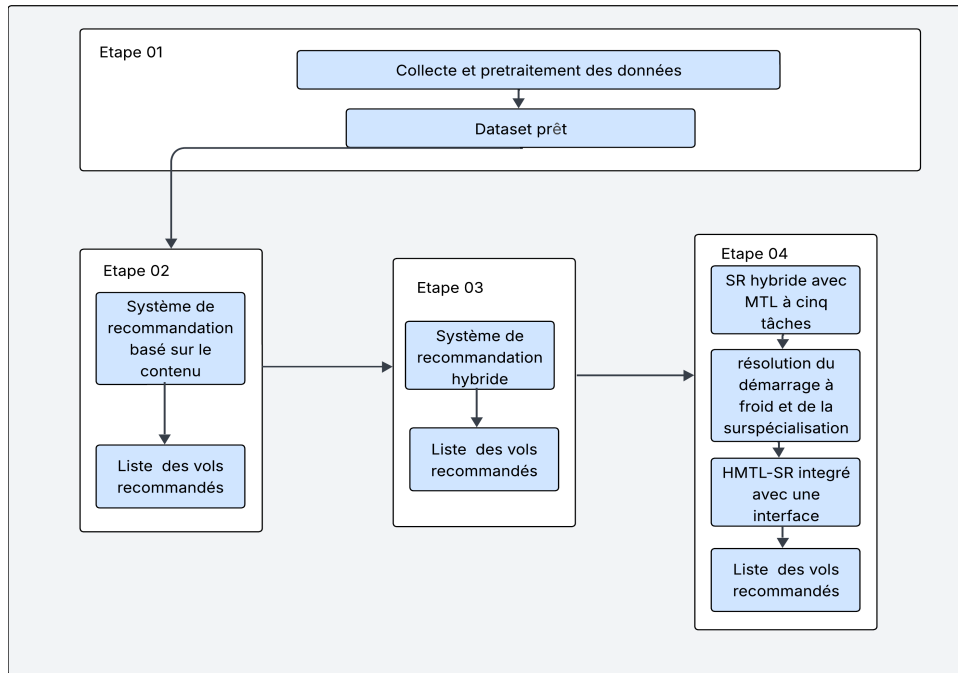


Fig. 3.1 – Schéma récapitulatif de l'architecture du HMTL-RS

### 3.2.1 Première étape : Collecte et prétraitement des données

Dans cette partie, nous nous concentrons sur la collecte de données et leur prétraitement.

#### 3.3.1.1 Collecte de données

En l'absence d'un dataset unique répondant pleinement aux besoins nous permettant de réaliser notre approche, nous avons exploité plusieurs datasets accessibles sur la plateforme Kaggle, présentant des informations pertinentes relatives aux passagers aériens et à leurs préférences de voyage. Ces datasets ont été fusionnés et harmonisés en s'appuyant sur des colonnes communes telles que l'âge, le genre, la classe du vol et le type de vol. L'objectif de cette collecte était d'enrichir les données initiales en combinant des informations objectives (retards, durées, distance) avec des informations subjectives (satisfaction, note des utilisateurs), afin de constituer un dataset représentatif et adapté aux exigences de notre approche. Ce processus a été réalisé à l'aide des bibliothèques pandas et numpy sous Python, assurant une manipulation efficace et cohérente des données tout au long du pipeline de préparation.

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

### 3.3.1.2 Prétraitement des données

Dans cette partie, nous allons présenter les différentes étapes de prétraitement des données utilisées dans HMTL-RS.

#### A. Sélection et renommage des colonnes

La première étape qui a été effectuée sur le dataset (Airline Dataset.csv) issu de Kaggle consiste à définir un dictionnaire de correspondance pour sélectionner les attributs utiles (tels que l'identifiant du passager, le nom, le prénom, le genre, l'âge, la date de départ, l'aéroport d'arrivée, le statut du vol, etc.) et à les renommer en français. La figure 3.2 représente le dataset initial avant le prétraitement.



Passenger ID	First Name	Last Name	Gender	Age	Nationality	Airport Name	Airport Country Code	Country Name	Airport Continent	Continents	Departure Date	Arrival Airport	Pilot Name	Flight Status	
0	10856	Edithe	Leggis	Female	62	Japan	Coldfoot Airport	US	United States	NAM	North America	6/28/2022	CXF	Edithe Leggis	On Time
1	43872	Elwood	Catt	Male	62	Nicaragua	Kugluktuk Airport	CA	Canada	NAM	North America	12/26/2022	YCO	Elwood Catt	On Time
2	42633	Darby	Felgate	Male	67	Russia	Grenoble-Isère Airport	FR	France	EU	Europe	1/18/2022	GNB	Darby Felgate	On Time
3	78493	Dominica	Pyle	Female	71	China	Ottawa / Gatineau Airport	CA	Canada	NAM	North America	9/16/2022	YND	Dominica Pyle	Delayed
4	82072	Bay	Pencost	Male	21	China	Gillespie Field	US	United States	NAM	North America	2/25/2022	SEE	Bay Pencost	On Time
5	39630	Lora	Durbann	Female	55	Brazil	Coronel Horácio de Mattos Airport	BR	Brazil	SAM	South America	6/10/2022	LEC	Lora Durbann	On Time
6	11940	Rand	Bram	Male	73	Ivory Coast	Duxford Aerodrome	GB	United Kingdom	EU	Europe	10/30/2022	QFO	Rand Bram	Cancelled
7	26470	Perceval	Dalosso	Male	36	Vietnam	Maestro Wilson Fonseca Airport	BR	Brazil	SAM	South America	4/7/2022	STM	Perceval Dalosso	Cancelled
8	29447	Aleda	Pigram	Female	35	Palestinian Territory	Venice Marco Polo Airport	IT	Italy	EU	Europe	8/20/2022	VCE	Aleda Pigram	On Time
9	75035	Burлие	Schustl	Male	13	Thailand	Vermilion Airport	CA	Canada	NAM	North America	4/6/2022	YVG	Burлие Schustl	On Time

Fig. 3.2 – Capture d'écran du dataset initial Airline Dataset.csv

#### B. Génération de variables synthétiques

Pour mieux adapter le dataset initial (Airline Dataset.csv) à nos besoins, certaines colonnes ont été générées aléatoirement à l'aide du module random() et de range(), telles que le nombre de places disponibles, l'heure de départ et d'arrivée, le prix du vol et le nom de la compagnie aérienne, ce qui génère le nouveau dataset nommé Airline\_Dataset\_Modifie.csv. La figure 3.3 illustre le dataset initial après les premières modifications.

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

[5]:

ID passager	Prenom	Nom de famille	Genre	Age	Airport Name	Code pays de Airport	Nom du pays	Continent de Airport	Date de depart	Arrival Airport	Statut du vol	Heure de depart	Heure d'arrivee	Prix du vol	Places disponibles	Nom de la compagnie	
0	10856	Edithe	Leggis	Female	62	Coldfoot Airport	US	United States	NAM	6/28/2022	CXF	On Time	18:21	13:00	1176	23	Air Horizon
1	43872	Elwood	Catt	Male	62	Kugluktuk Airport	CA	Canada	NAM	12/26/2022	YCO	On Time	20:23	01:06	910	188	SkyWorld
2	42633	Darby	Felgate	Male	67	Grenoble-Isère Airport	FR	France	EU	1/18/2022	GNB	On Time	00:41	19:38	1180	53	Air Horizon
3	78493	Dominica	Pyle	Female	71	Ottawa / Gatineau Airport	CA	Canada	NAM	9/16/2022	YND	Delayed	00:45	22:05	1145	52	AeroPlus
4	82072	Bay	Pencost	Male	21	Gillespie Field	US	United States	NAM	2/25/2022	SEE	On Time	03:49	05:50	1094	123	FlyNova
5	39630	Lora	Durbann	Female	55	Coronel Horácio de Mattos Airport	BR	Brazil	SAM	6/10/2022	LEC	On Time	00:42	12:58	171	167	FlyNova
6	11940	Rand	Bram	Male	73	Duxford Aerodrome	GB	United Kingdom	EU	10/30/2022	QFO	Cancelled	22:26	20:48	516	33	Vol Express
7	26470	Perceval	Dalosso	Male	36	Maestro Wilson Fonseca Airport	BR	Brazil	SAM	4/7/2022	STM	Cancelled	15:48	13:13	380	151	SkyWorld
8	29447	Aleda	Pigram	Female	35	Venice Marco Polo Airport	IT	Italy	EU	8/20/2022	VCE	On Time	12:42	03:11	137	71	Vol Express
9	75035	Burlie	Schustl	Male	13	Vermilion Airport	CA	Canada	NAM	4/6/2022	YVG	On Time	07:51	23:14	1173	128	FlyNova

Fig. 3.3 – Capture d’écran du dataset Airline\_Dataset\_Modifie.csv avec les variables synthétiques

### C. Fusion des données provenant de différentes sources

Afin d’améliorer le dataset Airline\_Dataset\_Modifie.csv, nous l’avons fusionné avec un second fichier train.csv. Cette fusion nous a permis d’intégrer à Airline\_Dataset\_Modifie.csv des colonnes supplémentaires telles que la classe du vol, la durée estimée du vol (calculée à partir de la distance), le type de vol (affaires ou personnel) et la note utilisateur (interprétée depuis la variable de satisfaction). Les correspondances entre les datasets ont été réalisées via des colonnes communes (âge, genre, classe du vol, type de vol), et les doublons ont été supprimés avec la méthode `drop_duplicates()` après la fusion, générant ainsi le nouveau dataset Airline\_Dataset\_Final.csv. La figure 3.4 représente le second dataset utilisé, tandis que la figure 3.5 elle, représente le dataset après la combinaison de train.csv et Airline\_Dataset\_Modifie.csv.

Unnamed: 0	id	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanlin	
0	0	70172	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	...	5	4	3	4	4	5
1	1	5047	Male	disloyal Customer	25	Business travel	Business	235	3	2	...	1	1	5	3	1	4
2	2	110028	Female	Loyal Customer	26	Business travel	Business	1142	2	2	...	5	4	3	4	4	4
3	3	24026	Female	Loyal Customer	25	Business travel	Business	562	2	5	...	2	2	5	3	1	4
4	4	119299	Male	Loyal Customer	61	Business travel	Business	214	3	3	...	3	3	4	4	3	3
5	5	111157	Female	Loyal Customer	26	Personal Travel	Eco	1180	3	4	...	1	3	4	4	4	4
6	6	82113	Male	Loyal Customer	47	Personal Travel	Eco	1276	2	4	...	2	3	3	4	3	5
7	7	96462	Female	Loyal Customer	52	Business travel	Business	2035	4	3	...	5	5	5	5	4	5
8	8	79485	Female	Loyal Customer	41	Business travel	Business	853	1	2	...	1	1	2	1	4	1
9	9	65725	Male	disloyal Customer	20	Business travel	Eco	1061	3	3	...	2	2	3	4	4	3

10 rows x 25 columns

Fig. 3.4 – Capture d’écran du second dataset train.csv

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

ID passager	Prenom	Nom de famille	Genre	Age	Airport Name	pays de Airport	Nom du pays	Continent de Airport	Date de depart	Arrival Airport	Statut du vol	Heure de depart	Heure d'arrivee	Prix du vol	Places disponibles	Nom de la compagnie	Classe du vol	
0	10856	Edithe	Leggis	Female	62	Coldfoot Airport	US	United States	NAM	6/28/2022	CXF	On Time	18:21	13:00	1176	23	Air Horizon	Eco Plus
1	43872	Elwood	Catt	Male	62	Kugluktuk Airport	CA	Canada	NAM	12/26/2022	YCO	On Time	20:23	01:06	910	188	SkyWorld	Business
2	42633	Darby	Felgate	Male	67	Grenoble-Isère Airport	FR	France	EU	1/18/2022	GNB	On Time	00:41	19:38	1180	53	Air Horizon	Business
3	78493	Dominica	Pyle	Female	71	Ottawa / Gatineau Airport	CA	Canada	NAM	9/16/2022	YND	Delayed	00:45	22:05	1145	52	AeroPlus	Business
4	82072	Bay	Pencost	Male	21	Gillespie Field	US	United States	NAM	2/25/2022	SEE	On Time	03:49	05:50	1094	123	FlyNova	Business
5	39630	Lora	Durbann	Female	55	Coronel Horácio de Mattos Airport	BR	Brazil	SAM	6/10/2022	LEC	On Time	00:42	12:58	171	167	FlyNova	Eco
6	11940	Rand	Bram	Male	73	Duxford Aerodrome	GB	United Kingdom	EU	10/30/2022	QFO	Cancelled	22:26	20:48	516	33	Vol Express	Eco
7	26470	Perceval	Dalosso	Male	36	Maestro Wilson Fonseca Airport	BR	Brazil	SAM	4/7/2022	STM	Cancelled	15:48	13:13	380	151	SkyWorld	Business
8	29447	Aleda	Pigram	Female	35	Venice Marco Polo Airport	IT	Italy	EU	8/20/2022	VCE	On Time	12:42	03:11	137	71	Vol Express	Business
9	75035	Burle	Schustl	Male	13	Vermillion Airport	CA	Canada	NAM	4/6/2022	YVG	On Time	07:51	23:14	1173	128	FlyNova	Eco

Fig. 3.5 – Capture d'écran du dataset après la fusion `Airline_Dataset_Final.csv`

### D. Estimation de la durée du vol

Une fonction de conversion a été utilisée pour estimer la durée du vol à partir de la distance parcourue, en supposant une vitesse moyenne équivalente à 500 km/h. Cela a permis de convertir une donnée brute en une information exploitable pour la recommandation.

### E. Conversion et génération de la satisfaction utilisateur

Après la fusion des datasets, seuls certains voyageurs disposent d'une évaluation binaire de l'expérience utilisateur (satisfait/neutre ou insatisfait), tandis que d'autres n'avaient aucune information liée à la satisfaction. Afin d'unifier cette variable et de pouvoir l'exploiter, nous avons d'abord converti les valeurs binaires en note utilisateur continue (entre 1.0 et 5.0); ensuite, nous avons généré aléatoirement des notes pour les utilisateurs qui ne disposent pas de retour exprimé. Cela permet d'avoir une variable continue à utiliser dans les tâches prédictives du MTL. Les figures 3.6 et 3.7 illustrent le dataset après l'ajout des nouvelles variables aléatoires.

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

ID passager	Prenom	Nom de famille	Genre	Age	Airport Name	Code pays de Airport	Nom du pays	Continent de Airport	Date de depart	Heure d'arrivee	Prix du vol	Places disponibles	Nom de la compagnie	Classe du vol	Type de vol	Numéro du vol	
0	10856	Edithe	Leggis	Female	62	Coldfoot Airport	US	United States	NAM	6/28/2022	13:00	1176	23	Air Horizon	Eco Plus	Personal Travel	FL5497
1	43872	Elwood	Catt	Male	62	Kugluktuk Airport	CA	Canada	NAM	12/26/2022	01:06	910	188	SkyWorld	Business	Business travel	FL6701
2	42633	Darby	Felgate	Male	67	Grenoble-Isère Airport	FR	France	EU	1/18/2022	19:38	1180	53	Air Horizon	Business	Business travel	FL8454
3	78493	Dominica	Pyle	Female	71	Ottawa / Gatineau Airport	CA	Canada	NAM	9/16/2022	22:05	1145	52	AeroPlus	Business	Business travel	FL4412
4	82072	Bay	Pencost	Male	21	Gillespie Field	US	United States	NAM	2/25/2022	05:50	1094	123	FlyNova	Business	Business travel	FL4579
5	39630	Lora	Durbann	Female	55	Coronel Horácio de Mattos Airport	BR	Brazil	SAM	6/10/2022	12:58	171	167	FlyNova	Eco	Personal Travel	FL7212
6	11940	Rand	Bram	Male	73	Duxford Aerodrome	GB	United Kingdom	EU	10/30/2022	20:48	516	33	Vol Express	Eco	Personal Travel	FL3173
7	26470	Perceval	Dalosso	Male	36	Maestro Wilson Fonseca Airport	BR	Brazil	SAM	4/7/2022	13:13	380	151	SkyWorld	Business	Business travel	FL3217
8	29447	Aleda	Pigram	Female	35	Venice Marco Polo Airport	IT	Italy	EU	8/20/2022	03:11	137	71	Vol Express	Business	Business travel	FL2746
9	75035	Burfi	Schusti	Male	13	Vermilion Airport	CA	Canada	NAM	4/6/2022	23:14	1173	128	FlyNova	Eco	Business travel	FL3047
10	29278	Porty	Jori	Male	39	Nuevo Casas Airport	MX	Mexico	NAM	5/27/2022	08:34	921	172	Vol Express	Eco	Business travel	FL4318

Fig. 3.6 – Capture d'écran du dataset Airline \_Dataset\_Final.csv avec les nouvelles variables aléatoires

Nom de famille	Genre	Age	Airport Name	Code pays de Airport	Nom du pays	Continent de Airport	Date de depart	Heure d'arrivee	Prix du vol	Places disponibles	Nom de la compagnie	Classe du vol	Type de vol	Numéro du vol	Type de trajet	Durée du vol	Note utilisateur
Leggis	Female	62	Coldfoot Airport	US	United States	NAM	6/28/2022	13:00	1176	23	Air Horizon	Eco Plus	Personal Travel	FL5497	Avec escale	0 h 55 min	2.0
Catt	Male	62	Kugluktuk Airport	CA	Canada	NAM	12/26/2022	01:06	910	188	SkyWorld	Business	Business travel	FL6701	Avec escale	0 h 28 min	1.4
Felgate	Male	67	Grenoble-Isère Airport	FR	France	EU	1/18/2022	19:38	1180	53	Air Horizon	Business	Business travel	FL8454	Direct	2 h 17 min	4.1
Pyle	Female	71	Ottawa / Gatineau Airport	CA	Canada	NAM	9/16/2022	22:05	1145	52	AeroPlus	Business	Business travel	FL4412	Direct	1 h 7 min	2.1
Pencost	Male	21	Gillespie Field	US	United States	NAM	2/25/2022	05:50	1094	123	FlyNova	Business	Business travel	FL4579	Avec escale	0 h 25 min	4.6
Durbann	Female	55	Coronel Horácio de Mattos Airport	BR	Brazil	SAM	6/10/2022	12:58	171	167	FlyNova	Eco	Personal Travel	FL7212	Direct	2 h 21 min	3.5
Bram	Male	73	Duxford Aerodrome	GB	United Kingdom	EU	10/30/2022	20:48	516	33	Vol Express	Eco	Personal Travel	FL3173	Direct	2 h 33 min	2.3
Dalosso	Male	36	Maestro Wilson Fonseca Airport	BR	Brazil	SAM	4/7/2022	13:13	380	151	SkyWorld	Business	Business travel	FL3217	Avec escale	4 h 4 min	4.8
Pigram	Female	35	Venice Marco Polo Airport	IT	Italy	EU	8/20/2022	03:11	137	71	Vol Express	Business	Business travel	FL2746	Avec escale	1 h 42 min	2.8
Schusti	Male	13	Vermilion Airport	CA	Canada	NAM	4/6/2022	23:14	1173	128	FlyNova	Eco	Business travel	FL3047	Direct	2 h 7 min	2.6
Jori	Male	39	Nuevo Casas Airport	MX	Mexico	NAM	5/27/2022	08:34	921	172	Vol Express	Eco	Business travel	FL4318	Direct	2 h 21 min	2.9

Fig. 3.7 – Capture d'écran de la suite du dataset Airline \_Dataset\_Final.csv contenant les variables aléatoires

### F. Combinaison et amélioration de données pour le dataset

Après avoir utilisé Airline\_Dataset\_Final.csv et entraîné notre modèle MTL avec ce dernier, nous avons rencontré un problème de surapprentissage (overfitting). Ce problème est principalement dû à la quantité limitée de données qui ne permettait pas au modèle de généraliser correctement. Pour y remédier, nous avons intégré un nouveau dataset supplémentaire nommé trainn.csv. L'ensemble des étapes précédentes (nettoyage, sélection des colonnes, renommage,

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

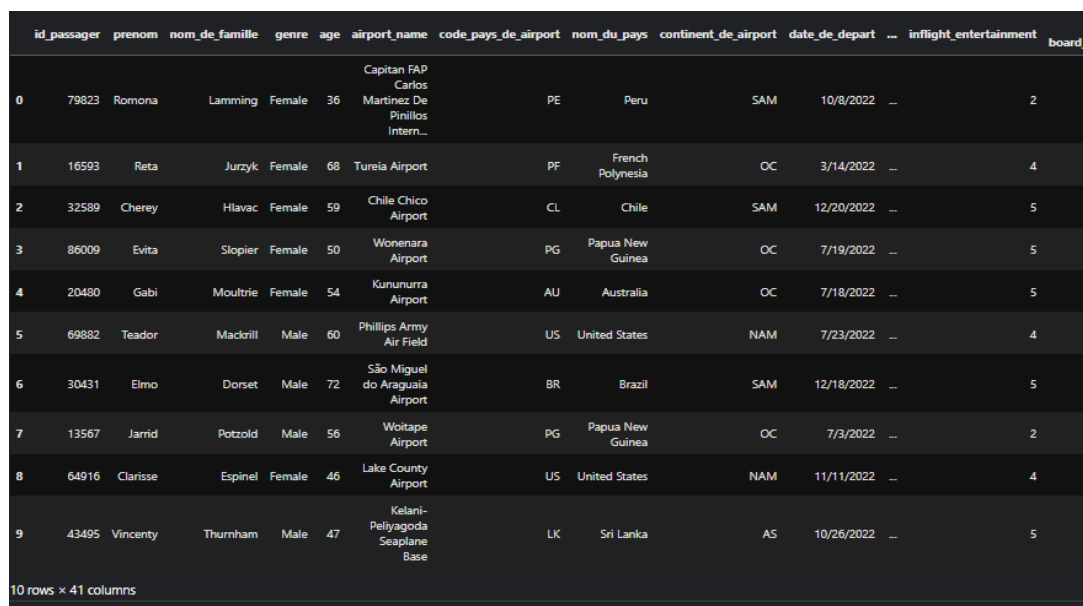
alignement, etc.) a été répété lors de cette intégration, générant ainsi un dataset nommé `dataset_fusionne_final.csv`.

### G. Normalisation et nettoyage

Après la dernière combinaison, certaines opérations de normalisation et de nettoyage ont été effectuées sur le dataset, notamment l'uniformisation des noms de colonnes (suppression des majuscules et des espaces), l'alignement des types de données entre les datasets fusionnés (exemple : conversion des types en chaînes de caractères pour certaines colonnes).

### H. Division du dataset en ensembles de données d'entraînement, de validation et de test

Le dataset final (`dataset_fusionne_final.csv`) a été divisé en trois sous-ensembles : 70 % pour l'entraînement du modèle illustré dans la figure 3.8, 15 % pour la validation représentée dans la figure 3.9, et 15 % pour le test (évaluation de la performance finale) illustré dans la figure 3.10. Cette séparation a été réalisée avec la fonction `train_test_split` de la bibliothèque `scikit-learn` et un paramètre `random_state()` fixé pour assurer la reproductibilité des résultats.



id_passager	prenom	nom_de_famille	genre	age	airport_name	code_pays_de_airport	nom_du_pays	continent_de_airport	date_de_depart	...	inflight_entertainment	board
0	79823	Romona	Lamming	Female	36	Capitan FAP Carlos Martinez De Pinillos Intern...	PE	Peru	SAM	10/8/2022	...	2
1	16593	Reta	Jurzyk	Female	68	Tureia Airport	PF	French Polynesia	OC	3/14/2022	...	4
2	32589	Cherey	Hlavac	Female	59	Chile Chico Airport	CL	Chile	SAM	12/20/2022	...	5
3	86009	Evita	Stopier	Female	50	Wonenara Airport	PG	Papua New Guinea	OC	7/19/2022	...	5
4	20480	Gabi	Moultrie	Female	54	Kununurra Airport	AU	Australia	OC	7/18/2022	...	5
5	69882	Teador	Mackrill	Male	60	Phillips Army Air Field	US	United States	NAM	7/23/2022	...	4
6	30431	Elmo	Dorset	Male	72	São Miguel do Araguaia Airport	BR	Brazil	SAM	12/18/2022	...	5
7	13567	Jarrid	Potzold	Male	56	Woitape Airport	PG	Papua New Guinea	OC	7/3/2022	...	2
8	64916	Clarisse	Espinel	Female	46	Lake County Airport	US	United States	NAM	11/11/2022	...	4
9	43495	Vincenty	Thurnham	Male	47	Kelani-Peliyagoda Seaplane Base	LK	Sri Lanka	AS	10/26/2022	...	5

10 rows x 41 columns

Fig. 3.8 – Capture d'écran de la première partie du dataset d'entraînement

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

id_passager	prenom	nom_de_famille	genre	age	airport_name	code_pays_de_airport	nom_du_pays	continent_de_airport	date_de_depart	...	inflight_entertainment	board
0	36667	Elfreda	Rutland	Female	13	Hazleton Municipal Airport	US	United States	NAM	5/29/2022	..	3
1	63089	Susana	Flagg	Female	34	Greenbrier Valley Airport	US	United States	NAM	11/2/2022	..	3
2	33402	Arley	Tickner	Male	17	Forestville Airport	CA	Canada	NAM	6/3/2022	..	5
3	34457	Gibb	Yarnton	Male	7	Oodnadatta Airport	AU	Australia	OC	7/12/2022	..	2
4	95567	Burty	Blurton	Male	33	Mariupol International Airport	UA	Ukraine	EU	8/9/2022	..	1
5	65487	Celia	Lavens	Female	30	Jefferson City Memorial Airport	US	United States	NAM	5/2/2022	..	2
6	55988	Ninon	Prier	Female	42	Del Caribe Santiago Mariño International Airport	VE	Venezuela, Bolivarian Republic of	SAM	9/8/2022	..	5
7	94954	Con	Bryer	Male	43	Ekibastuz Airport	KZ	Kazakhstan	AS	7/21/2022	..	4
8	15248	Sigvard	Wabb	Male	54	Fort Liard Airport	CA	Canada	NAM	4/15/2022	..	3
9	50328	Mathian	D'Aubney	Male	46	Blida Airport	DZ	Algeria	AF	1/4/2022	..	4

10 rows x 41 columns

Fig. 3.9 – Capture d’écran de la première partie du dataset de validation

id_passager	prenom	nom_de_famille	genre	age	airport_name	code_pays_de_airport	nom_du_pays	continent_de_airport	date_de_depart	...	inflight_entertainment	board
0	30340	Yolanda	Northidge	Female	52	Rouen Airport	FR	France	EU	8/16/2022	..	1
1	67252	Mike	Cahillane	Male	23	Staniel Cay Airport	BS	Bahamas	NAM	9/16/2022	..	5
2	54374	Marigold	McCaughey	Female	34	Silistra Polkovnik Lambinovo Air Base	BG	Bulgaria	EU	10/19/2022	..	1
3	66524	Heinrik	lowarch	Male	50	Abemama Atoll Airport	KI	Kiribati	OC	7/25/2022	..	5
4	55380	Phillipe	Mouch	Male	50	Belgrade Nikola Tesla Airport	RS	Serbia	EU	11/30/2022	..	5
5	43285	Alecia	Parcell	Female	44	Whakatane Airport	NZ	New Zealand	OC	12/26/2022	..	4
6	35063	Allie	Maddrell	Female	11	Sary Su Airport	KZ	Kazakhstan	AS	6/2/2022	..	4
7	39622	Dwight	Dietz	Male	51	Hope Bay Aerodrome	CA	Canada	NAM	12/15/2022	..	4
8	34588	Brett	Dybell	Male	29	Flying Cloud Airport	US	United States	NAM	12/14/2022	..	5
9	63937	Sumner	Bedding	Male	51	Verkhnevilyuisk Airport	RU	Russian Federation	AS	1/21/2022	..	2

10 rows x 41 columns

Fig. 3.10 – Capture d’écran de la première partie du dataset de test

### 3.2.2 Deuxième étape : Système de recommandation basé sur le contenu

Dans la première itération, nous avons conçu un système de recommandation exclusivement fondé sur le filtrage basé sur le contenu, exploitant les caractéristiques explicites des vols ainsi que les préférences saisies manuellement par l'utilisateur. À ce stade, l'historique d'interaction d'autres utilisateurs n'est pas encore utilisé. Chaque vol est représenté à l'aide d'un ensemble d'attributs, comprenant des variables catégorielles telles que la classe du vol, le nom de la compagnie ou encore le type de vol transformé en vecteur via un encodage one-hot, et des variables numériques telles que le prix, la durée normalisée à l'aide d'un scaler Min-Max. Les préférences recueillies des utilisateurs permettent de spécifier facilement des contraintes personnelles. Ces données sont ensuite transformées selon le même processus que les données des vols.

Un vecteur de profil utilisateur est alors généré par la concaténation des encodages des préférences. Pour le calcul de recommandations, ce vecteur est comparé à ceux des vols dispo-

nibles en utilisant la similarité cosinus qui mesure l'angle entre les deux vecteurs. Si les vecteurs pointent dans la même direction, la similarité est maximale (1). Si les vecteurs sont orthogonaux, la similarité est nulle (0) ce qui produit un score de pertinence compris dans l'intervalle  $[0,1]$  selon la formule(3.1). En sachant que :

- $A \cdot B$  : produit scalaire des vecteurs  $A$  et  $B$ .
- $\|A\|$  et  $\|B\|$  : normes (longueurs) des vecteurs  $A$  et  $B$ .
- Résultat : valeur comprise entre  $-1$  et  $1$  qui mesure l'angle entre les deux vecteurs.

$$\text{sim}(A, B) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \cos(A, B) \quad (3.1)$$

Les vols sont ensuite triés selon ce score, et les cinq meilleures correspondances sont présentées à l'utilisateur. Bien que simple et transparente, cette première version présente plusieurs limitations notables :

- Absence de mémoire utilisateur, ce qui signifie que le système ne tient pas compte de l'historique d'interaction ou des préférences implicites.
- Problème partiel de démarrage à froid, car les nouveaux utilisateurs peuvent être servis, mais les vols contenant des données manquantes sont ignorés.
- Manque de diversité, car les recommandations tendent à être redondantes et répétitives.
- Biais de l'importance des attributs, car malgré la normalisation, certains critères comme le prix ou la durée peuvent dominer le calcul de similarité.

Ces limitations nous ont conduit à faire évoluer notre système vers une version plus riche et hybride, décrite dans la contribution suivante.

### **3.2.3 Troisième étape : Système de recommandation hybride**

Dans une seconde version de notre approche, nous avons enrichi le SR initialement fondé exclusivement sur le filtrage basé sur le contenu en y intégrant le filtrage collaboratif, aboutissant ainsi à un système hybride conçu pour atténuer certaines limites du modèle initial. Dans ce cadre, chaque vol est représenté par ses caractéristiques (prix, classe, compagnie, etc.) encodées via un traitement de variables catégorielles par encodage one-hot et une normalisation des variables numériques (prix, durée, places disponibles) grâce à la méthode Min-Max, tout comme la version précédente étant donné que c'est une suite de notre travail. À partir de ces représentations, le système calcule un score de similarité cosinus entre le profil de l'utilisateur et l'ensemble de vols disponibles, générant ainsi un score basé sur le contenu. Ce score est ensuite combiné de manière pondérée avec la moyenne issue des données collaboratives selon la formule (3.2). En sachant que :

- $0.6 \times \text{Score}_{\text{contenu}}$  : contribution du score basé sur le contenu, pondérée à 60%.
- $0.4 \times \left( \frac{5}{\text{Note moyenne}} \right)$  : contribution du score basé sur la note moyenne, pondérée à 40%. L'inverse de la note moyenne est utilisé pour favoriser les éléments ayant une note plus faible (par exemple pour diversifier les recommandations ou équilibrer la popularité).
- Résultat : combinaison pondérée entre la recommandation basée sur le contenu et un facteur lié à la note moyenne.

:

$$\text{Score}_{\text{hybride}} = 0.6 \times \text{Score}_{\text{contenu}} + 0.4 \times \left( \frac{5}{\text{Note moyenne}} \right) \quad (3.2)$$

Cette fusion permet au système de gagner en robustesse face au démarrage à froid, d'améliorer la personnalisation tout en conservant une mémoire de l'historique utilisateur à travers les notes moyennes. Malgré ces améliorations, certaines limites subsistent :

- Le système demeure sensible au biais de popularité favorisant systématiquement les vols les mieux notés.
- Il n'intègre pas le contexte utilisateur.
- La représentation reste statique, elle ne s'adapte ni aux interactions passées ni à des dimensions comme le confort ou la satisfaction.
- La gestion temporelle est encore traitée comme une simple catégorie.

C'est pour adresser ces limites qu'on a intégré dans l'étape suivante l'apprentissage multi-tâche.

### **3.2.4 Quatrième étape : HMTL-RS**

Dans cette dernière partie, une nouvelle version de notre approche a été conçue afin d'apporter une dimension intelligente et adaptative à la recommandation. Cette évolution repose sur un modèle d'apprentissage multi-tâche capable de prédire simultanément plusieurs critères liés à la qualité, à la diversité et à la pertinence des vols.

#### **3.3.4.1 Première version : MTL à trois tâches**

Pour surmonter les limites de l'approche, nous avons mis en œuvre une première version du modèle MTL capable d'intégrer de façon conjointe la prédiction des notes utilisateur (régression), le niveau de nouveauté du vol (classification) ainsi que le score de diversité basé sur les différences caractéristiques entre vols (classification). Le cœur de ce modèle repose sur une architecture partagée composée de deux couches linéaires activées par ReLU et régularisées par dropout, suivies par trois têtes spécialisées, chacune dédiée à l'une des tâches mentionnées. Chaque vol est représenté par un vecteur de caractéristiques numériques préalablement normalisé à l'aide d'une fonction MinMaxScaler. L'apprentissage est supervisé par une fonction de perte pondérée à l'aide d'un optimisateur Adam tel que 60 % pour la note, 20 % pour la nouveauté et 20 % pour la diversité.

Cette version a permis d'améliorer la diversité des résultats, de réduire la surspécialisation et d'élargir la couverture, notamment pour les vols rares. Toutefois, elle présentait encore certaines limites : absence de prise en compte du contexte utilisateur, non-intégration au système hybride (ignorant les données implicites) et inefficacité face au problème du démarrage à froid, notamment pour les nouveaux utilisateurs.

### **3.3.4.2 Évolution vers HMTL-RS : un MTL à cinq tâches intégré à un système hybride**

Afin d’aboutir à une architecture de recommandation complète, nous avons fait évoluer notre modèle vers une version finale nommée HMTL-RS, reposant sur la fusion d’un modèle multitâche avancé et d’un système hybride enrichi. Cette version surmonte les limitations précédentes en proposant un modèle unifié à la fois explicatif, personnalisé et multi-objectifs, tout en respectant les contraintes du système hybride de base. L’architecture repose sur les modules suivants :

#### **Embedding utilisateur et vols**

Chaque entité est représentée par un vecteur latent appris automatiquement, modélisant les préférences implicites.

#### **Traitement de contenu explicite**

Les profils utilisateurs ainsi que les attributs de vols sont encodés, normalisés, puis traités par des réseaux de neurones distincts avant d’être fusionnés.

#### **Fusion neuronale**

Les embeddings implicites et les représentations explicites sont concaténés puis transformés via une couche de fusion afin de produire un vecteur commun enrichi.

#### **Têtes multi-tâches**

À partir du vecteur fusionné, cinq têtes sont entraînées simultanément :

- le `rating_head` pour la prédiction de la note utilisateurs.
- le `context_head` reconstruit les caractéristiques contextuelles initiales.
- le `novelty_head` pour l’estimation du caractère de nouveauté.
- le `diversity_head` l’évaluation de la diversité des recommandations.
- le `similarity_head` pour la projection de vols dans un espace de similarité vectorielle.

Les pertes sont pondérées et le modèle est entraîné avec l’optimiseur AdamW en utilisant un early stopping basé sur la perte de validation. La sortie du modèle MTL est ensuite réutilisée dans la logique de recommandation, permettant de calculer un score global issu de la combinaison des têtes. Les résultats sont ensuite filtrés selon les contraintes spécifiques de l’utilisateur. Cette approche HMTL-RS présente plusieurs avantages, notamment une personnalisation avancée, une recommandation multidimensionnelle, une résistance au biais de popularité et une adaptation au contexte utilisateur. Elle permet ainsi de répondre efficacement à deux des défis majeurs des SR, à savoir le démarrage à froid et la surspécialisation. Pour améliorer l’ergonomie et favoriser l’adoption du système, une interface utilisateur conviviale a été développée. Elle permet de visualiser les vols recommandés et d’appliquer des filtres personnalisés. Cette interface renforce l’expérience utilisateur et facilite l’interaction avec le système, constituant

## CHAPITRE 3. APPROCHE PROPOSÉE : SYSTÈME DE RECOMMANDATION HYBRIDE AVEC APPRENTISSAGE MULTI-TÂCHES (HMTL-RS)

ainsi un complément essentiel à l'approche HMTL-RS. La figure 3.11 représente l'interface utilisateur du HMTL-RS.

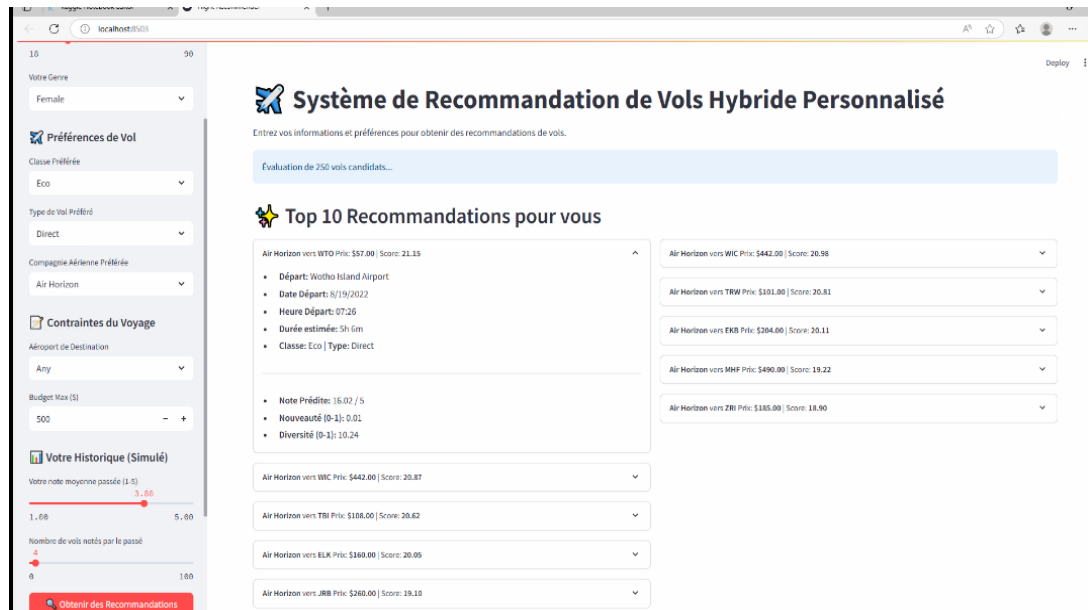


Fig. 3.11 – Capture d'écran de l'interface utilisateur

### 3.3 Conclusion

Dans ce chapitre, nous avons détaillé l'aspect méthodologique de l'approche développée dans le cadre de ce travail, laquelle combine un système de recommandation hybride avec une architecture d'apprentissage multi-tâche. Cette stratégie vise à exploiter les complémentarités entre ces deux approches, tout en favorisant l'apprentissage simultané de plusieurs tâches afin d'améliorer la personnalisation et la généralisation des recommandations. Le chapitre suivant sera entièrement consacré à l'évaluation de cette approche à travers l'application de différentes métriques permettant d'analyser les performances ainsi que son impact global sur la qualité des recommandations.

## Chapitre 4

# Évaluation des performances de HMTL-RS

### 4.1 Introduction

Ce chapitre est dédié à l'évaluation des performances de l'approche proposée pour la recommandation de vols. La première évaluation de HMTL-RS se base sur le calcul des valeurs de différentes métriques telles que la mean absolute error (MAE) ou le coefficient de détermination  $R^2$ . Ces métriques permettent de déterminer la capacité du système à fournir des recommandations adaptées aux préférences et aux besoins des utilisateurs. Pour une meilleure évaluation des performances de notre approche, nous avons implémenté un second RS nommé HSTL-RS où nous avons remplacé le MTL par le STL (Single-Task Learning). À la différence du MTL, le STL se concentre sur l'apprentissage d'une seule tâche à la fois. Cette comparaison permet de démontrer l'apport du MTL dans notre proposition. Les résultats obtenus ainsi qu'une discussion sont présentés dans la suite de ce chapitre.

### 4.2 Plateformes et outils de développement

Dans cette section, nous présentons l'ensemble des plateformes et des outils utilisés pour le développement de notre approche.

#### 4.2.1 Environnement de développement

**Kaggle** est une plateforme en ligne spécialisée en science des données, en IA et en ML. Elle permet de résoudre des problèmes de data science proposés par d'autres chercheurs à travers des compétitions ouvertes. Cette plateforme permet aux utilisateurs d'écrire, d'exécuter et de partager des codes Python dans un environnement en ligne. Elle offre également la possibilité de télécharger, explorer ou publier des datasets. En outre, elle favorise la collaboration et l'apprentissage grâce à une communauté active avec des échanges constants autour des projets. En résumé, Kaggle est un environnement cloud collaboratif dédié à l'analyse de données et à l'ap-

prentissage automatique, favorisant le partage de connaissance et l'amélioration continue entre chercheurs [54].

**VS Code** est un éditeur de code source gratuit qui est largement utilisé par les développeurs dans divers domaines, notamment le développement web, la data science et les applications logicielles. Cet éditeur se distingue par la prise en charge de nombreux langages de programmation tels que JavaScript, Python, C++, etc. Il dispose également d'un système d'extension puissant, d'une interface légère, rapide et personnalisable, ainsi que d'une compatibilité avec les environnements de développement locaux et distants [55].

### 4.2.2 Langage de programmation

**Python** est un langage de programmation puissant et facile à maîtriser grâce à une syntaxe claire et concise, conçu pour privilégier la lisibilité du code. Grâce à une large bibliothèque et à un vaste écosystème de modules tiers, Python est utilisé dans de nombreux domaines tels que la science des données, l'intelligence artificielle, le développement web et plus encore. Son interprétation dynamique et la richesse de ses bibliothèques font de lui un choix privilégié tant pour le prototypage rapide que pour le développement de solutions complexes en environnement de production [56, 57, 58].

**JavaScript** est un langage de programmation puissant et flexible, conçu pour créer des interfaces interactives et des expériences utilisateur dynamiques. Il fait partie des technologies fondamentales du web, aux côtés de HTML et de CSS. Sa richesse fonctionnelle est renforcée par un écosystème vaste de bibliothèques tel que React. De plus, avec Node.js, JavaScript s'étend au développement back-end, permettant de développer des applications entièrement avec JavaScript [59, 60].

### 4.2.3 Bibliothèques de Python

**NumPy** destiné aux calculs numériques et aux opérations sur les tableaux multidimensionnels. Elle est largement utilisée pour l'intelligence artificielle, le traitement scientifique et la modélisation mathématique [56].

**Pandas** est une bibliothèque open source spécifique à l'analyse et à la manipulation de données. Elle permet le traitement efficace de données structurées (CSV, Excel, SQL, JSON, etc.) [58].

**Matplotlib**, cette bibliothèque, est dédiée à la visualisation. Elle permet de générer une grande variété de graphiques [57].

**Random**, cette bibliothèque Python est utilisée pour générer différentes données aléatoirement [58].

**Sklearn**, cette bibliothèque, est spécialisée dans l'apprentissage automatique (machine learning). Elle offre une interface simple et efficace pour des tâches d'analyse prédictive [56].

**Torch** est une bibliothèque créée pour faciliter le développement des réseaux de neurones

et le calcul scientifique. Il est largement utilisé pour la recherche en IA et plus précisément dans l'apprentissage profond [57].

**Surprise** est une bibliothèque dédiée spécifiquement à l'évaluation des systèmes de recommandation (notamment pour évaluer et implémenter des algorithmes comme SVD, KNN, etc.) [56].

**Ipywidgets** est une bibliothèque Python très utilisée dans les environnements Jupyter qui permet de créer des interfaces utilisateur interactives pour explorer des données. Elle sert à créer des slides, des cases à cocher, des menus déroulants, des boutons, etc [58].

**Datetime** est une bibliothèque standard de Python (incluse d'office) qui permet de manipuler les dates et les heures [56].

### 4.3 Métriques d'évaluation de l'approche HMTL-RS

Après l'étude de l'état de l'art sur les RS, nous avons constaté l'utilisation fréquente de certaines métriques, ce qui a orienté notre choix vers les métriques mentionnées dans la table 4.1. En sachant que la légende des symboles utilisés dans la table 4.1 est comme suit :

- $y_i$  : valeur réelle (note ou classe observée) pour l'élément  $i$ .
- $\hat{y}_i$  : valeur prédite par le modèle pour l'élément  $i$ .
- $\bar{y}$  : moyenne des valeurs réelles  $y_i$ .
- $n$  : nombre total d'exemples (ou d'échantillons).
- $TP$  (True Positives) : nombre d'éléments positifs correctement prédits.
- $TN$  (True Negatives) : nombre d'éléments négatifs correctement prédits.
- $FP$  (False Positives) : nombre d'éléments négatifs incorrectement prédits comme positifs.
- $FN$  (False Negatives) : nombre d'éléments positifs incorrectement prédits comme négatifs.

Nom complet	Formule mathématique	Définition
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	Mesure l'écart-type des erreurs de prédiction ; pénalise davantage les grandes erreurs.
Mean Squared Error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	Moyenne des carrés des écarts entre les valeurs réelles et prédites ; elle amplifie l'impact des grandes erreurs en raison de l'élévation au carré.
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $	Moyenne des valeurs absolues des erreurs de prédiction.
Coefficient de détermination ( $R^2$ )	$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$	Proportion de la variance des données expliquée par le modèle.
Accuracy (ACC)	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	Le pourcentage global de bonnes prédictions.
Précision (Precision)	$Precision = \frac{TP}{TP + FP}$	La qualité des prédictions positives faites par le modèle.
Rappel (Recall)	$Recall = \frac{TP}{TP + FN}$	Capacité du modèle à retrouver tous les vrais positifs.
F1-score	$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$	Moyenne harmonique entre la précision et le rappel, le F1-score fournit une mesure globale de la performance .
Aire sous la courbe ROC (AUC)	Calculée à partir de la courbe ROC	Mesure la capacité à distinguer les classes ; 1 = parfait, 0,5 = aléatoire.

Tab. 4.1 – Résumé des principales métriques d'évaluation pour HMTL-RS

La section suivante présente une première évaluation de notre modèle HMTL-RS qui est basé sur le calcul des métriques présentées dans la table 4.1. Par la suite, nous présentons les résultats obtenus suivis d'une interprétation.

## 4.4 Évaluation 1 : Calcul de métriques pour HTML-RS

Afin d'évaluer les performances de notre approche HMTL-RS, nous avons entraîné le modèle sur plusieurs folds en utilisant une stratégie de validation croisée. Les résultats présentés dans cette section concernent à la fois l'évolution des performances au cours de l'apprentissage et les métriques finales obtenues à la meilleure époque selon le coefficient de détermination  $R^2$  pour la tâche principale (prédiction de la note).

### 4.4.1 Analyse des courbes d'apprentissage

La figure 4.1 présente l'évolution des pertes d'entraînement ainsi que des performances de validation sur les différentes tâches du modèle. On y observe :

- Une diminution progressive de la perte moyenne d'entraînement, indiquant une bonne convergence du modèle dès les premières époques.
- Sur la tâche principale (prédiction de note), le coefficient de détermination  $R^2$  atteint rapidement des valeurs élevées ( $> 0,9$ ), tandis que le RMSE diminue nettement avant de se stabiliser autour de 0,35, traduisant une bonne capacité d'approximation des préférences utilisateur.
- Les tâches secondaires (similarité, nouveauté, diversité, contexte) montrent également

des améliorations continues :

- Le F1-score de la nouveauté dépasse 0,80 et son AUC atteint 0,974.
- La similarité atteint un F1-score  $> 0,88$  et un AUC  $\approx 0,99$ , montrant une forte capacité discriminative.
- La diversité et le contexte présentent des RMSE en baisse constante, suggérant une qualité croissante des recommandations contextuelles et variées.
- Enfin, la faible dispersion entre les folds témoigne de la stabilité du modèle et de sa capacité de généralisation sur différents sous-ensembles de données.

Ces courbes confirment la capacité du modèle à apprendre efficacement des représentations utiles pour l'ensemble des tâches définies, sans signe apparent de surapprentissage.

Courbes d'Entraînement et de Validation Moyennes sur les Folds

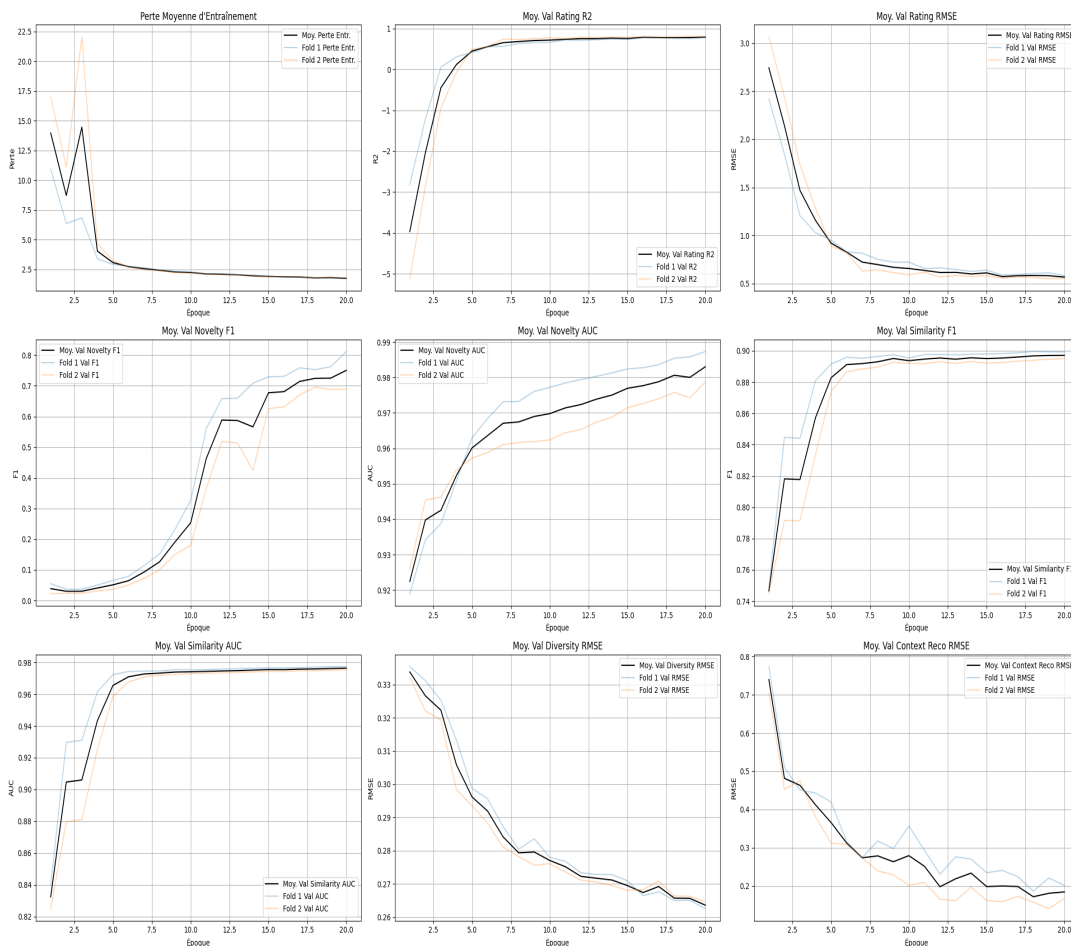


Fig. 4.1 – Présentation des courbes moyennes d'entraînement et de validation obtenues lors de l'évaluation du modèle HMTL-RS

### Discussion et interprétation des résultats

La table 4.2 illustre les résultats obtenus sur les meilleures époques des deux premiers folds parmi les cinq. Les performances globales montrent l'efficacité de notre approche multitâche.

En effet, pour la **prédiction de note**, le score  $R^2$  moyen est de 0,904, le RMSE moyen est à 0,354 et le MAE moyen est de 0,354. Pour la **similarité**, le F1-score est égal à 0,881 et l’AUC à 0,987. La **nouveauté** a un F1-score de 0.899 et une AUC de 0,974. Le score RMSE de la **diversité** est 0,263 et celui du **contexte** est 0,136.

Ces résultats confirment que **HMTL-RS** offre des prédictions précises, tout en intégrant efficacement la dimension contextuelle, la diversité et la nouveauté dans les recommandations.

Métrique	Fold 1	Fold 2
rating_mse	0.3403	0.3036
rating_rmse	0.5834	0.5510
rating_mae	0.4883	0.3542
rating_r <sup>2</sup>	0.7784	0.7986
novelty_acc	0.9975	0.9951
novelty_precision	0.8883	0.5989
novelty_recall	0.8155	0.8701
novelty_f1	0.8119	0.6881
novelty_auc	0.9873	0.9742
diversity_mse	0.0689	0.0709
diversity_mae	0.2169	0.2213
diversity_r <sup>2</sup>	0.1507	0.1435
diversity_rmse	0.2626	0.2663
similarity_acc	0.9004	0.8946
similarity_precision	0.9654	0.8991
similarity_recall	0.8932	0.8903
similarity_f1	0.8992	0.8946
similarity_auc	0.9775	0.9751
context_reco_mse	0.0406	0.0198
context_reco_r <sup>2</sup>	0.9964	0.9982
context_reco_mae	0.1535	0.1136
context_reco_rmse	0.2014	0.1407

Tab. 4.2 – Résultats obtenus sur les deux premiers folds

## 4.5 Évaluation 2 : Comparaison de HMTL-RS avec un modèle monotâche

Pour évaluer l’apport de l’apprentissage multitâche dans notre approche, nous l’avons comparée à un modèle fondé sur l’apprentissage monotâche, où une seule tâche est apprise indépendamment. Pour cela, nous avons implémenté une seconde version de notre approche où nous remplaçons le MTL par le STL. Notons que les deux modèles ont été entraînés sur les mêmes données et soumis et évalués sur les mêmes tâches. La figure 4.2 présente les résultats de comparaison des performances de deux modèles sur les cinq tâches du SR, notamment la prédiction

de note, la nouveauté, la diversité, la similarité et la prise en compte du contexte. Pour chacune de ces tâches, une métrique spécifique a été sélectionnée. Le paramètre  $R^2$  est calculé pour la prédiction de note, de diversité et de contexte, tandis que le F1-score est utilisé pour évaluer les tâches de nouveauté et de similarité.

Les résultats illustrés par la figure 4.2 montrent que le modèle HMTL-RS surpasse le modèle HSTL-RS, validant ainsi l'efficacité de l'approche multitâche. L'interprétation des résultats pour chaque tâche est détaillée dans ce qui suit :

- Concernant la prédiction de notes, le modèle HMTL-RS atteint un coefficient de détermination  $R^2$  de 0,791, alors que le modèle HSTL présente une valeur négative (-5,155), soulignant une instabilité significative et une faible capacité de modélisation dans cette tâche.
- Pour la tâche de nouveauté, HMTL-RS enregistre un F1-score de 0,689, bien supérieur à celui du modèle HSTL-RS (0,019), ce qui met en évidence sa meilleure aptitude à proposer des recommandations inédites et pertinentes.
- En ce qui concerne la diversité, HMTL-RS affiche également de meilleures performances, avec un  $R^2$  de 0,152 contre -0,194 pour STL, indiquant une meilleure capacité à générer des résultats variés.
- Pour la similarité, le modèle multitâche obtient un F1-score de 0,897, nettement supérieur à celui atteint par HSTL-RS (0,640).
- Enfin, pour la prise en compte du contexte, HMTL-RS maintient une légère supériorité avec un  $R^2$  de 0,996, contre 0,941 pour STL.

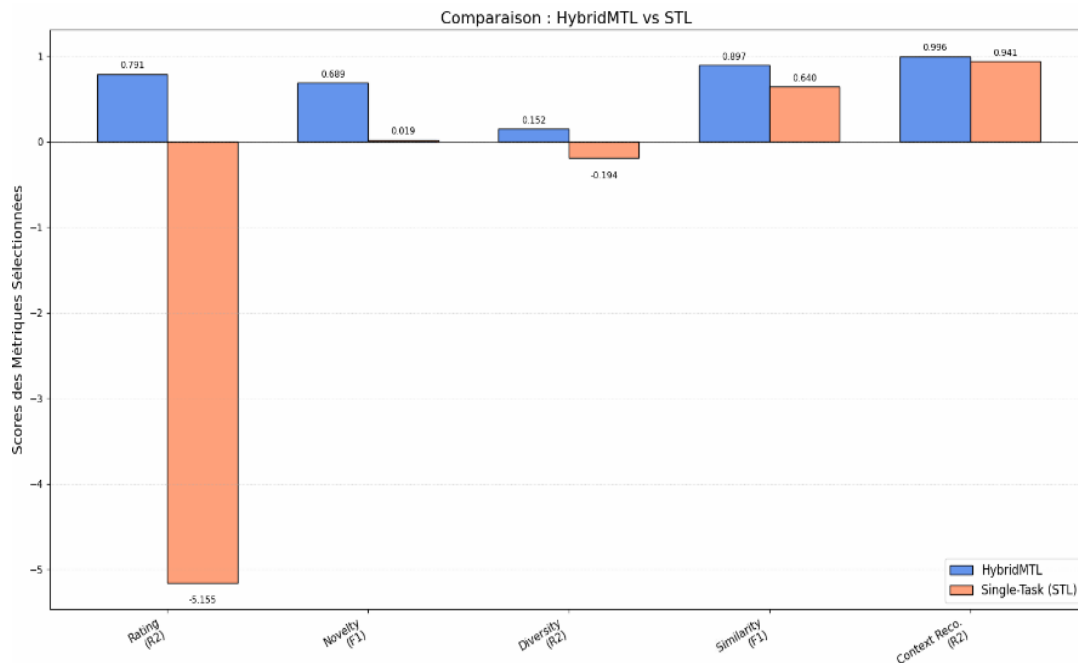


Fig. 4.2 – Histogramme comparatif des performances du HMTL-RS et du STL pour les différentes tâches

Les résultats présentés dans ce chapitre confirment les bonnes performances de HMTL-

RS, offertes par l'intégration de l'apprentissage simultané de tâches corrélées. Cette stratégie favorise le partage de représentations pertinentes entre les différentes tâches, ce qui contribue à améliorer significativement les performances globales du système de recommandation.

### 4.6 Conclusion

Dans ce chapitre, nous avons procédé à l'évaluation de notre système de recommandation HMTL-RS à l'aide de métriques standards. Ces indicateurs nous ont permis de quantifier la précision, la robustesse et la capacité de généralisation du modèle. Afin de mieux situer les performances de notre approche, nous l'avons également comparée à un modèle classique appliqué aux mêmes tâches et sur les mêmes données. Les résultats obtenus montrent clairement que notre approche surpasse le modèle de comparaison sur l'ensemble des métriques considérées, confirmant ainsi l'intérêt de l'approche multi-tâche dans le contexte de la recommandation de vols.

# Conclusion générale

Les systèmes de recommandation occupent aujourd'hui un rôle essentiel dans la personnalisation des services numériques, en particulier dans des environnements où les utilisateurs sont confrontés à une multitude de choix. Bien que les avancées en IA aient considérablement amélioré ces systèmes, des limitations persistent. Dans notre mémoire, nous avons proposé une approche nommée HMTL-RS (Hybrid and Multi-Task Learning-based Recommender System), destinée à la recommandation de vols. Cette approche combine les avantages du filtrage collaboratif, du filtrage basé sur le contenu et de l'apprentissage multitâches. L'intégration d'une architecture multitâche à cinq objectifs a permis de générer des recommandations personnalisées adaptables à plusieurs utilisateurs et d'apporter une réponse efficace aux défis du démarrage à froid et de la surspécialisation.

L'évaluation des performances de l'approche proposée a été menée à l'aide de métriques standards telles que le RMSE, le MAE, le  $R^2$ , la précision, le rappel et le F1-score. Ces indicateurs ont permis d'évaluer la qualité prédictive du modèle sur plusieurs dimensions. Pour mesurer l'apport du multitâche, HMTL-RS a été comparé à un modèle monotâche (HSTL-RS) entraîné sur les mêmes données. Les résultats obtenus montrent que notre approche surpasse nettement le modèle STL, confirmant l'intérêt de l'apprentissage multitâche pour améliorer la précision, la diversité et la personnalisation des recommandations.

Ce travail ouvre la voie à plusieurs pistes de recherche futures. Une première piste serait d'expérimenter sur des données réelles issues de plateformes de réservation de vols afin de tester le modèle dans des conditions plus proches du monde réel. Par ailleurs, le temps d'exécution et la complexité computationnelle du modèle, en particulier lors de l'entraînement multitâche, constituent des limitations qu'il conviendrait d'optimiser. Une autre évolution importante serait d'introduire une fonction de similarité plus sophistiquée. Enfin, une autre perspective utile serait le développement d'une interface utilisateur dédiée, car elle permettra d'améliorer l'expérience utilisateur et favorisera l'intégration du système dans une application réelle. En conclusion, l'approche HMTL-RS constitue une contribution prometteuse à la recherche sur les SR et offre une base solide pour des travaux futurs orientés vers l'intégration de solutions de recommandations dans des environnements réels.

## References

- [1] H. Chemingui, *Configuration guidée des lignes de produits basée sur les traces de processus*. PhD thesis, Université de Paris, 2021.
- [2] A. Grignola, “Ia / intelligence artificielle : Comment l’ia révolutionne le quotidien de notre monde?.” <https://www.data-bird.co/blog/intelligence-artificielle>, February 2025.
- [3] B. Belourghi, L. Houichi, and S. Heddami, “Réseaux de neurones artificiels pour la modélisation du dosage du coagulant dans les stations de traitements des eaux de surface à faible turbidité,” in *ATGRSR 2012 – II. Séminaire International Euro-Méditerranéen Aménagement du Territoire, Gestion des Risques et Sécurité Routière*, (Batna, Algérie), September 2012.
- [4] G. Ravikumar, T. P. Kumar, and K. Srinivas, “A pragmatic review on different approaches used in e-learning recommender systems,” *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 1571–1580, 2018.
- [5] H. B. P. Krupa Patel, “A state-of-the-art survey on recommendation system and prospective extensions,” *Computers and Electronics in Agriculture*, vol. 175, p. 105996, 2020.
- [6] W. Z. X. X. B. C. Shiwen Wu, Fei Sun, “Graph neural networks in recommender systems : A survey,” *arXiv preprint arXiv :2011.02260*, 2020.
- [7] B. H. F. Mehrdad Afsar, Trafford Crump, “Reinforcement learning based recommender systems : A survey,” *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–38, 2021.
- [8] P. S. Y. Lei Zheng, Vahid Noroozi, “Recommender systems : Challenges, solutions and research opportunities,” *Information Sciences*, vol. 418, pp. 117–132, 2021.
- [9] I. L. Z. Y. X. K. F. X. Xiaomei Bai, Mengyang Wang, “Scientific paper recommendation : A survey,” *arXiv preprint arXiv :2008.13538*, 2020.
- [10] X. W. K. Z. M. W. Lei Wu, Xiangnan He, “A survey on accuracy-oriented neural recommendation : From collaborative filtering to information-rich recommendation,” *ACM Transactions on Information Systems (TOIS)*, vol. 39, no. 1, pp. 1–42, 2021.

## REFERENCES

---

- [11] D. C. W. I. Islam Elnabarawy, Wei Jiang, “Survey of privacy-preserving collaborative filtering,” *arXiv preprint arXiv :2003.08343*, 2020.
- [12] S. I. D.-Y. K. D.-S. P. Sony Peng, Sophort Siet, “Integration of deep reinforcement learning with collaborative filtering for movie recommendation systems,” *Applied Sciences*, vol. 14, no. 3, p. 1155, 2024.
- [13] D. N. A. I. R. K. Zeshan Fayyaz, Mahsa Ebrahimian, “Recommendation systems : Algorithms, challenges, metrics, and business opportunities,” *Applied Sciences*, vol. 10, no. 21, p. 7748, 2020.
- [14] M. Singh, “Scalability and sparsity issues in recommender datasets : a survey,” *Knowledge and Information Systems*, vol. 62, no. 1, pp. 1–43, 2020.
- [15] S. S. Guy Aridor, Duarte Gonçalves, “Deconstructing the filter bubble : User decision-making and recommender systems,” in *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys 2020)*, pp. 82–91, ACM, 2020.
- [16] D. C. W. I. Islam Elnabarawy, Wei Jiang, “Survey of privacy-preserving collaborative filtering,” *arXiv preprint arXiv :2003.08343*, 2020.
- [17] A. S. Nath and E. Selvam, “A pragmatic review on different approaches used in e-learning recommender systems,” in *2018 IEEE International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, pp. 1–7, IEEE, 2018.
- [18] S. Najar, W. Jaziri, and F. Gargouri, “Comparaison des systèmes de recommandation,” in *Colloque Méditerranéen sur les Systèmes d’Information (CIMSIS)*, pp. 1–10, 2017.
- [19] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” *Journal of Big Data*, vol. 9, no. 1, p. 59, 2022.
- [20] H. Ko, S. Lee, Y. Park, and A. Choi, “A survey of recommendation systems : Recommendation models, techniques, and application fields,” *Electronics*, vol. 11, no. 1, p. 141, 2022.
- [21] Q. Zhang, J. Lu, and Y. Jin, “Artificial intelligence in recommender systems,” *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 439–457, 2021.
- [22] M. A. K. Sharifani and M. Amini, “A systematic review of machine learning and deep learning techniques in recommender systems,” *World Information Technology and Engineering Journal*, vol. 10, 2023.
- [23] K. Kieslich, M. Lünich, and F. Marcinkowski, “The threats of artificial intelligence scale (tai) : Development, measurement and test over three application domains,” *International Journal of Social Robotics*, vol. 13, no. 7, pp. 1563–1577, 2021.

## REFERENCES

---

- [24] G. Vilone and L. Longo, “Explainable artificial intelligence : A systematic review,” *arXiv preprint arXiv :2006.00093*, 2020.
- [25] J. Anuradha *et al.*, “Machine learning : Algorithms, real-world applications and research directions,” *SN Computer Science*, vol. 2, no. 3, pp. 1–14, 2021.
- [26] R.-G. U. A. Paleyes and N. D. Lawrence, “Challenges in deploying machine learning : A survey of case studies,” *arXiv preprint arXiv :2011.09926*, 2020.
- [27] X. H. S. H. M. Wang, W. Fu and X. Wu, “A survey on large-scale machine learning,” *arXiv preprint arXiv :2008.03911*, 2020.
- [28] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, pp. 685–695, 2021.
- [29] A. Mohammed and R. Kora, “A comprehensive review on ensemble deep learning : Opportunities and challenges,” *Journal of King Saud University – Computer and Information Sciences*, vol. 35, pp. 757–774, 2023.
- [30] I. H. Sarker, “Deep learning : A comprehensive overview on techniques, taxonomy, applications and research directions,” *SN Computer Science*, vol. 2, no. 6, pp. 1–20, 2021.
- [31] M. M. Taye, “Understanding of machine learning with deep learning : Architectures, workflow, applications and future directions,” *Computers*, vol. 12, no. 5, p. 91, 2023.
- [32] T. Wang and S.-C. Chen, “Multi-task local-global graph network for flight delay prediction,” in *2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)*, pp. 49–54, IEEE, 2022.
- [33] A. Navon, A. Shamsian, I. Achituve, H. Maron, K. Kawaguchi, G. Chechik, and E. Fetaya, “Multi-task learning as a bargaining game,” in *Proceedings of the 39th International Conference on Machine Learning*, vol. 162 of *Proceedings of Machine Learning Research*, pp. 16254–16276, PMLR, 2022.
- [34] M. Zhang, R. Yin, Z. Yang, Y. Wang, and K. Li, “Advances and challenges of multi-task learning method in recommender system : A survey,” *arXiv preprint arXiv :2305.13843*, 2023.
- [35] Y. Zhang and Q. Yang, “A survey on multi-task learning,” *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [36] S. Chen, Y. Zhang, and Q. Yang, “Multi-task learning in natural language processing : An overview,” *ACM Computing Surveys*, vol. 1, no. 1, pp. 1–31, 2024. Article publication date : April 2024.

## REFERENCES

---

- [37] R. K. Dina Nawara, "Context-aware recommendation systems in the iot environment (iot-cars)—a comprehensive overview," *IEEE Access*, vol. 9, pp. 144270–144284, 2021.
- [38] A. K. G. M. D. G. S. H. Vijayalaxmi N. Rathod, R. H. Goudar, "A survey on e-learning recommendation systems for autistic people," *IEEE Access*, vol. 12, pp. 11723–11732, 2024.
- [39] N. K. D. S. S. Sapna Jaiswal, Tejaswi Kharade, "Collaborative recommendation system for agriculture sector," *ITM Web of Conferences*, vol. 32, p. 03034, 2020.
- [40] Y. L. S. H. Z. O. Sizhe Yin, Yang Xia, "Fusing user preferences and spatiotemporal information for sequential recommendation," *IEEE Access*, vol. 10, pp. 89545–89554, 2022.
- [41] A. K. I. J. H. L. Kailash Chowdary Bodduluri, Francis Palma, "Exploring the landscape of hybrid recommendation systems in e-commerce : A systematic literature review," *IEEE Access*, vol. 12, pp. 28288–28294, 2024.
- [42] C. Gündüz, "Customised holiday experiences through artificial intelligence : Case studies from the aviation and hospitality sectors," *Journal of Aviation*, vol. 7, no. 3, pp. 337–345, 2023.
- [43] G. S. J. L. G. A. R. P. M. V. S. Siva Priyanka, M. Raju, "Iot based crop recommendation system using machine learning for smart agriculture," in *Proceedings of the Second International Conference on Emerging Trends in Engineering (ICETE 2023)*, pp. 893–904, Atlantis Press, 2023.
- [44] H. I. M. Z. F. S. C. L. Caiwen Li, Iskandar Ishak, "Deep learning-based recommendation system : Systematic review and classification," *IEEE Access*, vol. 11, pp. 113791–113830, 2023.
- [45] T. F. C. L. R. T. Amine Dadoun, Michael Defoin-Platel, "How recommender systems can transform airline offer construction and retailing," *Journal of Revenue and Pricing Management*, vol. 20, no. 3, pp. 301–315, 2021.
- [46] M. D.-P. G. A. S. Amine Dadoun, Raphaël Troncy, "Predicting your next trip : A knowledge graph-based multi-task learning approach for travel destination recommendation," in *Proc. of the International Workshop on Recommender Systems for Tourism (RecTour 2021)*, CEUR Workshop Proceedings, 2021.
- [47] K. C. Ho Yin Kan, Dennis Wong, "A personalized flight recommender system based on link prediction in aviation data," *IEEE Access*, vol. 12, pp. 36961–36973, 2024.
- [48] C. Tejas and S. M. Bargavi, "Ai-powered recommendation systems for trip planning," *International Journal of Advanced Research*, vol. 11, no. 3, pp. 671–678, 2023.

## REFERENCES

---

- [49] J. Schoonderbeek, “Predicting airline passengers with deep multi-task learning,” Master’s thesis, Eindhoven University of Technology, 2020.
- [50] X. S. S. D. C. L. X. L. G. C. Yue Ding, Zhiqing Chen, “Towards personalized federated multi-scenario multi-task recommendation,” *arXiv preprint arXiv :2401.02700*, 2024.
- [51] X. H. F. Gu, K. Jiang and J. Yang, “Deep learning-based image geolocation for travel recommendation via multi-task learning,” *Journal of Circuits, Systems and Computers*, vol. 31, no. 07, p. 2250127, 2022.
- [52] T. X. Z. Y. Y. Wu, R. Fu and F. Yin, “A user behavior-aware multi-task learning model for enhanced short video recommendation,” *Neurocomputing*, vol. 617, p. 129076, 2025.
- [53] K. D. S. D. Mohamed Bouni, Badr Hssina, “Towards an efficient recommender systems in smart agriculture : A deep reinforcement learning approach,” *Procedia Computer Science*, vol. 203, pp. 825–830, 2022.
- [54] Kaggle, “Kaggle Documentation.” <https://www.kaggle.com/docs>, 2023. En ligne. [Consulté le : 3-juin-2025]. Official documentation for Kaggle platforms, including notebooks, datasets, and competitions.
- [55] Microsoft, “Visual Studio Code Documentation.” <https://code.visualstudio.com/docs>, 2023. En ligne. [Consulté le : 3-juin-2025]. Official VS Code documentation covering installation, key features (debugging, extensions, Git integration), configuration (settings.json, keybindings), and language tutorials.
- [56] Python Software Foundation, “Python 3 Documentation.” <https://docs.python.org/3/>, 2023. En ligne. [Consulté le : 2-juin-2025].
- [57] Python Software Foundation, “Python Glossary.” <https://docs.python.org/3/glossary.html>, 2023. En ligne. [Consulté le : 2-juin-2025]. Définitions officielles des termes Python.
- [58] Real Python, “Real Python Tutorials.” <https://realpython.com/>, 2023. En ligne. [Consulté le : 2-juin-2025]. Tutoriels approfondis et articles sur Python.
- [59] ECMA International, “ECMA-262: ECMAScript Language Specification.” <https://www.ecma-international.org/publications-and-standards/standards/ecma-262/>, 2023. En ligne. [Consulté le : 3-juin-2025]. Standard officiel du langage JavaScript (ECMAScript).
- [60] Mozilla Developer Network, “JavaScript | MDN Web Docs.” <https://developer.mozilla.org/fr/docs/Web/JavaScript>, 2023. En ligne. [Consulté le : 3-juin-2025]. Documentation complète de JavaScript maintenue par Mozilla.

# RÉSUMÉ

Ce projet de fin d'études porte sur le développement d'un système de recommandation de vols basé sur une approche hybride et multitâche appelée HMTL-RS pour améliorer la personnalisation et résoudre deux problèmes fréquents dans les SR, à savoir le démarrage à froid et la surspécialisation. Notre solution combine le filtrage basé sur le contenu et le filtrage collaboratif, enrichie par une architecture d'apprentissage multitâches à cinq objectifs. Cette combinaison permet de générer des recommandations variées et adaptées aux profils des nouveaux utilisateurs.

Les résultats expérimentaux montrent une amélioration des performances par rapport aux approches classiques. Ce système offre une base solide pour des évolutions futures, notamment l'utilisation de données réelles, l'optimisation du temps d'exécution.

**Mots clés :** Recommandation, apprentissage multitâches, filtrage basé sur le contenu, filtrage collaboratif, filtrage hybride, surspécialisation, démarrage à froid.

# ABSTRACT

This final year project focuses on the development of a flight recommendation system based on a hybrid and multi-task learning approach called HMTL-RS, designed to enhance personalization and address two common issues in recommender systems : the cold-start problem and overspecialization.

Our solution combines content-based filtering and collaborative filtering, enriched by a multi-objective learning architecture with five tasks. This combination allows the generation of diverse recommendations tailored to the profiles of new users.

Experimental results show improved performance compared to classical approaches. This system provides a solid foundation for future developments, including the use of real-world data and execution time optimization.

**Keywords :** Recommendation, multi-task learning, content-based filtering, collaborative filtering, hybrid filtering, overspecialization, cold-start.

## المخلص

يركز مشروع السنة النهائية هذا على تطوير نظام توصية بالرحلات الجوية استناداً إلى نهج تعلم هجين ومتعدد المهام يسمى HMTL-RS مصمم لتعزيز التخصيص ومعالجة مشكلتين شائعتين في أنظمة التوصية: مشكلة البداية الباردة والتخصص الزائد. ويجمع حلنا بين التصنيف القائمة على المحتوى والتصنيف التعاونية، مع إرائته ببنية تعلم متعددة الأهداف مع خمس مهام. ويسمح هذا المزيج بتوليد توصيات متنوعة مصممة خصيصاً للمفاتيح تعريف المستخدمين الجدد. تُظهر النتائج التجريبية أداءً محسناً مقارنةً بالنهج الكلاسيكي. يوفر هذا النظام أساساً متيناً للتطورات المستقبلية، بما في ذلك استخدام بيانات العالم الحقيقي وتحسين وقت التنفيذ.

**الكلمات المفتاحية:** التوصية، والتعلم متعدد المهام، والتصنيف القائمة على المحتوى، والتصنيف التعاونية والتصنيف الهجينة، والتخصص الزائد، والبداية الباردة.