

## Chapitre 1.2. Les systèmes de recommandation

### 1.2.1. Introduction

Sur quelle page web me rendre pour trouver l'information que je recherche ? Quels sont les films que je n'ai pas encore vus et qui pourraient me plaire ? Quels articles devrais-je lire pour approfondir mes connaissances dans mon domaine de recherche ?

Aujourd'hui, ces questions sont devenues fréquentes chez les utilisateurs du Web. En effet, avec le développement rapide d'Internet et la numérisation croissante des ressources, l'information est de plus en plus accessible électroniquement. Toutefois, cette abondance entraîne un problème majeur : la surcharge informationnelle. Les utilisateurs se retrouvent souvent submergés par le volume de contenus disponibles, ce qui complique la recherche d'informations pertinentes et adaptées à leurs besoins.

Pour faire face à cette difficulté, les systèmes de recommandation ont été conçus. Leur objectif principal est de filtrer l'information et de proposer à chaque utilisateur un contenu personnalisé, en fonction de ses intérêts, de son comportement ou de son profil.

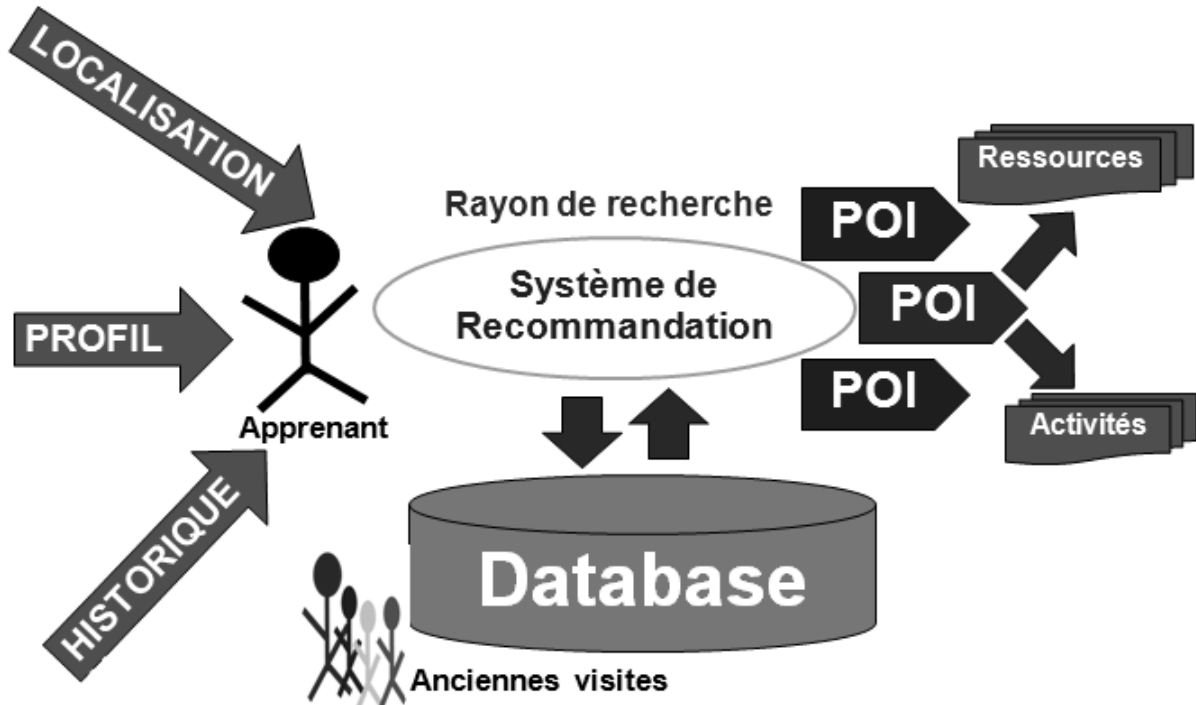
L'objectif de ce chapitre est d'examiner le fonctionnement des systèmes de recommandation, en proposant une définition rigoureuse, en présentant les approches principales utilisées, et en discutant leurs forces, faiblesses et limites

### 1.2.2 Définition d'un système de recommandation

« Un système de recommandation est une forme particulière de filtrage visant à présenter les éléments d'information (films, musique, livres, images, web, etc.) susceptibles d'intéresser l'utilisateur. » [21]

Un *système de recommandation* est un système intelligent d'aide à la décision, conçu pour proposer à un utilisateur des éléments jugés pertinents en fonction de ses préférences explicites ou implicites. Il s'appuie sur l'analyse de données issues du profil utilisateur (clics, achats, évaluations, historique de navigation, etc.) afin de prédire des éléments susceptibles de l'intéresser. L'objectif principal est de faciliter l'accès à l'information utile tout en réduisant l'effort de recherche.

Un système de recommandation peut être défini comme « *un système logiciel qui fournit des suggestions à un utilisateur sur des éléments qui peuvent l'intéresser* », en se basant sur des interactions passées, le profil utilisateur ou des similarités entre éléments. [21]



**Figure 1.2.1:** system de recommandation [22]

### 1.2.3 Approches utilisées pour la recommandation

Les systèmes de recommandation, également appelés systèmes de filtrage d'information, visent à aider les utilisateurs à faire face à l'abondance croissante de contenus disponibles, en sélectionnant et en proposant des informations pertinentes adaptées à leurs besoins spécifiques. Leur fonction principale est de filtrer un flux d'informations entrantes, en les collectant, en les analysant, puis en les triant pour ne suggérer que celles qui ont une forte probabilité d'intérêt pour un utilisateur donné [23].

Le principe de base repose sur une prédiction : le système évalue la probabilité qu'une information particulière soit jugée utile ou pertinente par l'utilisateur. Cette prédiction s'appuie généralement sur le profil de l'utilisateur, c'est-à-dire un ensemble de données représentant ses préférences, son comportement passé ou encore ses objectifs. En fonction de cette analyse, le système prend une décision : recommander ou non le contenu concerné [24].

### 1.2.3.1 Filtrage par contenus (content based filtering)

Les systèmes de filtrage à base de contenu recommandent des documents similaires à ceux que l'utilisateur a déjà appréciés. Ceci est calculé en comparant les centres d'intérêt des utilisateurs avec les métadonnées ou les caractéristiques utilisées pour représenter les ressources ou les produits, sans prendre en compte les avis et les informations concernant d'autres utilisateurs [25].

Deux techniques sont nécessaires à mettre en place pour ce type de filtrage. L'une pour représenter les items et une autre pour créer le profil utilisateur. Ces techniques sont implémentées de différentes manières dans les systèmes de recommandation basés sur le contenu. Deux fonctionnalités centrales ressortent, pour un système de filtrage cognitif : la sélection des documents pertinents vis-à-vis du profil, la mise à jour du profil en fonction du retour de pertinence fourni par l'utilisateur sur les documents qu'il a reçus ; la mise à jour se fait par intégration des thèmes abordés dans les documents jugés pertinents.

En effet, ce type de recommandation peut être appliqué pour la recommandation de pages web, de cours, de formation, etc. [26]

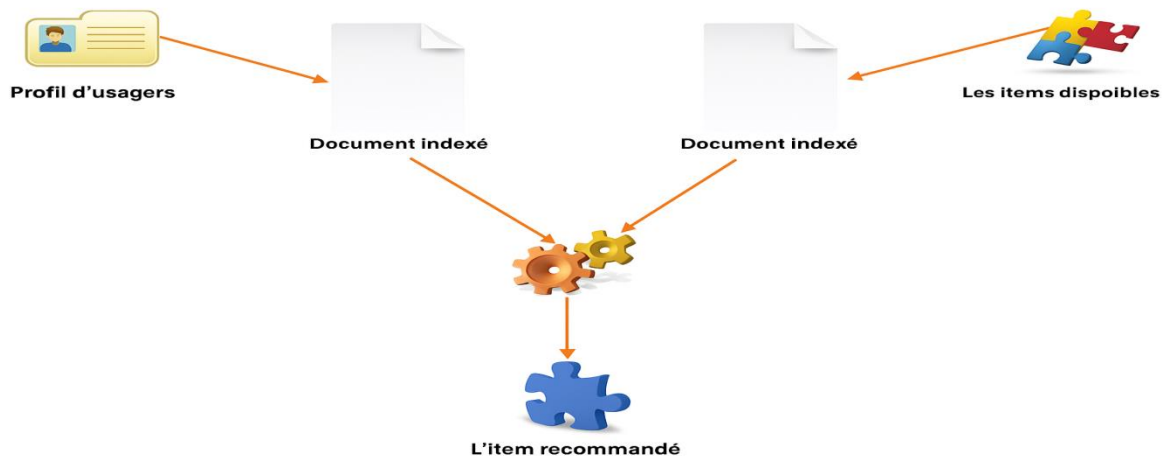


Figure 1.1.2 : Filtrage par contenus [27]

### 1.2.3.2 Le filtrage collaboratif (collaborative filtering)

Le filtrage collaboratif se fonde sur la recherche de proximité des utilisateurs en fonction de leur profile, il tente de prédire l'utilité d'items pour un utilisateur particulier en fonction des items précédemment évalués par d'autres utilisateurs qui présentent des similarités dans les préférences [28]

Cette technique est basée sur le partage d'opinions entre les utilisateurs. Bien que le terme n'ait été introduit que depuis moins de deux décennies, il implémente le principe du " bouche à-oreille" pratiqué depuis toujours par les humains pour se construire une opinion sur un produit ou un service [29].

Le filtrage collaboratif (*collaborative filtering*) constitue l'une des approches les plus répandues dans les systèmes de recommandation. Il repose sur l'analyse des comportements et des préférences d'un ensemble d'utilisateurs afin de formuler des recommandations pertinentes .Le filtrage collaboratif est très utilisé dans les environnements de type e-commerce dans lesquels les utilisateurs attribuent des scores à des produits qu'ils consultent ou achètent. L'approche consiste alors à faire des recommandations en recherchant des corrélations entre les produits « aimés » et « pas aimés » parmi les utilisateurs du système. Dans un système de recommandation de livres par exemple, le système recherchera les individus similaires à l'utilisateur ; et seuls les livres bien notés par ces individus seront recommandés à l'utilisateur.

Breese et al. Proposent une classification intéressante des techniques de filtrage collaboratif : les algorithmes basés « mémoire », et les algorithmes basés « modèle » [30].

- *Algorithmes basés « mémoire »*

Les algorithmes basés mémoire utilisent l'ensemble de la base de données des évaluations des utilisateurs pour faire les prédictions : les évaluations de l'utilisateur actif sont prédits à partir d'informations partielles concernant l'utilisateur actif, et un ensemble de poids calculés à partir de la base de données des évaluations des utilisateurs. [31] il y a trois types de ces algorithmes :

- i. Corrélation de Pearson
- ii. Cosinus des vecteurs
- iii. La distance de Spearman (utilisateurs)

- *Algorithmes basés modèle*

Les algorithmes basés « modèle » utilisent la base de données des évaluations des utilisateurs pour estimer ou apprendre un modèle qui est alors utilisé pour les prédictions. Il existe deux modèles probabilistes : le modèle à base de clusters, et le modèle à base de réseau bayésien.



**Figure 1.1.3** : filtrage collaboratif [32]

### 1.2.3.3 Filtrage hybride (hybrid filtering)

Le filtrage par contenus et le filtrage collaboratif étudiés précédemment possèdent chacun leurs avantages et leurs inconvénients. D'une part, le filtrage collaboratif pallie des inconvénients du filtrage par contenus tels que le manque de données subjectives (données explicites de l'utilisateur) ou de scores attribués par les utilisateurs eux-mêmes. D'autre part le filtrage par contenus pallie des inconvénients du filtrage collaboratif tel que le problème des données très éparpillées ou la complexité de calculs [33]. Le filtrage hybride vise intégrer ces deux types de filtrages en gardant les avantages de chacun [34]. Cette intégration peut se faire de différentes manières [35]:

- (i) implémenter chacune des méthodes séparément et intégrer leurs résultats ;
- (ii) rajouter certaines caractéristiques du filtrage collaboratif dans le filtrage par contenus
- (iii) rajouter certaines caractéristiques du filtrage par contenus dans le filtrage collaboratif
- (iv) construire un modèle unifié incorporant les caractéristiques de chacune des méthodes.

#### **1.2.4 Étude comparative entre les approches**

La notion du filtrage collaboratif est la base de la recommandation, tandis que le filtrage à base de contenu étant plutôt lié aux systèmes de recherche d'information dits personnalisés. Le filtrage collaboratif se fonde sur la recherche de proximité des utilisateurs en fonction de leur profil. Par contre aux systèmes de recommandation basée sur le contenu, les systèmes de recommandation filtrage collaboratif tentent de prédire l'utilité d'items pour un utilisateur particulier en fonction des items précédemment évalués par d'autres utilisateurs qui présentent des similarités dans les goûts de préférences [36]

<b>Critère</b>	<b>Filtrage collaboratif</b>	<b>Filtrage basé sur le contenu</b>	<b>Approche hybride</b>
<b>Principe</b>	Basé sur les interactions entre utilisateurs (ex. : "ceux qui ont aimé X ont aussi aimé Y")	Basé sur les caractéristiques des cours (mots-clés, thématiques, difficulté, etc.)	Combine les résultats des deux méthodes
<b>Dépendance aux autres utilisateurs</b>	Oui	Non	Oui (partielle)
<b>Données utilisées</b>	Historique des utilisateurs (notes, clics, achats)	Métadonnées des cours + historique personnel	Les deux types de données
<b>Avantages</b>	Recommandations personnalisées à partir de comportements collectifs	Bonnes suggestions même pour les nouveaux utilisateurs (cold Start)	Meilleure couverture et précision
<b>Limites</b>	Problèmes de démarrage à froid (cold Start), rareté des données	Moins de diversité dans les recommandations	Complexité computationnelle plus élevée
<b>Exemples de plateformes</b>	Coursera (comportement d'autres apprenants)	edX (basé sur les préférences et historique personnel)	Udemy, LinkedIn Learning
<b>Système de recommandation</b>	Collaborative Filtering, k-NN, Matrix Factorization	TF-IDF, Cosine Similarity, Embeddings	Hybrid (Weighted, Switching, Feature-level)
<b>Methods principaux</b>	ALS, SVD, User-based / Item-based kNN	Contenu textuel, profil utilisateur	Ensemble, réseaux de neurones, modèles hybrides
<b>Adaptability aux profile evolutes</b>	Faible (dépendance aux historiques)	Moyenne (profil évolutif possible)	Élevée (via mise à jour des poids entre méthodes)

Tableau1.2.1. : Comparaison entre les différentes approches

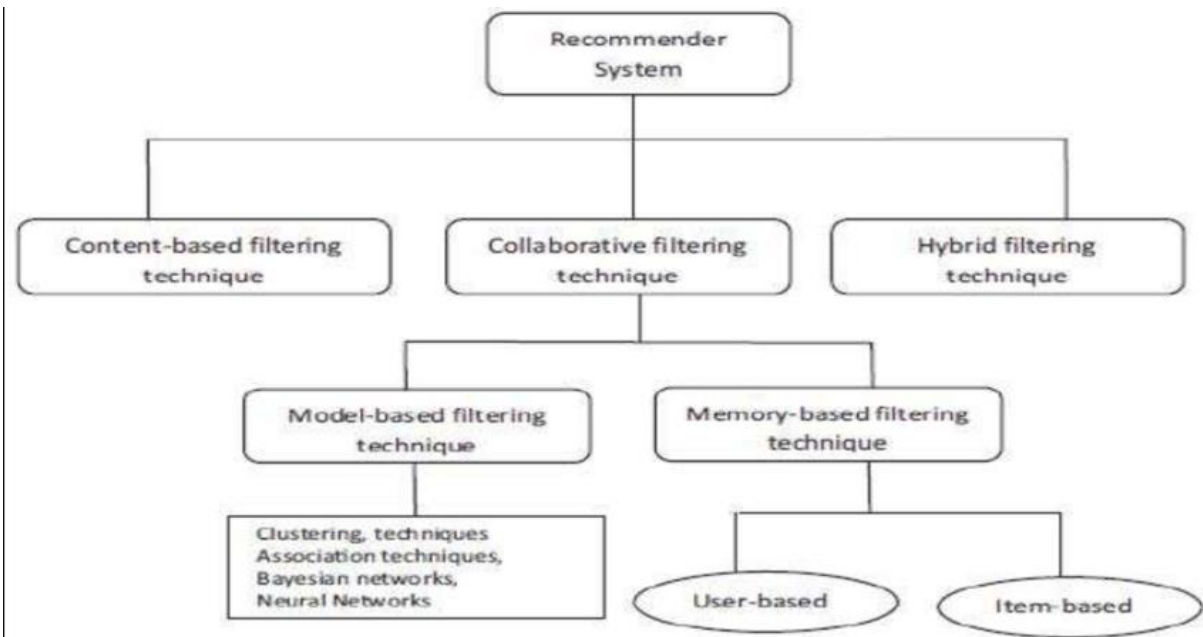


Figure1.1.2: Approches utilisées pour la recommandation [37]

### 1.2.5 Problèmes et limites des systèmes de recommandation

Les systèmes de recommandation, bien qu'efficaces pour personnaliser les contenus et améliorer l'expérience utilisateur, présentent plusieurs limitations notables comme :

#### 1.2.5.1 Le démarrage à froid (*cold Start*)

Est un problème courant dans les systèmes de recommandation, qui survient lorsque le système dispose de peu ou pas de données historiques sur les utilisateurs ou les éléments à recommander. Cela rend difficile la génération de recommandations pertinentes.

Il existe trois principaux types de démarrage à froid :

1. *Démarrage à froid utilisateur* : lorsqu'un nouvel utilisateur rejoint le système sans interactions préalables, rendant difficile la personnalisation des recommandations.

2. *Démarrage à froid élément* : lorsqu'un nouvel élément (produit, film, etc.) est ajouté au système sans évaluations ou interactions, compliquant son recommandation aux utilisateurs.
3. *Démarrage à froid système* : lors du lancement initial d'un système de recommandation, où il manque de données globales sur les utilisateurs et les éléments.

### 1.2.5.2 La sparsité des données

Désigne le fait qu'il y a très peu d'interactions enregistrées entre les utilisateurs et les éléments (produits, films, cours, etc.) dans un système de recommandation. Autrement dit, la majorité des utilisateurs n'ont évalué ou consulté qu'une petite partie des contenus disponibles.[38] *Par exemple* : Dans un site de e-Learning avec 1 000 cours et 10 000 utilisateurs, si chaque utilisateur a regardé seulement 10 cours, alors la plupart des cours n'ont reçu que très peu d'évaluations. Cela rend difficile de faire de bonnes recommandations, car il manque des données.

### 1.2.5.3 biais de popularité

Est un phénomène observé dans les systèmes de recommandation, où les éléments les plus populaires (ceux ayant reçu le plus de clics, de vues ou d'achats) sont systématiquement favorisés dans les suggestions faites aux utilisateurs. Cela conduit à une surreprésentation de ces éléments populaires, au détriment de contenus moins connus ou de niche. [39]

### 1.2.5.4 transparences des algorithmes

Désigne la capacité à comprendre, expliquer et vérifier le fonctionnement interne des algorithmes, notamment la manière dont ils traitent les données, prennent des décisions et les critères utilisés pour ces décisions. Cette transparence est essentielle pour garantir l'équité, la responsabilité et la confiance dans les systèmes basés sur l'intelligence artificielle [40]

### 1.2.5.5 protections de la vie privée

Désigne l'ensemble des mesures légales, techniques et organisationnelles visant à préserver la confidentialité, l'intégrité et la sécurité des informations personnelles d'un individu. Elle

permet à chacun de contrôler l'utilisation de ses données personnelles et de se prémunir contre les intrusions non autorisées dans sa sphère privée. [41]

Ces défis soulignent la nécessité d'une conception plus responsable et transparente des systèmes de recommandation, intégrant des mécanismes pour atténuer les biais, protéger la vie privée des utilisateurs et promouvoir la diversité des contenus proposés

## **1.2.6 Les techniques avancées basées sur l'intelligence artificielle et le machine Learning**

### ***1.2.6.1 Réseaux de neurones pour modéliser les préférences***

Les réseaux de neurones artificiels sont capables d'apprendre des représentations complexes des préférences des utilisateurs en analysant leurs interactions passées avec les contenus éducatifs. Par exemple, en utilisant des données telles que les évaluations, les clics ou le temps passé sur des ressources, ces réseaux peuvent prédire les contenus les plus pertinents pour chaque apprenant. Cette capacité à modéliser des relations non linéaires entre les utilisateurs et les items permet d'améliorer la personnalisation des recommandations. [42]

### ***1.2.6.2 Systèmes de type séquence-apprentissage (RNN, Transformer) pour suivre l'évolution du parcours de l'utilisateur***

Les modèles de type séquence-apprentissage, tels que les réseaux de neurones récurrents (RNN) et les Transformer, sont conçus pour traiter des données séquentielles. Dans le contexte de l'e-Learning, ils permettent de suivre l'évolution du parcours d'apprentissage d'un utilisateur en tenant compte de l'ordre des interactions. Par exemple, un RNN peut analyser la séquence des cours suivis pour prédire les prochaines étapes d'apprentissage les plus appropriées. Les Transformer, quant à eux, utilise des mécanismes d'attention pour capturer les dépendances à long terme dans les séquences, offrant ainsi des recommandations plus contextuelles. [43]

### **1.2 .6.3 Graphe neural pour modéliser les relations entre utilisateurs, cours et compétences**

Les Graph Neural Networks (GNN) sont des modèles conçus pour traiter des données structurées sous forme de graphes, où les nœuds représentent des entités (utilisateurs, cours, compétences) et les arêtes représentent les relations entre ces entités. Dans les systèmes de recommandation pour l'e-Learning, les GNN permettent de capturer les relations complexes entre les apprenants et les ressources pédagogiques. [44] Par exemple, en modélisant les interactions entre les utilisateurs et les cours, les GNN peuvent identifier des motifs d'apprentissage similaires et recommander des contenus adaptés. De plus, en intégrant des informations sur les compétences associées aux cours, les GNN peuvent aider à orienter les apprenants vers des parcours d'apprentissage cohérents avec leurs objectifs

### **1.2.7 Conclusion**

Un système de recommandation vise à réduire le temps de recherche des utilisateurs tout en leur proposant des ressources pertinentes qu'ils n'auraient peut-être pas découvertes autrement.

Dans ce chapitre, nous avons exploré différentes techniques de recommandation, notamment les approches basées sur le contenu, collaboratives, hybrides, entre autres. Nous avons également évoqué d'autres types de recommandation tels que la recommandation entre utilisateurs (people-to-people) ou encore celle basée sur les tags. Par ailleurs, nous avons mis en lumière les avantages et les inconvénients de chaque méthode, ce qui constitue une base essentielle pour orienter le choix de la technique la plus adaptée.