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

Mémoire de fin de cycle

En vue de l'obtention du diplôme de Master en Informatique
Option : Administration et Sécurité des Réseaux

Thème

Sécurisation cognitive des systèmes de recommandation e-commerce contre les attaques shilling

Réalisé par : M^{lle}. BAAR Hadjer M^{lle}. ANDJOUH Melissa

Encadré par : M^{me} BATTAT Nadia

Examinateurs : M^{me} AIT ABDELOUAHAB Karima
M DJEBARI Nabil
M^{me} CHIBANI Samia

Président : M^{me} CHABANE Sarah

Promotion 2024 - 2025

Remerciements

Avant tout, nous remercions Dieu Tout-Puissant de nous avoir donné la force, la santé et la patience pour aller jusqu'au bout de ce mémoire.

Un remerciement particulier à nos familles, qui nous ont toujours soutenues moralement, encouragées et motivées tout au long de notre cursus scolaire et universitaire.

Nous voudrions aussi remercier notre encadrante, M^{me} BATTAT, pour son aide, ses conseils et son orientation tout au long de ce travail.

Dédicaces

I humbly dedicate this work to :

***God**, for granting us strength, patience, and guidance throughout this journey.*

To my younger self

the little girl who once dreamed of this very moment —

To the girl who always gave her best to turn her dreams into reality.

To the one who struggled in silence, night after night, yet never gave up.

***I am proud of you** — proud of the woman you are becoming.*

Proud of the strength you showed through five challenging years.

Proud, because you did it.

And I will always, always be proud of you.

***To you mom,** for your strength, unconditional love, and constant support.*

***To you dad,** for your love , support and heartfelt duas.*

***To my sisters,** who have always been by my side with help, advice, and gentle care. To my **brother** Yasser for his love, care, and protection.*

*And to all **my friends,** who stood by me during moments of doubt, heartbreak, and hopelessness — thank you for never letting me fall.*

BAAR Hadjer

Dédicaces

À mon cher frère Massina et à mon oncle bien-aimé Nacer, qui nous ont quittés trop tôt.

À mes chers parents Rabiha et Lamri, pour leur amour et leurs sacrifices.

À tous mes oncles et tantes, et tout particulièrement à Madjid, Lounis, Tahar, Rachid, Djamel, Zoulikha, Djamila et Hakima, pour leur présence bienveillante.

À mes cousins : Syphax, Kouceila, Takfarinas, Micipsa , Karima, Tilelli et Siham, pour leur soutien fidèle malgré la distance.

ANDJOUH Melissa

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

ACL	Access Control List
CG	Cumulative Gain
DCG	Discounted Cumulative Gain
F1-score	Moyenne harmonique entre la précision et le rappel
FC	Filtrage Collaboratif
FN	False Negative
FP	False Positive
IA	Intelligence Artificielle
kNN	k-Nearest Neighbors (k plus proches voisins)
MAE	Mean Absolute Error (Erreur absolue moyenne)
MLP	Multi-Layer Perceptron
ML	Machine Learning (Apprentissage automatique)
MSE	Mean Squared Error (Erreur quadratique moyenne)
nDCG	Normalized Discounted Cumulative Gain
NLP	Natural Language Processing (Traitement du langage naturel)
Pandas	Python Data Analysis Library
RL	Reinforcement Learning
RMSE	Root Mean Squared Error (Erreur quadratique moyenne racine)
SR	Système de Recommandation

TP	True Positive
TSeMCCF	Trust-Semantic enhanced Multi-Criteria Collaborative Filtering
VSCode	Visual Studio Code

Introduction générale

Il y a encore quelques décennies, la relation entre le vendeur et le client était fondée sur une interaction directe et personnalisée. Les vendeurs connaissaient précisément les préférences, les besoins et les habitudes d'achat de leurs clients, ce qui leur permettait d'offrir des recommandations sur mesure et un service de qualité. Cette proximité favorisait non seulement la satisfaction des clients, mais aussi leur fidélité, les vendeurs étant capables d'anticiper et de répondre avec justesse aux attentes spécifiques de chacun [1].

Avec l'essor du commerce en ligne, cette relation de proximité s'est considérablement transformée. En 2024, le marché mondial du e-commerce a atteint 6,09 billions de dollars, enregistrant une croissance de 8,4% par rapport à l'année précédente, et cette tendance devrait se poursuivre dans les prochaines années [2]. Cependant, dans cet environnement numérique, les acheteurs n'ont plus accès aux conseils personnalisés d'un vendeur physique, et les commerçants rencontrent des difficultés à instaurer des relations de confiance individualisées. C'est dans ce contexte que les systèmes de recommandation (SR) ont émergé, avec pour objectif d'accompagner les utilisateurs en leur proposant des suggestions adaptées à leurs préférences et à leurs comportements d'achat.

Néanmoins, ces systèmes soulèvent des questions essentielles : les recommandations fournies sont-elles véritablement fiables et objectives ? Les utilisateurs peuvent-ils faire confiance à ces outils pour recevoir des suggestions pertinentes et sécurisées ?
Qui garantit l'intégrité, la transparence et la résistance de ces systèmes face aux manipulations potentielles ? En effet, au-delà de leur performance algorithmique, les systèmes de recommandation sont de plus en plus exposés à des menaces, notamment les attaques de type shilling, qui consistent à injecter de faux profils ou à manipuler des évaluations dans le but d'influencer les résultats.

Pour faire face à ces défis, plusieurs approches ont été développées afin de renforcer la sécurité des systèmes de recommandation, en particulier pour contrer les attaques ciblées. Toutefois, ces solutions présentent souvent des limites.

Dans le cadre de ce projet de fin cycle, nous avons mené une analyse approfondie des principales méthodes de sécurisation des systèmes de recommandation appliqués au e-commerce.

Cette étude a mis en évidence la nécessité de concevoir des approches plus robustes, adaptatives et intelligentes, capables de détecter et de contrer de manière proactive les attaques de type shilling.

Notre contribution s'inscrit dans cette perspective. Nous proposons une approche, fondée sur la confiance et intégrant une solution cognitive basée sur l'intelligence artificielle. Nous avons conçu un système capable d'apprendre de manière autonome à identifier et à neutraliser les tentatives de fraude, tout en assurant la transparence et la fiabilité des recommandations. Cette approche combine plusieurs technologies avancées, notamment l'apprentissage par renforcement, l'intelligence artificielle neuro-symbolique, la théorie des jeux et la blockchain. L'intégration de ces techniques permet au système de recommandation de devenir auto-apprenant, transparent, explicable et proactif dans la détection et la prévention des attaques de type shilling.

Pour évaluer l'efficacité de notre solution, nous avons simulé une attaque shilling par injection de faux profils afin d'analyser son impact sur la performance du système, mesurée notamment par l'augmentation de l'erreur quadratique moyenne (RMSE) et la diminution du score de confiance global. Nous avons ensuite appliqué notre mécanisme de défense cognitive, en utilisant des techniques de détection avancées et différents classificateurs. Les résultats obtenus démontrent que notre approche permet de détecter et d'identifier efficacement les comportements malveillants, contribuant ainsi à renforcer la sécurité et la fiabilité des systèmes de recommandation contre l'attaque de type shilling.

La structure de ce mémoire s'articule autour de trois chapitres :

- **Chapitre I** : Présentation générale des systèmes de recommandation.
- **Chapitre II** : Sécurité des systèmes de recommandation e-commerce.
- **Chapitre III** : Mise en place de notre approche cognitive de sécurisation.

Chapitre 1

Présentation générale des systèmes de recommandation

1.1 Introduction

Ce premier chapitre s'intéresse aux systèmes de recommandation. Nous allons expliquer leur fonctionnement, définir leurs objectifs et domaines d'application, les classifier puis réaliser une analyse comparative et évaluative.

1.2 Définitions de quelques notions de bases

Nous présentons ci-dessous quelques définitions liées au domaine de notre étude.

- **Entités :** les entités représentant le système de recommandation sont :
  - **L'utilisateur**, est la personne qui interagit avec le système, à qui le système recommande des ressources et à son tour donne son avis sur les ressources [3].
  - **Les items**, sont les éléments qui sont recommandés aux utilisateurs. Ils peuvent être d'une nature commerciale (produits à vendre), culturelle (films, chansons, presse), ou professionnelle (articles scientifiques) [4].
- **Communauté :** se définit simplement comme un groupe de personnes qui partagent un intérêt, une passion, un objectif ou une valeur commune. Elle repose sur des interactions régulières entre ses membres et sur un sentiment d'appartenance. Au-delà des échanges, une communauté est aussi un lieu de soutien, de co-construction, et souvent, de partage de ressources et de compétences [5].

— **U**ne **communauté en ligne** est un groupe de personnes qui communiquent régulièrement par l'intermédiaire de d'internet et du web. Leurs motivations de communication peuvent-être professionnelles, sociales, éducatives ou autres [6].

- **Profil utilisateur :** est un ensemble de données qui influencent le comportement d'un dispositif informatique en fonction des besoins ou des attentes de l'utilisateur ou utilisatrice.[7] Il peut contenir des données telles que, par exemple, identifiant, état civil, âge, sexe, préférences, compétences, historique des interactions avec le système...[8]
  En résumé, nos données, informations et connaissances explicitables sont alors intégrées au sein du système informatique à travers le profil utilisateur.
- **Personnalisation :** : Adaptation d'un produit, d'un service, d'un logement, etc., à la personnalité de celui à qui il est destiné [9].
  Dans le domaine des systèmes de recommandations, la **personnalisation** représente la part d'adaptation des recommandations aux goûts et aux besoins de l'utilisateur [10].

1.3 Système de recommandation

1.3.1 Définition

Un système de recommandation est un système de filtrage qu'on applique dans un contexte où il y a trop d'informations pour l'utilisateur. Celui-ci passerait beaucoup de temps à retrouver l'information ou l'item qui l'intéresse ce qui pourrait causer une certaine frustration. C'est aussi un moyen de proposer des produits inconnus de l'utilisateur mais susceptibles de l'intéresser [11].

1.3.2 Historique

Les systèmes de recommandation sont apparus au début des années **1990** juste après la création du **Web** en **1989**. La quantité de résultats renvoyés aux utilisateurs était si énorme au point qu'il leur était impossible de tous les consulter. Ces systèmes ont été donc créés pour remédier au problème de la surcharge d'information et pour faciliter aux utilisateurs de trouver ce qu'ils recherchent.
Les premiers moteurs de recherches se basaient sur le filtrage collaboratif.
L'année **1992** voit l'apparition du système de recommandation de documents **Tapestry** , ainsi que

la création du laboratoire de recherche **GroupLens**, qui travaille explicitement sur le problème de la recommandation automatique dans le cadre des forums de news de Usenet. Tapestry avait pour but de recommander à des groupes d'utilisateurs des documents issus des newsgroups susceptibles de les intéresser.

Les systèmes de filtrage collaboratif automatiques apparaissent ensuite. GroupLens utilise cette technique pour identifier les articles de Usenet susceptibles d'être intéressants pour un utilisateur donné. Les utilisateurs doivent seulement attribuer des notes ou effectuer d'autres opérations observables (par exemple, lire un article) ;

le système combine alors ces données avec les notes ou les actions d'autres utilisateurs pour fournir des résultats personnalisés.

Au cours de ces années, les systèmes de recommandation deviennent un sujet d'un intérêt croissant dans les domaines de l'interaction homme-machine, de l'apprentissage automatique ainsi que la recherche d'information. En **1995** apparaissent successivement **Ringo** un système de recommandation de musique, basé sur les appréciations des utilisateurs et **Bellcore** un système de recommandation de vidéos. La même année, **GroupLens** crée la société Net Perceptions dont le premier client a été Amazon [12].

Dès les années **2000**, la plateforme de commerce électronique a commencé à recommander des produits en se basant non seulement sur l'historique d'achat du client, mais également sur les comportements d'autres clients aux profils similaires. La dernière, **hybride** (Hybrid Recommendation System) est une combinaison des deux méthodes précédentes offrant des recommandations bien plus précises.

En 2006, **Netflix**, devenant un leader émergent du streaming vidéo a lancé le Netflix Prize, une compétition qui récompensait l'équipe de développeurs qui était capable d'améliorer son algorithme de recommandation d'au moins 10%. Les chercheurs devaient développer le meilleur algorithme de filtrage collaboratif en se basant uniquement sur leurs évaluations passées, sans disposer d'aucune autre information permettant d'identifier les utilisateurs. Pourtant, en **2009**, l'entreprise a annoncé la fin du concours. Elle avait décidé de ne pas rendre opérationnel l'algorithme gagnant et n'envisageait pas de le faire ultérieurement [13].

En milieu d'année **2012**, suite à l'intégration des systèmes de recommandation dans ses processus d'achat, la société **Amazon** a enregistré une augmentation de 29% des ventes pour un total de 12,83 milliards de dollars, contre 9,9 milliards de dollars

au cours de la même période en 2011 [14].

De même, la société **Netflix** estime en avril **2012** que 75% des activités de sélection de vidéos sur son site web, sont dûes à la présence des systèmes de recommandation [15].

Au milieu des années 2010, les algorithmes de recommandation ont connu une mutation significative impulsée par les avancées en matière d'intelligence artificielle, et notamment par le deep learning. Cette technologie, qui exploite la puissance des réseaux neuronaux, est une branche du machine learning, lui-même rattaché à l'intelligence artificielle.

YouTube a été parmi les premières à adopter les technologies de deep learning pour améliorer ses services. En effet, en **2016**, la plateforme a entrepris une refonte majeure de son algorithme de recommandation en y intégrant ces technologies avancées [13].

De nos jours, les systèmes de recommandation sont devenus des composantes incontournables pour la plupart des sites et plateformes numériques, ils s'améliorent constamment en parallèle avec les avancées de l'intelligence artificielle.

1.4 Processus de recommandation

Les systèmes de recommandation les plus efficaces opèrent généralement en **5 phases** [16] :

1. **Collecte de données**

   Les systèmes de recommandation s'appuient sur les données, c'est pourquoi la collecte de ces dernières constitue une première étape incontournable. Les deux principaux types de données à collecter sont les données explicites et les données implicites.

   Les données **explicites** englobent les actions et les activités des utilisateurs telles que les commentaires, les likes, les évaluations et les avis. Les données **implicites** comprennent le comportement des utilisateurs : historique de navigation, événements liés au panier, clics, achats antérieurs et historique des recherches.

   Les systèmes de recommandation utilisent également d'autres données sur les clients, comme les données démographiques (âge ou sexe) et psychographiques (centres d'intérêt ou style de vie), pour trouver des utilisateurs similaires, ainsi que des données sur les caractéristiques (par exemple, la fourchette de prix ou le type d'article), et identifier les produits ou services associés.

2. **Stockage**

   Une fois les données collectées, l'étape suivante consiste à les stocker. Le type de système de stockage utilisé dépendra du type de données collectées.

   Les **entrepôts de données** permettent d'agréger les données provenant de différentes sources pour faciliter leur analyse et l'apprentissage automatique, tandis que les **data lakes** permettent de stocker des données structurées et non structurées.

   Un data lakehouse regroupe les meilleurs aspects des entrepôts de données et des data lakes au sein d'une seule et même solution de gestion des données.

3. **Analyse**

   La phase d'analyse s'appuie sur des algorithmes de machine learning pour traiter et examiner les jeux de données. Ces algorithmes détectent les schémas, identifient les liaisons et évaluent la force de ces schémas et liaisons.

4. **Filtrage**

   La dernière étape consiste à filtrer les données pour afficher les éléments les plus pertinents de l'étape d'analyse précédente. Le filtrage des données consiste à appliquer certaines règles

et formules mathématiques aux données en fonction du type de moteur de recommandation utilisé.

5. **Affiner**

   Cette étape est **facultative**, elle permet d'évaluer régulièrement les sorties du système de recommandation et d'optimiser davantage le modèle pour améliorer en permanence sa précision et sa qualité.

1.5 Objectifs des systèmes de recommandation

Nous citons ci-dessous les principaux objectifs de ces systèmes.

- Les systèmes de recommandation aident les consommateurs à choisir des produits qu'ils vont probablement aimer et qu'ils sont susceptibles d'acheter, en fonction de leur navigation, recherches, achats et préférences [17].
- Augmentation du chiffre d'affaires : c'est une raison plus évidente, avec notamment les sites de e-commerce, des entreprises du retail, par exemple la Fnac, Amazon, ou Carrefour lorsque vous faites vos courses en ligne [18].
- Augmentation d'interaction entre utilisateurs : les réseaux sociaux vont vous suggérer des amis (Facebook, Linkedin) [18].
- Augmentation de l'engagement des utilisateurs : valable pour les sites de contenus tels que YouTube, Netflix, Spotify [18].
- Facilité d'accès aux contenus du système [4].
- Nouveauté : Proposer un contenu nouveau pour l'utilisateur [4].
- Diversité : Proposer une variété de suggestions [4].
- Accroître la satisfaction de l'utilisateur en personnalisant son expérience sur le site ou en lui envoyant des messages ciblés via des canaux tels que les courriels, les newsletters, ...[11].

1.6 Domaines d'application des systèmes de recommandation

De nos jours, les systèmes de recommandation sont appliqués dans presque tous les domaines du numérique. On peut mentionner les domaines suivants :

- **Commerce en ligne (E-commerce)** : Proposition de produits à l'utilisateur basée sur ses recherches, de ses achats précédents ou de ses évaluations. **Exemples** : Algerie Store, Jumia et Ouedkniss.
- **Réseaux sociaux** : Suggestion d'amis, de pages à suivre, de publications, de vidéos ou de contenus en fonction des interactions effectuées par l'utilisateur. **Exemples** : Facebook, Instagram et Twitter.
- **Plateformes de streaming** : Recommandation de vidéos, films et chansons populaires ou similaires à ceux que l'utilisateur a déjà visionnés ou écoutés. **Exemples** : YouTube, Netflix et Spotify.
- **Jeux vidéo** : Proposition de jeux adaptés aux habitudes du joueur. **Exemples** : Amazon Games, Epic Games Store et Steam.
- **Bibliothèque en ligne** : Recommandation de livres, romans, articles, guides, etc., selon les préférences du lecteur. **Exemples** : ArXiv, Google Livres et Gallica.
- **Tourisme et hôtellerie** : Suggestion de destinations, d'hôtels et de sites touristiques les plus pertinents. **Exemples** : Booking.com, Tripadvisor et Expedia.
- **Banque et Finance** : Proposition de services et produits financiers adaptés aux abonnements du client. **Exemples** : Banques en ligne, PayPal et les plateformes d'investissement.
- **Santé** : Recommandation de médecins spécialisés, de conseils médicaux, de régimes alimentaires et de pharmacies. **Exemples** : Health For You, Gluci Check et Santé.fr.

1.7 Prédiction dans les systèmes de recommandation

La prédiction consiste à estimer la note qu'un utilisateur pourrait attribuer à un item qu'il n'a pas encore vu ou évalué.

Dans la majorité des cas, les matrices d'évaluations utilisateur-item sont très creuses : seules quelques cellules contiennent des valeurs, tandis que la plupart restent vides ou sont initialisées à zéro, ce qui reflète l'absence d'interaction entre l'utilisateur et certains items.

Cette faible densité compromet la qualité des recommandations, car elle limite la quantité d'informations exploitables. Pour pallier ce problème, des méthodes de prédiction des évaluations manquantes sont utilisées.

Leur objectif est d'augmenter la densité de la matrice afin de permettre la génération de recommandations plus pertinentes et plus fiables [19].

1.7.1 Méthodes de prédiction

— **Prédiction basée utilisateur** : Le calcul de la prédiction se base sur l'utilisation des notes données par les profils similaires .

— **Prédiction basée item** :Attribuées aux items voisins de l'item test évalué par l'utilisateur actif c'est-à-dire Un item est recommandé à un utilisateur en fonction de sa similarité avec d'autres items déjà appréciés par cet utilisateur.

— **Prédiction basée modèle** : Donné par un modèle qui se base sur des algorithmes d'apprentissage automatique pour avoir une représentation de la matrice items [19].

1.8 Classification des systèmes de recommandation

1.8.1 Recommandation démographique

Les recommandations démographiques sont des recommandations simples qui fournissent des éléments liés aux données démographiques d'un utilisateur. Il divise les utilisateurs en plusieurs catégories ou groupes en fonction de données démographiques telles que le sexe, l'âge, la langue, le pays, etc. Le principe de cette méthode est que deux utilisateurs dans des environnements similaires partagent des goûts communs, d'autres n'ont pas partagé le même code car les deux personnes sont évoluées dans des environnements différents [3].

1.8.2 Recommandation à base de connaissances

Les recommandations sont faites à l'aide des connaissances spéciales, et certaines de ses caractéristiques de produit répondent aux préférences des utilisateurs. Si les données disponibles sont limitées, les systèmes basés sur les connaissances sont généralement plus fiables que les autres types de recommandations. Deux types de systèmes à base de connaissances : une recommandation à base de contraintes et le raisonnement à base des cas.

1. **Recommandation à base de contraintes**

   Les recommandations basées sur les contraintes utilisent une base de connaissances prédéfinie qui contient des règles claires sur la façon d'associer les besoins des clients aux fonctions de ressource. Par exemple, un utilisateur peut être intéressé par l'achat d'un produit avec un ensemble de caractéristiques spécifiques et dans une gamme de prix spécifique.

2. **Raisonnement à base des cas**

   Le raisonnement par cas exploite les régularités du monde réel pour résoudre des problèmes en trouvant des solutions à des cas similaires rencontrés et résolus dans le passé. ont été utilisés dans des systèmes de recommandation où les descriptions des problèmes des utilisateurs correspondent à des solutions aux problèmes basées sur des cas précédents [3].

1.8.3 Recommandation communautaire

Les systèmes de recommandation sociale (SocialRS) exploitent simultanément les interactions utilisateur-élément ainsi que les relations sociales utilisateur-élément pour générer des recommandations d'éléments aux utilisateurs.

De plus, l'exploitation des relations sociales est clairement efficace pour comprendre les goûts des utilisateurs en raison des effets de l'homophilie et de l'influence sociale. Pour cette raison, SocialRS a de plus en plus attiré l'attention [11].

— **Exemple des sites utilisant ces technologies**

1. **Facebook et Instagram : réseautage social intelligent**

   les réseaux sociaux Facebook et Instagram, opérés par Meta, utilisent également des algorithmes de recommandation pour améliorer l'engagement et la satisfaction des utilisateurs. Chaque utilisateur découvre son flux de publications orchestré sur mesure grâce à

l'analyse des interactions précédentes. Likes, commentaires, partages, et le temps passé sur différentes publications sont autant d'indicateurs pris en compte.

- Sélection de stories prioritaires.
- Publicités ciblées soigneusement choisies.
- Suggestions de groupes et d'amis potentiels.

2. **Spotify : la musique personnalisée à portée de clic**

   En matière de streaming musical, Spotify excelle dans la personnalisation grâce à ses algorithmes de recommandation. Ce service tire parti des données d'écoute pour façonner une expérience sonore unique pour chaque utilisateur. Parmi les fonctionnalités populaires proposées figurent les playlists comme « Découverte Hebdo » et « Discover Weekly ». Chaque semaine, Spotify génère une liste de chansons basée sur les genres musicaux récemment écoutés par l'utilisateur ainsi que les titres aimés. Dès qu'un utilisateur écoute une chanson ou créé une nouvelle playlist, l'algorithme détecte les préférences et propose des pistes similaires pour étoffer la sélection [20].

1.8.4 Filtrage basé sur le contenu

Le principe des systèmes de recommandation basé sur le contenu diffère du système collaboratif dans le fait qu'on utilise uniquement les préférences de l'utilisateur actif pour faire les recommandations. L'objectif est de retrouver des items similaires à ceux pour lesquels l'utilisateur a exprimé une préférence dans le passé. Chaque item possède des attributs (content) qui sont exploités pour faire la recommandation. Ces attributs constituent un descriptif de l'item qui se présente sous forme nominale (ex. : genre et auteur d'un livre) ou sous forme de texte libre.

Comment se fait la recommendation ?

Pour chaque utilisateur on définit un profil qui exprime les préférences de l'utilisateur. La méthode la plus simple est de construire le profil à partir d'attributs semblables à ceux des items. Le profil utilisateur est alors comparé aux attributs des items et les items les plus similaires sont recommandés à l'utilisateur. Pour faire le match entre le profil utilisateur et les items, on utilise des mesures de similarité telles que la similarité cosinus. Une méthode plus élaborée consiste à transformer le problème de recommandation en un problème de classification. Pour chaque utilisateur, on entraine un modèle de recommandation (profil) qui prend en entrée les attributs d'un item et

qui prédit l'intérêt de l'utilisateur pour cet item (like/dislike ou rating) [11]. La figure 1.1 ilustre ce processus.

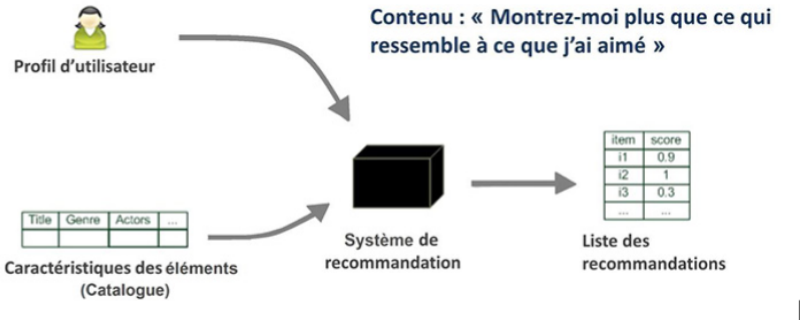

FIGURE 1.1 – Filtrage basé sur le contenu [21].

1.8.5 Filtrage collaboratif

Les systèmes basés sur le filtrage collaboratif produisent des recommandations en calculant la similarité entre les préférences d'un utilisateur et celles d'autres utilisateurs . De tels systèmes ne tentent pas d'analyser ou de comprendre le contenu des éléments à recommander. La méthode consiste à faire des prévisions automatiques sur les intérêts d'un utilisateur en collectant des avis de nombreux utilisateurs. L'hypothèse sous-jacente de cette approche est que ceux qui ont aimé un élément spécifique dans le passé auront tendance à aimer cet élément spécifique, ou un autre très « proche », à nouveau dans l'avenir. La figure 1.2 illustre ce processus.

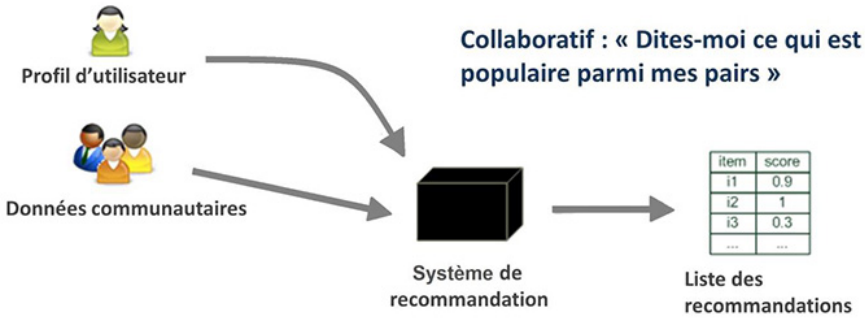

FIGURE 1.2 – Filtrage collaboratif [21].

L'idée des approches collaboratives est d'essayer de prédire l'opinion d'un utilisateur sur les différents éléments. La recommandation est basée sur les goûts et avis précédents de l'utilisateur et sur une mesure de similarité avec d'autres utilisateurs.

Les principales étapes de cette approche sont :

- De nombreuses préférences d'utilisateurs sont enregistrées .
- Un sous-groupe d'utilisateurs est repéré dont les préférences sont similaires à celles de l'utilisateur qui cherche la recommandation.
- Une moyenne des préférences pour ce groupe est calculée.
- La fonction de préférence qui en résulte est utilisée pour recommander des éléments à l'utilisateur qui cherche la recommandation.

Exemple Amazon : une personnalisation centrée sur les achats

L'une des plateformes pionnières dans l'utilisation des algorithmes de recommandation est Amazon. Ses systèmes sophistiqués sont conçus pour fournir des suggestions de produits très pertinentes, améliorant ainsi l'expérience d'achat en ligne [21].

1.8.6 Filtrage hybride

Un système de recommandation hybride utilise des composants de différents types d'approches de recommandation ou s'appuie sur leur logique . Par exemple, un tel système peut utiliser à la fois des connaissances extérieures et les caractéristiques des éléments, combinant ainsi des approches collaboratives et basées sur le contenu. La figure 1.3 ci-dessous illustre le fonctionnement d'un système de recommandation hybride combinant différentes approches.

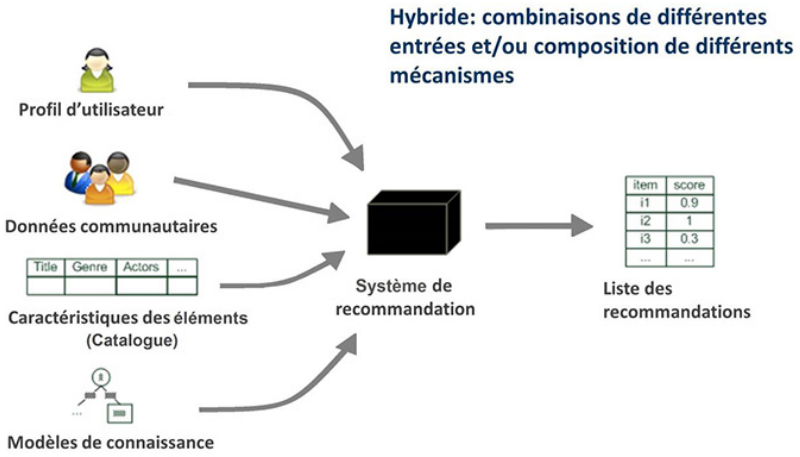

FIGURE 1.3 – Filtrage hybride [21].

Exemple : Netflix : révolutionne le divertissement par les recommandations Netflix a transformé le paysage du streaming vidéo avec ses algorithmes de recommandation avancés. Le géant du divertissement utilise plusieurs méthodes algorithmiques pour offrir une expérience utilisateur immersive et hyper-personnalisée. Contrairement à d'autres entreprises, Netflix adopte un système de filtrage hybride combinant filtrage collaboratif et filtrage basé sur le contenu. Cette approche multifacette maximise la pertinence des recommandations de contenus tels que films, séries et documentaires.

1. Analyse des historiques de visionnage.
2. Évaluation des notes données par les utilisateurs.
3. Utilisation de métadonnées descriptives des contenus.

Les utilisateurs de Netflix remarquent que leur page d'accueil est constamment mise à jour en fonction de leurs goûts personnels. Que ce soit via des recommandations générales, des genres spécifiques qu'ils aiment ou même des bandes-annonces adaptées, l'algorithme rend chaque session unique.

1.9 Analyse comparative des types de systèmes de recommandation

Le tableau 1.1 présente une synthèse des différents types de systèmes de recommandation. Chaque type est analysée selon son principe fonctionnement, ses principaux avantages ainsi que ses limites. Cette comparaison permet d'évaluer leur pertinence en fonction des besoins spécifiques d'une application.

TABLE 1.1 – Types de recommandations

Type de recommandation	**Description**	**Avantages**	**Inconvénients**
Recommandation démographique	Basée sur des critères démographiques comme l'âge, le sexe, la localisation, etc.	Simple à mettre en place, utile pour des recommandations génériques.	Manque de personnalisation, ne prend pas en compte les préférences individuelles.
Recommandation à base de connaissances	Exploite des règles spécifiques ou des connaissances expertes pour faire des recommandations.	Plus fiable lorsque les données utilisateurs sont limitées, utile dans les domaines nécessitant des connaissances précises.	Dépend fortement de la qualité des connaissances encodées, nécessite une maintenance régulière.
Recommandation à base de contraintes	Utilise des règles strictes pour filtrer les recommandations en fonction des besoins des utilisateurs.	Très précis et adapté à des critères spécifiques.	Moins flexible, ne s'adapte pas aux préférences évolutives des utilisateurs.
Raisonnement à base de cas	Recommandée en s'appuyant sur des cas similaires résolus précédemment.	Basé sur l'expérience, peut apprendre de nouvelles solutions.	Peut être limité si les cas passés ne couvrent pas suffisamment de situations.
Recommandation communautaire	Exploite les relations sociales et les interactions entre utilisateurs pour proposer du contenu.	Basé sur des préférences réelles, permet une meilleure découverte de contenu.	Peut biaiser les recommandations vers des tendances populaires, nécessite un grand volume de données.
Filtrage basé sur le contenu	Compare les attributs des items avec les préférences explicites de l'utilisateur.	Bonne personnalisation, fonctionne bien avec peu d'utilisateurs.	Sujet aux effets de surspécialisation, ne prend pas en compte les nouvelles tendances.
Filtrage collaboratif	Compare les préférences d'un utilisateur avec celles d'autres utilisateurs similaires.	Recommandations précises et pertinentes sur des tendances partagées.	Problème du démarrage à froid (nécessite beaucoup de données), peut être sensible aux biais.
Filtrage hybride	Combine plusieurs approches pour optimiser la recommandation.	Plus performant, réduit les inconvénients des autres méthodes.	Complexe à implémenter et coûteux en ressources.

1.10 Problèmes des systèmes de recommandation

1.10.1 Sécurité

Les systèmes de recommandation, même s'ils sont utiles, sont vulnérables à divers problèmes de sécurité qui peuvent affecter leur efficacité et leur fiabilité [22]. La figure 1.4 suivante représente les composantes essentielles qui constituent l'étude de la recommandation sécurisée.

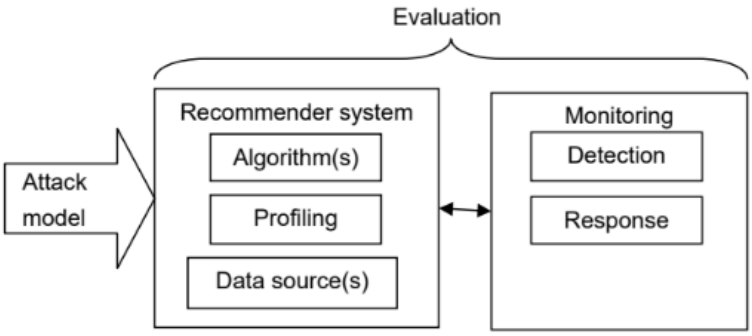

FIGURE 1.4 – Composantes clés de l'étude de la recommandation sécurisée [22].

1. **Modèles d'attaque**

   Les attaquants peuvent manipuler les systèmes de recommandation en injectant des profils biaisés ou en exploitant les interactions pour influencer les prédictions. Ces attaques incluent :

   - **Spam de moteur de recherche :** manipulation des métadonnées ou du contenu pour fausser les résultats.
   - **Attaques sophistiquées :** utilisation d'agents logiciels (softbots) pour simuler des comportements d'utilisateur.

2. **Sources de données**

   Les systèmes basés sur le contenu ou la collaboration sont exposés à des attaques ciblant leurs sources de données. Par exemple :

   - Les données falsifiées ou biaisées peuvent altérer les recommandations.
   - Les systèmes hybrides, combinant plusieurs types de données, offrent une meilleure défense contre ces attaques.

3. **Algorithmes**

   Certains algorithmes, comme le filtrage collaboratif basé sur les articles (kNN), présentent des avantages défensifs contre certaines attaques. Cependant :

   - Les algorithmes basés sur des modèles (réseaux bayésiens, arbres de décision, etc.) nécessitent une attention particulière pour éviter la manipulation.
   - Les approches hybrides, comme celles utilisées par Google, combinent différentes techniques pour réduire les biais.

4. **Profilage**

   Le profilage des utilisateurs peut être exploité par des attaquants via :

   - Des évaluations implicites (basées sur le comportement) qui sont plus difficiles à manipuler directement.
   - Des attaques sophistiquées nécessitant l'interaction avec le système pour créer des profils biaisés.

5. **Détection et réponse**

   La détection d'attaques repose sur :

   - L'identification de schémas d'activité suspects similaires aux approches classiques de détection d'intrusion.
   - La surveillance des changements dans la performance du système qui pourraient indiquer la présence de données biaisées.

   En réponse aux attaques :

   - Les profils suspects peuvent être supprimés si détectés individuellement.
   - Si le biais est global, le système doit compenser sans modifier la base de données des profils.

6. **Évaluation**

   Deux mesures clés permettent d'évaluer la sécurité des systèmes :

   - **Robustesse :** impact global d'une attaque sur le système.
   - **Stabilité :** mesure des changements dans les recommandations causés par une attaque.

1.10.2 Démarrage à froid

Les systèmes de filtrage collaboratif dépendent des évaluations des items par les utilisateurs. Ainsi, un nouvel item ne peut pas être recommandé tant qu'aucun utilisateur ne l'a évalué. Dans les systèmes de recommandation basés sur le filtrage collaboratif et les systèmes basés sur le contenu, il est impossible de prédire les préférences des utilisateurs sans connaître leurs historiques d'évaluations d'items. Ainsi, les nouveaux utilisateurs ne recevront pas de recommandations précises avant d'avoir évalué un certain nombre d'items [23].

1.10.3 Sparsity

Un système de recommandation souffre de la sparsity quand le nombre d'items évalués par les utilisateurs est très faible par rapport au nombre d'items total présent dans le système. Ce fait conduit à avoir une très faible densité dans la matrice d'évaluation utilisateurs/items.
Cela a des conséquences sur la capacité du système de recommandation à recommander toutes les items disponibles et sur l'exactitude des recommandations générées [23].

1.10.4 Problème du mouton gris

Les utilisateurs d'un système de recommandation peuvent avoir des goûts particuliers et des préférences très inhabituelles par rapport aux autres. Ces utilisateurs sont à la frontière entre deux ou plusieurs clusters d'utilisateurs. Il leur est donc difficile de trouver des utilisateurs similaires et des recommandations pertinentes [23].

1.11 Evaluation des systèmes de recommandation

La plupart des systèmes de recommandation ont été évalués en fonction de leur capacité à prédire avec précision les choix de l'utilisateur. Maintenant, il est largement admis que la précision

des prédictions est cruciale mais insuffisante pour déployer un bon système de recommandation. Les évaluations des systèmes de recommandations peuvent être effectuées en utilisant une analyse hors ligne (offline analysis) ou une expérimentation avec des utilisateurs réels (live user experiment).
Il existe une autre classification des méthodes d'évaluation des systèmes de recommandation. Ces méthodes d'évaluation sont classées en trois types : expérimentations offline, études avec des utilisateurs (user studies) et tests réels (real life testing). Ce dernier type est nommé expérimentations en ligne (Online experiments).

1.11.1 Types d'évaluation

1. **Études utilisateurs**

   Une étude utilisateurs est menée en recrutant un ensemble d'utilisateurs, et en leur demandant d'effectuer certaines tâches dans un environnement contrôlé pendant une courte période de temps. L'interaction entre les utilisateurs et le système de recommandation est observée et des informations sont enregistrées telles que le temps nécessaire pour terminer la tâche ou la qualité des résultats de la tâche. En plus de l'observation du comportement de l'utilisateur, il est possible de faire passer des questionnaires aux utilisateurs pour recueillir des données qui ne sont pas directement observables telles que l'appréciation de l'interface utilisateur ou de la pertinence des recommandations [23].

   - **Avantages :**
     - — Divers scénarios.
     - — Interfaces testable.
     - — Précision du retour.
   - **Inconvénients :**
     - — Utilisateurs averti que "c'est un test" (biais).
     - — Difficulté à recruter.
     - — Tester de nombreux utilisateurs. (représentativité de l'échantillon).[24]

2. **Évaluation à chaud("online")**

   L'évaluation online peut aussi recueillir le point de vue de l'utilisateur concernant le système de recommandation. Dans ce type d'évaluation, des utilisateurs réels utilisent le système dans des conditions réelles sur une longue période.

Ce type d'évaluation peut montrer les usages et les habitudes d'utilisation des utilisateurs, les problèmes et les besoins non satisfaits, et les problèmes que les chercheurs n'ont peut-être pas envisagés dans une étude utilisateurs. Avec ces tests réels sur le terrain, la plupart des objectifs centrés sur l'utilisateur peuvent être efficacement évalués, comme l'évaluation de l'expérience utilisateur, la satisfaction des utilisateurs ou la rétention des utilisateurs [23].

- **Avantages :**
  - Qualité des observations.
  - Lien avec apprentissage par renforcement et multi-armed bandits (optimisation gloutonne).
- **Inconvénients :**
  - Il faut en général beaucoup d'utilisateurs dans le système pour que ça marche.
  - Il faut la main sur le système pour mettre en place les tests.
  - Évaluation académique (reproductibilité) limitée.
  - Robustesse / généralisabilité réduites (manque de tests) [24].

3. **Évaluation à froid(Offline)**

Les évaluations offlines utilisent des ensembles de données (dataset) constitués d'actions des utilisateurs (principalement des évaluations de ressources). Les évaluations offlines simulent le processus de recommandation où un sous-ensemble des actions utilisateurs du dataset est caché et le système de recommandation prédit ces actions cachées. Le système de recommandation est évalué en fonction de sa capacité à prédire ces interactions cachées. Les résultats de ces prédictions sont analysés en utilisant une ou plusieurs métriques.

Deux types d'ensembles de données sont souvent utilisés dans ces évaluations :

- **Ensembles de données naturels :** ils sont constitués de données issues de l'historique des interactions d'utilisateurs réels dans un système donné sur une période donnée. De nombreux dataset sont disponibles pour mener des évaluations sur des algorithmes de recommandation. L'annexe III présente un comparatif entre les dataset les plus connus dans les EIAH.
- **Ensembles de données de synthèse :** ils sont construits de données artificielles. Ce type de dataset est habituellement utilisé pour tester comment les algorithmes de recommandation fonctionnent dans certaines conditions.

Les évaluations offlines ont l'avantage d'être rapide et économique et peuvent être réalisées sur plusieurs ensembles de données ou plusieurs algorithmes différents à la fois. Ce type d'évaluation est une évaluation objective des résultats de la prédiction. Aucune analyse hors ligne ne peut déterminer si les utilisateurs préfèrent un système particulier, soit en raison de ses prédictions, soit en raison d'autres critères moins objectifs tels que l'esthétique ou l'ergonomie de l'interface utilisateur [23].

1.11.2 Critères d'évaluation

1.11.2.1 Précision et méthodes basées sur les erreurs

a. **Erreur absolue moyenne (MAE) :** Mesure la différence moyenne entre les notes réelles et prédites. Plus le MAE est proche de zéro, plus le modèle est précis.

b. **Erreur quadratique moyenne (MSE) :** Quadrille les erreurs pour pénaliser davantage les écarts importants.

c. **Erreur quadratique moyenne racine (RMSE) :** Normalise le MSE pour le rendre comparable aux échelles de notation.

1.11.2.2 Méthodes d'aide a la décision

Les mesures d'aide à la décision aident à comprendre à quel point le recommandateur a été utile pour aider les utilisateurs à prendre de meilleures décisions en choisissant les bons articles et en évitant les mauvais articles. Deux des mesures les plus couramment utilisées sont la précision et le rappel [25].

a. **Précision**

La précision est le nombre d'éléments sélectionnés qui sont pertinents. Supposons donc que notre système de recommandation sélectionne 3 éléments à recommander aux utilisateurs, dont 2 sont pertinents, la précision sera de 66% la figure 1.5 ilustre ce processus.

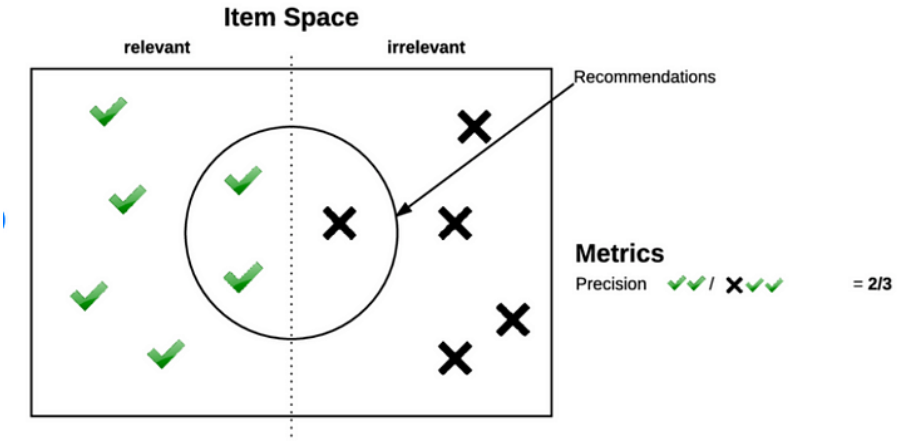

FIGURE 1.5 – La précision [25].

b. **Rappel**

Le rappel correspond au nombre d'éléments pertinents qui ont été recommandés. Supposons donc qu'il y ait 6 éléments pertinents parmi lesquels le recommandateur sélectionne 2 éléments pertinents, puis le rappel sera de 33%.

la figure 1.6 ilustre ce processus.

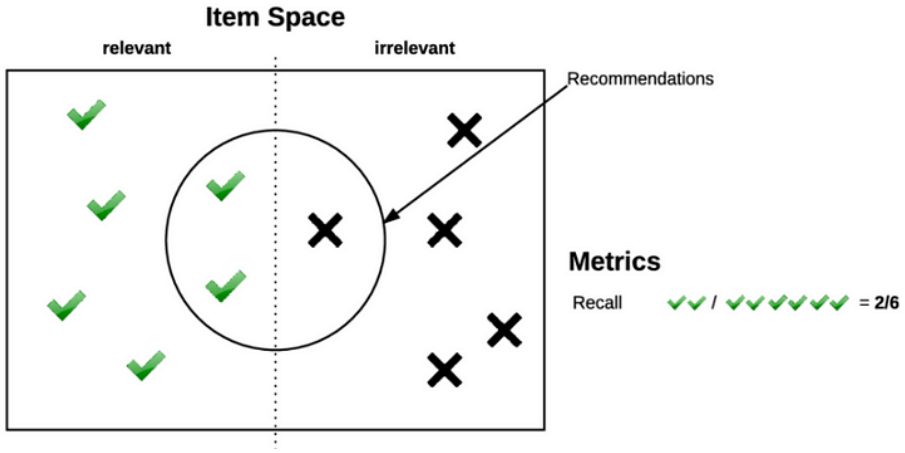

FIGURE 1.6 – Le Rappel [25].

1.11.2.3 Méthodes basées sur le classement méthodes

Les méthodes d'évaluation basées sur le classement nous aident à comprendre comment les éléments suggérés sont ordonnés en fonction de leur pertinence pour les utilisateurs. Ces méthodes nous permettent de mesurer la qualité du classement des articles [25].

a. **Normalized Discounted Cumulative Gain (nDCG)**

nDCG est une mesure qui évalue la qualité du classement des résultats d'un système de recommandation ou d'un moteur de recherche. Elle repose sur l'idée que les éléments les plus pertinents doivent apparaître en haut de la liste des recommandations.

- **Cumulative Gain (CG)**

  il traite du fait que les éléments les plus pertinents sont plus utiles que les éléments quelque peu pertinents qui sont plus utiles que les éléments non pertinents. Il additionne les éléments en fonction de sa pertinence, d'où le terme cumulatif. les figure 1.7 et figure 1.8 ilustre les méthodes de calcul.

$$\mathrm{CG_p} = \sum_{i=1}^{p} rel_i$$

FIGURE 1.7 – Cumulative Gain (CG) [25].

Items Ranking	Relevancy Score
Movie 1	1
Movie 3	2
Movie 2	2
Movie 5	0
Movie 4	1
CG =	6

FIGURE 1.8 – Résultat pour le gain cumulatif [25].

- **Le gain cumulatif actualisé (DCG)**

  DCG améliore le CG en pénalisant les éléments pertinents placés trop bas dans la liste.divisons le score de pertinence des éléments par le journal de leur classement dans la liste. les figure 1.9 et figure 1.10 ilustre les méthodes de calcul.

$$\mathrm{DCG_p} = \sum_{i=1}^{p} \frac{rel_i}{\log_2(i+1)}$$

FIGURE 1.9 – Discounted Cumulative Gain (DCG) [25].

Items Ranking	Relevancy Score
Movie 1	1
Movie 3	2
Movie 2	2
Movie 5	0
Movie 4	1
CG =	6
DCG =	12.1

FIGURE 1.10 – Résultat pour le gain cumulatif actualisé [25].

- **Normalized DCG (nDCG)** nDCG a normalisé les valeurs DCG des différents nombres de listes d'éléments. Pour ce faire, nous trions la liste d'articles par pertinence et calculons le DCG pour cette liste. Ce sera le score DCG parfait car les éléments sont triés en fonction de leur score de pertinence. Nous divisons tous les scores DCG de toute la liste que nous obtenons par ce DCG parfait pour obtenir le score normalisé pour cette liste.les figure 1.11 et figure 1.12 ilustre les méthodes de calcul.

$$\mathrm{nDCG_p} = \frac{DCG_p}{IDCG_p},$$

FIGURE 1.11 – nDCG [25].

Items Ranking	Relevancy Score
Movie 1	1
Movie 3	2
Movie 2	2
Movie 5	0
Movie 4	1
CG =	6
DCG =	12.1

Perfect Ranking	Relevancy Score
Movie 3	2
Movie 2	2
Movie 1	1
Movie 4	1
Movie 5	0
CG (p) =	6
DCG (p) =	13.9

nDCG = DCG/DCG (P) =	0.87

FIGURE 1.12 – Résultat pour Normalized DCG [25].

1.11.2.4 Autres méthodes

- **Nouveauté**

  La nouveauté permet de mesurer la capacité d'un système de recommandation à surprendre l'utilisateur, en suggérant des éléments qu'il ne connaît pas encore. Cela est particulièrement pertinent lors de la phase d'exploration du site, où les utilisateurs sont plus ouverts à de nouvelles découvertes. À l'inverse, dans des contextes comme la phase de paiement, les utilisateurs recherchent plutôt des articles similaires à ceux qu'ils viennent d'acheter ; dans ce cas, la nouveauté est moins utile [26].

- **Diversité**

  La diversité mesure à quel point les recommandations proposées couvrent des catégories ou des types d'articles variés. Elle est étroitement liée à la nouveauté, mais se concentre davantage sur l'hétérogénéité des suggestions fournies à un utilisateur [26].

  Un haut niveau de diversité permet aux utilisateurs d'être exposés à une plus grande variété de contenus, ce qui peut améliorer l'engagement, en particulier dans des domaines où l'on souhaite encourager la découverte (par exemple, les plateformes de streaming ou les boutiques en ligne proposant de larges catalogues) [26].

- **Couverture**

  La couverture fait référence à la proportion d'utilisateurs et d'articles pour lesquels le système peut générer des recommandations pertinentes. Un système ayant une couverture limitée peut négliger une grande partie du catalogue ou des utilisateurs, souvent en raison de données insuffisantes (ex. : articles peu évalués ou nouveaux utilisateurs).

  Des paramètres comme la taille du voisinage utilisé dans les algorithmes de filtrage collaboratif peuvent fortement influencer la couverture [26].

1.12 Conclusion

Ce chapitre a présenté les bases des systèmes de recommandation. Il a défini leurs concepts clés, leur fonctionnement, leurs objectifs et domaines d'application. Il a également présenté leur évolution et exposé les principales méthodes d'évaluation utilisées pour mesurer leur performance. Ces éléments fondamentaux serviront de base pour aborder les chapitres suivants.

Chapitre 2

Sécurité des systèmes de recommandation e-commerce

2.1 Introduction

Dans ce deuxiéme chapitre, nous étudions la sécurité des systèmes de recommandation appliqués à l'e-commerce. Nous commençons par présenter les caractéristiques des recommandations dans ce domaine. Nous définissons ensuite la notion de sécurité, en précisant ses objectifs, les principaux défis qu'elle soulève ainsi que les services qu'elle cherche à garantir. Par la suite, nous abordons la sécurité propre aux systèmes de recommandation en e-commerce, en présentant les approches existantes, leur analyse comparative. Enfin, nous proposons une approche innovante basée sur la confiance, intégrant des techniques avancées d'IA et de sécurité.

2.2 Recommandation en e-commerce

Les systèmes de recommandation dans l'e-commerce se transforment de simples innovations de quelques boutiques en ligne à des outils commerciaux sérieux qui remodèlent l'ensemble du paysage du commerce électronique.

De nombreux géants du commerce électronique utilisent des systèmes de recommandation pour aider leurs clients afin qu'ils puissent acheter des produits avec facilité et efficacité.
Dans ce domaine, on observe que les systèmes de recommandation sont en général plus approfondis. En plus de la possibilité d'afficher des « articles populaires », de plus en plus d'entreprises misent sur des recommandations hautement personnalisées.

Le plus souvent, plusieurs stratégies de recommandation sont prises en compte ; on notamment les **intérêts d'achat**, les **articles populaires** et d'autres facteurs tels que la **disponibilité des produits** et les **changements de prix**.
Concrètement, les produits peuvent être proposés sur la base de quelques analyses prenant en compte : les meilleures ventes sur la boutique en ligne, les **meilleures ventes** dans la catégorie spécifique que l'acheteur consulte, les **données démographiques** des clients, le **comportement du flux de clics (navigation)** ou d'**achats passés**.
Ainsi, ces analyses sont utilisées comme prédiction du comportement d'achat futur du client [26].

2.3 Sécurité

2.3.1 Définition

La sécurité informatique est l'ensemble des mesures, techniques, outils et ressources utilisés pour minimiser les vulnérabilités du système ou, dans la mesure du possible, pour protéger les systèmes contre les menaces accidentelles ou délibérées [28].

2.3.2 Objectifs

L'objectif de la sécurité informatique est d'assurer que les ressources matérielles et/ou logicielles d'un parc informatique sont uniquement utilisées dans le cadre prévu et par des personnes autorisées.

1. **Confidentialité**
   Seules les personnes habilitées doivent avoir accès aux données. Toute interception ne doit pas être en mesure d'aboutir, les données doivent être cryptées, seuls les acteurs de la transaction possèdent la clé de compréhension.
2. **Intégrité**
   Il faut garantir à chaque instant que les données qui circulent sont bien celles que l'on croit, qu'il n'y a pas eu d'altération (volontaire ou non) au cours de la communication. L'intégrité des données doit valider l'intégralité des données, leur précision, l'authenticité et la validité.

3. **Disponibilité**

   Il faut s'assurer du bon fonctionnement du système, de l'accès à un service et aux ressources à n'importe quel moment. La disponibilité d'un équipement se mesure en divisant la durée durant laquelle cet équipement est opérationnel par la durée durant laquelle il aurait dû être opérationnel.

4. **Non-répudiation**

   Une transaction ne peut être niée par aucun des correspondants. La non-répudiation de l'origine et de la réception des données prouve que les données ont bien été reçues. Cela se fait par le biais de c

5. **Authentification**

   Elle limite l'accès aux personnes autorisées. Il faut s'assurer de l'identité d'un utilisateur avant l'échange de données [29].

2.3.3 Challenges

Parmi les défis pertinents de la sécurité informatique, on peut citer les suivants :

— **L'utilisation de l'IA** : l'essor rapide de l'intelligence artificielle (IA) ces dernières années a des implications à la fois positives et négatives pour la cybersécurité des entreprises. Les cybercriminels vont faire de plus en plus appel à l'IA générative (une sous-catégorie de l'IA) pour perpétrer des **attaques extrêmement sophistiquées**, exploitant cette technologie à des fins malveillantes. "On pourrait parler d'une tendance « révolutionnaire » qui va transformer en profondeur le monde de la cybersécurité, démultiplier et rendre plus efficientes les attaques « traditionnelles ». L'IA offensive va cibler tant les entreprises que les particuliers" prévient Yannick Chatelain, professeur associé au sein de Grenoble École de Management, chercheur associé à la Chaire Dos [29].

— **Multiplication des attaques Zero-Day** : les cybercriminels intensifient leurs assauts en exploitant des failles de sécurité non corrigées, connues sous le nom d'attaques Zero-Day, mettant à l'épreuve les capacités de défense des systèmes informatiques [30].

— **Évolution des capacités** : les opérateurs de logiciels rançonneurs travaillent constamment à optimiser et à améliorer leurs attaques. L'**introduction du chiffrement intermittent**, des **attaques par violation uniquement** et des **techniques d'évasion avancées** rendent ces attaques plus difficiles à détecter et à arrêter avant que les dommages ne soient causés [31].

— **Violations de données** : les violations de données ont toujours été une préoccupation majeure pour les organisations. L'exposition de données sensibles des clients ou de l'entreprise peut nuire à la réputation d'une marque, réduire sa rentabilité ou entraîner des poursuites judiciaires ou réglementaires. Ces **dernières années**, il est devenu de plus en plus courant que les **violations de données** entraînent des **litiges**, des **amendes** et des **règlements** importants pour les organisations violées [31].

2.3.4 Services visés

- **Systèmes de Recommandation**

  Ces systèmes nécessitent une protection contre les attaques visant à manipuler les recommandations. La génération de recommandations dans les services en ligne dépend de données sensibles à caractère privé collectées auprès des utilisateurs. Les mécanismes traditionnels de protection des données se concentrent sur le contrôle d'accès et la transmission sécurisée [32].

- **Services de Protection des Données**

  La sécurité du stockage des données implique la protection des ressources de stockage des données qu'elles contiennent sur site, dans des centres de données externes et dans le cloud contre les dommages ou destructions accidentels ou délibérés, ainsi que contre les utilisateurs et utilisations non autorisés. En général, une bonne sécurité de stockage des données minimise le risque qu'une organisation subisse un vol de données, une divulgation non autorisée de données, une falsification de données, une corruption ou une destruction accidentelle, et cherche à garantir la responsabilité et l'authenticité des données ainsi que la conformité réglementaire et légale [33].

- **Réseaux et Communications**

  Un système d'information doit être sécurisé vis-à-vis des attaques extérieures. Un premier niveau de protection doit être assuré par des dispositifs de sécurité logique spécifiques tels que des routeurs filtrants (ACL), pare-feu, sonde anti intrusions, etc. Une protection fiable contre les virus et logiciels espions,les connexions entre les sites doivent s'effectuer de manière sécurisée, par l'intermédiaire des liaisons privées ou des canaux sécurisés par technique de « tunneling » ou VPN (réseau privé virtuel) [34].

- **Services d'Authentification et d'Accès**

  Les systèmes d'authentification, comme les annuaires Active Directory, doivent être sécurisés pour empêcher les accès non autorisés et garantir que seuls les utilisateurs autorisés puissent accéder aux ressources [35].

- **Services de Sécurité Spécialisés** Les services de détection d'intrusion, les systèmes de gestion des événements de sécurité (SIEM), et les services de supervision de sécurité (SOC) sont essentiels pour détecter et répondre aux menaces en temps réel [35].

2.4 Sécurité des systèmes de recommandation e-commerce

2.4.1 Présentation des approches existantes

De nombreux travaux ont abordé la problématique de la sécurité des systèmes de recommandation dans le domaine d'e-commerce. Nous en présentons ici quelques-unes.

- **Approche 1 : Détection des attaques « shilling » en utilisant les évaluations des utilisateurs**

  Les attaques de type « shilling » représentent une source importante de biais dans les systèmes de recommandation (SR). Le terme "shilling" vient de l'anglais et fait référence à une pratique où des individus ou des robots imitent les comportements d'utilisateurs légitimes en utilisant des comptes d'utilisateurs trompeurs. Ils manipulent ensuite les mécanismes de retour d'information des SR par des actions telles que la promotion de certains produits (attaques "push") ou la dévalorisation d'autres produits (attaques "nuke") [36].

  La recherche effectuée par **Chirita et al.** [37] sur la prévention des attaques de shilling dans les systèmes de recommandation en ligne a mené à l'identification de trois paramètres pour la détection de ces attaques utilisant **les évaluations des utilisateurs**.

  — La première mesure, connue sous le nom de « nombre de différences de prédiction », implique une comparaison des prédictions du système avant et après le retrait d'un utilisateur spécifique. Des disparités importantes dans les résultats des prédictions avant et après le retrait d'un utilisateur indiquent un biais dans le RS en faveur de cet utilisateur particulier.

  — La deuxième mesure, appelée « écart-type des évaluations des utilisateurs », se concentre sur la variabilité des évaluations fournies par un utilisateur. Cette mesure permet d'identifier des schémas suspects dans le comportement des utilisateurs, qui peuvent être révélateurs d'attaques de type « shilling ». Cette mesure évalue la disparité entre les évaluations des utilisateurs et les évaluations moyennes fournies par le système. Pour calculer ces écarts, on utilise les déviations standard des évaluations des utilisateurs. Ces valeurs d'écart standard permettent d'évaluer l'impact des biais individuels des utilisateurs.

  — Une autre statistique importante pour la détection des biais dans SR est le « degré

d'accord avec les autres utilisateurs », qui représente la moyenne de ces écarts types [36].

- **Approche 2 : Modèle de filtrage basé sur la confiance en trois dimensions**

  Ce modèle vise à détecter le bruit et les attaques dans les SR en calculant trois facteurs principaux :

  - **Importance (ζ) :** mesure le degré de conformité entre la valeur de notation et la tendance générale à noter ce même élément dans toutes les interactions provenant d'autres utilisateurs. Ce facteur met l'accent sur l'aspect social de la confiance, et a été intégré au modèle pour refléter le fait, observé dans le monde réel, que les évaluations très éloignées de la tendance générale historique pour un élément donné ne devraient pas influencer fortement le comportement de notation de l'algorithme de recommandation.
  - **Fréquence (γ) :** détermine à quelle fréquence un utilisateur participe aux activités de la communauté.

    Ce facteur englobe implicitement à la fois la longévité et les rôles d'interaction d'un utilisateur dans un SR. Cet élément du modèle de confiance formalisé cible le comportement de l'utilisateur qui a soumis la notation.
  - **Qualité (λ) :** constitue également un autre composant du modèle de confiance proposé. Elle évalue le degré d'excellence du comportement passé d'un utilisateur et de ses interactions, en lien avec le SR actuel [38].

- **Approche 3 : Développement d'une nouvelle attaque de shilling (Obscure attack)**

  Bien que les attaques de type **shilling** existantes puissent modifier dans une certaine mesure les plus proches voisins des utilisateurs cibles, elles ne réussissent pas dans tous les cas. l'attaque obfusquée, proposée par **Chad et al**, semble offrir les meilleures performances, bien que celles-ci ne soient pas optimales.

  — Si un article pertinent, répondant aux critères de qualité pour figurer dans les éléments **Top-N** recommandés, est mal noté ou pas noté du tout au départ par les utilisateurs, il sera progressivement retiré de la liste des recommandations.

  Dans ce cas, des **produits de qualité ou de niche** peuvent devenir invisibles aux consommateurs, car ils n'ont jamais été évalués ou n'ont pas eu la possibilité d'accéder à **la liste Top-N**.

— Un nouveau modèle d'attaque shilling peut être utilisé pour atténuer ce problème, également connu sous le nom de **problème de longue traîne** (long-tail problem) [39].

- **Approche 4 : TSeMCCF (Trust-Semantic enhanced Multi-Criteria Collaborative Filtering)**

  Cette approche exploite à la fois les relations de confiance et les évaluations multi-critères des utilisateurs, ainsi que les relations sémantiques entre les articles dans le cadre du filtrage collaboratif (FC). L'approche TSeMCCF utilise ces informations pour atténuer l'impact de la sparsité des données, du démarrage à froid des utilisateurs et des articles lorsqu'il n'y a pas assez d'informations disponibles.

  Elle combine deux méthodes :

  - **le FC basé sur la confiance implicite multi-critères entre utilisateurs :** cette méthode repose sur les caractéristiques inhérentes à la confiance et à sa propagation pour traiter les problèmes de sparsité et de nouveaux utilisateurs
  - **le FC sémantique basé sur les articles :** cette méthode exploite les relations sémantiques entre articles afin de réduire l'impact de la sparsité et du problème des nouveaux articles.

  Alors l'approche propose une solution intégrée qui renforce la sécurité et la fiabilité des systèmes de recommandation en e-Commerce, en s'appuyant sur la confiance, la richesse des données multi-critères, et la compréhension sémantique des items [40].

- **Approche 5 : Recommandation résistante aux attaques shilling basée sur les réseaux sociaux et la détection de communautés**

  Cette méthode proposée comprend quatre étapes : l'injection d'attaques de type « shilling », la création de réseaux sociaux d'utilisateurs, la détection de communautés et la recommandation d'utilisateurs.

  Lors de la première étape, de faux profils sont injectés dans le système afin de simuler une attaque de type « shilling ». Dans un deuxième temps, les réseaux sociaux d'utilisateurs et d'articles sont constitués sur la base des informations relatives à l'heure d'évaluation, de la matrice d'évaluation des articles par les utilisateurs, de la confiance entre les utilisateurs et du contexte des articles. Au troisième stade, les communautés dans les réseaux sociaux, les utilisateurs et les articles sont détectées en gérant le flux de données incrémentiel au fil du temps. Enfin, à la quatrième étape, en détectant et en ignorant les faux profils dans le voisinage des utilisateurs, l'un des N meilleurs articles est recommandé à l'utilisateur cible [41].

  La FIGURE 2.1 ci-dessous résume les étapes de cette méthode proposée.

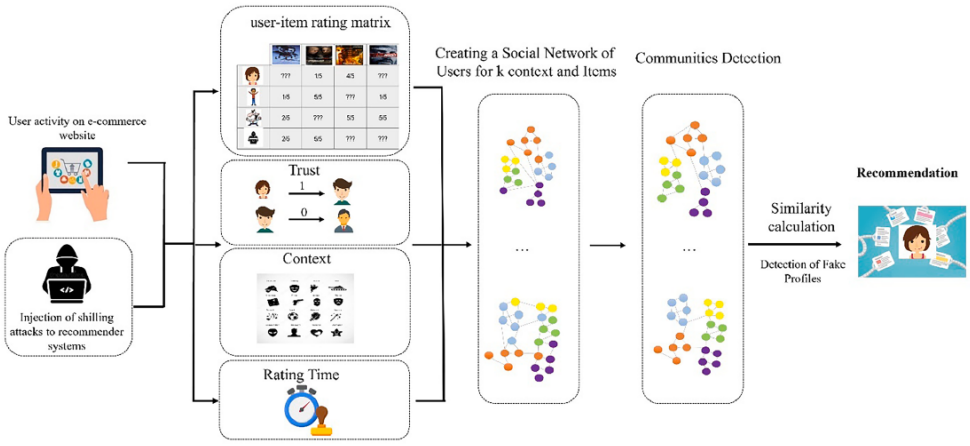

FIGURE 2.1 – Illustration des étapes de la méthode de recommandation proposée [41].

2.4.2 Analyse comparative des approches existantes

Le tableau 2.1 présente une synthèse des différentes approches proposées. Chaque approche est analysée en fonction du problème qu'elle cible, de la solution qu'elle apporte, ainsi que de ses points forts et limites

TABLE 2.1 – Tableau comparatif des approches présentées

Approche	**Problème ciblé**	**Solution proposée**	**Avantages**	**Limites**
Approche 1	Attaques shilling dans les évaluations utilisateurs	Utilisation de trois paramètres : différences de prédiction, écart-type des évaluations, degré d'accord avec les autres utilisateurs	Assure l'intégrité des SR en identifiant les comportements manipulateurs	Inefficace en cas de démarrage à froid ; ne prend pas en compte la fiabilité des sources (confiance absente)
Approche 2	Bruit et attaques dans les SR	Calcul dynamique basé sur l'importance , la fréquence et la qualité des utilisateurs	Détection en ligne possible ; adaptable aux utilisateurs	Requiert un historique suffisant d'interactions pour être fiable
Approche 3	Faiblesses dans les longues traînes	Simulation d'attaques masquées sur des articles peu notés	Révèle les vulnérabilités structurelles des SR ; outil d'évaluation utile	Ce n'est pas une méthode défensive mais un outil de test
Approche 4	Sparsité, démarrage à froid, absence de sémantique	Filtrage collaboratif enrichi par la confiance implicite, les critères multiples et les relations sémantiques	Améliore la précision ; réduit la sparsité et le démarrage à froid	Dépendance à des données bien structurées (confiance, multi-critères, sémantique)
Approche 5	Attaques shilling via faux profils sociaux	Détection de communautés dans les graphes sociaux et exclusion des faux profils	Résistant aux attaques groupées ; exploite les métadonnées temporelles et contextuelles	Nécessite des données sociales riches ; ne traite pas les aspects démographiques ou relationnels

2.4.3 Description de l'approche proposée

Face aux limites identifiées dans les approches traditionnelles, notamment leur incapacité à s'adapter aux attaques évolutives, leur forte dépendance à des données complexes et leur manque de robustesse face aux menaces sophistiquées, nous proposons une approche **cognitive hybride**, **dynamique** et **explicable**. Cette approche a pour objectif de renforcer la sécurité des systèmes de recommandation e-commerce contre les attaques de type shilling, tout en assurant la fiabilité des recommandations et la traçabilité sécurisée des interactions.

Notre contribution repose sur la conception d'un **pipeline de défense multi-couches**, articulé autour de quatre axes majeurs :

— Apprentissage actif des schémas d'attaque grâce à des simulations réalistes et progressives.

— Détection efficace et explicable des profils frauduleux en combinant des règles symboliques et des modèles de machine learning avancés.

— Anticipation et neutralisation des stratégies malveillantes via des mécanismes adaptatifs et évolutifs.

— Garantie d'intégrité et de traçabilité des avis utilisateurs grâce à des technologies de sécurisation robustes.

Contrairement aux approches classiques, souvent statiques et peu réactives, notre solution met en œuvre des mécanismes d'apprentissage continus et adaptatifs permettant au système de :

— S'ajuster dynamiquement face à des attaques sophistiquées et répétées.

— Prévenir les comportements frauduleux émergents.

— Fournir des décisions transparentes et explicables, renforçant la confiance des utilisateurs dans le système de recommandation.

Notre approche se distingue par l'intégration synergique de quatre technologies cognitives complémentaires, permettant de construire un système **résilient**, **intelligent** et **auto-évolutif** :

- **Apprentissage par renforcement** : Nous mettons en place une compétition dynamique entre un agent attaquant, simulant des comportements frauduleux, et un agent défenseur, capable d'apprendre à détecter et neutraliser ces menaces. Ce cadre d'auto-apprentissage permet au système d'anticiper les attaques et d'adapter ses seuils de détection en temps réel, même face à des stratégies évolutives.

- **IA neuro-symbolique** : Nous combinons des règles symboliques explicables (par exemple, des seuils de variance et de nombre d'avis) avec des réseaux de neurones multicouches (MLP) pour généraliser la détection à des profils suspects plus complexes. Cette hybridation permet au système de fournir des décisions interprétables, tout en améliorant la précision et la généralisation de la détection des fraudeurs.
- **Théorie des jeux** : Nous modélisons les interactions entre l'attaquant et le défenseur sous la forme d'un jeu répété, dans lequel chaque acteur adapte sa stratégie au fil du temps. Cette approche permet au système de prédire les tactiques de fraude émergentes et de s'adapter de manière proactive, garantissant une réponse efficace même dans un contexte de confrontation évolutive.
- **Blockchain** : Pour assurer la traçabilité et l'intégrité des avis utilisateurs, nous intégrons une structure blockchain simplifiée. Chaque avis est haché et enregistré de manière décentralisée, rendant la falsification des évaluations pratiquement impossible et sécurisant l'historique des interactions.

En combinant ces techniques avancées, le SR devient **auto-apprenant**, **transparent** et **proactif** dans la prévention de la fraude. Il s'adapte constamment aux nouvelles menaces, fournit des explications claires et logiques pour ses décisions et assure l'authenticité des interactions avec l'utilisateur. Cette approche cognitive transforme les SR en écosystèmes **intelligents** et **autodéfensifs** qui préservent l'**équité**, la **sécurité** et la **confiance** dans les plateformes d'e-commerce.

2.5 Conclusion

La sécurité des systèmes de recommandation est cruciale face aux attaques sophistiquées comme le shilling. Les approches actuelles utilisent diverses techniques, mais présentent des limites. Une solution combinant IA cognitive, apprentissage par renforcement, théorie des jeux et blockchain offre un potentiel prometteur pour renforcer la défense et la fiabilité des recommandations e-commerce.

Chapitre 3

Simulation de notre approche cognitive de sécurisation.

3.1 Introduction

Ce dernier chapitre présente la mise en œuvre concrète de notre stratégie de défense contre les menaces et attaques rendant les SR moins fiables et moins sûrs, en particulier les attaques de type «shilling». Nous présentons en premier les outils, langages et l'environnement utilisés pour le développement. Ensuite, nous expliquons l'architecture structurelle et comportementale du système proposé. Nous détaillons par la suite le processus d'implémentation, de la préparation des données issues du dataset Amazon, jusqu'à la mise en place de notre stratégie de défense. Enfin, nous analysons les résultats obtenus pour évaluer les performances et l'efficacité de notre solution.

3.2 Outils et technologies utilisés

Nous présentons dans cette section les différents langages, outils et environnements pour la mise en œuvre de notre simulation.

3.2.1 Langages de développement

Python

Python est un langage de programmation interprété, orienté objet et de haut niveau, doté d'une sémantique dynamique. Ses structures de données intégrées de haut niveau, combinées à un typage et une liaison dynamiques, le rendent particulièrement attractif pour le développement

rapide d'applications, ainsi que pour une utilisation comme langage de script ou de liaison pour connecter des composants existants. Sa syntaxe simple et facile à apprendre favorise la lisibilité et réduit ainsi les coûts de maintenance des programmes. Python prend en charge les modules et les packages, ce qui favorise la modularité des programmes et la réutilisation du code.
L'interpréteur Python et sa vaste bibliothèque standard sont disponibles gratuitement au format source ou binaire pour toutes les principales plateformes et peuvent être distribués librement [42].
La version Python utilisée dans le cadre de notre étude est la **3.11.9**.

3.2.2 Environnement de développement

Visual Studio Code (VSCode)

Visual Studio Code (VS Code) est un éditeur de code open-source développé par Microsoft, compatible avec Windows, Linux et macOS. Il prend en charge plusieurs langages de programmation (Python, C++, Java...) et propose des fonctionnalités telles que la coloration syntaxique, l'auto-complétion, la gestion Git intégrée et l'extension via des plugins. La version utilisée dans ce travail est la **1.101.2 (x64)**, sous Windows 10 (64 bits) [43].

3.2.3 Bibliothèques

Scikit-learn (sklearn)

Sklearn est une bibliothèque Python dédiée au machine learning. Elle a été créée en 2007 par l'ingénieur David Cournapeau. Son interface simple et unifiée permet d'accéder à de nombreux algorithmes de classification, de régression, de clustering, de réduction de dimensionnalité et de sélection de modèles. Scikit-Learn est basée sur les bibliothèques NumPy, SciPy et matplotlib, qui sont des outils indispensables pour la manipulation de données et la visualisation en Python [44]. La version utilisée est la **1.6.1**.

Pandas

Pandas est une bibliothèque Python utilisée pour travailler avec des ensembles de données. Il dispose de fonctions d'analyse, de nettoyage, d'exploration et de manipulation des données. Le nom « Pandas » fait référence à la fois à « Panel Data » et à « Python Data Analysis » et a été créé par Wes McKinney en 2008 [45].
La version utilisée est la **2.2.3**.

NumPy

NumPy est le package fondamental pour le calcul scientifique en Python. Cette bibliothèque Python fournit un objet tableau multidimensionnel, divers objets dérivés (tels que des tableaux masqués et des matrices), ainsi qu'un ensemble de routines pour des opérations rapides sur les tableaux, notamment mathématiques, logiques, de manipulation de formes, de tri, de sélection, d'E/S, de transformées de Fourier discrètes, d'algèbre linéaire de base, d'opérations statistiques de base, de simulation aléatoire et bien plus encore. NumPy est donc une dépendance de Pandas [46]. La version utilisée est la **1.24.4**.

Matplotlib

Matplotlib est une bibliothèque en Python qui permet aux utilisateurs de générer des visualisations telles que des histogrammes, des nuages de points, des graphiques à barres, des graphiques à secteurs et bien plus encore [47].

La version utilisée est la **3.10.1**.

Seaborn

Seaborn est une bibliothèque de visualisation basée sur Matplotlib. Elle fournit des visualisations de données généralement plus esthétiques et statistiquement plus sophistiquées [47]. La version utilisée est la **0.13.2**

Surprise

Surprise est un scikit Python permettant de créer et d'analyser des systèmes de recommandation qui traitent des données d'évaluation explicites [48]. Cette bibliothèque propose un large éventail d'algorithmes de filtrage collaboratif prêts à l'emploi, ainsi que des outils permettant d'évaluer et de comparer leurs performances [49]. La version utilisée est **surprise-0.1**

3.3 Implémentation de la stratégie de défense

3.3.1 Préparation des Données

Dans notre étude, nous avons utilisé le sous-ensemble du dataset **Amazon Product Data (2018)** [50]. Un **dataset** est un ensemble de données généralement organisées en tableaux ou formats spécifiques, tels que CSV ou JSON, pour faciliter la recherche et l'analyse. Il est essentiel à l'analyse de données, l'apprentissage automatique (ML), l'intelligence artificielle (IA) et à d'autres applications nécessitant des données fiables et accessibles [51].

Ce sous-ensemble de données que nous avons utilisé, appelé "Appliances", regroupe des avis et évaluations d'utilisateurs sur des «**Produits électroménagers** » vendus sur Amazon. Dans cette catégorie de produits, les avis des consommateurs jouent un rôle fondamental dans la décision d'achat car ces produits représentent souvent un investissement important.
Avant d'utiliser les données, nous avons nettoyé le jeu de données pour le rendre prêt à l'analyse et garantir sa fiabilité : nous avons supprimé les enregistrements contenant des valeurs manquantes ou incomplètes, par exemple,les avis sans identifiant d'utilisateur, sans identifiant de produit, ou sans date.
Puis, nous avons procédé à la simulation d'une attaque shilling en en injectant des faux profils frauduleux (10 faux utilisateurs) dans le dataset original. Ces faux utilisateurs attribuent la note 5 sur 5 à des produits ciblé afin de les promouvoir artificiellement (Push Attack).

3.3.2 Présentation du processus de défense

Afin de repérer et d'atténuer les attaques de type shilling infiltrées dans le système de recommandation e-commerce, nous avons conçu une pipeline de défense cognitive. Ce pipeline implique un ensemble de méthodes de détection avancées qui s'inspirent de l'intelligence artificielle hybride et des approches défensives adaptatives.

- **Étape 01 : Apprentissage par renforcement (RL)**

  Dans cette première étape, nous avons simulé un **agent attaquant** construit à partir d'un modèle de Q-learning et qui interagit avec un environnement simulé (ShillingEnv) dans lequel il apprend à sélectionner les actions (évaluations d'items) les plus stratégiques pour promouvoir un item cible, comme dans une attaque shilling de type push. Au fur et à mesure que cet agent s'entraîne pendant plusieurs épisodes, il apprend à concentrer ses attaques sur des items spécifiques, rendant ainsi l'attaque plus subtile et difficile à détecter par les méthodes de défense classiques. En parallèle, un **agent de défense**, également fondé sur l'apprentissage par renforcement, est mis en place. Son rôle est d'ajuster dynamiquement un seuil de détection afin d'identifier les produits suspects tout en limitant les faux positifs. En interagissant avec un environnement de détection simulé (DefenderEnv), il apprend, au fil des épisodes, à reconnaître les schémas d'attaque les plus probables et à adapter sa stratégie de filtrage en conséquence.

  Cette étape permet un apprentissage actif du système, renforçant sa capacité à anticiper et

à prévenir les comportements malveillants.

- **Étape 02 : Détection neuro-symbolique**

  Ensuite, nous sommes passées à l'application des **règles symboliques** explicites afin d'identifier des utilisateurs suspects. Quatre critères ont été utilisés :une moyenne de notes égale à 5, une variance très faible (inférieure à 0,1), un nombre d'avis supérieur ou égal à 5, et des avis tous postés le même jour. Les utilisateurs répondant à toutes ces critères ont été marqués comme suspects selon la logique symbolique.

  Nous avons entraîné un **réseau de neurones** (classifieur MLP – Multi-Layer Perceptron) pour apprendre à reconnaître des profils similaires, en utilisant comme variables d'entrée : la moyenne des notes attribuées, la variance des notes, le nombre total d'avis et le nombre de dates distinctes. Ce modèle a été entraîné à partir des résultats de la partie symbolique, lui permettant ainsi de généraliser la détection à d'autres cas moins évidents.

  Enfin, les utilisateurs détectés par le classifieur MLP ont été fusionnés avec ceux identifiés par les règles symboliques, afin d'obtenir un ensemble final de profils suspects.

- **Étape 03 : Analyse par la théorie des jeux**

  Puis, avons appliqué des concepts de théorie des jeux en considérant deux acteurs : **un attaquant** et un **défenseur**. L'attaquant simule un comportement frauduleux en attribuant systématiquement des notes maximales (5 sur 5) avec une faible variance des notes et une seule date d'évaluation, dans le but d'influencer les évaluations. Le défenseur ajuste dynamiquement ses seuils de détection pour identifier ce comportement suspect.

  Le modèle simule un jeu répété dans lequel chaque itération correspond à une interaction utilisateur-système. Lorsqu'un utilisateur dont la moyenne est égale à 5.0 et la variance inférieure à un seuil (0.1) et une seule date de notation, il est considéré comme hautement suspect et si ce profil n'avait pas encore été détecté par les méthodes précédentes, le défenseur l'ajoute à la liste des utilisateurs suspects.

  L'attaquant gagne (+1 point) s'il réussit à agir sans être détecté et le défenseur gagne (+1 point) s'il parvient à détecter un nouvel attaquant.

  Cette étape permet de simuler l'évolution stratégique les deux acteurs tout en renforçant le système de détection en identifiant de nouveaux profils frauduleux potentiels.

- **Étape 04 : Suppression des utilisateurs suspects**

  Enfin, nous avons appliqué un filtrage qui retire les utilisateurs identifiés comme suspects du dataset final, obtenant ainsi une base de données **de confiance** adaptée à la génération de recommandations fiables.

- **Étape 05 : Simulation Blockchain**

  Aprés la défense, nous avons simulé un dispositif de type **blockchain**. Chaque avis utilisateur a été haché à l'aide d'une fonction cryptographique (SHA-256), associant l'identifiant de l'utilisateur, l'identifiant du produit, la note attribuée ainsi que la date de l'avis ce qui garantit l'**intégrité** des avis et leur **traçabilité** générant une empreinte numérique unique pour chaque enregistrement, permettant ainsi de détecter toute altération des avis dans le futur.

Nous résumons dans la figure 3.1 suivante les étapes détaillées précédemment de la stratégie de défense mise en place.

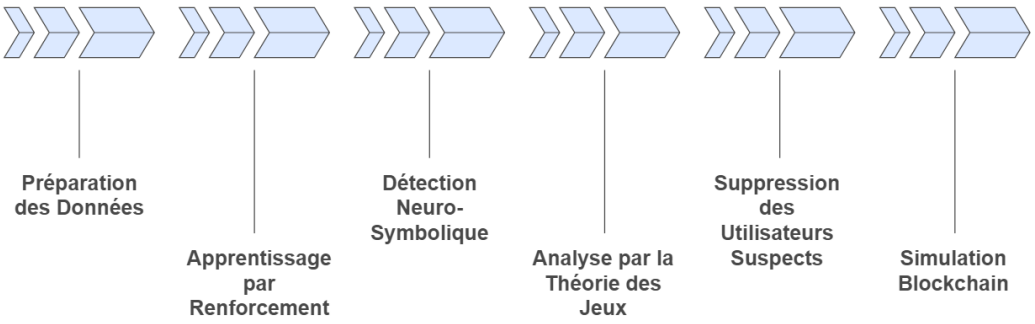

FIGURE 3.1 – Schéma récapitulatif de la stratégie de défense contre les attaques shilling

3.4 Évaluation et analyse des résultats

3.4.1 Les mesures d'évaluation des modèles

Pour évaluer les performances du système de détection (Random Forest, SVM, Gradient Boosting, Régression Logistique) nous avons utilisé un ensemble de métriques classiques :

- **Accuracy (Précision globale) :** Pourcentage de prédictions correctes sur l'ensemble des instances il est donné par :

$$Pr = \frac{TP + TN}{TP + TN + FP + FN}$$

ou TP, TN, FP, et FN représentent respectivement les vrais positifs, vrais négatifs, faux positifs, et faux négatifs.

- **Précision (Pr) :** le pourcentage de prédictions correctes parmi tous les exemples prédictés comme attaque, il est donné par :

$$Pr = \frac{TP}{TP + FP}$$

- **Rappel (Recall) :** le taux de vrais positifs (TVP), ou la proportion d'attaquants correctement identifiés est un modèle parfait hypothétique ne comporterait aucun faux négatif et aurait donc un rappel (TPR) de 1, 0, c'est-à-dire un taux de détection de 100 %. il est définie par :

$$TVP = \frac{TP}{TP + FN}$$

- **F1-Score(F1) :** la moyenne harmonique pondérée de précision et de rappel (Recall), il est donné par :

$$F_1 = \frac{2 \times (Pr \times TVP)}{Pr + TVP}$$

- **ROC-AUC (Area Under Curve) :** L'aire sous la courbe ROC (AUC) elle constitue une mesure de la capacité du modèle à distinguer les classes (utilisateurs légitimes vs attaques) à travers différents seuils de classification. L'AUC varie de 0 à 1, une valeur de 1 indiquant une discrimination parfaite.

- **RMSE (Root Mean Squared Error) :** Erreur quadratique moyenne entre les prédictions et les valeurs réelles, utilisée pour évaluer la qualité des recommandations avant et après attaque, définie par :

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

où y_i représente la valeur réelle, $\hat{y}_i$ la valeur prédite, et n le nombre total de prédictions. Ces métriques ont été calculées à l'aide de **Scikit-learn** et visualisées via des **matrices de confusion**.

- **Matrice de confusion :** la matrice de confusion est une matrice qui mesure la qualité d'un système de classification, elle montre rapidement si un système de classification parvient à classifier correctement.

3.4.2 Résultats obtenus

Dans cette section, nous présentons les résultats obtenus tout au long de notre expérimentation, en examinant l'impact de l'attaque shilling et l'efficacité de la défense cognitive sur le système de recommandation :

- **Impact de l'attaque shilling**

  L'attaque a été simulée par l'injection de 10 faux profils via une **attaque push**. Le **RMSE**, qui mesure la qualité des prédictions, est passé de **0.2480** avant l'attaque à **0.2258** après l'attaque shilling, soit une réduction de 8.95 %. Cette diminution, bien qu'en apparence bénéfique, est en réalité trompeuse. Elle reflète l'impact de profils artificiels soigneusement construits pour tromper l'algorithme. En injectant des données cohérentes mais manipulées, le système devient plus précis en apparence, mais sur une base faussée. Cela introduit un biais dans les recommandations, favorisant des items ciblés sans justification réelle et compromettant la fiabilité du système.

TABLE 3.1 – Impact de l'attaque shilling

Métrique	**Sans attaque**	**Avec attaque**
RMSE	0.2480	0.2258
Valeur de confiance	/	0.4608
Nombre d'utilisateurs	47	57 (47 + 10 faux)

- **Performances des classifieurs**

  Les classifieurs traditionnels (tels que Random Forest, SVM, Gradient Boosting,Régression Logistique.) ont identifié que **3** utilisateurs **suspects** sur les **10 faux** utilisateurs injectés. Ce résultat met en évidence les **limites** de ces méthodes face à des attaques subtiles, où les faux utilisateurs sont conçus pour imiter des comportements légitimes.
  Seuls 3 des 10 attaquants ont été détectés, ce qui montre que des **techniques plus avancées** sont **nécessaires** pour une détection exhaustive. Les métriques sont résumées dans le tableau 3.2 suivant :

TABLE 3.2 – Performances comparées des classifieurs pour la détection des utilisateurs suspects

Modèle	**Accuracy**	**Précision**	**Rappel**	**F1-Score**	**ROC-AUC**
Random Forest	1.00	1.00	1.00	1.00	1.00
SVM	0.894	0.039	1.00	0.076	1.00
Gradient Boosting	1.00	1.00	1.00	1.00	1.00
Régression Logistique	0.889	0.038	1.00	0.073	0.970

  — **Random Forest et Gradient Boosting :** Ces deux modèles ont affiché des performances parfaites, avec une accuracy de 1.0, une précision de 1.0, un rappel de 1.0 et un F1-score de 1.0. Cela signifie qu'ils ont correctement identifié les 3 utilisateurs suspects sans aucune erreur (pas de faux positifs ni de faux négatifs).

  — **SVM et Régression Logistique :** Ces modèles ont obtenu des résultats moins impressionnants. Leur accuracy était plus faible (0.8937 pour SVM et 0.8894 pour Régression Logistique), et bien qu'ils aient un rappel de 1.0 , leur précision était très basse (0.0395 pour SVM et 0.0380 pour Régression Logistique). Cela indique un grand nombre de faux positifs.

Les performances des classifieurs dans ce cas dépendent de leur **capacité** à gérer des relations **complexes** et un déséquilibre marqué entre les classes. **Random Forest** et **Gradient Boosting** ont surpassé les autres grâce à leur approche non linéaire .
En revanche, **SVM** et **Régression Logistique**, plus limités par leur sensibilité au déséquilibre, ont généré beaucoup de FP, ce qui a réduit leur précision.
Pour une analyse plus fiable, nous avons défini un pipeline de défense cognitive structuré.
Les matrices de confusion suivantes illustrent les performances identiques nos quatre classifieurs sur le jeu de données :

1. **Random Forest**

TABLE 3.3 – Matrice de confusion pour Random Forest

	Prédit	
Réel	**Légitime**	**Attaquant**
Légitime	684 (TN)	0 (FP)
Attaquant	0 (FN)	3 (TP)

2. **SVM**

TABLE 3.4 – Matrice de confusion pour SVM

	Prédit	
Réel	**Légitime**	**Attaquant**
Légitime	611 (TN)	73 (FP)
Attaquant	0 (FN)	3 (TP)

3. **Gradient Boosting**

TABLE 3.5 – Matrice de confusion pour Gradient Boosting

	Prédit	
Réel	**Légitime**	**Attaquant**
Légitime	684 (TN)	0 (FP)
Attaquant	0 (FN)	3 (TP)

4. **Régression Logistique**

TABLE 3.6 – Matrice de confusion pour Régression Logistique

Prédit		
Réel	**Légitime**	**Attaquant**
Légitime	608 (TN)	76 (FP)
Attaquant	0 (FN)	3 (TP)

- **Efficacité de la défense cognitive**

  L'approche neuro-symbolique, qui combine des règles symboliques explicites avec des modèles d'apprentissage automatique, a permis de détecter **13** utilisateurs **suspects** (3 par le classifieur **MLP** et 10 par des règles logique).
  Ce nombre dépasse les 10 faux utilisateurs injectés, car cette méthode est conçue pour identifier tout comportement anormal, qu'il provienne des attaquants ou d'utilisateurs légitimes. Ainsi, les 13 utilisateurs détectés incluent :

  - Les 10 faux utilisateurs injectés.
  - 3 utilisateurs légitimes présentant des anomalies (par exemple, des variances de notes inhabituelles ou des schémas de notation suspects).

  La théorie des jeux, appliquée comme une couche supplémentaire de défense, a identifié **10** utilisateurs **suspects** additionnels. Cette méthode simule des interactions stratégiques entre attaquants et défenseurs pour anticiper et neutraliser les **comportements malveillants** qui auraient pu échapper aux approches précédentes. Les 10 utilisateurs détectés ici peuvent correspondre à :

  — Des attaquants subtils non identifiés par les classifieurs ou l'approche neuro-symbolique.
  — Des utilisateurs légitimes avec des comportements anormaux non encore repérés.
  — Éventuellement, des faux positifs dus à la sensibilité de la méthode.

  Le total de **23** utilisateurs **suspects** (**13** via l'approche neuro-symbolique + **10** via la théorie des jeux) dépasse les 10 faux utilisateurs injectés, ce qui est cohérent avec l'objectif de sécurisation globale du système.

La phase finale du pipeline a permis de supprimés 23 profils utilisateurs malveillants avec leur 2089 avis frauduleux. Afin d'assurer la transparence des données restantes, les 198 enregistrements restants considérés comme fiables ont été **hachés** via la **blockchain**, constituant ainsi un registre **fiable** des interactions utilisateurs.
Le **score de confiance moyen** du système a légèrement diminué, passant de **0.4608** à **0.4306**. Cette baisse est attendue, car le système devient plus strict en écartant les profils douteux, renforçant ainsi la fiabilité des données conservées.

En effet, la simulation vise à protéger le système contre toutes les anomalies potentielles, pas seulement contre les attaquants injectés. Ce résultat démontre que l'approche multicouche adoptée permet de :

- Détecter les faux utilisateurs injectés.
- Identifier des utilisateurs légitimes présentant des comportements suspects ou anormaux.
- Anticiper des menaces potentielles grâce à des techniques avancées comme la théorie des jeux.

3.5 Conclusion

Dans ce chapitre, nous avons vu comment déployer une défense interactive pour les systèmes de recommandation afin de détecter les attaques de shilling. Après avoir présenté le contexte et les outils (Python, Scikit-learn, Surprise), nous avons mis en œuvre une solution basée sur un pipeline de détection hybride. Ensuite, nous avons détaillé les étapes clés du projet, à savoir la génération du dataset, l'entrainement des modèles, la détection et l'analyse des résultats. Notre médiation atteint ses objectifs si l'on en croit les métriques présentées dans ce chapitre.
En effet, les algorithmes Random Forest et Gradient Boosting ont détecté tous les faux profils sans se tromper.

Conclusion générale

La sécurité des systèmes de recommandation dans le domaine du e-commerce constitue aujourd'hui un enjeu crucial. Avec la multiplication des attaques, en particulier celles visant à manipuler les résultats de recommandation, la protection de ces systèmes devient une priorité incontournable. Ces menaces affectent non seulement la qualité des recommandations mais compromettent également la confiance des utilisateurs et l'intégrité des plateformes.

Dans ce travail, nous avons proposé une approche cognitive pour renforcer la sécurité des systèmes de recommandation face aux attaques de type shilling, qui consistent à injecter des profils frauduleux dans le but de fausser les prédictions. Après avoir analysé des méthodes existantes et identifié leurs limites, nous avons développé un pipeline de défense multi-couches, capable d'apprendre activement les schémas d'attaque, de détecter efficacement les profils suspects et de s'adapter à des menaces évolutives.

Nous avons testé cette solution sur un jeu de données provenant de la plateforme Amazon (catégorie Appliances), en simulant des attaques de type shilling afin de mesurer leur impact sur les performances du système. Les expérimentations ont montré que notre approche cognitive, intégrant des techniques avancées telles que l'apprentissage par renforcement, l'IA neuro-symbolique, la théorie des jeux et la blockchain, permet de détecter efficacement les profils frauduleux, tout en maintenant une qualité des recommandations pour les utilisateurs légitimes.

Les résultats obtenus sont encourageants et montrent que notre solution contribue à renforcer la sécurité et la confiance dans les systèmes de recommandation. Cependant, nous reconnaissons que les performances sont partiellement influencées par la nature et la structure des jeux de données utilisés.

Perspectives

Dans la continuité de ce travail, des pistes d'amélioration et d'approfondissement peuvent être envisagées :

Validation sur des jeux de données variés et réels : Afin d'évaluer la robustesse et la généricité de notre solution sur différentes plateformes, contextes et types de produits.

Bibliographie

[1] Affde, "Systèmes de recommandation dans le commerce électronique : le moyen le plus rapide d'augmenter les ventes," Affde Marketing, 18 mai 2021. [En ligne]. Disponible sur : https://www.affde.com/fr/recommender-systems-in-ecommerce.html. [Consulté le : 10 mars 2025].

[2] "Global Ecommerce Sales Growth Report," Shopify, Oct. 20, 2024. [En ligne]. [En ligne]. Disponible sur : https://www.shopify.com/blog/global-ecommerce-sales. [Consulté le : 10 mai 2025].

[3] S. Afghoul et E. H. Abdellali, "Système de recommandation basé sur les préférences temporelles," Projet de fin d'études, Université Abdelhamid Ibn Badis – Mostaganem, 2022.

[4] R. Fournier-S'niehotta, "Systèmes de recommandation – 1e partie," CNAM Paris, RCP217, 2020–2021.

[5] Mydid, "Qu'est-ce qu'une communauté et pourquoi sont-elles essentielles ?," Medium. [En ligne]. Disponible sur : https://mydid.medium.com/quest-ce-qu-une-communaut%C3%A9-et-pourquoi-sont-elles-essentielles-dcf64be25a02. [Consulté le : 05 février 2025].

[6] Visiplus Academy, "Communauté en ligne - Définition," Visiplus Academy. [En ligne]. Disponible sur : https://academy.visiplus.com/ressources/definition/communaute-en-ligne. [Consulté le : 05 février 2025].

[7] S. K. Bhatia, J. S. Deogun et V. V. Raghavan, "Profils utilisateurs pour la recherche d'information," in *ISMIS*, vol. 542, Springer, 1991, pp. 102-111.

[8] E. Negre, *Systèmes de recommandation - Introduction*. ISTE, 2015.

[9] Larousse, "Définition de 'Personnalisation'," Larousse, 2025. [En ligne]. Disponible sur : https://www.larousse.fr. [Consulté le : 05 février 2025].

[10] G. Bonnin, *Vers des systèmes de recommandation robustes pour la navigation Web*, Thèse de doctorat, Université Nancy II, 2010. [En ligne]. Disponible sur : https://theses.hal.science/tel-00581331v1. [Consulté le : 05 février 2025].

[11] K. Fokou, "Introduction aux systèmes de recommandation," Smals Research, 28 juin 2022. [En ligne]. Disponible sur : https://www.smalsresearch.be/introduction-aux-systemes-de-recommandation/. [Consulté le : 08 février 2025].

[12] I. Benouaret, *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*, Thèse de doctorat, Université de Technologie de Compiègne, 2017. [En ligne]. Disponible sur : https://theses.hal.science/tel-01767997v1. [Consulté le : 10 février 2025].

[13] M. Touati, *Les systèmes de recommandation des plateformes UGC et PGC*, Mémoire de Master, Sorbonne Université, 2023. [En ligne]. Disponible sur : https://dumas.ccsd.cnrs.fr/dumas-04441655v1. [Consulté le : 10 février 2025].

[14] Fortune, "Le secret des recommandations d'Amazon," 30 juillet 2012. [En ligne]. Disponible sur : http://fortune.com/2012/07/30/amazons-recommendation-secret/. [Consulté le : 12 février 2025].

[15] Netflix Tech Blog, "Recommandations Netflix : au-delà des 5 étoiles," avril 2012. [En ligne]. Disponible sur : http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html. [Consulté le : 12 février 2025].

[16] R. Caballar et C. Stryker, "Qu'est-ce qu'un moteur de recommandation ?," IBM, 19 juin 2024. [En ligne]. Disponible sur : https://www.ibm.com/fr-fr/think/topics/recommendation-engine. [Consulté le : 13 février 2025].

[17] M. Ménard, "Systèmes de recommandation de biens culturels," *Les Cahiers du Numérique*, vol. 10, no. 1, pp. 69–94, 2014. DOI : 10.3166/LCN.10.1.69-94. [Consulté le : 13 février 2025].

[18] B. Bahija et E.-R. Zohayr, "Introduction aux systèmes de recommandation," SlideShare. [En ligne]. Disponible sur : https://fr.slideshare.net/slideshow/introduction-aux-systmes-de-recommandationpptx/267365081. [Consulté le : 18 février 2025].

[19] M. Maatallah, *Une Technique Hybride pour les Systèmes de Recommandation*, Thèse, Université d'Annaba, 2015. [En ligne]. Disponible sur : https://biblio.univ-annaba.dz/wp-content/uploads/2016/11/These-Maatallah-Majda.pdf. [Consulté le : 25 janvier 2025].

[20] Antoine, "Quels sont les principaux sites web qui utilisent des algorithmes de recommandation ?," Prospection Ciblée, 9 septembre 2024. [En ligne]. Disponible sur : https://www.prospection-ciblee.com/quels-sont-les-principaux-sites-web-qui-utilisent-des-algorithmes-de-recommandation/. [Consulté le : 18 février 2025].

[21] Christine, "Les systèmes de recommandation : une catégorisation," Interstices, 27 janvier 2020. [En ligne]. Disponible sur : https://interstices.info/les-systemes-de-recommandation-categorisation/collaborative. [Consulté le : 18 février 2025].

[22] R. Burke, B. Mobasher, R. Zabicki, et R. Bhaumik, "Identifying Attack Models for Secure Recommendation," DePaul University, janvier 2004. [En ligne]. Disponible sur : http://facweb.cs.depaul.edu/mobasher/research/papers/sp-iui05.pdf. [Consulté le : 01 mars 2025].

[23] M. Tadlaoui, *Système de recommandation de ressources pédagogiques*, Thèse, Université de Tlemcen, 2018. [En ligne]. Disponible sur : http://dspace.univ-tlemcen.dz/bitstream/112/13027/1/Systeme-de-recommandation-de-ressources.pdf. [Consulté le : 20 février 2025].

[24] R. Fournier-S'niehotta, "Systèmes de recommandation 2e partie," CNAM Paris, 2021. [En ligne]. Disponible sur : https://cedric.cnam.fr/vertigo/cours/RCP217/docs/RCP217-RecSys2.pdf.

[25] ICHI.PRO, "Une liste exhaustive de méthodes pour évaluer les systèmes de recommandation," 27 novembre 2020. [En ligne]. Disponible sur : https://ichi.pro/fr/une-liste-exhaustive-de-methodes-pour-evaluer-les-systemes-de-recommandation-183670080 [Consulté le : 01 mars 2025].

[26] H. Zarzouni, *La création d'un système de recommandation explicatif basé sur les Tags*, Thèse, Université Badji Mokhtar - Annaba, 2021. [En ligne]. Disponible sur :

https://biblio.univ-annaba.dz/ingeniorat/wp-content/uploads/2022/02/Zarzouni-Hemza.pdf. [Consulté le : 10 mars 2025].

[27] IONOS, "Systèmes de recommandation en e-Commerce," 30 décembre 2020. [En ligne]. Disponible sur : https://www.ionos.fr/digitalguide/web-marketing/vendre-sur-internet/les-systemes-de-recommandation-pour-le-e-commerce/. [Consulté le : 2 mars 2025].

[28] A. Bouadjemi, *Polycopié de cours Sécurité Informatique*, Université de Relizane, 2022-2023.

[29] R. G. Yende, *Support de Cours de Sécurité Informatique et Crypto*, Facultés Africaine Bakhita, 2018. [En ligne]. Disponible sur : https://hal.science/cel-01965300v1. [Consulté le : 3 mars 2025].

[30] R. Charbonnier, "Cybersécurité 2024 : les 11 prédictions et 7 défis," Guardia Cybersecurity School, 20 décembre 2024. [En ligne]. Disponible sur : https://guardia.school/le-lab/cybersecurite-2024-les-11-predictions-et-7-defis.html. [Consulté le : 3 mars 2025].

[31] Check Point Software, "Biggest cyber security challenges in 2024," 2 avril 2024. [En ligne]. Disponible sur : https://www.checkpoint.com/fr/cyber-hub/cyber-security/what-is-cybersecurity/biggest-cyber-security-challenges-in-2024/. [Consulté le : 3 mars 2025].

[32] A. PA et S. M. Santhosh, "A Secure Schema for Recommendation Systems," *International Journal On Cybernetics & Informatics*, vol. 5, no 2, pp. 99-107, avril 2016. [En ligne]. Disponible sur : https://doi.org/10.5121/ijci.2016.5211. [Consulté le : 10 mars 2025].

[33] P. Rubens, "Data Storage Security : Best Practices for Security Teams," eSecurity Planet, 28 janvier 2021. [En ligne]. Disponible sur : https://www.esecurityplanet.com/cloud/data-storage-security-best-practices-for-security-teams/. [Consulté le : 10 mars 2025].

[34] CNIL, "10 conseils pour la sécurité de votre système d'information." [En ligne]. Disponible sur : https://www.cnil.fr/fr/10-conseils-pour-la-securite-de-votre-systeme-dinformation. [Consulté le : 15 mars 2025].

[35] CERT-FR, "Scans et services." [En ligne]. Disponible sur : https://cert.ssi.gouv.fr/scans/. [Consulté le : 15 mars 2025].

[36] O. A. S. Ibrahim et al., "Revisiting recommender systems : an investigative survey," *Neural Computing and Applications*, vol. 37, pp. 2145-2173, 2025. DOI : 10.1007/s00521-024-10828-5. [Consulté le : 20 mars 2025].

[37] P. A. Chirita, W. Nejdl, and C. Zamfir, "Preventing shilling attacks in online recommender systems," in *Proc. 7th ACM Int. Workshop on Web Information and Data Management*, 2005, pp. 67-74. DOI : 10.1145/1097047.1097061. [Consulté le : 25 mars 2025].

[38] E. Bagheri et A. A. Ghorbani, "Exploiting Trust and Suspicion for Real-time Attack Recognition in Recommender Applications," in *Springer eBooks*, 2007, pp. 239-254. DOI : 10.1007/978-0-387-73655-6-16. [Consulté le : 25 mars 2025].

[39] P. K. Singh et al., "Generating A New Shilling Attack for Recommendation Systems," *Computers, Materials & Continua*, vol. 71, no 2, pp. 2827-2846, 2021. DOI : 10.32604/cmc.2022.020437. [Consulté le : 25 mars 2025].

[40] Q. Y. Shambour, N. M. Turab, et O. Y. Adwan, "An Effective e-Commerce Recommender System Based on Trust and Semantic Information," *Cybernetics And Information Technologies*, vol. 21, no 1, pp. 103-118, 2021. DOI : 10.2478/cait-2021-0008. [Consulté le : 30 mars 2025].

[41] H. Hamidi et R. Moradi, "Design of a dynamic and robust recommender system," *Journal Of King Saud University - Computer And Information Sciences*, vol. 36, no 2, 2024. DOI : 10.1016/j.jksuci.2024.101964. [Consulté le : 5 avril 2025].

[42] Python.org, "What is Python ? Executive Summary." [En ligne]. Disponible sur : `https://www.python.org/doc/essays/blurb/`. [Consulté le : 5 mai 2025].

[43] BILITY, "Définition Visual Studio Code." [En ligne]. Disponible sur : `https://bility.fr/definition-visual-studio-code/`. [Consulté le : 5 mai 2025].

[44] IA School, "Qu'est-ce que Scikit-Learn en machine learning ?," 5 février 2024. [En ligne]. Disponible sur : `https://www.intelligence-artificielle-school.com/ecole/technologies/quest-ce-que-scikit-learn-en-machine-learning/`. [Consulté le : 5 mai 2025].

[45] W3Schools.com, "Pandas Introduction." [En ligne]. Disponible sur : `https://www.w3schools.com/python/pandas/pandas_intro.asp`. [Consulté le : 5 mai 2025].

[46] NumPy, "NumPy documentation." [En ligne]. Disponible sur : https://numpy.org/doc/stable/. [Consulté le : 5 mai 2025].

[47] S. Pierre, "Python Data Visualization With Seaborn and Matplotlib," Built In, 16 février 2023. [En ligne]. Disponible sur : https://builtin.com/data-science/data-visualization-tutorial. [Consulté le : 5 mai 2025].

[48] N. Hug, "Home," Surprise. [En ligne]. Disponible sur : https://surpriselib.com/. [Consulté le : 5 mai 2025].

[49] Oscarprietoalvarez, "Exploring the Surprise library in Python," Medium, 21 juin 2023. [En ligne]. Disponible sur : https://medium.com/@oscarprietoalvarez/exploring-the-surprise-library-in-python-uses-and-applications-in-data-science-877fa91 [Consulté le : 5 mai 2025].

[50] J. Ni, "Amazon review data." Disponible sur : https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/. [Consulté le : 27 avril 2025]

[51] Badman et Kosinski, « Dataset » , Qu'est-ce qu'un ensemble de données ? , 23 avril 2025. Disponible sur : https://www.ibm.com/think/topics/dataset. [Consulté le : 5 mai 2025].

Résumé

Avec l'expansion rapide de l'univers numérique et la diversification des services proposés sur le World Wide Web, les systèmes de recommandation sont devenus des outils incontournables pour accompagner les utilisateurs dans la navigation et la sélection de produits au sein de vastes catalogues. Toutefois, ces systèmes demeurent exposés à des menaces sérieuses, notamment les attaques de type shilling, où des utilisateurs malveillants injectent de faux profils dans le but de manipuler les recommandations à leur avantage. L'analyse des approches existantes met en évidence la nécessité de développer des stratégies de défense plus robustes, capables de protéger efficacement les systèmes de recommandation contre ces formes d'attaques sophistiquées.

Dans ce cadre, notre contribution consiste à proposer une approche de **sécurité cognitive**, visant à détecter et à atténuer les attaques de type shilling de manière proactive et adaptative. Les simulations réalisées ont permis de valider la pertinence de notre approche et ont démontré sa robustesse ainsi que sa capacité à renforcer la fiabilité des systèmes de recommandation face aux tentatives de manipulation.

Mots clés : Systèmes de recommandation (SR), E-commerce, Sécurité, Attaques shilling.

Abstract

With the rapid expansion of the digital landscape and the growing diversity of services offered through the World Wide Web, recommendation systems have emerged as indispensable tools for assisting users in navigating and selecting items from extensive product catalogs. Nevertheless, these systems remain susceptible to significant threats, particularly shilling attacks, wherein malicious users inject fabricated profiles with the objective of manipulating recommendation outcomes to their advantage. A critical analysis of existing defense mechanisms underscores the pressing need for more robust and resilient strategies to safeguard recommendation systems against such sophisticated adversarial behaviors. In this context, the present work introduces a cognitive security approach designed to proactively and adaptively detect and mitigate shilling attacks. The experimental simulations conducted validate the effectiveness of the proposed method, demonstrating its robustness and its potential to enhance the reliability and integrity of recommendation systems in the face of manipulative threats.

Keywords : Recommendation systems, E-commerce, Shilling attacks, Security.