

**Louvain School of Management**

# **Influence de la Diversité dans les Systèmes de Recommandation sur la Consommation Intentionnelle des Utilisateurs**

Auteur : Teo Baldi  
Promoteur : Corentin Vande Kerckhove  
Co-promotrice : Camille Charles  
Année académique 2023-2024  
Master [120] Ingénieur de gestion, à finalité spécialisée

# Declaration Regarding AI Tool Usage in Master's Thesis

We recognize that AI tools might be valuable aids during the master's thesis work, but they are not infallible. Remember that transparency fosters trust, and acknowledging AI's role enhances the credibility of your work.

Therefore, when deciding to use such a tool, you need to adhere to the following principles of responsible use of AI.

## 1. Critical Evaluation :

- We critically assessed the AI-generated output, ensuring its alignment with our research objectives.
- Any modifications or corrections were made based on our expertise and domain knowledge.

## 2. Transparency :

- We acknowledge the use of ChatGPT transparently, emphasizing that it contributed to our work but did not replace human judgment.
- Our commitment to transparency ensures the integrity of this thesis.

## 3. Ethical Considerations :

- We actively monitored for biases or unintended consequences introduced by the AI tool.
- Our ethical responsibility guided our decisions throughout the research process.

# Declaration

During the preparation of this master's thesis, the author(s) utilized ChatGPT for the following purpose:

**1. Phrase/Idea Reformulation:** The tool was used to reformulate phrases and ideas to improve clarity and coherence.

**2. Python Code Debugging:** The tool assisted in debugging Python code, particularly for the construction of the Web app.

**3. Clarification of Theoretical Points:** The tool provided clarifications on certain theoretical points necessary for the general understanding of a scientific article.

After using ChatGPT, the author(s) diligently reviewed and edited the content produced by the tool. We take full responsibility for the final content presented in this thesis.

By signing this declaration, we affirm that the content of this master's thesis reflects our original work, augmented by the responsible use of AI.

BALDI Teo

16-05-2024





## RÉSUMÉ

Les systèmes de recommandation (SdR) occupent une place grandissante dans notre société. Ces algorithmes s'appuient sur les préférences, les choix passés et les attitudes des utilisateurs pour leur offrir des recommandations personnalisées. En outre, ils se sont révélés être des outils de décision efficaces pour aider les utilisateurs à trouver plus facilement et rapidement les produits qu'ils préfèrent, en particulier sur les plateformes de e-commerce. Ces dernières années, de nombreux critères, autres que la précision, ont suscité un intérêt croissant dans le domaine de l'évaluation des SdR. C'est le cas de la diversité qui est devenue l'un des critères les plus essentiels. À ce sujet, dans la littérature scientifique, de nombreux chercheurs ont montré que les utilisateurs étaient attirés par des listes d'articles diversifiées et qu'une recommandation diversifiée pouvait accroître la probabilité de répondre à leurs divers besoins. De plus, l'offre d'un contenu diversifié contribue au développement et à l'élargissement des préférences de l'utilisateur, réduisant ainsi le risque qu'il soit piégé dans des « bulles de filtre ». D'autres études ont également prouvé l'effet positif de la diversité sur la satisfaction des utilisateurs.

En revanche, aucune étude n'a tenté de comprendre quelle était l'influence de l'inclusion de la diversité dans les listes de recommandations sur le volume de consommation intentionnelle des utilisateurs. Par conséquent, ce mémoire tentera de répondre à cette question de recherche et représentera un premier effort dans l'étude de cette relation. Pour y parvenir, nous avons, tout d'abord, décidé d'investiguer le filtrage collaboratif et, plus précisément, l'approche basée sur la mémoire (*item-based*) à partir de la base de données MovieLens. Pour diversifier les recommandations de cette approche, nous nous sommes penchés sur le modèle de reclassement nommé *Maximal Marginal Relevance* (MMR). Cette méthode de diversification combine linéairement la pertinence et la diversité en se basant sur un paramètre ajustable qui contrôle le niveau de diversité. Ensuite, nous avons déployé une expérience participative qui a été menée dans une conception *between-subjects* et à partir d'une application en ligne réaliste. Les résultats obtenus auprès de 111 participants, répartis en quatre groupes qui se distinguent par le niveau de diversité inclus dans les recommandations (à savoir 0%, 40%, 60% ou 100%), démontrent l'importance capitale d'inclure une diversité modérée, entre 40% et 60%, dans les recommandations. En effet, cette diversification est cruciale pour garantir un équilibre optimal entre diversité et pertinence mais également pour améliorer la satisfaction des utilisateurs, accroître la consommation intentionnelle et promouvoir une diversification de la consommation à la fois individuelle et globale.

## REMERCIEMENTS

Je tiens à exprimer mes plus sincères remerciements à toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire.

Tout d'abord, je suis profondément reconnaissant envers le professeur Corentin Vande Kerckhove pour son rôle de promoteur et son suivi régulier. Ses conseils avisés ont contribué à formuler ma question de recherche de manière pertinente, à orienter mon travail dans la bonne direction et à émettre des critiques constructives qui ont enrichi ma réflexion et m'ont permis de progresser tout au long du processus de recherche. En outre, son assistance précieuse dans le paramétrage et le déploiement de l'application en ligne utilisée pour l'expérience a été d'une importance capitale.

Ensuite, je souhaite exprimer ma gratitude envers Camille Charles, assistante à l'UCLouvain Mons, pour son soutien. Ses relectures attentives et ses remarques constructives sur les procédés de recherche expérimentale ont grandement amélioré mon travail.

Je remercie également Colin Timmers, chercheur à l'UCLouvain, pour son aide et sa disponibilité au sujet du paramétrage de l'application en ligne et du partage de l'expérience sur Prolific.

De plus, je tiens à adresser des remerciements particuliers à mes parents, ma fiancée, ma sœur, et mon beau-frère ainsi qu'à tous mes proches. Leur soutien indéfectible ainsi que leur contribution à la relecture et à la correction de ce mémoire ont été d'une aide inestimable.

Enfin, je souhaite exprimer ma profonde reconnaissance envers toutes les personnes qui ont participé à l'expérience en ligne. Leur engagement a été essentiel pour la réalisation de cette recherche.

## TABLE DES MATIERES

LISTE DES FIGURES.....	V
LISTE DES TABLEAUX.....	VI
LISTE DES ANNEXES .....	VII
LISTE DES ABRÉVIATIONS .....	VIII
INTRODUCTION.....	1
CHAPITRE 1 : REVUE DE LITTÉRATURE .....	3
1. Systèmes de recommandation .....	3
1.1. Définition d'un système de recommandation.....	3
1.2. Fonctionnement d'un système de recommandation .....	4
1.3. Différents types de systèmes de recommandation .....	5
1.3.1. Approche basée sur le filtrage collaboratif .....	6
1.3.1.1.Calcul de similarité.....	6
1.3.1.2.Approche basée sur la mémoire .....	7
1.3.1.3.Approche basée sur le modèle.....	8
1.4. Évaluation d'un système de recommandation.....	10
1.4.1. Évaluation en ligne.....	10
1.4.2. <i>User study</i> .....	10
1.4.3. Évaluation hors-ligne .....	11
1.4.3.1.Évaluation de l' <i>accuracy</i> .....	11
1.4.3.2.Nouveauté, sérendipité et diversité .....	12
2. Diversité, dimension importante des systèmes de recommandation .....	13
2.1. Les limites de l' <i>accuracy</i> .....	13
2.2. L'adoption de la diversité .....	14
2.3. Diversité objective et subjective .....	15
3. Influence sur la consommation .....	17

3.1.	Comportement d'achat des consommateurs.....	17
3.2.	Volume des ventes.....	18
4.	Conclusion.....	20
<b>CHAPITRE 2 : DESIGN EXPÉRIMENTAL.....</b>		<b>21</b>
1.	Objectifs de la recherche.....	21
2.	Base de données MovieLens.....	22
3.	Algorithmes de recommandation.....	23
4.	Diversification des listes de recommandations.....	25
5.	Hypothèses à tester.....	28
6.	Procédure.....	30
7.	Participants.....	33
7.1.	Profil des répondants.....	34
7.2.	Habitudes et préférences cinématographiques.....	34
7.3.	Connaissance et utilisation des systèmes de recommandation.....	35
8.	Conclusion.....	35
<b>CHAPITRE 3 : RÉSULTATS ET ANALYSES.....</b>		<b>36</b>
1.	Satisfaction des consommateurs.....	36
2.	Consommation intentionnelle.....	37
2.1.	Volume de consommation intentionnelle.....	38
2.2.	Diversification de la consommation intentionnelle.....	39
3.	Diversité perçue dans les recommandations.....	40
4.	Pertinence perçue par les consommateurs face à leurs préférences.....	41
5.	Conclusion.....	41
<b>CONCLUSION GÉNÉRALE.....</b>		<b>43</b>
<b>BIBLIOGRAPHIE.....</b>		<b>45</b>
<b>ANNEXES.....</b>		<b>51</b>

## **LISTE DES FIGURES**

Figure 1. Les différents types de systèmes de recommandation étudiés dans le cadre de ce mémoire.....	6
Figure 2. Analyse de la satisfaction par groupe de participants .....	36
Figure 3. Résultats obtenus dans le cadre de l'analyse de la consommation intentionnelle des participants .....	37
Figure 4. Analyse de la diversité perçue par groupe de participants .....	40
Figure 5. Analyse de la pertinence perçue par groupe de participants .....	41

## LISTE DES TABLEAUX

Tableau 1. Aperçu numérique de la base de données finale.....	23
Tableau 2. Valeurs d'accuracy (MAE, RMSE) des meilleurs modèles <i>user-based</i> , <i>item-based</i> et SVD.....	24
Tableau 3. Variation de la diversité des recommandations en fonction du pourcentage de diversification et de l'algorithme de recommandation de départ.....	27
Tableau 4. Résultats des tests de Student unilatéraux pour des échantillons indépendants mesurant la significativité statistique (p-valeurs) des différentes hypothèses testées.....	37

## LISTE DES ANNEXES

Annexe 1 : Bases de données les plus populaires dans le domaine des systèmes de recommandation (Beel et Brunel, 2019) .....	51
Annexe 2 : Codes Python implémentés pour les trois algorithmes de recommandation <i>Item-Based</i> , <i>User-Based</i> et SVD .....	52
Annexe 3 : Valeurs d' <i>accuracy</i> (MAE, RMSE) des modèles <i>user-based</i> .....	54
Annexe 4 : Valeurs d' <i>accuracy</i> (MAE, RMSE) des modèles <i>item-based</i> .....	55
Annexe 5 : Valeurs d' <i>accuracy</i> (MAE, RMSE) des modèles SVD .....	56
Annexe 6 : Codes Python implémentés pour l'algorithme de diversification MMR .....	58
Annexe 7 : Code Python implémenté pour le calcul de l'Intra-List Distance .....	61
Annexe 8 : Visualisation de l'application en ligne et des différentes étapes de la procédure ..	62
Annexe 9 : Échelle de mesure de la pertinence des recommandations .....	67
Annexe 10 : Échelle de mesure de la satisfaction des utilisateurs face aux listes de recommandations .....	68
Annexe 11 : Échelle de mesure de la diversité perçue par les utilisateurs dans les listes de recommandations .....	69
Annexe 12 : Tableau récapitulatif des caractéristiques des participants .....	70
Annexe 13 : Résultats obtenus dans le cadre de l'analyse sur la satisfaction des utilisateurs ..	72
Annexe 14 : Comparaisons de moyennes (test de Student) dans le cadre de l'analyse sur la satisfaction moyenne des utilisateurs .....	73
Annexe 15 : Comparaisons de moyennes (test de Student) dans le cadre de l'analyse sur le volume moyen de consommation intentionnelle des utilisateurs .....	74
Annexe 16 : Comparaisons de moyennes (test de Student) dans le cadre de l'analyse sur la diversification de la consommation intentionnelle de l'utilisateur .....	75
Annexe 17 : Résultats obtenus dans le cadre de l'analyse de la diversité perçue par les participants .....	76
Annexe 18 : Comparaisons de moyennes (test de Student) dans le cadre de l'analyse sur la diversité moyenne perçue par les utilisateurs .....	77
Annexe 19 : Résultats obtenus dans le cadre de l'analyse de la pertinence perçue par les participants .....	78
Annexe 20 : Comparaisons de moyennes (test de Student) dans le cadre de l'analyse sur la pertinence moyenne perçue par les utilisateurs .....	79

## LISTE DES ABRÉVIATIONS

**SdR** : Système(s) de recommandation

$r_{ui}$  : Le *rating* que l'utilisateur  $u$  assigne à l'*item*  $i$

**CF** : Filtrage Collaboratif – ou *Collaborative Filtering*

**CBF** : Filtrage Basé sur le Contenu – ou *Content-Based Filtering*

**ILD** : Distance Intra-Liste – ou *Intra-List-Distance*

**SVD** : Décomposition en Valeurs Singulières – ou *Singular Value Decomposition*

**TF-IDF** : Fréquence des termes-Fréquence inverse du document – ou *Term Frequency-Inverse Document Frequency*

**MMR** : Pertinence Marginale Maximale – ou *Maximal Marginal Relevance*

## INTRODUCTION

Le volume d'informations en ligne ne cesse de croître chaque jour avec l'accès à un nombre de plus en plus conséquent de produits et services en ligne. Face à ce contexte de surcharge d'informations, il devient ardu pour les utilisateurs d'accéder à la bonne information au bon moment et de traiter sans aide toutes les informations et produits auxquels ils ont accès (Wu et al., 2022). À ce sujet, Herbert Simon (1947) a introduit le concept de rationalité limitée. Il souligne que les acteurs économiques ne peuvent traiter qu'un nombre limité d'informations et disposent de capacités cognitives restreintes pour optimiser leurs choix. Ce concept met donc en évidence le besoin pour les êtres humains d'avoir des outils pour les assister dans la prise de décision. Les systèmes de recommandation (SdR) en ligne en sont alors des solutions puissantes puisqu'ils tentent de résoudre ce problème de surcharge d'informations en offrant aux utilisateurs un moyen plus intelligent et personnalisé de trouver plus rapidement et facilement des informations ou articles (dits *items* en anglais) (Bradley et Smyth, 2001). En outre, les objectifs principaux d'un SdR sont, d'une part, d'améliorer l'expérience vécue par les utilisateurs sur la plateforme de recommandation et, par extension, leur satisfaction. D'autre part, un SdR permet d'optimiser les objectifs commerciaux de l'entreprise tels que les ventes ou l'engagement des utilisateurs (Kim et al., 2021).

Néanmoins, le critère qui prédomine depuis de nombreuses années dans le domaine de la recommandation est l'*accuracy* (Wu et al., 2022). Cette dernière mesure la capacité du SdR à identifier avec précision les *items* considérés comme les plus « pertinents » pour l'utilisateur en fonction de ses préférences individuelles, de ses choix passés, mais aussi des préférences des autres utilisateurs (Ziegler et al., 2005 ; Wu et al., 2022). Dès lors, au cours ces dernières années, de nombreux critères, autres que l'*accuracy*, ont suscité un intérêt croissant dans le domaine des SdR. C'est le cas de la diversité qui est devenue l'un des critères les plus essentiels (Wang et al., 2020). La diversité d'un SdR se réfère à sa capacité à proposer des *items* diversifiés (Ziegler et al., 2005). D'ailleurs, la notion de diversité a été définie par Smyth et McClave (2001) comme étant la dimension opposée à la similarité. Castagnos et al. (2013, p. 4) ont complété cette notion en la définissant « *comme une mesure quantifiant la dissimilarité dans un ensemble d'items* ». Toutefois, il convient de reconnaître que l'intégration de la diversité dans les SdR est un sujet sur lequel de nombreux chercheurs se sont penchés. Néanmoins, dans le cadre de ce mémoire, nous chercherons à aller plus loin et à enrichir la littérature en tentant de répondre à la question de recherche suivante : *Comment l'inclusion de la diversité dans les systèmes de recommandation, évalués online, influence-t-elle le volume de consommation*

*intentionnelle des utilisateurs ?* Pour atteindre cet objectif, nous aurons l'occasion de faire un état de l'art des SdR dans le premier chapitre. De plus, nous décrirons le point de vue des chercheurs quant à l'inclusion de la diversité dans les SdR et nous détaillerons les avantages et conséquences qu'elle peut générer pour les consommateurs et les fournisseurs d'*items*.

À la suite de cette revue de la littérature, nous serons en mesure, dans le deuxième chapitre, de réaliser notre propre expérience participative, via une application en ligne, afin de vérifier s'il existe un lien entre diversité et volume de consommation intentionnelle des utilisateurs. Pour y parvenir, nous commencerons, tout d'abord, par introduire la base de données sur laquelle repose notre recherche. Ensuite, à partir de cette base de données, nous évaluerons, en matière d'*accuracy*, la performance des différents algorithmes de recommandation basés sur le filtrage collaboratif (*user-based*, *item-based* et SVD). Pour diversifier les listes de recommandations, nous pencherons également sur la méthode de diversification *Maximal Marginal Relevance* (MMR), qui sera fondée sur l'algorithme de recommandation le plus précis parmi ceux qui auront été évalués. Finalement, nous préciserons la procédure expérimentale que nous avons menée via notre application en ligne, dans une conception *between-subjects*, afin de tester nos hypothèses et tenter de répondre à notre question de recherche.

En plus de vérifier l'influence des recommandations diversifiées à différents niveaux sur l'intention de consommation des utilisateurs, l'analyse des résultats obtenus aura également pour objectif d'évaluer d'autres variables dépendantes, à savoir la diversification de la consommation intentionnelle, la perception par les consommateurs de la pertinence des recommandations face à leurs préférences, la diversité perçue dans les recommandations, et la satisfaction des consommateurs vis-à-vis des listes de recommandations. En somme, dans le troisième chapitre, nous serons en mesure de tirer des conclusions particulièrement intéressantes pour les entreprises développant ou utilisant des SdR.

# CHAPITRE 1 : REVUE DE LITTÉRATURE

Tout d'abord, nous nous intéresserons aux systèmes de recommandation (SdR). Nous tenterons de les définir et d'expliquer leur fonctionnement. En outre, nous étudierons les différents types de SdR et nous nous concentrerons sur la manière dont ils sont évalués. Ensuite, nous nous attarderons plus précisément à l'importance de la notion de diversité dans les recommandations en analysant ce que les chercheurs ont déjà pu découvrir à ce sujet au travers de leurs études. Enfin, en analysant la littérature, nous nous attarderons sur le comportement d'achat des utilisateurs et l'effet positif des SdR sur le volume des ventes des entreprises.

## 1. Systèmes de recommandation

### 1.1. Définition d'un système de recommandation

Certains chercheurs, comme Castagnos et al. (2014), définissent les SdR comme des technologies qui visent à faciliter la recherche et l'accès à l'information en proposant des *items* adaptés aux besoins et aux préférences des utilisateurs. Helberger et al. (2018) complètent cette perspective en soulignant que les recommandations peuvent dépendre des préférences individuelles de l'utilisateur, de ses choix passés et des préférences des autres utilisateurs. C'est pourquoi, le concept de personnalisation est omniprésent. En effet, la personnalisation consiste à guider le consommateur vers des *items* correspondants au mieux à ses préférences (Chandra et al., 2022). Dans sa forme la plus courante, le SdR peut aussi se définir comme un système de filtrage capable de prédire les préférences d'un utilisateur et donc les notes (en anglais *ratings*) que cet utilisateur attribuerait aux *items* qu'il n'a pas encore considérés. En effet, cette approche exclut les *items* que l'utilisateur avait déjà l'intention de consommer, car, dans un tel cas, le SdR n'exercerait aucune influence. Une fois les *ratings* prédits, le SdR a pour but de recommander à l'utilisateur les *items* ayant les *ratings* estimés les plus élevés (Adomavicius et Tuzhilin, 2005).

Les SdR sont implémentés sur des plateformes de diffusion de contenus multimédias (Netflix, YouTube, Spotify...), de vente en ligne (Zalando, Amazon...), sur des réseaux sociaux (Facebook, Twitter...), etc. Quant aux *items* cités précédemment, ceux-ci peuvent être de tous types : films, musique, livres, pages web, produits alimentaires... (Bradley et Smyth, 2001 ; Candillier et al., 2009 ; Gunawardana et Shani, 2015).

Les SdR sont désormais incontournables dans de multiples aspects de notre quotidien. En qualité de consommateurs, ils nous guident dans nos choix d'achat, de divertissement tels que les films ou la musique. En tant qu'utilisateurs d'Internet, ils dirigent notre navigation vers des contenus pertinents à notre recherche. Par ailleurs, ils peuvent aussi influencer sur notre vie sociale en facilitant de nouvelles connexions sur les réseaux sociaux (centres d'intérêts

similaires entre utilisateurs, groupes ou connexions en commun) ou en identifiant les profils à recruter sur base d'une offre d'emploi (Du, 2021). Les entreprises mondiales du commerce électronique telles qu'Amazon, Netflix et Google, quant à elles, tirent aussi parti des SdR pour maintenir un avantage concurrentiel durable. En offrant des produits et services qui correspondent aux préférences des clients, elles simplifient la recherche d'offres pour ces derniers, ce qui a une influence positive sur la satisfaction client et les ventes (Kim et al., 2021). Le processus de recommandation peut aussi permettre de créer de la demande (proposer d'autres *items* pertinents) et de l'activité sur le site (Jannach et Adomavicius, 2016).

## 1.2. Fonctionnement d'un système de recommandation

Tout d'abord, il est important de préciser qu'un SdR repose sur un ensemble d'utilisateurs  $u \in U$  et d'un catalogue d'*items*  $i \in I$ . En ce qui concerne le fonctionnement des SdR, il se distingue en 3 étapes (Castagnos et al., 2013).

La première étape consiste en la collecte de données d'un individu ou d'un groupe. En effet, pour être capable de calculer des prédictions pour les utilisateurs, il est indispensable d'avoir à sa disposition des informations les concernant. De plus, il est nécessaire de rassembler un très grand nombre de données afin de pouvoir recommander avec pertinence. Ces données peuvent-être explicites ou implicites (Castagnos et al., 2013). Les données explicites sont celles transmises consciemment et volontairement par l'utilisateur. Par exemple, les *ratings* attribués aux produits et services, les réponses à des sondages, etc. En règle générale, les *ratings* (ou  $r_{ui}$ ), indiquant le niveau d'appréciation de l'utilisateur  $u$  par rapport à l'*item*  $i$ , sont souvent représentés sur une échelle graduée. Ils peuvent être des valeurs discrètes (comme, des entiers allant de 1 à 5), des valeurs continues (par exemple, n'importe quelle valeur entre -10 et 10), des catégories ordonnées (comme « Pas d'accord », « Neutre » et « D'accord » qui seront, ensuite, transformées en valeurs entières), ou des valeurs binaires (« J'aime », « J'aime pas ») (Aggarwal, 2016). Les données implicites, quant à elles, n'exigent pas la participation active des utilisateurs, car elles sont collectées par l'entreprise sans que l'utilisateur ne s'en rende compte réellement. Le nombre de clics (sur des liens, des boutons, des images, des vidéos, des articles...), l'historique de recherche, la durée passée sur un site ou les achats réalisés sont des exemples de données implicites (Zanker et al., 2019). Contrairement aux évaluations explicites, les données implicites sont généralement représentées par des évaluations unaires. Ces dernières permettent à un utilisateur d'indiquer qu'il apprécie un *item*, mais il n'existe pas de mécanisme pour signaler qu'il ne l'apprécie pas (Aggarwal, 2016). Par exemple, l'achat d'un

produit par un utilisateur suggère qu'il l'apprécie, tandis que le fait de ne pas acheter un produit ne signifie pas nécessairement que l'utilisateur ne l'aime pas (Wu et al., 2021).

La seconde étape est la classification et le traitement des données. Il s'agit d'analyser et de transformer les données de manière à pouvoir les exploiter. Les informations récoltées, qu'elles soient explicites ou implicites, constituent une base de données pouvant être représentée par une matrice, nommée  $R$ . À titre d'exemple, si les *items* sont des films, chaque ligne de la matrice représentera un utilisateur et chaque colonne un film. À l'intersection entre une ligne et une colonne, soit entre un utilisateur  $u$  et un film  $i$ , se trouve, s'il s'agit de données explicites, le *rating*  $r_{ui}$  que l'utilisateur  $u$  a attribué au film  $i$  (Castagnos et al., 2013).

Enfin, puisque l'objectif ultime d'un SdR est de proposer à l'utilisateur une courte liste ordonnée et composée des *items* en fonction de ses préférences et ses choix précédents, la dernière étape consiste donc d'abord à établir un système prédisant les évaluations (sous forme de *ratings*) de chaque utilisateur concernant les *items* qu'il n'a pas encore considérés ou évalués. Après vient la phase de recommandation, où le SdR ordonne les *items* en fonction des *ratings* prédits pour ne recommander qu'une liste des  $n$  *items* (top- $n$ ) ayant les prédictions les plus élevées (Du, 2021). Pour permettre cette dernière étape, il existe plusieurs types d'algorithmes.

### 1.3. Différents types de systèmes de recommandation

L'étude scientifique des algorithmes de recommandation en a apporté des centaines pouvant être classés en fonction de leur approche (Aggarwal, 2016) :

- **l'approche basée sur le contenu**, dite *content-based filtering (CBF)* qui repose sur la forte probabilité que les utilisateurs apprécient les *items* qui sont similaires, en termes de contenu, à ceux qu'ils ont consommé, et évalué positivement, dans le passé ;
- **l'approche basée sur le contexte** qui prend en compte le contexte actuel de l'utilisateur, comme l'heure de la journée, la localisation, le dispositif utilisé pour fournir des recommandations plus pertinentes ;
- **l'approche démographique** qui organise les produits en fonction des caractéristiques démographiques de l'individu comme l'âge, le genre et la classe sociale, et recommande ceux qui ont été populaires auprès de consommateurs présentant des profils démographiques similaires ;
- **l'approche basée sur la connaissance** par laquelle le processus de recommandation s'effectue sur la base des similitudes entre les exigences des clients (budget défini, préférences, des caractéristiques fonctionnelles...) et les descriptions des articles ;

- **l'approche basée sur la popularité** qui recommande des éléments populaires ou tendance à tous les utilisateurs ;
- **l'approche hybride** qui utilise différentes approches de recommandation pour une même tâche dans le but de combiner leurs meilleurs aspects et ainsi optimiser les recommandations.

Au travers de ce mémoire et comme démontré dans la *Figure 1*, nous nous concentrerons exclusivement sur une autre approche : le filtrage collaboratif.

### 1.3.1. Approche basée sur le filtrage collaboratif

Dans le cadre du filtrage collaboratif – ou *collaborative filtering* (CF), l'idée essentielle repose sur l'hypothèse que les opinions exprimées par d'autres utilisateurs peuvent être employées pour établir une prédiction raisonnable quant à la préférence de l'utilisateur cible pour un *item* qu'il n'a pas encore évalué. Par conséquent, les données d'entrée du système sont les évaluations des *items* par les utilisateurs, soit la matrice des *ratings*  $R$  construite après la récolte des données (Aggarwal, 2016). Deux catégories d'approches basées sur le CF se distinguent : celle basée sur la mémoire et celle basée sur un modèle.

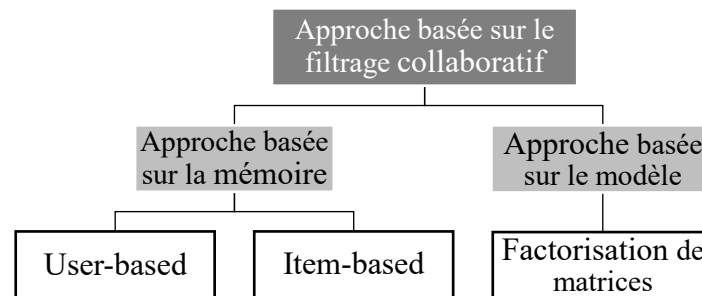


Figure 1. Les différents types de systèmes de recommandation étudiés dans le cadre de ce mémoire

#### 1.3.1.1. Calcul de similarité

Avant d'explorer ces algorithmes de recommandation, il est essentiel de déterminer comment mesurer la similarité entre les utilisateurs et entre les *items*. La littérature propose diverses mesures de similarité. Dans le cadre de ce mémoire, nous nous attarderons sur la similarité cosinus.

Chaque utilisateur est représenté par un vecteur (provenant de la matrice des *ratings*  $R$ ) où chaque composante correspond à l'évaluation par l'utilisateur d'un *item* particulier. De manière similaire, chaque *item* est également représenté par un vecteur où chaque composante correspond au *rating* de cet *item* par différents utilisateurs. Néanmoins, dans le cadre d'un calcul de similarité entre deux utilisateurs, seuls les *items* évalués en commun par ces deux utilisateurs (notés  $u$  et  $v$ ) sont considérés. À l'inverse, pour le calcul de similarité entre deux *items*, seuls

les utilisateurs qui ont évalué ces deux *items* (notés  $i$  et  $j$ ) sont pris en compte. Par conséquent, l'ensemble des *items* évalués par les utilisateurs  $u$  et  $v$  est donné par  $I_u \cap I_v$ . Par exemple, si l'utilisateur  $u$  a évalué les *items* 1, 3 et 5 et l'utilisateur  $v$  les quatre premiers *items*, alors  $I_u = \{1, 3, 5\}$ ,  $I_v = \{1, 2, 3, 4\}$  et  $I_u \cap I_v = \{1, 3, 5\} \cap \{1, 2, 3, 4\} = \{1, 3\}$ . De même,  $U_i \cap U_j$  représente l'ensemble des utilisateurs qui ont évalué à la fois l'*item*  $i$  et  $j$  (Aggarwal, 2016).

- *Similarité cosinus*

La similarité cosinus calcule la similarité entre des utilisateurs ou des *items* sur base du cosinus de l'angle entre deux vecteurs de même dimension (Du, 2021). Cette similarité peut varier entre -1 et 1. Une valeur de 0 signifie qu'il n'existe aucune similarité, contrairement à une valeur de 1 qui signifie une similarité parfaite. Une valeur de -1, en revanche, indique que les vecteurs sont parfaitement opposés (Wang et al., 2024). La similarité cosinus entre les utilisateurs  $u$  et  $v$  et entre les *items*  $i$  et  $j$  sont données par les équations (1) et (2), où  $r_{ui}$  représente, par exemple, le *rating*  $r$  donné par l'utilisateur  $u$  à l'*item*  $i$  (Aggarwal, 2016).

$$\begin{array}{c} \text{Sim}(u, v) \\ = \frac{\sum_{i \in I_u \cap I_v} r_{ui} \times r_{vi}}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{ui})^2} \times \sqrt{\sum_{i \in I_u \cap I_v} (r_{vi})^2}} \\ (1) \end{array} \quad \left| \quad \begin{array}{c} \text{Sim}(i, j) \\ = \frac{\sum_{u \in U_i \cap U_j} r_{ui} \times r_{uj}}{\sqrt{\sum_{u \in U_i \cap U_j} (r_{ui})^2} \times \sqrt{\sum_{u \in U_i \cap U_j} (r_{uj})^2}} \\ (2) \end{array} \right.$$

### 1.3.1.2. Approche basée sur la mémoire

- *Voisinage basé sur les utilisateurs*

Grâce à la matrice des *ratings*  $r_{ui}$  et au calcul de similarité entre les utilisateurs, une comparaison entre ces derniers peut être réalisée sur base de leur appréciation commune des *items*, ce qui crée la notion de voisinage des utilisateurs, ou de *peer group* en anglais. Des utilisateurs qui ont évalué des *items* de la même manière dans le passé auront beaucoup de chances de produire la même évaluation dans le futur. L'équation (3) permet de calculer la note prédite  $\widehat{r}_{uj}$  de l'utilisateur cible  $u$  pour l'*item* donné  $j$  à partir de son voisinage  $P_u(j)$  (soit à partir de l'ensemble de ses  $k$  voisins les plus similaires qui ont également évalué l'*item*  $j$ ) :

$$\widehat{r}_{uj} = \bar{r}_u + \frac{\sum_{v \in P_u(j)} \text{Sim}(u, v) \times (r_{vj} - \bar{r}_v)}{\sum_{v \in P_u(j)} |\text{Sim}(u, v)|} \quad (3)$$

Tout d'abord, il est nécessaire de calculer la moyenne pondérée des *ratings* fournis par les individus du voisinage. En règle générale, la manière dont les utilisateurs notent les *items* est biaisée par leur personnalité, leur état d'esprit ou par le contexte dans lequel ils se trouvent

(Du, 2021). C'est pourquoi, pour prendre en compte la disparité des systèmes de notation des utilisateurs, les *ratings* doivent être, ensuite, centrés sur la moyenne de la manière suivante :  $(r_{vj} - \bar{r}_v)$ . De plus, comme démontré dans l'équation (3),  $(r_{vj} - \bar{r}_v)$  est multipliée par le degré de similarité entre l'utilisateur cible et l'individu du voisinage  $Sim(u, v)$ . Cela permet de donner plus d'importance aux *ratings* des utilisateurs fort similaires à  $u$ . Enfin, la note moyenne de l'utilisateur cible  $\bar{r}_u$  est ensuite additionnée à cette prédiction pour fournir une prédiction de note « brute »  $\widehat{r}_{uj}$  de l'utilisateur cible  $u$  pour l'item  $j$  (Aggarwal, 2016).

- *Voisinage basé sur les items*

Les *items* peuvent également être comparés en fonction de l'appréciation partagée des utilisateurs, ce qui crée la notion de voisinage des *items*. Si plusieurs utilisateurs donnent des évaluations similaires à deux *items* distincts, cela signifie qu'un utilisateur évaluera ces deux *items* de manière similaire. En d'autres termes, lorsque les évaluations données par différents utilisateurs à deux *items* sont les mêmes ou très proches, cela indique une similitude entre ces *items* du point de vue de ces utilisateurs (Du, 2021). Parallèlement au voisinage basé sur les utilisateurs, l'équation (4) permet de calculer la note prédite  $\widehat{r}_{uj}$  de l'utilisateur  $u$  envers l'item  $j$  où  $P_j(u)$  est l'ensemble des  $k$  *items* les plus proches de l'item  $j$  ayant été notés par  $u$  :

$$\widehat{r}_{uj} = \bar{r}_j + \frac{\sum_{i \in P_j(u)} Sim(j, i) \times (r_{ui} - \bar{r}_i)}{\sum_{i \in P_j(u)} |Sim(j, i)|} \quad (4)$$

### 1.3.1.3. Approche basée sur le modèle

Les approches basées sur des modèles utilisent la matrice des *ratings*  $R$  pour entraîner un modèle avec des techniques d'apprentissage automatique (*machine learning*) ou d'exploration de données (*data mining*) dans le but d'accroître la performance des algorithmes de recommandation (Kim et al., 2021). Dans ce type de modèle, il existe donc une claire distinction entre la phase d'entraînement et la phase de prédiction. Autrement dit, pendant la phase d'entraînement, ces algorithmes extraient de l'information des données présentes dans la matrice  $R$ . En effet, leur objectif premier est d'apprendre les relations présentes dans les données. Ensuite, une fois la phase d'entraînement terminée, les algorithmes utilisent l'information extraite comme un modèle pour effectuer des prédictions et donc générer des recommandations sans nécessiter la réutilisation de l'ensemble de données à chaque fois. Ces méthodes ont donc l'avantage d'être rapides lors de la phase de prédiction, tout en étant plus précises sur la capacité à prédire les *ratings* ou les préférences des utilisateurs. Lorsque le modèle est paramétré, ses paramètres sont ajustés sur la base des *ratings* disponibles dans le jeu de données, en optimisant une fonction objectif (Aggarwal, 2016).

Concentrerons-nous plus précisément sur les méthodes de factorisation matricielle, telles que le SVD (*Singular Value Decomposition*). À la suite de la compétition lancée par Netflix en 2006 visant à trouver le meilleur algorithme de CF, des avancées dans le domaine du CF ont été observées et les méthodes qui s’inspirent de la décomposition matricielle ont gagné en popularité. Le modèle de factorisation matricielle SVD en est l’un des plus populaires dans le domaine du CF. L’idée centrale d’un tel modèle est de représenter la matrice des *ratings*  $R \in \mathbb{R}^{m \times n}$  comme une estimation basée sur le produit de deux matrices de dimension inférieure, à savoir celle des facteurs latents des utilisateurs  $U \in \mathbb{R}^{m \times k}$  par la matrice transposée des facteurs latents des *items*  $V^T \in \mathbb{R}^{k \times n}$ . Une fois que les matrices  $U$  et  $V$  ont été estimées, l’ensemble de la matrice des *ratings*  $R$  peut être approximée telle que  $UV^T$ , ce qui permet de prédire tous les *ratings* manquants (Aggarwal, 2016) :

$$R \approx UV^T \quad (5)$$

En réalité, la factorisation matricielle permet de représenter à la fois les utilisateurs et les *items* dans un espace latent commun de dimension  $k$ . Dans cet espace, les interactions entre les utilisateurs et les *items* sont modélisées comme des produits scalaires. Cet espace latent vise à expliquer les *ratings* en caractérisant tant les utilisateurs que les *items*. Par exemple, prenons le cas de la recommandation de films où les concepts latents pourraient représenter des dimensions évidentes telles que la comédie, la romance, les films d’action, etc. Ils pourraient également inclure des dimensions moins intuitives, comme la durée du film, voire des dimensions non directement interprétables - celles que le modèle a appris au cours de l’entraînement. Plus la correspondance entre les facteurs latents d’un utilisateur et d’un *item* est grande, plus il est probable que le film corresponde aux goûts de l’utilisateur (Du, 2021).

Pour en revenir à l’équation (5), l’erreur de cette approximation est égale à  $\|R - UV^T\|^2$  où  $\|\cdot\|^2$  représente la somme des carrés des entrées de la matrice résiduelle  $(R - UV^T)$ . Cette quantité est également appelée norme de Frobenius (au carré) de la matrice résiduelle.  $\|R - UV^T\|^2$  mesure ainsi la distance entre chaque *item* de la matrice  $R$  et son équivalent dans l’approximation  $UV^T$ . Tel que démontré dans l’équation d’optimisation (6), l’objectif est de minimiser cette distance, c’est-à-dire de trouver les valeurs optimales pour les matrices  $U$  et  $V$  de telle sorte que la différence entre la matrice  $R$  et son approximation soit réduite au maximum. In fine, une erreur minimale permettra une meilleure précision dans la prédiction des *ratings* ou des préférences des utilisateurs.

$$\text{Min } J = \frac{1}{2} \|R - UV^T\|^2 \quad (6)$$

## 1.4. Évaluation d'un système de recommandation

Comme décrit dans les sections précédentes, le cœur des SdR réside dans l'algorithme de recommandation lui-même. L'évaluation des performances de ces algorithmes est fondamentale pour sélectionner les meilleurs et optimiser leurs fonctionnalités. Elle permet aux chercheurs d'analyser et de comprendre les avantages et les inconvénients de différentes approches. Elle est également cruciale pour l'optimisation continue des systèmes existants. En règle générale, tout nouveau SdR doit passer par trois étapes d'évaluation des performances avant d'être pleinement déployé : l'analyse hors ligne, l'étude des utilisateurs et l'expérimentation en ligne (Chen et Liu, 2017).

### 1.4.1. Évaluation en ligne

L'évaluation en ligne (*online*) implique la réalisation de tests à grande échelle sur des SdR déjà déployés et opérationnels dans un environnement réel. Elle permet d'évaluer ou de comparer divers SdR en fonction des tâches réelles effectuées par de vrais utilisateurs. Parmi les méthodes d'évaluation, l'évaluation *online* est celle qui permet d'obtenir les résultats les plus réels. Cette approche d'évaluation présente l'avantage d'analyser l'ensemble des performances des SdR. En effet, elle permet de comprendre comment certains paramètres tels que la précision des prédictions ou la diversité des recommandations influencent les performances globales du système (par exemple, sur le volume de consommation intentionnelle des utilisateurs). Dans de nombreuses applications de recommandation réelles, les concepteurs de ces systèmes cherchent à influencer les comportements des utilisateurs. Par conséquent, lorsque les utilisateurs interagissent avec des SdR basés sur différents algorithmes, les concepteurs cherchent à évaluer via des expériences *online* comment ces SdR influencent les comportements des utilisateurs (Chen et Liu, 2017).

### 1.4.2. User study

L'étude des utilisateurs (*user studies*) est une méthode essentielle pour évaluer les SdR. Contrairement à l'évaluation *online*, cette méthode implique le recrutement d'un ou plusieurs groupe(s) de volontaires qui sont invités à effectuer des tâches spécifiques en utilisant les SdR (Alchiekh Haydar, 2014). Elle a pour but de tester l'interaction entre les utilisateurs et les SdR, ainsi que d'analyser l'influence de ces systèmes sur les comportements des utilisateurs. De plus, l'étude des utilisateurs peut servir à recueillir des données qualitatives, par exemple des informations sur l'interface utilisateur ou la complexité des tâches, qui revêtent une grande importance pour compléter et expliquer les résultats quantitatifs obtenus (Chen et Liu, 2017).

### 1.4.3. Évaluation hors-ligne

De manière générale, les principaux objectifs de l'évaluation hors-ligne (*offline*) sont de comparer les performances des algorithmes de recommandation en fonction de certaines mesures, de filtrer les algorithmes non-performants et de conserver les algorithmes candidats les plus performants (Chen et Liu, 2017). L'évaluation *offline* peut se baser sur les prédictions et/ou les listes des top-n *items* recommandés (Du, 2021). Par conséquent, les SdR seront comparés en utilisant des jeux de données préalablement collectées et sur base de métriques d'évaluation standardisées, telles que la précision, la nouveauté, la sérendipité ou la diversité. L'avantage de l'évaluation hors-ligne est qu'elle n'exige aucune interaction avec des utilisateurs réels et permet donc de comparer un large éventail d'algorithmes candidats à un faible coût. L'inconvénient des évaluations hors-ligne réside dans leur incapacité à mesurer la propension réelle de l'utilisateur à réagir face au SdR (Gunawardana et Shani, 2015).

#### 1.4.3.1. Évaluation de l'*accuracy*

Afin de calculer la précision (*accuracy*) d'un SdR, l'ensemble de données (ou matrice de *ratings*  $R$ ) est divisé en deux parties : l'ensemble d'apprentissage (*trainset*) et l'ensemble de test (*testset*). Tout d'abord, l'ensemble des *items* repris dans le *trainset* permettra de construire et d'entraîner le modèle de recommandation. Cette étape permet au modèle de comprendre les préférences des utilisateurs afin de prédire leurs futurs choix. Ensuite, les *items* appartenant au *testset* seront utilisés par le modèle pour réaliser des prédictions et évaluer sa performance en comparant ces prédictions aux vrais *ratings* des utilisateurs. Cette méthode utilisée est la technique de validation croisée *Hold Out* (*cross-validation Hold Out*) (Chen et Liu, 2017).

En ce qui concerne l'évaluation de l'*accuracy* sur base de la prédiction des *ratings* et puisque les *ratings* sont généralement exprimés sous forme de valeurs numériques, l'*accuracy* représente l'erreur de prédiction, c'est-à-dire l'écart entre le *rating* prédit et le *rating* réel. Il est donc souhaitable d'obtenir un écart aussi petit que possible afin de générer, in fine, des recommandations en pertinence avec les préférences des utilisateurs. Trois mesures permettent d'évaluer l'*accuracy* sur base de la prédiction des *ratings* :

<p style="text-align: center;">Mean Absolute Error (MAE)</p> $= \frac{1}{ Q } \sum_{(u,i) \in Q}  r_{ui} - \widehat{r}_{ui} $ <p style="text-align: center;">(7)</p>	<p style="text-align: center;">Mean Square Error (MSE)</p> $= \frac{1}{ Q } \sum_{(u,i) \in Q} (r_{ui} - \widehat{r}_{ui})^2$ <p style="text-align: center;">(8)</p>	<p style="text-align: center;">Root Mean Squared Error (RMSE)</p> $= \sqrt{\frac{1}{ Q } \sum_{(u,i) \in Q} (r_{ui} - \widehat{r}_{ui})^2}$ <p style="text-align: center;">(9)</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

$Q$  est le *testset*,  $r_{ui}$  représente les vrais *ratings* de l'utilisateur  $u$  et  $\widehat{r}_{ui}$  représente les *ratings* prédits par le SdR pour l'utilisateur  $u$ . MAE est la plus simple, elle calcule la moyenne des valeurs absolues des écarts entre les *ratings* réels et les *ratings* prédits. Néanmoins, elle ne tient pas compte de la direction de l'erreur (erreur positive ou négative). MSE calcule la moyenne des carrés des écarts entre les *ratings* réels et les *ratings* prédits. Elle pénalise davantage les erreurs importantes mais l'erreur au carré n'a pas de signification intuitive. Par conséquent, RMSE qui est simplement la racine carrée du MSE s'avère être plus largement utilisée pour calculer l'*accuracy* de prédiction du SdR (Chen et Liu, 2017).

Nous pouvons également distinguer l'évaluation de l'*accuracy* basée sur la liste de top- $n$  recommandations. Cette dernière ne considère que les *items* présents dans la liste des *items* recommandés afin d'en mesurer l'utilité pour l'utilisateur concerné (Du, 2021). La précision peut être une mesure adéquate à ce type d'évaluation :

$$\text{Précision} = \frac{N_{rs}}{N_s} \quad (10)$$

où  $N_{rs}$  est le nombre d'*items* recommandés et préférés par l'utilisateur et  $N_s$ , le nombre d'*items* recommandés. La précision de la liste de recommandation décrit donc la proportion d'*items* que l'utilisateur préfère (Chen et Liu, 2017).

#### 1.4.3.2. Nouveauté, sérendipité et diversité

Outre l'objectif d'*accuracy*, il est également important de considérer d'autres objectifs complémentaires lors de l'évaluation des recommandations générées par un algorithme :

- **la nouveauté** : la nouveauté d'un SdR réside dans sa capacité à recommander des *items* que l'utilisateur ne connaît pas encore (Gunawardana et Shani, 2015) ;
- **la sérendipité** : la sérendipité, étant un niveau supérieur de nouveauté, permet de mesurer la surprise dans les recommandations proposées à l'utilisateur (Aggarwal, 2016) ;
- **la diversité** : la diversité d'un SdR se réfère à sa capacité à proposer des *items* diversifiés (Ziegler et al., 2005). Par exemple, si un utilisateur se voit recommander cinq films, et que ces cinq films sont tous du même genre et mettent en scène les mêmes acteurs, alors ces recommandations manquent de diversité. D'ailleurs, la notion de diversité a été définie par Smyth et McClave (2001) comme étant la dimension opposée à la similarité dans les SdR. Castagnos et al. (2013, p. 4) ont complété cette notion en la définissant « *comme une mesure quantifiant la dissimilarité dans un ensemble d'items* ».

## 2. Diversité, dimension importante des systèmes de recommandation

### 2.1. Les limites de l'*accuracy*

L'objectif principal des SdR réside dans l'amélioration de la satisfaction des utilisateurs en leur proposant des recommandations personnalisées. Depuis longtemps, le critère dominant dans le domaine de la recommandation est l'*accuracy* (MAE, MSE, précision...) décrite dans la sous-section précédente et mesurée par la capacité du système à identifier avec précision les *items* considérés comme les plus « pertinents » pour l'utilisateur. Par conséquent, la priorité est donnée à la similarité (Bradley et Smyth, 2001 ; Ziegler et al., 2005 ; Wang et al., 2020 ; Wu et al., 2022). Néanmoins, cela peut impliquer une diminution de la qualité des recommandations finales et altérer l'expérience vécue par l'utilisateur sur la plateforme de recommandations (Bradley et Smyth, 2001). Par ailleurs, la satiété et la réduction de la satisfaction se produisent lorsque nous consommons de manière répétée un produit ou une caractéristique de produit (Coombs et Avrunin, 1977). De plus, comme démontré par McNee et al. en 2002, la satisfaction des utilisateurs ne coïncide pas nécessairement avec une grande précision dans les SdR.

En fait, les mesures de précision jouent un rôle significatif dans le domaine des SdR en fournissant un moyen de comparer les différents algorithmes et ne sont, par conséquent, pas à négliger. Cependant, il est important de ne pas les utiliser seules pour évaluer les performances des SdR. En ce sens, les auteurs précisent : « *Nous devons à présent nous intéresser de près aux utilisateurs des systèmes de recommandation. Ils ne se soucient pas d'utiliser un algorithme qui a obtenu un meilleur score sur une métrique, ils veulent une recommandation significative.* » (McNee et al., 2006). Deux ans plus tôt, Herlocker et al. (2004, p. 39) annonçaient dans la même direction que : « *Les systèmes de recommandation ne doivent pas fournir uniquement de la précision, mais aussi de l'utilité.* » En d'autres termes, les SdR les plus précis, en matière d'*accuracy*, ne sont donc pas automatiquement les plus bénéfiques pour les utilisateurs finaux puisqu'ils peuvent engendrer des répercussions négatives tant pour ces utilisateurs que pour les fournisseurs<sup>1</sup> (d'*items*) (Wu et al., 2022). En effet, ils peuvent entraîner, chez les utilisateurs, la création d'une « bulle de filtre » et nuire à leur satisfaction à long terme. La « bulle de filtre », définie par Pariser (2012) dans son ouvrage *The Filter Bubble : How the New Personalized Web Is Changing What We Read and How We Think*, fait référence au fait que les utilisateurs sont principalement exposés à des informations qui correspondent à leurs croyances et intérêts existants. Les utilisateurs se retrouvent donc enfermés dans leurs préférences dominantes, ce

---

<sup>1</sup> Le processus commence par les fournisseurs qui fournissent des *items* à une ou plusieurs plateformes de vente, qui à leur tour les proposent aux clients/utilisateurs finaux.

qui restreint leur capacité à découvrir de nouveaux éléments (Pariser, 2012). Imaginons un SdR de films qui ne suggère que des films de science-fiction à un utilisateur, simplement parce qu'il a adoré le film « Terminator ». Cette pratique peut conduire l'utilisateur à rester enfermé dans un seul genre de films, ce qui ne reflète pas nécessairement ses goûts réels, puisque le SdR n'explorera au final jamais ses divers centres d'intérêt. Cette absence de diversité peut donc provoquer de la frustration et compromettre l'expérience utilisateur du client (Zhang et Hurley, 2008). Les fournisseurs, d'autre part, sont souvent confrontés à des inégalités en matière de visibilité en raison du phénomène bien connu de l'économie des « *Superstars* », tel que décrit par Rosen en 1981. Dans cette économie, un infime nombre d'articles et de fournisseurs extrêmement populaires accaparent une part disproportionnée de l'exposition aux clients. Cette injustice peut entraîner une désillusion chez les fournisseurs méconnus qui ont du mal à attirer des clients et finissent par quitter les plateformes en ligne (Wu et al., 2022).

## 2.2. L'adoption de la diversité

Ainsi, ces dernières années, de nombreux critères, autres que l'*accuracy*, ont suscité un intérêt croissant dans le domaine des SdR. C'est le cas de la diversité qui est devenue l'un des critères les plus essentiels et l'une des premières motivations à biaiser les SdR pour diversifier les recommandations. En effet, l'objectif n'est plus de proposer uniquement des produits en adéquation avec les préférences du consommateur, mais d'y joindre aussi de la diversité en ajustant la fonction objective des algorithmes (Wang et al., 2020).

Il a été reconnu que les utilisateurs étaient attirés par des listes d'articles diversifiées (Willemsen et al., 2016 ; Ferwerda et al., 2017) et qu'une recommandation diversifiée pouvait accroître la probabilité de répondre aux divers besoins des utilisateurs. En effet, l'offre d'un contenu diversifié optimise les chances qu'au moins un *item* plaise à l'utilisateur (Castells, 2011), ce qui est essentiel pour améliorer l'expérience utilisateur. En outre, la diversification des recommandations s'est révélée indispensable pour contrer la difficulté perçue par l'utilisateur lorsqu'il doit faire face à un ensemble non-diversifié d'*items* (Willemsen et al., 2016). Concernant les fournisseurs moins populaires, la diversité peut aussi jouer un rôle crucial dans l'augmentation de la visibilité de leurs articles (Wu et al., 2022). De plus, d'après les travaux de Helberger et al. en 2018, il est à ajouter que la promotion de la diversité au sein des SdR peut faciliter la lutte contre les « bulles de filtre » parce qu'elle permet de contribuer au développement et à l'élargissement des préférences de l'utilisateur.

Pour répondre à cette nécessité de diversité dans les recommandations, bon nombre d'études ont été menées en collaboration avec des utilisateurs pour évaluer l'influence de

l'inclusion de la diversité dans les recommandations. Ziegler et al. ont identifié, dans leurs recherches en 2005, qu'Amazon.com recommandait plusieurs livres du même auteur créant collectivement une liste de recommandations très peu diversifiée. C'est pourquoi, ils ont introduit le concept de diversification des thèmes (*topic diversification*), un algorithme visant à accroître la diversité d'une liste top-n de produits recommandés (à savoir des livres). Leurs résultats ont démontré que l'appréciation globale des listes de recommandations par les utilisateurs ne se limitait pas à la précision, mais incluait également des facteurs tels que la diversité<sup>2</sup> perçue dans la liste. Ces résultats empiriques ont donc renforcé l'idée que les listes de recommandations ne sont pas de simples agrégations de recommandations individuelles, mais qu'elles apportent une valeur ajoutée intrinsèque. Par ailleurs, une autre expérience menée auprès de 250 utilisateurs par Castagnos et al. en 2013 a corroboré l'effet positif de la diversité (de films) sur la satisfaction<sup>3</sup> des utilisateurs. En outre, cette satisfaction se traduit de manière indirecte par des avantages pour l'entreprise, comme un renforcement de la fidélisation et de l'engagement de sa clientèle (Castells et al., 2011). Cependant, Castagnos et al. (2013) ont également souligné la nécessité d'utiliser la diversité avec parcimonie. En effet, une utilisation excessive pourrait engendrer de la méfiance et de l'incompréhension chez l'utilisateur car, en règle générale, l'augmentation de la diversité est associée à une réduction de la précision des recommandations (Bradley et Smyth, 2001 ; Zhou et al., 2008 ; Adomavicius et Kwon, 2012 ; Wang et al., 2020). Dès lors, accorder une importance excessive à la diversité dans la liste des *items* recommandés pourrait compromettre considérablement la pertinence, ce qui n'est pas une option acceptable. L'équilibre entre *accuracy* et diversité représente donc un réel enjeu pour les SdR (Adomavicius et Kwon, 2012).

### 2.3. Diversité objective et subjective

Il convient de préciser que Ge et al. (2011) ont avancé l'idée de décomposer le concept de diversité en deux aspects distincts afin de mieux la définir : la diversité inhérente (objective) et la diversité perçue (subjective).

La diversité inhérente se réfère à la diversité basée sur la dissimilitude entre les recommandations. Elle peut être subdivisée en deux sous-catégories : la diversité individuelle et la diversité globale/agrégée (Ge et al., 2011). Tandis que la diversité individuelle se réfère à la diversité au sein d'une liste de recommandations spécifique à un utilisateur donné, la diversité agrégée, quant à elle, mesure la capacité globale d'un SdR (pour l'ensemble des utilisateurs) à

---

<sup>2</sup> Les utilisateurs ont exprimé leur préférence pour des listes diversifiées à environ 30% (Ziegler, 2005).

<sup>3</sup> Castagnos et al. (2013) ont calculé la satisfaction sur base des *ratings* que les participants ont attribués aux *items* recommandés.

recommander des *items* variés. La distance intra-liste (*Intra-List Distance* ou ILD) peut être considérée comme la véritable mesure de la diversité objective. C'est pourquoi, la diversité d'un ensemble d'*items* recommandés est définie comme la distance moyenne entre les couples d'*items* de l'ensemble :

$$ILD = \frac{1}{|R|(|R| - 1)} \sum_{i \in R} \sum_{j \in R} d(i, j) \quad (11)$$

Dans l'équation (11),  $d(i, j)$  est une mesure de distance<sup>4</sup> et  $|R|$  est le nombre de recommandations faites à un utilisateur. Par défaut, la définition de cette métrique est pour un utilisateur cible. Néanmoins, la métrique peut aussi être calculée en moyenne pour tous les utilisateurs (Castells et al., 2011). En considérant le compromis entre la précision et la diversité, plusieurs chercheurs (Ziegler et al., 2005 ; Zhang et Hurley, 2008) ont développé des algorithmes visant à accroître la diversité individuelle, même si cela peut entraîner une diminution de la précision. L'objectif de leurs travaux était d'optimiser l'équilibre entre la précision et la diversité afin de maintenir un certain niveau de précision tout en favorisant la diversité. D'autre part, d'autres chercheurs (Zhou et al., 2010 ; Adomavicius et Kwon, 2012) se sont concentrés sur l'amélioration de la diversité globale dans le but de résoudre ce dilemme précision-diversité. Cependant, une étude de Fleder et Hosanagar (2007) a mis en évidence que la diversité individuelle et la diversité globale ne sont pas nécessairement interdépendantes.

Quant à la diversité perçue, elle fait référence à la diversité ressentie par l'utilisateur. L'intérêt de la diversité perçue réside dans sa capacité à saisir l'opinion de l'utilisateur concernant la diversité (Ge et al., 2011). Toutefois, bien que les utilisateurs puissent percevoir la diversité de manière différente, de nombreuses études ont prouvé que la diversité perçue constitue un élément central dans le domaine des recommandations. En effet, les conclusions de l'étude menée par Ziegler et al. (2005) ont indiqué que la satisfaction générale des utilisateurs à l'égard des listes de recommandations ne repose pas uniquement sur la précision des recommandations, mais englobe également des aspects tels que la diversité perçue par les utilisateurs. De plus, Knijnenburg et al. ont introduit, en 2012, un cadre pour évaluer l'expérience utilisateur, et notamment la diversité perçue. Ils ont également mis en lumière que lorsque les utilisateurs perçoivent les recommandations comme diversifiées, cela facilite leur processus de sélection, entraînant ainsi une satisfaction globale plus accrue à l'égard du système. Ekstrand et al. (2014), après avoir adapté les questions utilisées par Knijnenburg et al. (2012) pour évaluer la diversité perçue, ont aussi constaté cette corrélation positive entre la satisfaction

---

<sup>4</sup>  $d(i, j) = 1 - Sim(i, j)$ .

des utilisateurs à l'égard des recommandations de films et leur diversité perçue. Enfin, dans une étude portant sur 97 participants, où chacun devait évaluer trois listes de recommandations de films avec différents niveaux de diversité (faible, moyen et élevé), Willemsen et al. (2016) ont mis en évidence que plus la diversification des listes de recommandations est élevée, plus la diversité perçue par les utilisateurs est élevée.

### **3. Influence sur la consommation**

#### **3.1. Comportement d'achat des consommateurs**

Les SdR, largement utilisés dans le domaine des achats en ligne, exercent une influence significative sur les décisions d'achat des consommateurs liées à deux facteurs critiques : le niveau de connaissance du produit par le consommateur et la pression du temps pendant l'expérience d'achat. Pour les personnes ayant une connaissance limitée ou inexistante des produits, les recommandations personnalisées et les informations d'évaluation agissent comme une heuristique précieuse. Dans ce scénario, ces aides à l'achat orientées servent à simplifier le processus de prise de décision, en atténuant la difficulté perçue associée à l'évaluation des différentes options. Les consommateurs peu informés sont plus susceptibles d'adopter ces aides, ce qui augmente la probabilité d'achat et la satisfaction, même si, en contrepartie, ils sont moins confiants dans leur décision finale. De même, les consommateurs pressés par le temps trouvent également une utilité significative à ces aides à l'achat. Confrontés à des contraintes de temps, ils utilisent les recommandations comme des raccourcis pratiques pour rationaliser le processus de décision. Par conséquent, ce groupe est plus enclin à effectuer un achat et exprime une plus grande satisfaction. En outre, les coûts d'évaluation sont moins élevés en raison de l'efficacité de ces aides (Chang et Kukar-Kinney, 2011).

En ce qui concerne les environnements décisionnels caractérisés par la complexité, les individus n'ont souvent pas la capacité d'évaluer de manière approfondie toutes les alternatives disponibles en raison de limitations cognitives. Payne (1982) propose un processus de prise de décision en deux étapes pour remédier à ce problème, à savoir une sélection initiale et une comparaison approfondie. Les SdR jouent alors un rôle crucial dans ces deux étapes en facilitant la sélection des alternatives et l'évaluation détaillée de l'ensemble de considérations, ce qui permet de réduire le temps de décision et d'améliorer la qualité de la décision. Par conséquent, le nombre de recommandations présentées influe sur l'effort et la qualité de la décision. Il a également constaté qu'une liste triée de recommandations personnalisées augmentait la tendance des consommateurs à s'engager dans une comparaison heuristique de l'utilité locale, ce qui se traduisait par une meilleure qualité de décision (Xiao et Benbasat, 2007). Aksoy et al.

(2001) ont observé une meilleure qualité de décision lorsque les recommandations étaient ordonnées en fonction des préférences des consommateurs.

Enfin, la familiarité avec les recommandations, la quantité d'informations et les explications sur la génération de recommandations ont une incidence sur l'évaluation des SdR par les utilisateurs et leur confiance. En effet, le fait d'intégrer des recommandations peu familières parmi celles qui le sont déjà améliore leur attrait (Xiao et Benbasat, 2007).

### 3.2. Volume des ventes

Pour rappel, les objectifs principaux d'un SdR sont d'améliorer l'expérience utilisateur, mais aussi d'optimiser les objectifs commerciaux de l'entreprise tels que les ventes ou l'engagement des utilisateurs (Kim et al., 2021). D'ailleurs, comme nous le précise Helberger et al. (2018), il est fondamental de garder à l'esprit que des décisions stratégiques et commerciales interviennent dans les algorithmes de recommandations mis en place par les entreprises.

Les SdR jouent un rôle crucial dans l'amélioration des ventes du e-commerce de trois manières distinctes :

- **Conversion des visiteurs en acheteurs.** Sur un site web, de nombreux visiteurs naviguent sans effectuer d'achat. Les SdR aident les clients à découvrir les produits ce qui transforme régulièrement les visiteurs en acheteurs.
- **Vente croisée.** Les SdR encouragent la vente croisée en suggérant des produits supplémentaires aux clients. Des recommandations efficaces peuvent entraîner une augmentation du montant moyen des commandes. Par exemple, lors du processus d'achat, un site peut recommander des produits complémentaires en fonction de ceux déjà présents dans le panier.
- **Fidélisation.** Les SdR favorisent la fidélisation en créant une relation de valeur entre le site et le client. Les clients reviennent sur les sites qui répondent à leurs besoins, et plus ils interagissent avec le SdR, plus ils deviennent fidèles. Même si des concurrents offrent les mêmes fonctionnalités, les clients préfèrent rester fidèles au site d'origine, ce qui renforce leur engagement et les ventes (Ben Schafer et al., 1999).

À ce sujet, la littérature est claire concernant les effets positifs des SdR sur les ventes. Par exemple, des rapports industriels ont signalé une nette augmentation des revenus (et donc des ventes) ou de l'engagement des utilisateurs grâce aux SdR, comme illustré par les cas d'Amazon et de Netflix. 80 % de ce qui est regardé par les utilisateurs de Netflix provient des recommandations qui leur ont été faites (Gomez-Uribe, 2016) et 35 % des achats des clients

d'Amazon proviennent des recommandations de produits (Brent et al., 2017). De plus, les résultats de l'expérience menée par Hostler et al. (2012) indiquent qu'une approche de recommandation basée sur le CF s'avère efficace pour augmenter l'attrait perçu des produits recommandés par les consommateurs en ligne. Leur étude révèle également que l'augmentation de la satisfaction des clients et de leur fidélité entraîne des résultats inestimables à long terme. Ainsi, une mise en œuvre réussie d'un SdR peut établir une boucle de rétroaction positive, les clients revenant au fil du temps, devenant plus fidèles et partageant activement leurs préférences en matière de produits avec le marchand. Ce cycle, à son tour, facilite une promotion encore plus efficace des produits et une plus grande satisfaction des clients. Une autre étude menée par Dias et al. (2008) sur des supermarchés a montré que les SdR exercent une influence positive à la fois sur les revenus directs<sup>5</sup> et les revenus indirects<sup>6</sup>, ces derniers étant générés par les comportements d'achat croisé des consommateurs. Quant à l'étude de cas réalisée par Jannach et Hegelich (2009) dans le domaine des applications mobiles, elle a aussi révélé que l'intégration de SdR augmentait les taux de clics et les ventes globales. En outre, De et al. (2010) ont démontré que l'intégration d'un SdR implique une influence positive sur les ventes de produits, qu'ils soient promus ou non. En comparaison avec les moteurs de recherche classiques, ceux-ci n'augmentent les ventes que pour les produits promus, et ont un effet négatif sur les ventes des produits non promus. De manière similaire, Hinz et Eckert (2010) ont mis en évidence que l'utilisation de SdR peut contribuer à accroître les bénéfices des détaillants. En fait, cette augmentation du volume des ventes peut s'expliquer par le fait que les consommateurs tendent à acheter davantage après avoir reçu des recommandations personnalisées, comme l'ont aussi souligné Hosanagar et al. (2014).

C'est ici, en parallèle aux sections 2 et 3 du chapitre 1, que notre question de recherche prend tout son sens. En effet, qu'en est-il de l'influence de recommandations diversifiées sur l'intention de consommation des utilisateurs ? Au travers de la littérature, il a été reconnu que la diversification des listes de recommandations permet notamment d'accroître la satisfaction des utilisateurs (McNee et al., 2002 ; Ziegler et al., 2005 ; Castagnos et al., 2013) et que les utilisateurs étaient attirés par des listes d'articles diversifiées (Willemsen et al., 2016 ; Ferwerda et al., 2017). Ces constatations devraient se traduire de manière indirecte par des avantages pour une entreprise, tels qu'un renforcement de la fidélisation ou de l'engagement de sa clientèle

---

<sup>5</sup> Les revenus directs sont ceux directement liés à l'achat d'un *item* recommandé par les SdR pendant la session d'achat en cours (Dias et al., 2008).

<sup>6</sup> Un achat particulier ne contribue aux revenus indirects que si l'*item* acheté n'a pas été recommandé à l'acheteur pendant la session d'achat en cours (Dias et al., 2008).

(Castells et al., 2011). En conséquence, il est envisageable de croire que l'inclusion de la diversité dans les listes de recommandations se traduise par une augmentation du volume de consommation des utilisateurs. Cette conclusion serait particulièrement intéressante pour les entreprises développant ou utilisant des SdR, offrant ainsi des perspectives prometteuses pour améliorer l'efficacité et la rentabilité de leurs services de recommandation.

## 4. Conclusion

Au sein de ce premier chapitre, nous avons réalisé une revue de la littérature en examinant les études et les opinions des chercheurs dans le domaine des SdR.

Tout d'abord, nous avons décrit le fonctionnement ainsi que les différents types de SdR classés en fonction de leur approche. Nous avons également pu décrire les avantages et inconvénients de certaines de ces approches.

Ensuite, nous avons présenté les principales métriques permettant d'évaluer la performance des SdR. À ce sujet, nous avons appris que la pertinence (*accuracy*) n'est plus le seul critère important dans l'évaluation des SdR. La diversité des recommandations est désormais essentielle parce qu'elle permet d'éviter que les utilisateurs s'enferment dans leurs préférences dominantes, d'améliorer l'expérience utilisateur et d'accroître la satisfaction.

De plus, nous avons découvert qu'un équilibre entre précision et diversité était, tout de même, nécessaire, car une diversité excessive peut grandement diminuer la pertinence et créer l'incompréhension ou la méfiance chez les utilisateurs. Les chercheurs ont, d'ailleurs, tenté de développer des algorithmes pour optimiser cet équilibre tout en prenant en compte soit la diversité individuelle (celle au sein d'une liste de recommandations spécifique à un utilisateur donné), soit la diversité agrégée (celle pour l'ensemble des utilisateurs du système).

Les SdR sont aussi cruciaux pour les achats en ligne, surtout pour les consommateurs peu informés ou pressés par le temps, en simplifiant le processus décisionnel et en réduisant les coûts d'évaluation. Dans des environnements décisionnels complexes, les SdR facilitent la sélection et la comparaison des alternatives, améliorant ainsi la qualité et la rapidité des décisions.

Enfin, nous avons mis en lumière par le biais d'études ultérieures que les SdR visent également à augmenter le volume des ventes. En effet, il a été démontré que les SdR ont un effet positif sur les ventes de produits promus ou non, qu'ils favorisent les revenus directs et indirects et que les consommateurs tendent à acheter davantage après avoir reçu des recommandations personnalisées. En réalité, les SdR favorisent une boucle de rétroaction positive, encourageant le retour des clients, leur fidélisation et la promotion active des produits.

## CHAPITRE 2 : DESIGN EXPÉRIMENTAL

Après avoir posé le cadre théorique et les différents concepts relevés au sein de la littérature, nous définirons, tout d'abord, les objectifs de notre recherche. Ensuite, nous détaillerons la base de données que nous emploierons en mettant en lumière ses caractéristiques essentielles et sa pertinence pour notre recherche. Nous présenterons également les différents algorithmes de recommandation choisis pour notre étude, et justifierons leur sélection en fonction de leur *accuracy*, c'est-à-dire leur capacité à identifier avec précision les *items* considérés comme les plus « pertinents » pour l'utilisateur. De plus, nous explorerons le modèle de reclassement nommé *Maximal Marginal Relevance* (MMR). Cette méthode de diversification vise à améliorer la diversité en sélectionnant des *items* qui maximisent la pertinence marginale (*marginal relevance*) tout en minimisant la similarité avec les *items* déjà sélectionnés. Après avoir exposé clairement les hypothèses à tester, nous décrirons minutieusement la procédure expérimentale qui sera suivie via une application en ligne développée à cet effet. Enfin, nous terminerons par une analyse détaillée des participants recrutés.

### 1. Objectifs de la recherche

Pour rappel, ce mémoire tente de répondre à la question de recherche suivante : *Comment l'inclusion de la diversité dans les systèmes de recommandation, évalués online, influence-t-elle le volume de consommation intentionnelle des utilisateurs ?* Autrement dit, l'objectif est d'enrichir la littérature existante en vérifiant si cet enjeu de diversité est bien compatible avec un SdR et s'il encourage les consommateurs à consommer davantage. De plus, ce mémoire vise à mesurer et analyser d'autres variables dépendantes :

- **la satisfaction des utilisateurs face aux listes de recommandations.** Est-ce que la diversification des recommandations favorise la satisfaction des utilisateurs ?
- **la diversification de la consommation intentionnelle.** Est-ce que l'inclusion de la diversité dans les listes de recommandations incite les utilisateurs à consommer de manière plus diversifiée ?
- **la diversité perçue dans les recommandations.** Est-ce que les utilisateurs perçoivent l'inclusion de la diversité dans les listes de recommandations ?
- **la perception par les utilisateurs de la pertinence des recommandations face à leurs préférences.** Comment les utilisateurs perçoivent la pertinence des recommandations face à leurs préférences lorsque la diversité est favorisée au détriment de la pertinence ?

## 2. Base de données MovieLens

L'Annexe 1 évoque les jeux de données les plus populaires dans le domaine des SdR, à savoir MovieLens, Amazon ou encore Yelp. Cette popularité a été calculée en fonction du nombre d'apparitions de chaque base de données dans 88 articles scientifiques publiés lors de la conférence ACM<sup>7</sup> sur les SdR en 2017 et 2018 (Beel et Brunel, 2019). En analysant de plus près l'Annexe 1, nous pouvons remarquer que la base de données MovieLens<sup>8</sup>, portant sur la cinématographie, est fréquemment utilisée par les chercheurs. C'est pourquoi, dans le cadre de ce mémoire, nous utiliserons le jeu de données MovieLens le plus récent<sup>9</sup> qui donne accès à 100 836 *ratings* explicites et 9 742 films. Notre choix s'est également porté sur le domaine cinématographique car il présente de multiples avantages. Tout d'abord, il offre la possibilité de rassembler un grand volume d'informations sur les films, créant ainsi un contexte réaliste pour l'expérience en ligne que nous déploierons. En outre, les films sont fréquemment notés par les utilisateurs, ce qui en constitue un choix idéal pour comparer divers algorithmes. Enfin, le cinéma est un domaine populaire et familier, ce qui augmente la probabilité que les utilisateurs connaissent les films recommandés et puissent exprimer leurs préférences.

Néanmoins, nous avons adapté cette base de données pour différentes raisons : faciliter le déploiement de l'expérience en ligne, conserver un niveau de popularité et de récence des films suffisamment corrects. Cette démarche vise notamment à éviter que les participants, lors de cette expérience en ligne, se voient recommander des films trop peu connus - de niche - qu'ils ne choisiront pas *in fine*. Une telle situation biaiserait les résultats finaux de l'expérience. Par conséquent, nous avons conservé les 125 films les plus notés pour chacune des quatre périodes de temps suivantes : entre 1970 et 1985, entre 1985 et 2000, entre 2000 et 2010 et après 2010. Le *Tableau 1* est un aperçu numérique et statistique de la base de données finale que nous emploierons dans le cadre de ce mémoire.

Nombre de films <sup>10</sup>	500 sortis entre 1970 et 2018	Nombre moyen de ratings par film	76.01
-------------------------------	----------------------------------	-------------------------------------	-------

<sup>7</sup> Association for Computing Machinery qui a pour but de « faire progresser l'informatique en tant que science et profession » (<https://www.acm.org>).

<sup>8</sup> Cette base de données a été créée par le GroupLens (<https://grouplens.org>), un groupe de recherche en informatique de l'Université du Minnesota. Leur objectif était de recueillir des évaluations de films auprès d'utilisateurs pour la recherche dans le domaine des SdR. Le site web de GroupLens rend public une collection de données contenant des évaluations d'utilisateurs pour divers films (*ratings* de 0.5 à 5 étoiles, tags), ainsi que des informations sur les films eux-mêmes (titre, année de sortie, genre) au travers de différents jeux de données MovieLens. Les jeux proposés varient selon leur taille et leur date de mise à jour (GroupLens, 2024).

<sup>9</sup> <https://grouplens.org/datasets/movielens/latest/>.

<sup>10</sup> Appartenant à différents genres bien connus tels que comédie, animation, romance, drame, crime, action, horreur, thriller, aventure, science-fiction, mystère, fantastique, etc.

Nombre d'utilisateurs uniques	608	Nombre moyen de ratings par utilisateur	62.5
Nombre total de ratings	38 003	Nombre médian de ratings par utilisateur	38

Tableau 1. Aperçu numérique de la base de données finale

Par ailleurs, nous avons également recueilli, à partir du site IMDb, une vaste gamme d'informations concernant chaque film de notre base de données, à savoir la bande-annonce en anglais, l'affiche, les acteurs, le(s) réalisateur(s), la langue d'origine et la durée. Lors de l'expérience en ligne, toutes ces données permettront aux participants de se remémorer ou de découvrir le contenu des différents films face auxquels ils seront confrontés. En outre, certaines de ces données ont permis de calculer la similarité cosinus entre tous les films. Pour ce faire, nous avons utilisé la méthode d'encodage *One-Hot*<sup>11</sup> pour les genres et les années de sortie, ainsi que la méthode TF-IDF<sup>12</sup> (*Term Frequency-Inverse Document Frequency*) pour vectoriser le contenu textuel des résumés des films.

### 3. Algorithmes de recommandation

Sur base de notre jeu de données, nous avons décidé d'étudier le filtrage collaboratif (CF) en développant deux approches basées sur la mémoire (*user-based* et *item-based*) et une approche basée sur le modèle (SVD). Pour implémenter ces trois types d'algorithmes<sup>13</sup> de recommandation, nous avons utilisé le langage de programmation Python et, plus précisément, le *package Surprise* (Hug, 2015). Il s'agit d'un *scikit*<sup>14</sup> Python pour la construction et l'analyse de SdR qui traitent des données explicites.

Dans le contexte de ces algorithmes de recommandation basés sur l'utilisateur et l'*item*, « *min\_support* », « *k* » et « *min\_k* » sont des paramètres indispensables et prédéfinis qui contrôlent le fonctionnement de l'algorithme. Dans le cadre du calcul de similarité, le terme « *min\_support* » indique le nombre minimum d'*items* évalués en commun par deux utilisateurs (pour l'*user-based*) pour qu'ils soient considérés comme similaires. Dans le cadre de l'*item-based*, ce terme représente le nombre minimum d'utilisateurs qui ont évalué en commun les

<sup>11</sup> L'encodage *One-Hot* permet la transformation de chaque film en un vecteur binaire pour lequel chaque position correspond à une caractéristique spécifique du film (genre ou année de sortie dans notre cas).

<sup>12</sup> La méthode TF-IDF permet d'évaluer l'importance relative d'un terme dans un résumé par rapport à la collection de tous les résumés. Cette approche permet de mettre en évidence les mots-clés spécifiques à un résumé et de les utiliser pour en représenter le contenu de manière plus précise (Aggarwal, 2016).

<sup>13</sup> Les codes Python qui ont été implémentés pour ces trois types d'algorithmes de recommandation sont détaillés en *Annexe 2*.

<sup>14</sup> Les *SciKits* (abréviation de *SciPy Toolkits*) sont des paquets complémentaires pour la bibliothèque *SciPy*, hébergés et développés séparément et indépendamment de la distribution principale de *SciPy*.

deux *items* comparés. Si le « *support* » réel est inférieur à « *min\_support* », la similarité est évaluée comme nulle et aucun voisin ne sera considéré. Ensuite, le terme « *k* » représente le nombre de voisins les plus similaires à prendre en compte lors de la prédiction d'un *rating*. Enfin, le terme « *min\_k* » est le nombre minimum de voisins requis pour effectuer une prédiction. Si le nombre réel de voisins est inférieur à « *min\_k* », l'algorithme ne peut pas fournir de prédiction et renvoie la note moyenne (Hug, 2015).

Quant au modèle SVD, il dépend, lui aussi, de divers hyperparamètres. C'est pourquoi, nous en avons ciblé trois importants : « *n\_factors* », « *n\_epochs* » et « *reg\_all* ». Le premier détermine le nombre de facteurs latents utilisés dans le modèle. Le second spécifie le nombre d'itérations d'entraînement pour ajuster les paramètres du modèle, influençant ainsi sa précision tout en optimisant la prédiction. Le troisième contrôle la régularisation globale du modèle afin d'éviter l'ajustement excessif en réduisant les valeurs des paramètres, ce qui contribue à une meilleure généralisation des prédictions (Hug, 2015).

En ce qui concerne les valeurs choisies pour ces différents paramètres, nous avons effectué une analyse expérimentale afin de déterminer quelles valeurs offraient le meilleur modèle *user-based*, *item-based* et SVD en termes d'*accuracy*. Pour y parvenir, la technique de validation croisée *Hold Out* a été utilisée, divisant la matrice des *ratings* de notre base de données en un *trainset* (75 %) et un *testset* (25 %). Pour les modèles *user-based* et *item-based*, nous avons donc choisi d'étudier un « *k* » = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45 et 50 combiné à un « *min\_k* » toujours égal à « *k-1* » et à un « *min\_support* » toujours égal à « *k* ». Pour le modèle SVD, nous avons décidé d'évaluer cet algorithme lorsque « *n\_epochs* » appartient à [10, 20, 30, 40, 50] pour un « *n\_factors* » variant entre 10 et 100 (par saut de 10). Concernant « *reg\_all* », nous l'avons toujours fixé à 0.1. Cette valeur a été jugée appropriée pour prévenir le surajustement. En *Annexes 3, 4 et 5*, vous découvrirez respectivement les valeurs obtenues dans le cadre de cette évaluation *offline* de l'*accuracy* (MAE, RMSE) des modèles *user-based*, *item-based* et SVD. Sur base de ces annexes, nous pouvons remarquer que les modèles *user-based* et *item-based* avec un « *k* » de 10 et que l'algorithme SVD fixant un « *n\_factors* » à 50 et un « *n\_epochs* » à 50 sont les plus précis en matière de MAE et RMSE.

	MAE	RMSE
user-based( <i>k</i> = 10, <i>min_k</i> = 9, <i>min_support</i> = 10)	0.6328	0.8333
item-based( <i>k</i> = 10, <i>min_k</i> = 9, <i>min_support</i> = 10)	0.6266	0.8256
SVD( <i>n_factors</i> = 50, <i>n_epochs</i> = 50, <i>reg_all</i> = 0.1)	0.6107	0.8033

Tableau 2. Valeurs d'*accuracy* (MAE, RMSE) des meilleurs modèles *user-based*, *item-based* et SVD

Le *Tableau 2* présente les valeurs de MAE et RMSE obtenues pour chacun des meilleurs modèles *user-based*, *item-based* et SVD. Nous avons trouvé judicieux de nous intéresser uniquement aux algorithmes de recommandation les plus performants en matière d'*accuracy*. En effet, dans le cadre de notre expérience, nous utiliserons un algorithme de diversification pour inclure un certain niveau de diversité dans les recommandations finales proposées à l'utilisateur. Comme expliqué dans la section suivante, cet algorithme de diversification utilisera les prédictions de *ratings* initialement calculées par un algorithme de recommandation, d'où l'importance de leur précision.

#### 4. Diversification des listes de recommandations

Les méthodes permettant de diversifier les recommandations peuvent être divisées en trois catégories : le prétraitement (*pre-processing*), le traitement en cours (*in-processing*) et le post-traitement (*post-processing*). Les méthodes de prétraitement sont établies avant le processus d'apprentissage du modèle. Les méthodes en cours de traitement sont, quant à elles, directement adoptées lors de l'apprentissage du modèle. Les méthodes de post-traitement sont, pour leur part, utilisées dès que le modèle est bien entraîné afin de reclasser les listes finales de recommandations d'*items* (Wu et al., 2022).

Afin de diversifier les recommandations, nous avons décidé, dans le cadre ce mémoire, de nous pencher sur le modèle de reclassement (*re-ranking*) nommé *Maximal Marginal Relevance* (MMR). Cette méthode de diversification vise à améliorer la diversité en sélectionnant des *items* qui maximisent la pertinence marginale (*marginal relevance*) tout en minimisant la similarité avec les *items* déjà sélectionnés. En réalité, un *item* a une pertinence marginale élevée s'il est à la fois pertinent pour l'utilisateur et peu similaire aux *items* précédemment sélectionnés. Cette méthode combine donc linéairement la pertinence et la diversité pour améliorer les recommandations finales en réponse au fait que les besoins des utilisateurs incluent non seulement la pertinence mais aussi la diversité (Carbonell et Goldstein, 1998). Nous avons choisi cette méthode *post-processing* parce qu'elle est parfaitement adaptée à la diversification des SdR et parce qu'elle se base notamment sur un paramètre ajustable contrôlant le niveau de diversité. Ce paramètre<sup>15</sup> autorise donc des variations allant de résultats totalement pertinents à des résultats totalement diversifiés. Pour formaliser notre approche, nous avons résumé ci-dessous notre modèle de reclassement dans l'*Algorithme 1*.

---

<sup>15</sup> Ce paramètre nous sera très utile dans la mesure où nous déploierons notre expérience sur plusieurs groupes de participants se distinguant par le niveau de diversité qui sera inclus dans leur liste de recommandations.

### Algorithmme 1 MMR re-ranking model<sup>16</sup>

#### Input :

- Un utilisateur cible,  $u$
- Une liste d'*items* candidats<sup>17</sup> pour l'utilisateur cible,  $M$
- Un ensemble des prédictions<sup>18</sup> de *ratings* pour chaque *item* présent dans  $M$ ,  $P$
- Une matrice de similarité<sup>19</sup> entre les *items* candidats,  $S$
- Un pourcentage de diversification dans les recommandations,  $d \in [0, 1]$
- Le nombre d'*items* à recommander,  $k$

#### Output :

- Une liste reclassée des *items*,  $Y$

#### Process :

1. **Initialize** :  $Y = \emptyset$
2. **While**  $|Y| < k$  **do**
3.  $y = \text{MMR}(u, M, P, S, d)$
4.  $Y \cup y$
5.  $M = M - y$
6. **End while**
7. **Return**  $Y$

La fonction  $\text{MMR}(u, M, P, S, d)$ , utilisée dans l'Algorithmme 1, fonctionne en deux temps. Premièrement, pour chaque *item* (*iid*) de la liste  $M$ , un score MMR est calculé comme suit :

$$\text{score\_mmr} = (1 - d) \times P[\text{iid}] - d \times \max([S[\text{iid}, y] \text{ for } y \text{ in } Y - \{\text{iid}\}], \text{default} = 0)$$

Deuxièmement, lorsqu'un score MMR a été calculé pour chaque *item* (restant) de la liste  $M$ , seul l'*item* affichant le score MMR le plus élevé est ajouté à la liste de recommandations  $Y$ . Ce processus est répété jusqu'à ce que la liste  $Y$  atteigne la taille  $k$ .

En ce qui concerne le pourcentage de diversification (nommé  $d$ ), son objectif est de contrôler l'importance relative de la diversité par rapport à la pertinence dans le processus de recommandation. Lorsque  $d = 0$ , la diversité n'exerce aucune influence sur le score MMR. Les recommandations reposent entièrement sur la pertinence des *items*, ce qui signifie que les *items* les plus pertinents (sur base des prédictions) sont recommandés en premier. Quand  $d = 1$ , la pertinence n'exerce aucune influence sur le score MMR. Les recommandations sont entièrement basées sur la diversité, ce qui signifie que les *items* les moins similaires à ceux déjà recommandés seront privilégiés, indépendamment de leur pertinence. Si  $0 < d < 1$ , l'objectif est d'établir un compromis entre la pertinence et la diversité des recommandations. En effet, en ajustant la valeur de  $d$ , le degré de diversification des recommandations est contrôlé. Une valeur

<sup>16</sup> Le code Python qui a été implémenté pour cet algorithme de diversification MMR a été détaillé en Annexe 6.

<sup>17</sup> Les *items* candidats font référence aux *items* potentiels qui pourraient être recommandés à un utilisateur donné.

<sup>18</sup> Dans notre cas, ces prédictions ont été calculées grâce à notre algorithme *item-based*.

<sup>19</sup> Dans notre cas, nous avons calculé la similarité cosinus entre tous les *items* à partir des genres, des années de sortie et du contenu textuel des résumés des films.

plus élevée favorisera la diversité au détriment de la pertinence, tandis qu'une valeur plus faible favorisera la pertinence au détriment de la diversité.

L'étude qu'ont menée Castagnos et al. (2010) a permis de démontrer que le niveau de diversité à inclure devait se situer entre 40% et 60% pour préserver une similarité raisonnable entre les recommandations et renforcer considérablement la confiance des utilisateurs. C'est pourquoi, dans le cadre de notre expérience en ligne, nous testerons quatre pourcentages de diversité fixés à 0 (minimum) pour le groupe de contrôle et 40% (moyen faible), 60% (moyen élevé) et 100% (maximum) pour les groupes de traitement.

Le *Tableau 3* examine les variations de diversité dans les listes de dix<sup>20</sup> recommandations (mesurées par l'ILD moyen) faites aux utilisateurs, ayant noté entre trente-cinq et quarante-cinq<sup>21</sup> films chacun, en fonction des différents pourcentages de diversification (d). L'objectif est de comprendre comment la diversité moyenne des listes de recommandations évolue en fonction de ce paramètre d.

Pourcentage de diversification	Distance Intra-Liste <sup>22</sup> (ILD)		
	Item-based	User-based	SVD
<b>d = 0</b>	0.77	0.81	0.77
<b>d = 40%</b>	0.84	0.86	0.86
<b>d = 60%</b>	0.87	0.89	0.89
<b>d = 100%</b>	0.89	0.89	0.89

*Tableau 3. Variation de la diversité des recommandations en fonction du pourcentage de diversification et de l'algorithme de recommandation de départ*

Il est logiquement observable que lorsque le pourcentage de diversification augmente, l'ILD croît également. Par conséquent, cela indique une augmentation de la diversité dans les listes de recommandations fournies par l'algorithme MMR pour chaque utilisateur. Le *Tableau 3* indique aussi que les méthodes *item-based* et SVD produisent initialement (d = 0%) les listes

<sup>20</sup> Ce choix de dix recommandations est lié au nombre de films que les participants à notre expérience en ligne se verront recommander lors de la phase de recommandation (étape 3 de la procédure expérimentale, décrite dans la section 6 du chapitre 2).

<sup>21</sup> Cet intervalle de *ratings* (35-45) a été choisi parce que les participants à notre expérience en ligne devront noter quarante films au cours de la phase de *rating* (étape 1 de la procédure expérimentale, décrite dans la section 6 du chapitre 2). En effet, en choisissant cet intervalle de *ratings* (35-45), nous nous assurons que les listes de recommandations générées pour l'évaluation de l'ILD à partir de notre base de données sont comparables aux listes que les participants à notre expérience en ligne pourraient potentiellement recevoir.

<sup>22</sup> Pour rappel, la formule de l'Intra-List Distance (ILD) pour un seul utilisateur est définie à la sous-section 2.3. du chapitre 1. Dans cette formule, nous avons calculé la distance (= 1 - similarité cosinus) entre chaque paire de films. Pour rappel, dans ce mémoire, la similarité cosinus entre tous les films a été établie à partir de leurs genres, leur année de sortie et le contenu textuel de leur résumé. Le code Python qui a été implémenté pour calculer l'ILD est détaillé en *Annexe 7*.

de recommandations les moins diversifiées. De plus, pour les algorithmes *user-based* et SVD, les variations de l'ILD demeurent presque inchangées lorsque le pourcentage de diversification passe de 60% à 100%. Pour toutes ces raisons, l'algorithme *item-based* est celui qui permet de percevoir le mieux, sur base de l'ILD, les augmentations de diversité dans les listes de recommandations, notamment lorsque le pourcentage de diversification devient très élevé. Puisque ce concept de diversité perçue sera très important et testé dans le cadre de notre expérience en ligne, nous avons décidé, au travers de ce mémoire, de nous concentrer sur l'algorithme de recommandation *item-based* ( $k = 10, \min\_k = 9, \min\_support = 10$ ).

## 5. Hypothèses à tester

Ce mémoire adopte une approche hypothético-déductive. C'est pourquoi, un ensemble d'hypothèses, décrites ci-dessous, a été formulé à partir de la revue de littérature scientifique. Ces hypothèses sont des affirmations provisoires qui seront testées au travers d'une expérience en ligne et d'une analyse des données obtenues. En effet, les résultats qui seront analysés dans le troisième chapitre permettront de confirmer ou d'infirmer les hypothèses suivantes (Uzunidis, 2007) :

- **Hypothèse 1 (H1a).** *La diversification à 40% et 60% des listes de recommandations permet une satisfaction moyenne des utilisateurs plus élevée que lorsque le pourcentage de diversité est nul.*

**Hypothèse 1b (H1b).** *La diversification à 100% des listes de recommandations permet une satisfaction moyenne des utilisateurs plus faible que lorsque le pourcentage de diversité est nul.*

Les résultats du post-questionnaire (étape 4 de la procédure expérimentale<sup>23</sup>) au sujet de la satisfaction des utilisateurs face aux listes de recommandations devraient mettre en lumière l'effet positif de la diversité sur la satisfaction des utilisateurs (Castagnos et al., 2013), en particulier pour les groupes<sup>24</sup> de participants affectés à un pourcentage de diversité de 40% et 60%. En effet, comme démontré par McNee et al. (2002), la satisfaction des utilisateurs ne coïncide pas nécessairement avec une grande précision dans les SdR. Cependant, puisque nous estimons qu'une inclusion excessive de diversité pourrait engendrer de la méfiance et de l'incompréhension chez l'utilisateur, un pourcentage de diversité de 100% devrait donc réduire la satisfaction des utilisateurs affectés à ce groupe (Castagnos et al., 2013).

---

<sup>23</sup> La procédure expérimentale est décrite à la section 6 du chapitre 2.

<sup>24</sup> Les participants de l'expérience en ligne sont répartis de manière entièrement aléatoire et transparente en quatre groupes qui se distinguent par le pourcentage de diversité inclus dans les recommandations (0 pour le groupe de contrôle et 40%, 60% ou 100% pour les groupes de traitement).

- **Hypothèse 2a (H2a).** *La diversification à 40% et 60% des listes de recommandations permet un volume moyen de consommation intentionnelle des utilisateurs plus élevé que lorsque le pourcentage de diversité est nul.*

**Hypothèse 2b (H2b).** *La diversification à 100% des listes de recommandations permet un volume moyen de consommation intentionnelle des utilisateurs plus faible que lorsque le pourcentage de diversité est nul.*

Le nombre de films que les utilisateurs ajouteront à leur *watchlist*<sup>25</sup> devrait croître au fur à mesure que le pourcentage de diversité augmente jusque 60%. En effet, puisque la diversification à 40% et 60% des listes de recommandations devrait accroître la satisfaction des utilisateurs, cette dernière devrait se traduire de manière indirecte par des avantages pour une entreprise, tels qu'un renforcement de la fidélisation ou l'engagement de sa clientèle (Castells et al., 2011). En conséquence, nous anticipons une augmentation du volume moyen de consommation intentionnelle des utilisateurs à la suite de l'inclusion de la diversité. Néanmoins, cette conclusion ne devrait pas être observée lorsque le pourcentage de diversité vaut 100% pour les mêmes raisons que celles évoquées pour H1.

**Hypothèse 2c (H2c).** *La diversification à 40% et 60% des listes de recommandations permet une consommation intentionnelle de chaque utilisateur plus diversifiée, en moyenne, que lorsque le pourcentage de diversité est nul.*

**Hypothèse 2d (H2d).** *La diversification à 100% des listes de recommandations permet une consommation intentionnelle de chaque utilisateur moins diversifiée, en moyenne, que lorsque le pourcentage de diversité est nul.*

La diversification à 40% et 60% des listes de recommandations expose chaque utilisateur à une gamme modérément plus large de contenus, ce qui devrait donc les inciter à explorer et consommer une variété plus grande de films. En revanche, une diversification excessive (100%), suscitant de la méfiance auprès chaque utilisateur, ne devrait pas entraîner une consommation intentionnelle plus variée comparativement à un scénario sans diversité incluse.

- **Hypothèse 3 (H3).** *L'inclusion de la diversité dans les listes de recommandations permet une perception moyenne plus élevée de la diversité par les utilisateurs que lorsque le pourcentage de diversité est nul.*

---

<sup>25</sup> La *watchlist* est une liste personnelle dans laquelle les utilisateurs pourront ajouter les films qu'ils prévoient de regarder à l'avenir parmi ceux recommandés. Elle est très similaire à un panier d'achats en ligne.

Les résultats du post-questionnaire (étape 4 de la procédure expérimentale) concernant la perception de la diversité dans les listes de recommandations devraient confirmer cette tendance, en particulier pour les groupes de participants faisant face à des recommandations diversifiées (avec un pourcentage de diversité de 40%, 60% ou 100%). À ce sujet, l'étude menée par Willemsen et al. (2016) a montré que plus la diversification des listes de recommandations est élevée, plus la diversité perçue par les utilisateurs est élevée.

- **Hypothèse 4 (H4).** *L'inclusion de la diversité dans les listes de recommandations permet aux utilisateurs de percevoir la pertinence des recommandations face à leurs préférences plus faiblement que lorsque le pourcentage de diversité est nul.*

Les résultats du post-questionnaire (étape 4 de la procédure expérimentale) au sujet de la pertinence des recommandations devraient donc refléter cette tendance, en particulier pour les groupes de participants affectés à un pourcentage de diversité élevé, c'est-à-dire de 60% ou 100%. Puisque notre algorithme de diversification des recommandations (MMR) se base sur un pourcentage de diversité ( $d$ ), des variations allant de résultats totalement pertinents à des résultats totalement diversifiés sont autorisées. Une valeur de  $d$  plus élevée favorisera la diversité au détriment de la pertinence, tandis qu'une valeur plus faible favorisera la pertinence au détriment de la diversité (Carbonell et Goldstein, 1998). En effet, en règle générale, l'augmentation de la diversité est associée à une réduction de la pertinence des recommandations (Zhou et al., 2008 ; Adomavicius et Kwon, 2012 ; Wang et al., 2020). C'est pourquoi, nous estimons avec les hypothèses H3 et H4 que les participants seront capables de percevoir une augmentation de la diversité au désavantage de la pertinence des recommandations face à leurs préférences.

## 6. Procédure

Afin de tenter de répondre à notre question de recherche et de tester nos hypothèses, nous avons réalisé une expérience participative permettant de plonger les participants dans un environnement réaliste. Pour ce faire, nous avons développé une application en ligne sophistiquée en utilisant le *framework* Python *Flask* qui permet de créer des applications Web avec des fonctionnalités avancées. Le choix de *Flask* s'est imposé en raison de sa flexibilité et de sa robustesse, facilitant ainsi le développement de notre application complexe qui a nécessité une planification minutieuse et une implémentation soignée. Puisque, dans notre cas, nous cherchons à évaluer l'influence de recommandations diversifiées (entièrement ou en partie<sup>26</sup>)

---

<sup>26</sup> Cela dépendra du niveau de diversification (40%, 60% ou 100%).

sur le volume de consommation intentionnelle, l'interface utilisateur de l'application est demeurée identique pour chaque participant. Il s'agit d'une variable de contrôle essentielle évitant d'introduire tout biais.

Pour commencer, une brève introduction a été réalisée sur la page d'accueil<sup>27</sup> afin de présenter le contexte de notre étude et permettre au participant de s'identifier (via un identifiant et un mot de passe généré aléatoirement). En effet, notre application en ligne utilise *Flask-Login* pour gérer l'authentification des utilisateurs de manière sécurisée et fiable. Au cours de cette introduction, nous n'avons pas divulgué les objectifs de l'expérience aux participants, pour éviter toute influence sur leurs comportements et réponses, comme une acceptation involontaire d'un plus grand nombre de recommandations (Chen et Liu, 2017). Ensuite, chaque participant a été invité à suivre une procédure en cinq étapes que nous détaillons ci-dessous. D'un point de vue pratique, notre expérience en ligne a été conçue pour une durée totale d'environ quinze minutes par participant.

- **Étape 1.** Au cours de cette première étape, nommée phase de *rating*, nous avons demandé aux participants de donner leur avis sur des films en les notant, sans limite de temps, via une échelle à étoiles allant de 1 (« Je déteste ») à 5 (« J'adore »). Pour des raisons de complexité et dans le but de préserver la motivation des participants, nous avons délibérément limité à quarante<sup>28</sup> le nombre de films à noter. Cette approche reste pertinente pour la pratique et offre une image suffisamment différenciée des préférences cinématographiques de chaque participant. En effet, ces *ratings* ont permis de compléter le profil de préférences du participant qui a servi de base à l'élaboration des recommandations dans l'étape 3. Il s'agit d'une stratégie également approuvée dans la littérature scientifique (Hostler et al., 2012) et utilisée par certains sites web, tels que Netflix (Netflix, 2023). Dans notre cas, les *ratings* fournis par l'utilisateur lors de cette étape ont directement été ajoutés à la base de données MovieLens qui, pour rappel, contenait déjà les *ratings* de 608 *users* différents. De cette manière, nous avons pu entraîner en ligne notre modèle de recommandation *item-based* avant de pouvoir générer des recommandations.

Quant à la sélection des quarante films, identiques pour chaque participant, nous avons tenté de couvrir une large variété de genres dans le but de tenir compte des différences de goûts

---

<sup>27</sup> Des captures d'écran de l'application en ligne et des différentes étapes de la procédure (introduction, questionnaire socio-démographique, présentation des films à noter, présentation d'exemples de recommandations, présentation de la *watchlist*, post-questionnaire et remerciements) se trouvent en *Annexe 8*.

<sup>28</sup> Nous avons également limité à quarante le nombre de films à noter car, dans notre base de données de 500 films, le nombre médian de *ratings* par utilisateur est égal à 38.

entre les participants. C'est pourquoi, nous avons choisi les dix genres les plus représentés dans notre base de données, à savoir Comédie, Thriller, Action, Romance, Aventure, Science-Fiction, Horreur, Drame, Fantastique et Animation. En outre, pour éviter les biais dus à la récence ou non des films, chacun des participants a reçu<sup>29</sup>, pour chaque genre, le film le plus noté de notre base de données sorti entre 1970 et 1985, entre 1985 et 2000, entre 2000 et 2010 et après 2010. Nous avons considéré les films les plus notés afin de s'assurer de la disponibilité d'un groupe d'utilisateurs de référence (*peer group*) lors de la phase de recommandation.

Concernant l'affichage<sup>30</sup> dans notre application des quarante films sélectionnés, le système affiche systématiquement et de manière structurée les informations suivantes : le titre, l'affiche, l'année de sortie, le(s) genre(s), les acteurs principaux, le lien de la bande-annonce et le résumé en anglais.

- **Étape 2.** Un premier questionnaire a été établi afin de recueillir des caractéristiques socio-démographiques, telles que le sexe, l'âge, la nationalité, le niveau d'éducation et la catégorie socio-professionnelle de chaque participant, ainsi que ses habitudes cinématographiques (Willemsen et al., 2016), comprenant le nombre de films moyen qu'il visionne en *streaming* par mois et ses genres cinématographiques favoris. Nous avons également cherché à savoir si le participant était familier avec le concept de SdR dans le domaine du e-commerce, s'il était conscient de la personnalisation des recommandations en fonction de ses préférences et dans quelle mesure il suivait les recommandations. Ces questions avaient pour unique objectif de rassembler des données statistiques sur les participants et, le cas échéant, de contrôler voire exclure ceux dont les réponses ne seraient pas pertinentes.
- **Étape 3.** Une fois le profil de chaque participant initialisé, les participants, peu importe leurs caractéristiques, ont été répartis de manière entièrement aléatoire et transparente en quatre groupes, de telle sorte qu'ils n'aient pas conscience de cette répartition, et que chaque groupe compte un nombre de participants assez similaire. Cette répartition aléatoire permettra, lors de l'analyse des résultats, une comparaison (statistique) équitable, pertinente et fiable entre les divers groupes. Notre expérience en ligne a donc été menée dans une conception *between-subjects*.

---

<sup>29</sup> Tout en évitant de proposer plus d'un film appartenant à une même série/saga.

<sup>30</sup> Des captures d'écran de l'application en ligne concernant cette phase de *ratings* se trouvent en *Annexe 8*.

Ensuite, le système est entré dans une phase de recommandation en ligne afin d'afficher une liste de dix recommandations de films à chacun des participants. Ces recommandations ont été calculées à partir de l'algorithme MMR et affichées<sup>31</sup> de la même manière pour tous les participants de l'expérience. Seul le niveau de diversité inclus pouvait varier. Pour rappel, chaque groupe se distingue par le pourcentage de diversité inclus dans les recommandations (0 pour le groupe de contrôle et 40%, 60% ou 100% pour les groupes de traitement). De plus, nous avons explicitement informé le participant qu'il s'agissait de recommandations, sans détailler le fonctionnement de l'algorithme.

Face à ses recommandations, l'utilisateur a ensuite été invité à ajouter (ou non) à sa *watchlist* tous les films<sup>32</sup> recommandés qu'il prévoyait de regarder dans un futur proche. Par conséquent, cette troisième étape nous permettra, lors de l'analyse des résultats, d'évaluer le volume de consommation intentionnelle (variable dépendante) de chaque groupe de participants et donc pour chaque pourcentage de diversité (facteur expérimental).

- **Étape 4.** Un post-questionnaire a été établi afin de mesurer et analyser d'autres variables dépendantes (relatives aux hypothèses H1, H3 et H4). En réalité, nous avons demandé à chaque participant de noter sur une échelle de Likert en cinq points<sup>33</sup> les facteurs suivants : la pertinence des recommandations avec ses préférences, la diversité perçue des films recommandés, et sa satisfaction face à la liste de recommandations. Les concepts de pertinence perçue et de satisfaction ont été évalués à partir de l'adaptation d'échelles<sup>34</sup> de mesure utilisées par Hostler et al. (2012) alors que le concept de diversité perçue a été évalué à partir des échelles<sup>35</sup> de mesure adaptées des études de Knijnenburg et al. (2012) et Willemsen et al. (2016).
- **Étape 5.** Cette étape finale consistait à remercier les participants de leur collaboration et investissement tout en leur expliquant, finalement, leur contribution réelle à notre étude.

## 7. Participants

Après avoir réalisé une phase pilote convaincante auprès de cinq personnes, nous avons recruté les participants à notre expérience en ligne via la plateforme Profilic. Cette dernière met

---

<sup>31</sup> Des captures d'écran de l'application en ligne pour cette phase de recommandations se trouvent en *Annexe 8*.

<sup>32</sup> Puisque nous nous concentrons sur le volume de consommation intentionnelle, nous ne fixons pas de limite d'articles dans la *watchlist*.

<sup>33</sup> Les notes variant de -2 (« Pas du tout d'accord ») à +2 (« Tout à fait d'accord »), 0 étant neutre.

<sup>34</sup> Ces échelles de mesure adaptées sont respectivement détaillées en *Annexes 9 et 10*.

<sup>35</sup> Cette échelle de mesure adaptée est détaillée en *Annexe 11*.

en relation des chercheurs et des participants du monde entier pour aider à la recherche scientifique et aux études marketing. Chaque participant a été rémunéré à hauteur de £2.25.

### **7.1. Profil des répondants**

L'étude a impliqué 111 participants, répartis aléatoirement en 4 groupes (numérotés de G1 à G4 et détaillés en *Annexe 12*). La majorité des participants sont des hommes (50.5%), suivis de près par les femmes (48.6%), et une petite proportion (0.9%) se déclarant comme « autre ». Les tranches d'âge les plus représentées sont les 25-34 ans (42.3%), les 18-24 ans (30.6%) et les 35-44 ans (19.8%). Les participants viennent du monde entier, mais principalement d'Europe<sup>36</sup> (67.5%) et d'Afrique<sup>37</sup> (17.5%). Au sujet du niveau d'éducation, 53.2% des participants ont un diplôme de niveau bachelier, 23.4% un master, et 20.7% un diplôme de secondaire. Quant à la catégorie socio-professionnelle, la majorité des répondants travaillent à temps plein (44.1%), suivis des étudiants (22.5%), des travailleurs à temps partiel (16.2%), des indépendants (9.9%) et des personnes sans emploi (7.2%). En bref, ces données offrent un aperçu significatif de la diversité des participants à l'étude et offre un échantillon de consommateurs représentatif qui garantit la généralisation des résultats.

### **7.2. Habitudes et préférences cinématographiques**

En analysant les habitudes cinématographiques des participants (détaillées en *Annexe 12*), nous pouvons remarquer que 46.8% des participants se décrivent comme des amateurs de cinéma, 36% affirment regarder beaucoup de films par rapport à leurs pairs, 2.7% se considèrent experts en cinéma, et 14.4% admettent ne connaître que quelques films. En ce qui concerne le nombre de films visionnés par mois, la catégorie la plus représentée est celle regardant entre 5 et 10 films (40.5%), suivie de près par ceux regardant moins de 5 films (34.2%). Les tranches allant au-delà de 10 films représentent des proportions moins importantes. En outre, il est intéressant de noter les genres préférés des participants, qui ont été choisis parmi une liste variée incluant le drame (52), la comédie (65), l'action (68), l'aventure (54), l'animation (52), la fantasy (54), le crime (45), la romance (40), la science-fiction (57) le thriller (60), et l'horreur (35). Les données montrent que chaque genre a été sélectionné par un nombre significatif de participants, mettant en évidence la diversité des préférences cinématographiques (sur base du genre) au sein de l'échantillon.

---

<sup>36</sup> Royaume-Uni, Grèce, Portugal, Espagne, Italie, France, Belgique, Hongrie, Pologne, Estonie, Allemagne, République Tchèque.

<sup>37</sup> Afrique du Sud, Nigéria, Zimbabwe.

### 7.3. Connaissance et utilisation des systèmes de recommandation

De manière générale, les participants démontrent une certaine familiarité et acceptation des SdR sur les plateformes en ligne. En effet, les résultats (détaillés en *Annexe 12*) révèlent une connaissance étendue du concept de SdR par les participants, avec une grande majorité (91%) déclarant en avoir déjà entendu parler et étant conscients que les recommandations sont générées par des algorithmes. Quant au concept de personnalisation, dans le domaine du cinéma, 89.2% des participants affirme avoir déjà remarqué des recommandations adaptées à leurs préférences cinématographiques. Toutefois, les attitudes des participants envers le suivi des recommandations sont plus variées : 15.3% des participants les suivent scrupuleusement, 58.6% occasionnellement, 18.9% rarement, et 7.2% ne les suivent jamais.

## 8. Conclusion

Dans ce deuxième chapitre, nous avons entrepris une étude approfondie des techniques de filtrage collaboratif, en nous concentrant spécifiquement sur les approches *user-based*, *item-based* et SVD. En complément de ces algorithmes de recommandation, nous avons exploré le modèle de reclassement *Maximal Marginal Relevance* (MMR) pour diversifier les recommandations. Cette méthode de diversification vise à améliorer la diversité en sélectionnant des *items* qui maximisent la pertinence marginale tout en minimisant la similarité avec les *items* déjà sélectionnés. Elle dépend aussi d'un pourcentage de diversification qui permet de contrôler le degré de diversification des recommandations. À ce sujet, l'étude qu'ont menée Castagnos et al. (2010) a permis de démontrer que le niveau de diversité à inclure devait se situer entre 40% et 60% pour préserver une similarité raisonnable entre les recommandations et renforcer considérablement la confiance des utilisateurs. C'est pourquoi, dans le cadre de notre expérience en ligne, nous avons décidé de tester quatre pourcentages de diversité fixés à 0 pour le groupe de contrôle et 40%, 60% et 100% pour les groupes de traitement.

Au travers de notre question de recherche et de nos hypothèses, notre objectif est clair : évaluer l'influence de recommandations diversifiées à différents niveaux sur l'intention de consommation des utilisateurs et analyser d'autres variables dépendantes clés, à savoir la diversification de la consommation, la pertinence des recommandations face aux préférences des utilisateurs, la diversité perçue et la satisfaction des utilisateurs. Pour y parvenir, nous avons développé une application en ligne réaliste, permettant de mener une expérience participative dans une conception *between-subjects*. 111 participants ont été recrutés via la plateforme Prolific, assurant un échantillon représentatif de consommateurs.

## CHAPITRE 3 : RÉSULTATS ET ANALYSES

Dans ce troisième chapitre, nous analyserons les résultats de notre expérience en ligne menée auprès de 111 participants. Nous testerons également nos différentes hypothèses formulées dans le deuxième chapitre et tenterons de répondre à notre question de recherche. À travers ce dernier chapitre, nos objectifs principaux sont d'enrichir la littérature existante et offrir des conclusions managériales pertinentes pour les entreprises impliquées dans le développement ou l'utilisation de SdR.

### 1. Satisfaction des consommateurs

Les résultats du post-questionnaire<sup>38</sup>, en *Figure 2*, montrent une influence positive de la diversité sur la satisfaction des utilisateurs, corroborant ainsi l'hypothèse H1a selon laquelle une diversification à 40% et 60% des recommandations entraîne une satisfaction moyenne des utilisateurs plus élevée que lorsque le pourcentage de diversité est nul. En effet, les groupes de participants G2 et G3 affectés à un pourcentage de diversité de 40% et 60% présentent des niveaux de satisfaction plus élevés que

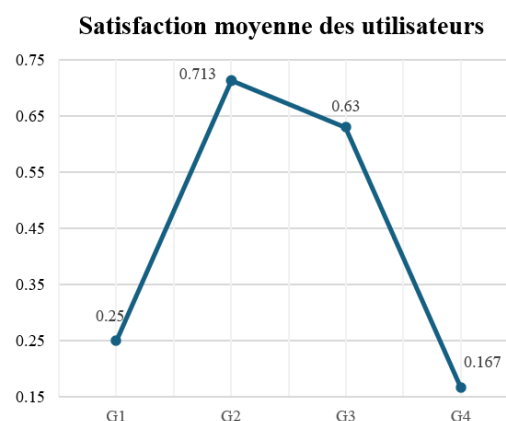


Figure 2. Analyse de la satisfaction par groupe de participants

le groupe G1 sans diversité (0%) et le groupe G4 avec une diversité totale (100%). Cette tendance confirme la conclusion de McNee et al. (2002), selon laquelle la satisfaction des utilisateurs ne coïncide pas nécessairement avec une grande précision dans les SdR, et souligne l'effet positif des recommandations diversifiées sur la satisfaction des utilisateurs, comme l'ont également constaté Castagnos et al. (2013).

En ce qui concerne le groupe G4 avec une diversité totale, ce dernier présente les scores de satisfaction les plus bas. Cette observation valide l'hypothèse H1b, qui suggère qu'une diversification à 100% des recommandations entraîne une satisfaction moyenne des utilisateurs plus faible que lorsque le pourcentage de diversité est nul. En bref, l'inclusion excessive de diversité entraîne de la méfiance et de l'incompréhension chez les utilisateurs, comme le soulignent les recherches antérieures (Castagnos et al., 2013). Cette conclusion renforce aussi l'importance cruciale d'un équilibre entre diversité et pertinence pour garantir une satisfaction optimale des utilisateurs. (Adomavicius et Kwon, 2012).

<sup>38</sup> Les résultats obtenus pour chaque question posée au sujet de la satisfaction des utilisateurs se trouvent en *Annexe 13*.

De plus, l'analyse isolée de la question 1 (dont les résultats se trouvent en *Annexe 13*) souligne que l'inclusion modérée de diversité dans les recommandations permet aux utilisateurs de découvrir de nouveaux contenus cinématographiques. En effet, les résultats mettent en évidence que les niveaux de diversité observés dans les G2 et G3 (40% et 60% respectivement), favorisent une satisfaction élevée grâce à la découverte de nouveaux films. À ce sujet, l'étude d'Helberger et al. (2018) affirme que la diversité contribue au développement et à l'élargissement des préférences de l'utilisateur.

En conclusion, une diversité de 40% ou 60% contribue positivement à la satisfaction globale des utilisateurs en leur offrant une expérience enrichissante et variée, tandis qu'une inclusion excessive de diversité affiche des effets négatifs sur la satisfaction des utilisateurs. D'un point de vue statistique, les tests de Student unilatéraux entre G1 versus G2 et G3 dont les résultats se trouvent dans le *Tableau 4* et en *Annexe 14*, montrent que ces conclusions sont significatives à 95%. Néanmoins, le test-t entre G1 versus G4 n'est pas significatif à 95%.

Variable dépendante	Hypothèse	p-valeur			Conclusion
		G1 vs G2	G1 vs G3	G1 vs G4	
Satisfaction	H1a	0.02	0.04	-	Non rejet de H1a
	H1b	-	-	0.36	Rejet de H1b
Volume de consommation intentionnelle	H2a	0.05	0.05	-	Non rejet de H2a
	H2b	-	-	0.35	Rejet de H2b
Diversification de la consommation	H2c	0.001	0.05	-	Non rejet de H2c
	H2d	-	-	0.24	Rejet de H2d
Diversité perçue	H3	0.007	0.001	< 0.001	Non rejet de H3
Pertinence perçue	H4	0.05	0.05	< 0.001	Non rejet de H4

Tableau 4. Résultats des tests de Student unilatéraux pour des échantillons indépendants mesurant la significativité statistique (p-valeurs) des différentes hypothèses testées

## 2. Consommation intentionnelle

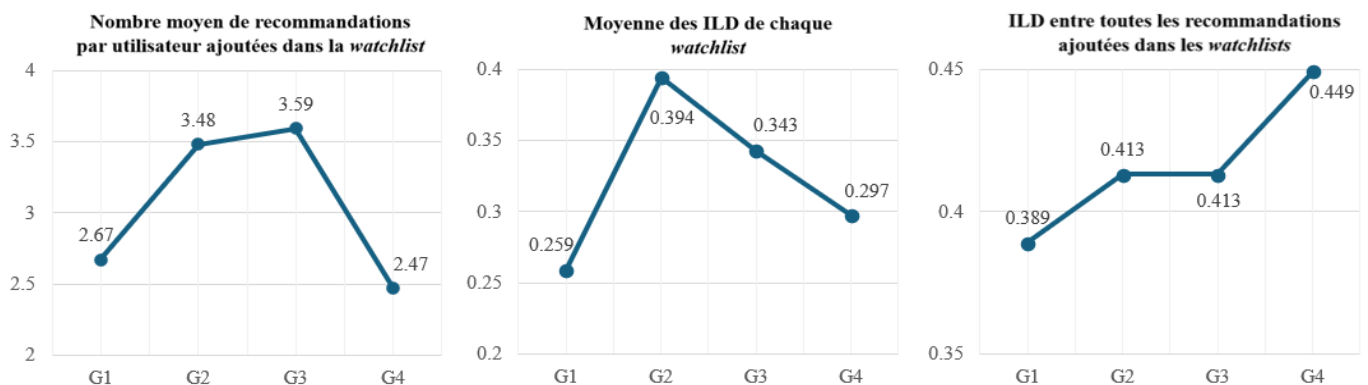


Figure 3. Résultats obtenus dans le cadre de l'analyse de la consommation intentionnelle des participants

## 2.1. Volume de consommation intentionnelle

D'après nos analyses précédentes, il existe une tendance montrant que l'utilisateur est plus satisfait lorsque ses recommandations sont modérément diversifiées. Mais, a-t-il pour autant l'intention de les suivre ? La littérature est claire concernant les effets positifs des SdR sur les ventes et donc le volume de consommation (Gomez-Uribe, 2016 ; Brent et al., 2017). D'ailleurs, l'étude d'Hostler et al. (2012) révèle également que l'augmentation de la satisfaction des clients entraîne des résultats inestimables à long terme.

Les résultats de notre étude, résumés en *Figure 3*, révèlent une augmentation du **nombre moyen de recommandations ajoutées dans les *watchlists*** à mesure que le pourcentage de diversité augmente de 0 à 60%. Cette conclusion est cohérente avec l'hypothèse H2a, suggérant qu'une diversification à 40% et 60% des recommandations encourage une consommation intentionnelle plus élevée de la part des utilisateurs que lorsque le pourcentage de diversité est nul. Cela permet de répondre à notre question de recherche et appuie aussi l'idée que la diversité dans les recommandations stimule l'engagement des utilisateurs. Cependant, il convient de noter que cette augmentation n'est pas nécessairement linéaire et peut diminuer, lorsque le pourcentage de diversité devient très élevé, comme observé pour le groupe G4 avec une diversité de 100%. Cette diminution observée avec le groupe G4 suggère qu'il peut exister une limite à l'effet positif de la diversité sur l'intention de consommation des utilisateurs. En effet, une inclusion excessive de diversité engendre de la méfiance et de l'incompréhension chez l'utilisateur (Castagnos et al., 2013). Par conséquent, la validité de l'hypothèse H2b, selon laquelle la diversification à 100% des listes de recommandations permet un volume moyen de consommation intentionnelle des utilisateurs plus faible que lorsque le pourcentage de diversité est nul, se vérifie totalement.

En conclusion, ces résultats soulignent l'importance de trouver un équilibre entre la diversité des recommandations et les préférences individuelles des utilisateurs pour optimiser leur engagement et consommation. C'est ce que démontrent les groupes G2 et G3 pour qui les niveaux de diversité sont respectivement de 40% et 60%. De plus, il est important de souligner que parmi les groupes étudiés, c'est le groupe G3 avec 60% de diversité qui se démarque en matière de nombre moyen de recommandations ajoutées dans les *watchlists*. D'un point de vue statistique, les tests de Student unilatéraux entre G1 versus G2 et G3, dont les résultats sont repris dans le *Tableau 4* et en *Annexe 15*, affichent une significativité à 95%. Néanmoins, le test-t entre G1 versus G4 n'est pas significatif à 95%.

## 2.2. Diversification de la consommation intentionnelle

En analysant les résultats de l'étude, résumés en *Figure 3*, plusieurs tendances émergent quant à la diversification de la consommation intentionnelle des utilisateurs en fonction du pourcentage de diversité dans les recommandations.

Une première question se pose : est-ce que recevoir des recommandations diversifiées (entièrement ou en partie) permet une consommation intentionnelle plus diversifiée de l'utilisateur ? En examinant **la moyenne des ILD<sup>39</sup> de chaque watchlist**, nous pouvons valider l'hypothèse H2c, mais rejeter H2d puisque l'inclusion de la diversité dans les recommandations est toujours associée à une intention de consommation plus diversifiée de chaque utilisateur par rapport au groupe G1 où aucune diversité n'est incluse dans les recommandations. Cela suggère qu'un utilisateur est donc plus enclin à explorer une gamme plus large de contenus lorsque ses recommandations sont diversifiées. Néanmoins, cette relation n'est pas nécessairement linéaire car il existe des variations dans la manière dont cette augmentation apparaît. En effet, il est intéressant de noter que dans les groupes G3 et G4, bien que le pourcentage de diversité dans les recommandations soit élevé (60% et 100% respectivement), la moyenne des ILD de chaque *watchlist* diminue par rapport au groupe G2. Il n'est donc pas vrai d'affirmer que plus la diversité augmente, plus la consommation intentionnelle de chaque utilisateur devient diversifiée. En ce qui concerne les tests de Student unilatéraux entre G1 versus G2 et G3 dont les résultats se trouvent dans le *Tableau 4* et en *Annexe 16*, ils affichent une significativité à 95%. Cependant, le test-t entre G1 versus G4 n'est pas statistiquement significatif à 95%.

Pour aller plus loin, une seconde question peut être posée : est-ce que fournir des recommandations diversifiées (entièrement ou en partie) aux utilisateurs permet une intention de consommation globalement plus diversifiée ? En analysant **l'ILD entre toutes les recommandations ajoutées dans les watchlists**, les résultats indiquent une diversification globale constante, voire croissante, de la consommation intentionnelle à mesure que le pourcentage de diversité augmente. Cette conclusion, notamment pertinente pour les fournisseurs d'*items*, signifie que les recommandations diversifiées encouragent une consommation plus variée dans l'ensemble du groupe de consommateurs, ce qui peut contribuer à enrichir l'expérience utilisateur et à favoriser l'exploration de nouveaux contenus.

---

<sup>39</sup> La distance intra-liste (*Intra-List Distance* ou ILD) peut être considérée comme la véritable mesure de la diversité objective. C'est pourquoi, la diversité d'un ensemble d'*items* recommandés est définie comme la distance moyenne entre les couples d'*items* de l'ensemble. Pour calculer la distance entre les paires d'*items*, nous avons utilisé la similarité cosinus entre les *items* que nous avons calculé à partir des genres, des années de sortie et du contenu textuel des résumés des films. Pour rappel, la distance(i,j) = 1 - similarité(i,j).

En conclusion, ces deux constatations fournissent des éclairages complémentaires de l'influence de la diversité des recommandations sur les comportements de consommation des utilisateurs. Une diversification des recommandations encourage non seulement une consommation intentionnelle plus diversifiée au niveau individuel, mais elle favorise également une diversification globale de l'intention de consommation dans l'ensemble du groupe.

### 3. Diversité perçue dans les recommandations

Les résultats du post-questionnaire<sup>40</sup>, en *Figure 4*, montrent une corrélation positive et significative entre le niveau de diversité inclus dans les recommandations et la perception de la diversité par les utilisateurs, confirmant l'hypothèse H3. En effet, cette dernière prévoit que l'inclusion de la diversité dans les recommandations conduit à une perception moyenne plus élevée de la diversité par les utilisateurs que lorsque le pourcentage de diversité est nul. Les résultats confirment cette hypothèse puisque le groupe G1 avec un pourcentage de diversité nul affiche les scores les plus bas, comparativement aux groupes G2, G3 et G4. Par conséquent,

plus le pourcentage de diversité est élevé, plus la diversité moyenne perçue par les utilisateurs est élevée. D'ailleurs, en *Annexe 17*, nous remarquons que le groupe G4, avec une diversité totale, affiche logiquement les scores les plus élevés pour l'ensemble des questions relatives à la diversité perçue, contrairement au groupe G1 avec un pourcentage de diversité nul. Quant aux tests de Student unilatéraux entre G1 versus G2, G3 et G4, dont les résultats sont repris dans le *Tableau 4* et en *Annexe 18*, montrent une significativité à 95%.

En conclusion, nos résultats corroborent l'effet positif de l'inclusion de la diversité dans les recommandations sur la perception accrue des utilisateurs de la variété et de la différenciation des *items* proposés, répondant ainsi aux attentes des utilisateurs en matière de diversité dans les recommandations. Nos résultats s'alignent aussi à l'étude menée par Willemsen et al. (2016) qui a montré que plus la diversification des listes de recommandations est élevée, plus la diversité perçue par les utilisateurs est élevée.

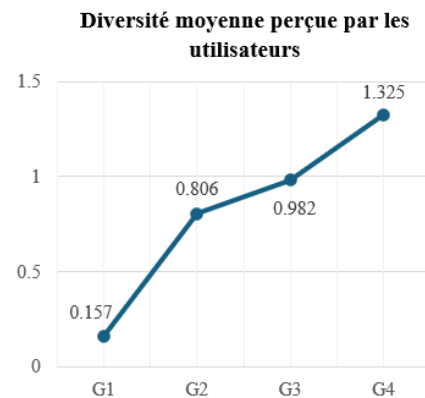


Figure 4. Analyse de la diversité perçue par groupe de participants

<sup>40</sup> Les résultats obtenus pour chaque question posée au sujet de la diversité perçue se trouvent en *Annexe 17*.

## 4. Pertinence perçue par les consommateurs face à leurs préférences

Les résultats du post-questionnaire<sup>41</sup>, en *Figure 5*, valident l'hypothèse H4 selon laquelle l'inclusion de la diversité dans les recommandations conduit les utilisateurs à percevoir la pertinence des recommandations face à leurs préférences plus faiblement que lorsque le pourcentage de diversité est nul. En effet, les scores les plus bas pour la pertinence perçue sont observés dans les groupes G2, G3 et G4 avec des niveaux plus élevés de diversité (de 40% à 100%), contrairement au groupe G1 avec un pourcentage de diversité nul. Cela suggère que les utilisateurs

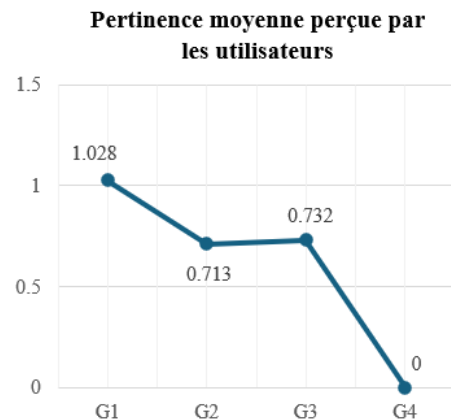


Figure 5. Analyse de la pertinence perçue par groupe de participants

perçoivent une baisse de la pertinence des recommandations lorsque de la diversité est incluse dans les recommandations. Cette conclusion est cohérente avec le fonctionnement de notre algorithme de diversification des recommandations (MMR), où un pourcentage de diversité (d) plus élevé favorise la diversité au détriment de la pertinence, tandis qu'une valeur plus faible de d favorise la pertinence au détriment de la diversité (Carbonell et Goldstein, 1998). Concernant les tests de Student unilatéraux entre G1 versus G2, G3 et G4, dont les résultats se trouvent dans le *Tableau 4* et en *Annexe 20*, affichent une significativité à 95%.

De plus, les scores de pertinence perçue diminuent généralement à mesure que le pourcentage de diversité augmente. D'ailleurs, en *Annexe 19*, nous pouvons remarquer que le groupe G4, avec une diversité totale, affiche logiquement les scores les plus faibles pour chacune des questions. Cependant, le résultat du groupe G3 suggère une nuance dans cette relation, indiquant que d'autres facteurs pourraient également jouer un rôle. Cela ne permet donc pas d'affirmer que plus le pourcentage de diversité est élevé, plus les utilisateurs perçoivent faiblement la pertinence des recommandations face à leurs préférences.

## 5. Conclusion

Dans ce troisième chapitre consacré à l'analyse des résultats obtenus, nous avons, tout d'abord, mis en lumière que l'inclusion de la diversité, à des niveaux modérés (40% et 60%), a un effet positif et significatif sur la satisfaction des utilisateurs, prouvant ainsi que la satisfaction des utilisateurs ne coïncide pas nécessairement avec une grande précision dans les SdR. Nos résultats ont également souligné que la diversité contribue à élargir les préférences des

<sup>41</sup> Les résultats obtenus pour chaque question posée au sujet de la pertinence perçue se trouvent en *Annexe 19*.

utilisateurs en leur permettant de découvrir de nouveaux contenus cinématographiques. Afin d'optimiser la satisfaction moyenne des consommateurs, il convient d'opter pour un pourcentage de diversité de 40%.

Ensuite, nous avons découvert que les recommandations diversifiées à 40% et 60% encouragent une consommation intentionnelle significativement plus élevée de la part des consommateurs que lorsque le pourcentage de diversité est nul. Cette constatation permet de répondre à notre question de recherche et suggère un lien entre la satisfaction des utilisateurs et leur propension à suivre les recommandations diversifiées. En matière de volume moyen de consommation intentionnelle, c'est un pourcentage de diversité de 60% qui est le meilleur. Néanmoins, une diversification excessive (à 100%) a montré une satisfaction des consommateurs et un volume moyen de consommation intentionnelle plus faibles que lorsque le pourcentage de diversité est nul, corroborant ainsi l'importance cruciale d'un équilibre entre diversité et pertinence.

De plus, nous avons observé que l'inclusion de la diversité dans les recommandations conduit non seulement à une diversification de la consommation intentionnelle de chaque utilisateur, mais également à une diversification globale de la consommation intentionnelle de l'ensemble du groupe de consommateurs. La diversification individuelle, montrant que l'utilisateur est significativement plus enclin à explorer une gamme plus large de contenus lorsque ses recommandations sont diversifiées, contribue donc à une diversification globale de la consommation intentionnelle de tous les utilisateurs.

Enfin, nos résultats ont démontré une corrélation positive et significative entre le niveau de diversité inclus dans les recommandations et la perception de la diversité par les utilisateurs. D'ailleurs, plus le pourcentage de diversité est élevé, plus la diversité moyenne perçue par les utilisateurs est élevée. Cependant, cette inclusion de la diversité dans les recommandations mène aussi les utilisateurs à percevoir, de manière significative, la pertinence des recommandations face à leurs préférences plus faiblement que lorsque le pourcentage de diversité est nul. C'est pourquoi, les pourcentages de diversité de 100% et 0% sont, respectivement et logiquement, les plus performants en termes de diversité moyenne perçue et de pertinence moyenne perçue par les consommateurs.

En conclusion, nos résultats mettent en lumière l'importance d'intégrer une diversité modérée (40% à 60%) dans les recommandations pour améliorer la satisfaction des utilisateurs, encourager une consommation intentionnelle plus élevée et favoriser une diversification tant individuelle que globale de la consommation. Toutefois, il reste crucial de trouver un équilibre entre diversité et pertinence pour garantir une expérience utilisateur optimale.

## CONCLUSION GÉNÉRALE

Les SdR occupent une place grandissante dans notre société et sur les plateformes de e-commerce. En plus d'être des outils de décision efficaces pour aider les utilisateurs à trouver plus facilement et rapidement les produits qu'ils préfèrent, les SdR visent également à augmenter le volume des ventes. En effet, il a été démontré que les consommateurs tendent à acheter davantage après avoir reçu des recommandations personnalisées.

Néanmoins, dans ce domaine de la recommandation, l'*accuracy* a longtemps été le critère dominant, mesurant la capacité des SdR à identifier avec précision les *items* les plus pertinents pour chaque utilisateur en fonction de ses préférences, de ses choix passés et des préférences des autres utilisateurs. Cependant, au fil du temps, d'autres critères, tels que la diversité, ont gagné en importance. La diversité des recommandations est devenue essentielle pour éviter l'enfermement des utilisateurs dans leurs préférences dominantes, améliorer leur expérience, et accroître leur satisfaction. Par conséquent, il est devenu crucial de trouver un équilibre entre précision et diversité, car une diversité excessive peut grandement diminuer la pertinence et créer l'incompréhension ou la méfiance chez les utilisateurs.

Notre étude, réalisée à travers une expérience participative rigoureuse et via une application en ligne réaliste, a permis d'apporter des éclairages significatifs sur l'intention de consommation des utilisateurs, la diversification de la consommation, la pertinence perçue par les consommateurs face à leurs préférences, la diversité perçue par les consommateurs et la satisfaction des consommateurs lorsque de la diversité est incluse dans les recommandations, et ce à différents niveaux (40%, 60% et 100%).

Nos résultats nous ont permis de confirmer que l'inclusion de la diversité, notamment à des niveaux modérés (40% et 60%), a un effet positif et significatif sur la satisfaction des utilisateurs et contribue à élargir les préférences des utilisateurs. Cette diversification (à 40% et 60%) encourage également une consommation intentionnelle significativement plus élevée, répondant ainsi à la question de recherche de ce mémoire et soulignant le lien entre satisfaction et engagement des utilisateurs envers les recommandations diversifiées. Cependant, nos résultats ont aussi révélé que pousser la diversification à son maximum, soit à 100%, entraîne une diminution de la satisfaction des consommateurs ainsi que du volume moyen de consommation intentionnelle, comparativement à une absence de diversité. Cette conclusion prouve donc l'importance cruciale d'un équilibre entre diversité et pertinence. De plus, nos résultats ont révélé que la diversification des recommandations provoque non seulement une diversification individuelle de la consommation intentionnelle de chaque utilisateur, mais

également une diversification globale de la consommation intentionnelle au sein du groupe de consommateurs.

Enfin, notre étude a mis en lumière une corrélation positive et significative entre le niveau de diversité inclus dans les recommandations et la diversité perçue par les utilisateurs. Cependant, cette diversification affecte significativement la perception de la pertinence des recommandations par les consommateurs par rapport à leurs préférences.

En conclusion, cette recherche enrichit la littérature sur les SdR. Elle offre des conclusions significatives qui sont particulièrement intéressantes pour les entreprises développant ou utilisant des SdR. En effet, elle met en lumière l'importance capitale d'inclure une diversité modérée, entre 40% et 60%, dans les recommandations. Cette diversification est cruciale pour garantir un équilibre optimal entre diversité et pertinence, améliorer la satisfaction des utilisateurs, accroître la consommation intentionnelle, et promouvoir une diversification de la consommation à la fois individuelle et globale.

Toutefois, il demeure essentiel de reconnaître que malgré les efforts déployés dans cette étude, plusieurs limites subsistent, ouvrant ainsi la voie à de futures recherches et explorations.

- **Base de données MovieLens :**

L'utilisation exclusive de la base de données MovieLens, qui se concentre principalement sur les *ratings* de films, peut limiter la généralisation des résultats à d'autres domaines de recommandation tels que la musique, les livres ou les produits de consommation courante. Il serait intéressant d'explorer comment la diversité influence la consommation intentionnelle dans ces divers contextes de recommandation.

- **Algorithme de recommandation et de diversification :**

Bien que le filtrage collaboratif et l'algorithme *item-based* soient puissants et largement utilisés dans les SdR, d'autres approches algorithmiques pourraient offrir des perspectives différentes sur l'effet de la diversité sur la consommation intentionnelle. En outre, bien que l'algorithme MMR soit un choix pertinent pour diversifier les recommandations, d'autres techniques de diversification pourraient être étudiées et comparées à l'approche MMR.

- **Aspect intentionnel de la consommation :**

Il est essentiel de noter que la conception de l'expérience ne permet pas de vérifier si les utilisateurs regardent effectivement les films après les avoir ajoutés à leur *watchlist*. Par conséquent, bien que nous puissions étudier l'intention de consommation des utilisateurs, nous ne disposons pas d'une mesure directe de leur comportement réel. Il serait bénéfique pour des études futures d'explorer d'autres méthodes permettant de suivre le comportement réel des utilisateurs après avoir reçu des recommandations diversifiées.

## BIBLIOGRAPHIE

ADOMAVICIUS, G. & KWON, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, 24(5), 896–911.

ADOMAVICIUS, G., & TUZHILIN, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749.

AGGARWAL, C. C. (2016). *Recommender Systems: The Textbook*. Cham: Springer. <https://doi.org/10.1007/978-3-319-29659-3>

AKSOY, L., BLOOM, P. N., LURIE, N. H. & COOIL, B. (2006). Should Recommendation Agents Think Like People ? *Journal of Service Research*, 8, 297-315. <https://doi.org/10.1177/1094670506286326>

ALCHIEKH HAYDAR, C. (2014). *Les systèmes de recommandation à base de confiance*. Thèse de doctorat, Université de Lorraine. En ligne [https://hal.univ-lorraine.fr/tel-01751172/file/DDOC\\_T\\_2014\\_0203\\_ALCHIEKH\\_HAYDAR.pdf](https://hal.univ-lorraine.fr/tel-01751172/file/DDOC_T_2014_0203_ALCHIEKH_HAYDAR.pdf), consulté le 9 septembre 2023.

BEEL, J., & BRUNEL, V. (2019). Data Pruning in Recommender Systems Research: Best-Practice or Malpractice ? *ACM Conference on Recommender Systems (RecSys)*. En ligne <http://ceur-ws.org/Vol-2431/paper6.pdf>

BEN SCHAFER, J., KONSTAN J., & RIEDL, J. (1999). Recommender systems in e-commerce. Proceedings of the 1st ACM conference on Electronic commerce (EC '99). Association for Computing Machinery, 158–166. doi:10.1145/336992.337035

BRADLEY, K. & SMYTH, B. (2001). Improving Recommendation Diversity. *Proceedings of the 12th National Conference in Artificial Intelligence and Cognitive Science*, 75–84.

BRENT, S. & LINDEN, G. (2017). Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing*, 21(3), 12–18. doi:10.1109/mic.2017.72

CANDILLIER, L., KRIS, J., FESSANT, F. & MEYER, F. (2009). State-of-the-Art Recommender Systems. *Collaborative and Social Information Retrieval and Access – Techniques for Improved User Modeling*. IGI Global, Hershey.

CARBONELL, J. G. & GOLDSTEIN, J. (1998). The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries”, *SIGIR*, pp. 335–336.

CASTAGNOS, S., BRUN, A., & BOYER, A. (2013). Utilité et perception de la diversité dans les systèmes de recommandation. *CORIA 2013 - Conférence en Recherche d'Informations et Applications*. En ligne <https://inria.hal.science/hal-00783329>

CASTAGNOS, S., BRUN, A., & BOYER, A. (2014). La diversité : Entre besoin et méfiance dans les systèmes de recommandation. *Revue I3 - Information Interaction Intelligence*. <https://inria.hal.science/hal-01108998/document>

CASTAGNOS, S., JONES, N. & PU, P. (2010). Eye-Tracking Product Recommenders' Usage. *Proceedings of the 4th ACM Conference on Recommender Systems*, 29-36. <https://doi.org/10.1145/1864708.1864717>

CASTELLS, P., VARGAS, S. & WANG, J. (2011). Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. *Proceedings of International Workshop on Diversity in Document Retrieval (DDR)*, 29-37.

CHANDRA, S., VERMA, S., LIM, W. M., KUMAR, S., & DONTU, N. (2022). Personalization in personalized marketing : Trends and ways forward. *Psychology & Marketing*, 39(8), 1529-1562. doi: 10.1002/mar.21670

CHANG, A. C. & KUKAR-KINNEY, M. (2011). The effects of shopping aid usage on consumer purchase decision and decision satisfaction, *Asia Pacific Journal of Marketing and Logistics*, 23(5), 745-754. <https://doi.org/10.1108/13555851111183110>

CHEN, M. & LIU, P. (2017). Performance Evaluation of Recommender Systems. *International Journal of Performability Engineering*, 13(8), 1246-1259. doi:10.23940/ijpe.17.08.p7.12461256

COOMBS, C. & AVRUNIN, G.S. (1977). Single peaked preference functions and theory of preference. *Psychological Review*, 84(2), 216–230. <https://doi.org/10.1037/0033-295X.84.2.216>

DE, P., Hu, Y. J. & Rahman, M. S. (2010). Technology usage and online sales: An empirical study. *Management Science*, 56(11), 1930–1945. <https://ssrn.com/abstract=1829798>

DIAS, M. B., LOCHER, D., LI, EL-DEREDY, M. W. & LISBOA, P. J. (2008). The value of personalised recommender systems to e-business: a case study. *Proceedings of the 2008 ACM conference on Recommender systems*, 291–294. <https://doi.org/10.1145/1454008.1454054>

DU, Y. (2021). *Des données aux connaissances : vers des recommandations plus pertinentes, diversifiées et transparentes*. Thèse de doctorat, IMT - MINES ALES - IMT - Mines Alès Ecole Mines - Télécom. En ligne <https://theses.hal.science/tel-03611956/document>, consulté le 17 septembre 2023.

EKSTRAND, M. D., HARPER, F. M., WILLEMSSEN, M. C. & KONSTAN, J. A. (2014). User perception of differences in recommender algorithms. *Proceedings of the 8th ACM Conference on Recommender Systems*. 161–168. <https://doi.org/10.1145/2645710.2645737>

FERWERDA, B., GRAUS, M. P., VALL, A., TKALCIC, M. & SCHEDL, M. (2017). How item discovery enabled by diversity leads to increased recommendation list attractiveness. *Proceedings of the Symposium on Applied Computing*, 1693–1696.

FLEDER, D. & HOSANAGAR, K. (2007). Recommender Systems and their Impact on Sales Diversity. *Proceedings of the 8th ACM Conference on Electronic Commerce*, 192-199.

GE, M., DELGADO-BATTENFELD, C., & JANNACH, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. *Proceedings of the 2010 ACM Conference on Recommender Systems*. doi: [10.1145/1864708.1864761](https://doi.org/10.1145/1864708.1864761)

GE, M., GEDIKLI, F. & JANNACH, D. (2011). Placing high-diversity items in top-N recommendation lists. *Proceedings of the 9th Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at IJCAI 2011*, 65-68.

GOMEZ-URIBE, C. & HUNT, N. (2016). The Netflix Recommender System. *ACM Transactions on Management Information Systems*, 4(6), 1–19. doi:10.1145/2843948

GUNAWARDANA, A. & SHANI, G. (2015). *Recommender Systems Handbook*. Boston, MA: Springer. [https://doi.org/10.1007/978-0-387-85820-3\\_8](https://doi.org/10.1007/978-0-387-85820-3_8)

HELBERGER, N., KARPPINEN, K., & D'ACUNTO, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2), 191-207. doi:10.1080/1369118x.2016.1271900

HERLOCKER, J., KONSTAN, J., TERVEEN, L., & RIEDL, J. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53. doi:10.1145/963770.963772

HINZ, J. D. O. & Eckert, D.-K. J. (2010) The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering*, 2(2), 67–77. <https://doi.org/10.1007/s12599-010-0092-x>

HOSANAGAR, K., FLEDER, D., LEE, D. & BUJA, A. (2014). Will the Global Village Fracture into Tribes: Recommender Systems and Their Effects on Consumers. *Management Science*, 60(4), 805–823. <http://dx.doi.org/10.2139/ssrn.1321962>

HOSTLER, R. E., YOON, V. Y., & GUIMARAES, T. (2012). Recommendation agent impact on consumer online shopping: The Movie Magic case study. *Expert Systems with Applications*, 39(3), 2989–2999. <https://doi.org/10.1016/J.ESWA.2011.08.160>

HU, R. & PU, P. (2011). Enhancing recommendation diversity with organization interfaces. *Proceedings of the 16th international conference on Intelligent user interfaces (IUI '11)*, 347-350. <https://doi.org/10.1145/1943403.1943462>

HUG, N. (2015). *Surprise Documentation. Surprise*. En ligne <https://surprise.readthedocs.io/en/stable/>, consulté le 06 octobre 2023.

JANNACH, D. & HEGELICH, K. (2009). A case study on the effectiveness of recommendations in the mobile internet. *Proceedings of the third ACM conference on Recommender systems*, 205–208. <https://doi.org/10.1145/1639714.1639749>

JANNACH, J. & ADOMAVICIUS, G. (2016). Recommendations with a purpose. *Proceedings of the 10th ACM Conference on Recommender Systems*, 7–10. <http://doi.acm.org/10.1145/2959100.2959186>

KIM, J., CHOI, I. & LI, Q. (2021). Customer Satisfaction of Recommender System: Examining Accuracy and Diversity in Several Types of Recommendation Approaches. *Sustainability 2021*, 13(11), 6165. <https://doi.org/10.3390/su13116165>

KNIJNENBURG, B. P., WILLEMSSEN, M. C., GANTNER, Z., SONCU, H. & NEWELL, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22(4–5), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>

MATT, C., HESS, T., & WEISS, C. (2019). A factual and perceptual framework for assessing diversity effects of online recommender systems. *Internet Res.*, 29, 1526-1550.

MCNEE, S. M., RIEDL, J., & KONSTAN, J. A. (2006). Being accurate is not enough : How accuracy metrics have hurt recommender systems. *Proceedings of the Conference on Human Factors in Computing Systems*, 1097–1101.

MCNEE, S.M., ALBERT, I., COSLEY, D., GOPALKRISHNAN, P., LAM, S.K., RASHID, A.M., KONSTAN, J.A. & RIEDL, J. (2002). On the Recommending of Citations for Research Papers. *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, 116 - 125.

Netflix (2023). *Fonctionnement du système de recommandations de Netflix*. Netflix. En ligne <https://help.netflix.com/fr/node/100639>, consulté le 22 octobre 2023.

PARISER, E. (2011). *The filter bubble : what the Internet is hiding from you*. New York: Penguin press.

PARISER, E. (2012), *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. USA: Penguin Books.

PAYNE, J. W. (1982). Contingent Decision Behavior, *Psychological Bulletin*, 92(2), 382-402. <https://doi.org/10.1037/0033-2909.92.2.382>

ROSEN, S. (1981). The Economics of Superstars. *The American Economic Review*, 71(5), 845–858. <http://www.jstor.org/stable/1803469>

SIMON, H. A. (1947). *Administrative behavior; a study of decision-making processes in administrative organization*. Macmillan.

SMYTH B. & MCCLAVE P. (2001). Similarity vs. Diversity. *Proceedings of the 4th International Conference on Case-Based Reasoning*, 347-361.

UZUNIDIS, D. (2007, 11). De la méthode de recherche économique. *Marché & organisations*, 5(3), 101-106. doi: 10.3917/maorg.005.0101

WANG, L., LUO, J., DENG, S. & GUO X. (2024). RoCS: Knowledge Graph Embedding Based on Joint Cosine Similarity. *Electronics*, 13(1):147. <https://doi.org/10.3390/electronics13010147>

WANG, Y., ZHANG, X., LIU, Z., DONG, Z., FENG, X., TANG, R., & HE, X. (2020). Personalized Re-ranking for Improving Diversity in Live Recommender Systems. *ArXiv*. <https://doi.org/10.48550/arXiv.2004.06390>

WILLEMSSEN, M.C., GRAUS, M.P. & KNIJNENBURG, B.P. (2016). Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction*, 26, 347–389. <https://doi.org/10.1007/s11257-016-9178-6>

WU, H., ZHANG, Y., MA, C., LYU, F., HE, B., MITRA, B., & LIU, X. (2022). Result Diversification in Search and Recommendation: A Survey. *ArXiv*. <https://doi.org/10.48550/arXiv.2212.14464>

XIAO, B. & BENBASAT, I. (2007). E-commerce product recommendation agents: use, characteristics, and impact. *MIS Quarterly*, 31(1), 137–209. <https://doi.org/10.2307/25148784>

ZANKER, M., ROOK, L., & JANNACH, D. (2019). Measuring the impact of online personalisation: Past, present and future. *International Journal of Human-Computer Studies*, 131, 160-168. doi:10.1016/j.ijhcs.2019.06.006

ZHANG, M., & HURLEY, N.J. (2008). Avoiding monotony: improving the diversity of recommendation lists. *Proceedings of the 2nd ACM Recommender Systems*, 123-130.

ZHANG, Y., CALLAN, J., & MINKA, T.P. (2002). Novelty and redundancy detection in adaptive filtering. *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 81-88. <https://doi.org/10.1145/564376.564393>

ZHOU, T., KUSCSIK, Z., LIU, J. G., MEDO, M., WAKELING, J. R., & ZHANG, Y. C. (2008). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4511–4515.

ZIEGLER, C. N., MCNEE, S. M., KONSTAN, J. A., & LAUSEN, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web*, 22-32. doi:10.1145/1060745.1060754