**REPUBLIQUE DEMOCRATIQUE DU CONGO**

**« R.D.C »**

**ENSEIGNEMENT SUPERIEUR ET UNIVERSITAIRE**

**«E.S.U»**

**UNIVERSITE DE L’ASSOMPTION AU CONGO**

**« U.A.C »**

**E-mail: contact@uaconline.edu.cd**

Site: www.uaconline.edu.cd

**B.P. :104 BUTEMBO/NORD-KIVU**

**FACULTE DES SCIENCES APPLIQUEES**

**DEPARTEMENT DE GENIE INFORMATIQUE**

**SYSTÈME DE RECOMMANDATION COLLABORATIF POUR OPTIMISER LES VENTES ET LA PROMOTION DES PRODUITS À BUTEMBO**

**Par**

**KAVIRA MBULINYOLO Marguerite**

*Mémoire présenté et défendu en vue de l’obtention du diplôme de licence en Sciences Appliquées*

**Option :** Génie Informatique

**Directeur:**Dr. NSENGE MPIAHéritier, PhD

**ANNEE ACADEMIQUE: 2024-2025**

**Premier Chapitre : INTRODUCTION GÉNÉRALE**

* 1. **Contexte de l'étude**

Dans le monde actuel, marqué par une compétition commerciale croissante et une surabondance de choix pour les consommateurs, les systèmes de recommandation se sont imposés comme des outils incontournables pour améliorer l’expérience utilisateur et optimiser les ventes (Alfaifi, 2024). Ces systèmes, qui reposent sur des algorithmes avancés, permettent de personnaliser les interactions entre les utilisateurs et les plateformes, en leur proposant des produits ou services qui correspondent à leurs préférences ou besoins présumés (Mpia et al., 2023). Les modèles collaboratifs, en particulier, exploitent les données partagées entre utilisateurs pour identifier des tendances et fournir des recommandations précises.

Dans la ville commerciale de Butembo, située dans l’est de la République démocratique du Congo (RDC), le commerce constitue une activité économique majeure. Cette ville se caractérise par un grand nombre de petites et moyennes entreprises (PME) qui offrent une large gamme de produits (Menomavuya, 2018). Cependant, la gestion efficace des stocks et la satisfaction des besoins des clients restent des défis importants. Les clients se retrouvent souvent confrontés à des problèmes de surabondance de choix, tandis que les commerçants peinent à anticiper les préférences des consommateurs (Ngangoue, 2016). Dans un tel contexte, l’introduction d’un système de recommandation basé sur le filtrage collaboratif pourrait transformer les pratiques commerciales locales en améliorant à la fois la satisfaction des clients et la gestion des stocks.

La particularité de la ville de Butembo, en tant que hub commercial, offre une opportunité unique pour la mise en œuvre d’un tel système. Il s’agit non seulement de résoudre des problèmes locaux, mais aussi de créer un modèle pouvant être répliqué dans des contextes similaires à travers l’Afrique et ailleurs. Les systèmes de recommandation ont été largement adoptés dans des environnements numériques à forte densité de données, mais leur potentiel dans des écosystèmes commerciaux locaux reste sous-exploré (Zhang et Chen, 2020).

* 1. **Problématique**

Malgré la popularité et l’efficacité prouvées des systèmes de recommandation dans les économies avancées, leur mise en œuvre dans des contextes locaux comme celui de Butembo pose plusieurs défis. Premièrement, les PME locales génèrent souvent des volumes de données relativement faibles, ce qui limite l’efficacité des approches traditionnelles de filtrage collaboratif, qui s'appuient sur des bases de données riches pour produire des recommandations fiables (Sarwar et al., 2021). Deuxièmement, la gestion des stocks et des produits reste essentiellement manuelle pour la plupart des commerçants de la ville, résultant en des pertes économiques dues à des surstocks ou à des ruptures de stock.

Un autre problème clé est l’absence de personnalisation dans les interactions entre les commerçants et les clients. Actuellement, la majorité des achats sont réalisés sans qu’il y ait une analyse préalable des préférences des clients. Cela limite la fidélisation de la clientèle et la capacité des entreprises à cibler efficacement leurs offres. En conséquence, les commerçants de Butembo ont peu de moyens pour améliorer leur compétitivité face à des marchés extérieurs de plus en plus numérisés.

La question principale que cette étude cherche à résoudre est la suivante : comment concevoir et développer un système de recommandation collaboratif adapté aux spécificités du marché local de Butembo afin d’améliorer la gestion des stocks et la satisfaction des clients ? Pour répondre à cette question, il sera nécessaire d’intégrer des approches collaboratives et des techniques d’analyse de données dans un système qui soit à la fois efficace et simple à utiliser par les commerçants locaux.

* 1. **Objectifs de l’étude**
     1. **Objectif général**

L’objectif global de cette étude a été de développer un système de recommandation collaboratif pour optimiser les ventes et la promotion des produits à Butembo.

* + 1. **Objectifs spécifiques**

De façon spécifique, cette étude a eu pour objectifs de :

1. Analyser les besoins des commerçants et des clients
2. Développer un algorithme de recommandation collaboratif
3. Intégrer des modules de gestion des stocks
4. Evaluer le système
5. Déployer le modèle dans un environnement Web
   1. **Questions de recherche**

Se basant sur les objectifs ci-haut, ces questions de recherche ont guidé cette recherche :

1. Quels sont les besoins spécifiques des commerçants et des clients de Butembo pour améliorer la gestion des stocks et la satisfaction client ?
2. Quel algorithme de recommandation collaboratif est le plus adapté pour fournir des suggestions personnalisées dans le contexte de Butembo ?
3. Comment concevoir un module de gestion des stocks qui s'intègre efficacement au système de recommandation collaboratif ?
4. Quels indicateurs peuvent être utilisés pour évaluer l’efficacité et l’impact du système sur la satisfaction des clients et les performances commerciales ?
5. Comment garantir un déploiement efficace et accessible du système de recommandation dans un environnement Web pour les utilisateurs locaux ?
   1. **Choix et intérêt du sujet**

Le choix de ce sujet s’appuie sur le besoin de résoudre des problèmes réels auxquels font face les commerçants et les clients dans la ville commerciale de Butembo. Les commerçants rencontrent souvent des difficultés dans la gestion des stocks, ce qui entraîne des ruptures ou des surplus inutiles, tandis que les clients éprouvent des difficultés à trouver des produits qui correspondent réellement à leurs attentes. Ce projet vise à combler ces insuffisances en développant une solution technologique innovante : un système de recommandation collaboratif, associé à des outils de gestion des stocks.

D’un point de vue scientifique, ce projet s’intéresse à l’adaptation des technologies de recommandation à un environnement où les données disponibles sont limitées, un aspect peu exploré dans la recherche actuelle. Bien que Butembo soit un pôle commercial important, les outils numériques y sont encore peu utilisés pour améliorer les relations entre commerçants et clients. La mise en place d’un algorithme de recommandation collaboratif capable de s’adapter à ces contraintes spécifiques pourrait apporter une contribution significative à la recherche, tout en tenant compte des particularités des économies locales. Sur le plan pratique, cette initiative aidera les commerçants à mieux prévoir les attentes des clients et à rationaliser la gestion des stocks, réduisant ainsi les pertes et améliorant leur rentabilité. De leur côté, les clients bénéficieront d’une expérience d’achat plus fluide et personnalisée. Le déploiement du système sur une plateforme Web garantira également une large accessibilité et encouragera son adoption.

Ainsi, ce projet se distingue par un double intérêt : il explore des problématiques théoriques liées aux limites des algorithmes collaboratifs dans des environnements spécifiques et propose une solution concrète pour améliorer les pratiques commerciales locales. Il s’agit d’une démarche qui allie recherche académique et impact pratique tangible.

* 1. **Délimitation du sujet**

Ce projet se concentre sur la conception et le développement d’un système de recommandation collaboratif adapté au contexte de la ville commerciale de Butembo. Le but principal est d’améliorer la gestion des stocks des commerçants locaux tout en offrant aux clients des recommandations personnalisées pour les aider à trouver les produits répondant à leurs besoins.

Les principales limites du sujet sont définies comme suit : (i) Zone géographique : L’étude et le développement du système se concentrent uniquement sur la ville de Butembo, en tenant compte des spécificités économiques, culturelles et technologiques locales. (ii) Type de système : Le projet se limite à un système de recommandation collaboratif, exploitant les interactions et évaluations des utilisateurs pour générer des suggestions. Les approches basées sur le contenu ou hybrides ne seront pas prises en compte dans cette phase. (iii) Modules intégrés : Le système inclut des fonctionnalités spécifiques, à savoir la gestion des stocks, la recommandation des produits et le déploiement dans un environnement Web. Les aspects tels que la gestion des ressources humaines ou la logistique des livraisons ne font pas partie du périmètre de cette étude. (iv) Technologies utilisées : Le développement de l’algorithme repose principalement sur l’utilisation de la librairie Surprise et sur d’autres outils compatibles pour le déploiement Web tel que Flask. Les comparaisons avec d'autres bibliothèques ou frameworks ne constituent pas l’objet de ce travail. (v) Évaluation : L’évaluation du système se limite à des indicateurs comme la précision des recommandations, la satisfaction des utilisateurs et l’efficacité de la gestion des stocks. Les impacts à long terme sur l’économie locale ou l’évolution des comportements des consommateurs ne seront pas abordés en profondeur.

En résumé, ce projet se concentre sur un système de recommandation collaboratif conçu pour répondre aux besoins immédiats des commerçants et clients de Butembo, en restant dans un cadre technologique précis et un périmètre géographique restreint.

* 1. **Justification de l’étude**

Le choix de mener cette étude repose sur plusieurs raisons à la fois pratiques, scientifiques et socio-économiques, qui démontrent sa pertinence et son utilité. (i) Problèmes pratiques à résoudre. Dans la ville commerciale de Butembo, les commerçants rencontrent d’importants défis liés à la gestion des stocks, souvent caractérisée par des ruptures ou des surplus. Parallèlement, les clients peinent à trouver des produits adaptés à leurs préférences, ce qui limite leur satisfaction et fidélité. Ce projet propose une solution qui combine un système de recommandation collaboratif et des outils de gestion des stocks, offrant ainsi des avantages tangibles aux commerçants comme aux consommateurs. (ii) Pertinence scientifique  
Les systèmes de recommandation, largement étudiés dans des environnements riches en données, restent peu explorés dans des contextes à faible densité de données, comme celui de Butembo. Ce projet ambitionne d'adapter ces technologies à un tel cadre en tenant compte des spécificités locales. L’utilisation de la librairie *Surprise* pour développer un algorithme collaboratif a permis d’élargir le champ des recherches en s’intéressant à des environnements sous-représentés dans la littérature académique. (iii) Impact socio-économique  
Butembo est un important centre économique où le commerce joue un rôle clé dans la vie quotidienne. Un système de recommandation bien conçu pourra contribuer à améliorer la rentabilité des commerçants, en optimisant leurs processus de gestion des stocks, tout en offrant aux clients une expérience d’achat plus fluide et personnalisée. En outre, l’utilisation de la technologie comme levier pour résoudre des problèmes locaux constitue un pas important vers la modernisation du secteur commercial dans la région. (iv) Avantage technologique et innovation. Le développement et le déploiement d’un tel système dans un environnement Web constituent une approche innovante pour introduire la transformation numérique dans le commerce local. Cette étude ouvre également des perspectives pour l’adoption de solutions similaires dans d’autres contextes régionaux ou internationaux partageant des défis similaires.

* 1. **Subdivision du travail**

Ce travail est structuré en cinq chapitres. Le premier chapitre introduit le cadre général de l’étude, en présentant le contexte, la problématique, les objectifs, les questions de recherche, ainsi que l’intérêt et la justification du sujet. Le deuxième chapitre est consacré à la revue de littérature, où sont explorés les concepts clés, les approches existantes en matière de systèmes de recommandation, et les outils technologiques pertinents. Le troisième chapitre détaille la méthodologie adoptée, en décrivant les techniques, outils, et données utilisés pour concevoir et développer le système. Le quatrième chapitre présente les résultats obtenus, avec une analyse approfondie des performances du système de recommandation et des modules de gestion des stocks, ainsi qu’une évaluation basée sur des indicateurs définis. Enfin, le cinquième chapitre synthétise les conclusions de l’étude, discute les implications pratiques et scientifiques des résultats, et propose des perspectives pour de futures recherches ou améliorations.

**Deuxième Chapitre : REVUE DE LITTÉRATURE**

**2.1. Introduction**

La revue de littérature est une étape essentielle dans toute recherche scientifique, car elle permet de situer le travail dans le contexte des connaissances existantes. Ce chapitre vise à explorer les concepts clés, les approches théoriques et les outils technologiques liés aux systèmes de recommandation, en particulier ceux basés sur le filtrage collaboratif. L'objectif est de fournir une base solide pour la conception et le développement du système de recommandation collaboratif proposé dans cette étude.

En outre, ce chapitre examine les défis et les limites des systèmes de recommandation, ainsi que les technologies disponibles pour leur mise en œuvre. Enfin, une revue des études empiriques permettra de comprendre comment ces systèmes ont été appliqués dans des contextes similaires, notamment dans le commerce local et international.

**2.2. Revue de littérature théorique**

**2.2.1. Systèmes de recommandation : Définition et classification**

Les systèmes de recommandation sont des algorithmes intelligents qui analysent les données comportementales (historique d’achats, notations, consultations) pour prédire et proposer des items pertinents à chaque utilisateur. Selon Zhang & Chen (2020), leur application couvre :

* Le commerce électronique (recommandations produits sur Amazon)
* Les réseaux sociaux (suggestions d’amis sur Facebook)
* Le streaming (sélections personnalisées sur Netflix)

Ils peuvent être classés en trois grandes catégories : les systèmes basés sur le contenu, les systèmes collaboratifs, et les systèmes hybrides (Mpia et al., 2023).

***2.2.1.1. Systèmes basés sur le contenu***

Ces systèmes exploitent les métadonnées des produits (catégorie, prix, marque) pour établir des similarités. Par exemple, si un utilisateur a acheté un smartphone Samsung, le système recommandera d’autres appareils électroniques partageant des caractéristiques techniques similaires (Lops et al., 2021). Cette approche est particulièrement utile lorsque les données sur les interactions entre utilisateurs sont limitées (Zhang & Chen 2020). Cependant, ces systèmes peuvent souffrir d’un manque de diversité dans les recommandations, car ils tendent à proposer des produits très similaires à ceux déjà consommés par l’utilisateur (Alfaifi, 2024).

***2.2.1.2. Systèmes collaboratifs***

Ces systèmes exploitent les interactions entre utilisateurs pour générer des recommandations. Ils peuvent être basés sur les similarités entre utilisateurs (user-user) ou entre produits (item-item). Par exemple, si deux utilisateurs ont noté plusieurs produits de manière similaire, le système peut recommander à l'un des utilisateurs un produit que l'autre a apprécié (Sarwar et al., 2021). Cette approche est particulièrement efficace dans les environnements où les utilisateurs interagissent fréquemment avec les produits, comme les plateformes de commerce électronique.

* + - 1. **Systèmes hybrides**

Ces systèmes combinent les approches basées sur le contenu et collaboratives pour améliorer la précision des recommandations. Par exemple, un système hybride peut utiliser les caractéristiques des produits pour initialiser les recommandations, puis ajuster ces recommandations en fonction des interactions entre utilisateurs (Burke et al., 2021). Les systèmes hybrides sont particulièrement utiles pour surmonter les limitations des systèmes basés sur le contenu ou collaboratifs pris individuellement (Zhang & Chen, 2020).

**2.2.2. Recommandation collaborative : principes et fonctionnement**

La recommandation collaborative repose sur l'idée que les utilisateurs ayant des préférences similaires dans le passé auront tendance à avoir des préférences similaires à l'avenir (Sarwar et al., 2021). Cette approche utilise des matrices de notation (ratings) pour identifier des patterns de comportement parmi les utilisateurs. Par exemple, si deux utilisateurs ont noté plusieurs produits de manière similaire, le système peut recommander à l'un des utilisateurs un produit que l'autre a apprécié. Il existe deux principales méthodes de filtrage collaboratif :

1. Filtrage collaboratif basé sur l’utilisateur(user-user) : cette méthode identifie les utilisateurs similaires et recommande des produits appréciés par ces utilisateurs similaires. Par exemple, si un utilisateur A a noté plusieurs produits de manière similaire à un utilisateur B, le système peut recommander à l'utilisateur A des produits que l'utilisateur B a appréciés (Sarwar et al., 2021).
2. Filtrage collaboratif basé sur le produit(item-item) : cette méthode identifie des produits similaires et recommande des produits qui sont souvent appréciés par les mêmes utilisateurs. Par exemple, si un produit A est souvent acheté avec un produit B, le système peut recommander le produit B à un utilisateur qui a acheté le produit A (Saraei et al., 2023).

**2.2.3. Gestion des stocks : définitions, importance, et relation avec les systèmes de recommandation**

La gestion des stocks est un aspect crucial pour les entreprises, car elle permet de minimiser les coûts liés aux surplus ou aux ruptures de stock (Chopra & Meindl, 2020). Une gestion efficace des stocks repose sur la prévision de la demande, l'optimisation des niveaux de stock, et la réduction des délais de livraison. Les systèmes de recommandation peuvent jouer un rôle clé dans la gestion des stocks en prédisant les préférences des clients et en ajustant les niveaux de stock en conséquence. Par exemple, un système de recommandation peut identifier les produits les plus susceptibles d'être achetés, permettant ainsi aux commerçants de maintenir des niveaux de stock optimaux (Zhang & Chen, 2020).

* + 1. **Techniques et approches des systèmes de recommandation**

Les systèmes de recommandation utilisent diverses techniques pour fournir des suggestions personnalisées aux utilisateurs. Ces approches peuvent être classées en trois catégories principales : les méthodes basées sur le contenu, les méthodes collaboratives et les méthodes hybrides. Chacune de ces approches présente des avantages et des limites, et leur choix dépend du contexte d’application, de la nature des données disponibles et des objectifs spécifiques du système. Dans cette section, nous explorerons en détail ces techniques, en mettant l’accent sur leurs principes fondamentaux, leurs applications pratiques et leurs défis.

* + - 1. **Approches basées sur le contenu**

Les approches basées sur le contenu génèrent des recommandations en analysant les caractéristiques intrinsèques des produits ou des services. Par exemple, si un utilisateur a acheté plusieurs livres d'un même auteur, le système pourra recommander d'autres ouvrages du même auteur ou traitant de sujets similaires (Lops et al., 2021). Cette méthode repose sur l'idée que les préférences passées d'un utilisateur sont un indicateur fiable de ses préférences futures.

**Fonctionnement :**

Le système extrait des métadonnées décrivant les produits (par exemple, catégorie, prix, marque, mots-clés). Il construit un profil utilisateur en analysant les interactions passées (achats, consultations, évaluations). Les recommandations sont générées en comparant le profil utilisateur avec les caractéristiques des produits disponibles.

Cette approche est particulièrement efficace dans les contextes où les données utilisateur sont limitées, car elle ne dépend pas des interactions entre utilisateurs (Zhang & Chen, 2020).

**Avantages :**

* Indépendance des interactions utilisateur : contrairement aux méthodes collaboratives, cette approche ne nécessite pas de données volumineuses sur les interactions entre utilisateurs.
* Évitement du problème de "cold start" pour les nouveaux produits : tant que les métadonnées des produits sont disponibles, le système peut générer des recommandations (Alfaifi, 2024).

**Limites :**

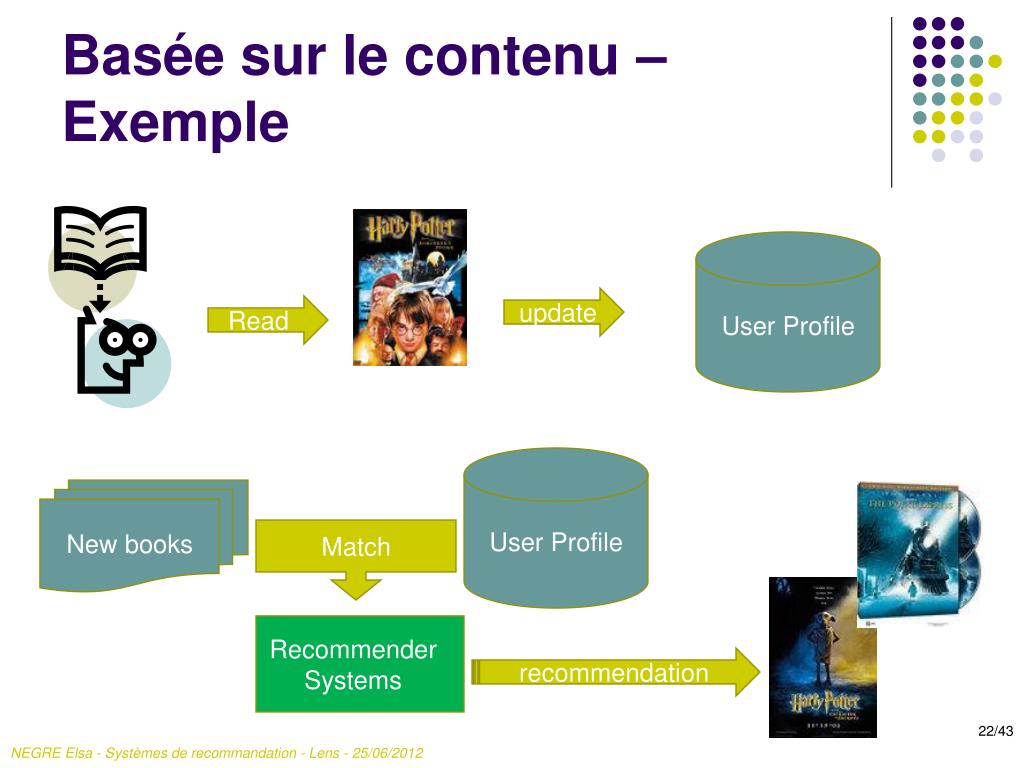
* Risque de surspécialisation : les recommandations peuvent manquer de diversité, se limitant à des produits très similaires à ceux déjà connus de l'utilisateur (Alfaifi, 2024).
* Nécessité des métadonnées riches et bien structurées : la qualité des recommandations dépend fortement de la disponibilité et de la structuration des métadonnées des produits (Lops et al., 2021).

**Applications :**

Cette approche est couramment utilisée dans les plateformes de streaming (recommandation de films ou de musique partageant des genres similaires) ou les librairies en ligne (recommandation de livres par auteur ou thème). Par exemple, Netflix utilise des algorithmes basés sur le contenu pour suggérer des films similaires à ceux que l’utilisateur a déjà regardés (Gomez-Uribe & Hunt, 2020).

**Illustration :**

La figure ci-dessous schématise le fonctionnement d’un système de recommandation basé sur le contenu. Le système analyse les caractéristiques des produits et les compare avec le profil utilisateur pour générer des suggestions personnalisées.



*Figure1 : Fonctionnement d’un système de recommandation basé sur le contenu (source : lops et al., 2021)*

Cette approche, bien que limitée par sa tendance à la surspécialisation, reste une solution robuste pour les environnements où les données collaboratives sont rares. Elle constitue une base essentielle pour les systèmes hybrides, qui combinent ses avantages avec ceux des méthodes collaboratives pour améliorer la diversité et la pertinence des recommandations (Burke et al., 2021).

* + - 1. **Approches collaboratives : utilisateur/utilisateur et produit/produit**

Les approches collaboratives, comme mentionné précédemment, se basent sur les interactions entre utilisateurs pour générer des recommandations. Ces approches sont particulièrement efficaces dans les environnements où les utilisateurs interagissent fréquemment avec les produits, comme les plateformes de commerce électronique (Sarwar et al., 2021).

* + - 1. **Approches hybrides : synthèse des avantages et mécanismes**

Les approches hybrides représentent une solution optimale dans le domaine des systèmes de recommandation en combinant les forces des méthodes basées sur le contenu et collaboratives, tout en atténuant leurs limites respectives. Cette fusion permet de surmonter les limitations inhérentes à chaque méthode prise isolément, offrant ainsi des recommandations plus précises, diversifiées et adaptives. Selon Burke et al. (2021), ces systèmes améliorent significativement la précision, la robustesse des recommandations, notamment dans des contextes où les données sont peu denses ou hétérogènes.

**Fondements théoriques**

1. **Définition :**

Un système hybride intègre simultanément :

* L’analyse des caractéristiques intrinsèques des produits (approche basée sur le contenu)
* Les schémas d’interaction entre utilisateurs (approche collaborative)

1. **Principe de fonctionnement**

Un système hybride intègre généralement deux stratégies principales :

1. **La combinaison parallèle** : les résultats des deux approches sont calculés séparément puis fusionnés. Pour dire que les recommandations issues des approches collaboratives et basée sur le contenu sont générées indépendamment.
2. **L’intégration séquentielle** : cette méthode sert à initialiser les recommandations (ex. contenu pour résoudre le cold start), puis l’autre affine les résultats (ex. filtrage collaboratif pour personnalisation).

*Exemple concret :*

* Netflix utilise une architecture hybride où :
* Les métadonnées des films (genre, acteurs) alimentent un modèle basé sur le contenu.
* Les interactions utilisateurs (historique de visionnage) alimentent un modèle collaboratif.
* Les deux sorties sont combinées pour proposer des suggestions "Parce que vous avez regardé X" et “Tendances similaires” (Gomez-Uribe & Hunt, 2020).

Images ……………………….

**Avantages principaux**

|  |  |  |
| --- | --- | --- |
| **Caractéristiques** | **Avantage** | **Exemple** |
| Réduction du clod start | Les métadonnées des produits(contenu) compense le manque initial des données utilisateurs | Nouveaux produits recommandés grâce à leurs caractéristiques |
| Diversité accrue | Evite la surspécialisation des recommandations purement basées sur le contenu. | Suggestions couvrant différents aspects des préférences utilisateur |
| Robustesse aux données clairsemées | L’approche collaborative compense les lacunes des métadonnées incomplètes (Zhang & Chen, 2020). | Performances stables dans les petites communautés d’utilisateurs |

**Applications concrètes**

1. **Plateformes majeures :**

* Amazon : combine l’historique d’achat (collaboratif) avec les caractéristiques produits (contenu)
* Spotify : utilise à la fois les métadonnées musicales et les écoutes similaires entre utilisateurs

1. Contexte local (Butembo) :

* Adaptation possible pour les PME locales :
* Données produits : catégories, prix, disponibilité
* Données utilisateurs : achats récurrents, paniers typiques
* Résultat : recommandations adaptées au volume limité de données (Menomavuya, 2018).

**Défis et solutions**

Complexité algorithmique : l’intégration de multiples modèles augmente la charge computationnelle. Des frameworks comme TensorFlow Recommenders (TFRS) optimisent cette étape (Adomavicius & Tuzhilin, 2022). Il faut l’utilisation des frameworks spécialisés. Quant à l’optimisation, ici, on fait la sélection des caractéristiques les plus pertinentes.

**Perspectives futures**

* Intégration de l’IA explicable pour plus de transparence
* Adaptation aux contraintes des marchés émergents
* Combinaison avec l’analyse des émotions utilisateurs

Les systèmes hybrides offrent une solution optimale pour les environnements complexes comme Butembo, combinant la robustesse technique et l’adaptabilité contextuelle. Leur capacité à tirer parti des données hétérogènes en fait des outils particulièrement prometteurs pour la transformation numérique du commerce local.

**2.2.5. Défis et limites des systèmes de recomandation**

Dans ce point, nous allons parcourir certains defis que d’autres auteurs avant nous ont rencontré et voir les moyens d’y apporter solution si possible pour etre différent des autres travaux existantes avant nous. C’est pourquoi, nous allons dans notre travail, analyser les problèmes récents et voir les moyens d’y apporter une durable solution et prevenir les futurs chercheurs sur les solutions possibles en cas d’un traitement du même sujet ou problème.

* + - 1. **Faible densité de données (sparse data)**

L'un des principaux défis des systèmes de recommandation est la faible densité des données, c'est-à-dire le manque d'interactions entre les utilisateurs et les produits. Dans de nombreux cas, les utilisateurs ne notent ou n'évaluent qu'un petit nombre de produits, ce qui rend difficile l'identification de patterns de comportement (**Koren et al., 2023**).

* + - 1. **Froid utilisateur ou produit (cold start)**

Le **problème de cold start** (démarrage à froid) est l’un des défis majeurs rencontrés par les systèmes de recommandation, notamment ceux basés sur le filtrage collaboratif. Ce problème survient lorsqu’un nouvel utilisateur ou un nouveau produit est introduit dans le système sans historique d’interactions. Le manque de données empêche alors le système de générer des recommandations personnalisées et pertinentes (Schein et al., 2022 ; Adomavicius & Tuzhilin, 2022).

**Illustration mathématique du problème**

Dans les systèmes de recommandation utilisant la **factorisation matricielle**, la prédiction de la note  attribuée par un utilisateur u à un item  s’exprime généralement ainsi :

où :

* sont les biais de l’utilisateur et de l’item,
* sont les vecteurs de facteurs latents de l’utilisateur et de l’item.

Lorsque  ou  sont nouveaux, leurs paramètres ( ou sont inconnus, ce qui rend la prediction impossible ou peu fiable (Koren, Bell & Volinsky, 2023).

**Types de cold start**

* **Cold start utilisateur :** un nouvel utilisateur n’a pas encore interagi avec le système, ce qui empêche de cerner ses préférences.
* **Cold start produit**: un nouvel item n’a reçu aucune interaction, il est donc difficile
* **Cold start système** : lors du lancement d’une nouvelle plateforme, aucune donnée n’est disponible.

**Stratégies**

Pour surmonter le problème du démarrage à froid, plusieurs approches sont utilisées :

* **Méthodes hybrides :** elles combinent filtrage collaboratif et filtrage basé sur le contenu pour exploiter les attributs des utilisateurs ou des items (Burke et al., 2021).
* **Utilisation des métadonnées :** les informations démographiques, contextuelles ou les descriptions d’items permettent de générer des recommandations initiales (Adomavicius & Tuzhilin, 2022).
* **Apprentissage actif :** le système sollicite explicitement les préférences de l’utilisateur lors de l’inscription (Schein et al., 2022).

**Exemple schématique**

|  |  |  |
| --- | --- | --- |
| **Situation** | **Données disponibles** | **Solution recommandée** |
| Nouvel utilisateur | Métadonnées, préférences | Approche hybride ou basée sur le contenu |
| Nouvel item | Description, tags | Similarité de contenu ou hybridation |
| Utilisateur/item existant | Historique d’interactions | Filtrage collaboratif classique |

En fait, le problème du démarrage à froid limite la capacité de personnalisation des systèmes de recommandation, mais il peut être atténué par l’utilisation de méthodes hybrides et l’exploitation de données contextuelles ou explicites (Schein et al., 2022 ; Burke et al., 2021).

* + - 1. ***Problèmes de scalabilité et d’adaptabilité***

Les systèmes de recommandation modernes sont confrontés à deux défis majeurs : **la scalabilité** (capacité à gérer de grands volumes de données et d’utilisateurs) et **l’adaptabilité** (capacité à s’ajuster rapidement aux changements dans les données et les préférences). Ces problématiques sont cruciales pour garantir la performance, la pertinence et la réactivité des recommandations dans des environnements réels, souvent dynamiques et massifs (Alfaifi, 2024 ; Adomavicius & Tuzhilin, 2022).

1. **Scalabilité**

**D éfinition**

La scalabilité désigne la capacité d’un système de recommandation à maintenir des performances acceptables lorsque le nombre d’utilisateurs, d’items ou d’interactions croît de façon exponentielle. Par exemple, des plateformes comme Netflix ou Amazon doivent traiter des millions d’utilisateurs et de produits, générant des milliards de données d’interactions (Gomez-Uribe & Hunt, 2020).

**Defis principaux**

* **Complexite algorithmique :**

Les algorithmes de filtrage collaboratif, notamment la factorisation matricielle, impliquent la manipulation de matrices d’interactions très larges et très creuses (sparse). La complexité de calcul augmente rapidement avec la taille du système (Koren, Bell & Volinsky, 2023).

* **Cout mémoire et temps reel :** stocker et traiter de grandes matrices ou graphes peut devenir prohibitif en termes de mémoire et de temps de réponse.
* **Mises à jour en temps réel :** l’ajout constant de nouvelles interactions nécessite des mises à jour rapides des modèles, ce qui est difficile à grande échelle (Sarwar et al., 2021).

1. **Adaptabilité**

**Définition**

L’adaptabilité est la capacité d’un système à intégrer rapidement les nouvelles données (interactions, utilisateurs, produits) et à ajuster ses recommandations en conséquence (Alfaifi, 2024).

**Défis principaux**

* **Dynamisme des préférences :** les goûts des utilisateurs évoluent dans le temps, rendant les modèles statiques rapidement obsolètes.
* **Arrivée des nouvelles données :** les systèmes doivent pouvoir apprendre en continu (apprentissage incrémental) pour intégrer efficacement les nouvelles interactions (Mpia, Mburu & Mwendia, 2023).
* **Gestion du concept drift :** les distributions de données peuvent changer (concept drift), nécessitant une réactualisation fréquente des modèles.

**Exemples concrets**

* **Netflix** doit adapter ses recommandations à chaque nouvelle sortie de film ou à chaque changement de comportement d’un utilisateur, tout en maintenant la réactivité pour des millions de requêtes simultanées (Gomez-Uribe & Hunt, 2020).
* **Amazon** doit gérer l’ajout quotidien de milliers de nouveaux produits et l’évolution rapide des tendances d’achat.

Les défis de **scalabilité** et **d’adaptabilité** sont au cœur de la recherche et du développement des systèmes de recommandation modernes. Leur résolution repose sur des avancées algorithmiques, des architectures distribuées et des techniques d’apprentissage en ligne, afin de garantir la pertinence et la fraîcheur des recommandations dans des environnements massifs et dynamiques (Alfaifi, 2024 ; Adomavicius & Tuzhilin, 2022 ; Koren, Bell & Volinsky, 2023).

**2.2.6. Technologies et outils pour les systèmes de recommandation**

***2.2.6.1. Présentation de la librairie Surprise***

La librairie ***Surprise*** (Simple Python RecommendatIon System Engine)est un outil open-source développé en Python, spécifiquement conçu pour la construction, l’expérimentation et l’évaluation de systèmes de recommandation, en particulier ceux basés sur le filtrage collaboratif (Hug, 2020).

**Rôle de la librairie Surprise**

Surprise joue un rôle central dans la recherche et le développement de systèmes de recommandation en fournissant :

* Un cadre unifié pour la manipulation des jeux de données d’évaluations utilisateurs,
* L’implémentation et la comparaison rapide des divers algorithmes de recommandation,
* Des outils d’évaluation rigoureux pour mesurer la performance des modèles selon plusieurs métriques (comme RMSE, MAE, etc.).

Elle est particulièrement adaptée à l’expérimentation académique et au prototypage industriel, car elle permet de tester facilement différentes approches et de valider leur efficacité sur des jeux de données standards ou personnalisés.

**Fonctionnement et avantages**

* **Richesse algorithmique** : Surprise propose un vaste ensemble d’algorithmes de filtrage collaboratif, notamment la factorisation de matrices (SVD, SVD++), les méthodes de voisinage (KNNBasic, KNNWithMeans, KNNBaseline), ainsi que des algorithmes de base (BaselineOnly, NormalPredictor).
* **Simplicité d’utilisation**: grâce à une API inspirée de scikit-learn, la prise en main est rapide, même pour les utilisateurs débutants en machine learning.
* **Flexibilité des données**: elle permet de charger facilement des jeux de données standards (comme MovieLens) ou des jeux de données personnalisés, facilitant ainsi l’adaptation à divers contextes d’application.
* **Extensibilité**: les utilisateurs peuvent implémenter leurs propres algorithmes ou protocoles d’évaluation et les intégrer aisément dans le framework existant.
* **Documentation claire et communauté active** : la documentation détaillée et la communauté de développeurs contribuent à la robustesse et à l’évolution continue de la librairie.

**Limites**

Il est important de noter que Surprise se concentre principalement sur les systèmes de recommandation utilisant des évaluations explicites (notes attribuées par les utilisateurs), et ne prend pas en charge les données implicites (clics, achats, etc.) ni le filtrage basé sur le contenu (Hug, 2020).

En bref, la librairie Surprise constitue un outil puissant et flexible pour la conception, l’expérimentation et l’évaluation de systèmes de recommandation collaboratifs. Elle se distingue par sa simplicité d’utilisation, la richesse de ses fonctionnalités et sa capacité à accélérer le développement de prototypes fiables dans le domaine de la recommandation.

* + - 1. ***Comparaison avec d’autres outils courants***

Le domaine des systèmes de recommandation s’appuie sur divers outils et bibliothèques, chacun ayant des spécificités, des avantages et des limites qui les rendent plus ou moins adaptés selon les besoins du projet. Parmi les plus utilisés figurant **Surprise**, **TensorFlow** et **Scikit-Learn.**

**Surprise**

Surprise est une bibliothèque Python spécialisée dans la construction de systèmes de recommandation, notamment ceux basés sur le filtrage collaboratif. Elle se distingue par sa simplicité d’utilisation, la richesse de ses algorithmes intégrés (SVD, KNN, NMF, etc.) et ses outils d’évaluation performants. Elle est idéale pour la recherche, le prototypage rapide et l’expérimentation sur des jeux de données structurés contenant des évaluations explicites (Hug, 2020).

* **Avantages**
* Spécialisation dans le filtrage collaboratif.
* Prise en main rapide et documentation claire.
* Outils intégrés pour la validation croisée et la recherche d’hyperparamètres.
* **Limites**
* Peu adaptée aux données implicites (clics, achats).
* Ne prend pas en charge les modèles complexes de deep learning ni le filtrage basé sur le contenu.

**TensorFlow**

TensorFlow est une plateforme de machine learning open-source développée par Google, principalement axée sur le deep learning. Elle permet de construire des modèles complexes, y compris des réseaux de neurones profonds, des architectures hybrides et des systèmes de recommandation avancés (Alfaifi, 2024).

* **Avantages :**
* Extrêmement flexible et puissante pour le développement de modèles personnalisés.
* Supporte le traitement de grandes quantités de données, y compris les données implicites et multimodales (texte, image, etc.).
* Permet d’implémenter des architectures de pointe (deep learning, réseaux de neurones convolutifs, modèles séquentiels, etc.).
* Large communauté et écosystème riche (TensorFlow Recommenders, TensorFlow Lite, etc.)
* **Limites**
* Courbe d’apprentissage plus élevée, nécessite des connaissances avancées en machine learning et en programmation.
* Moins adaptée pour des expérimentations rapides sur de petits jeux de données structurés.

**Scikit-Learn**

Scikit-Learn est une bibliothèque généraliste d’apprentissage automatique en Python, largement utilisée pour l’implémentation d’algorithmes classiques de machine learning (régression, classification, clustering, etc.) (Pedregosa et al., 2023). Bien qu’elle ne soit pas spécialisée dans les systèmes de recommandation, elle offre des outils pour le prétraitement, l’évaluation et la modélisation qui peuvent être utiles dans ce contexte.

* **Avantages**
* Large éventail d’algorithmes de machine learning classiques.
* API simple et cohérente, idéale pour le prototypage rapide.
* Outils puissants pour la validation croisée, la sélection de modèles et le prétraitement des données.
* **Limites**
* Ne propose pas nativement d’algorithmes de recommandation spécialisés (pas de SVD, KNN adaptés aux matrices utilisateur-objet, etc.)
* Moins adaptée pour le traitement de données massives ou non structurées

|  |  |  |  |
| --- | --- | --- | --- |
| **Critère** | **Surprise** | **TensorFlow** | **Scikit-Learn** |
| Spécialisation | Recommandation collaborative | Deep learning, modèles avancés | Machine learning généraliste |
| Facilité d’utilisation | Très simple | Complexe (avancé) | Simple |
| Algorithmes intégrés | SVD, KNN, NMF, etc. | Réseaux de neurones, hybrides, etc. | Régression, classification, etc. |
| Données supportées | Évaluations explicites | Explicites et implicites, multimoda | Principalement tabulaires |
| Evolutivité | Moyenne | Très élevée | Moyenne |
| Cas d’usage typique | Prototypage, recherche | Production à grande échelle, IA | Analyse exploratoire, prototypage |

Alors le choix de l’outil dépend fortement des objectifs du projet. **Surprise** est recommandé pour des expérimentations rapides et des études académiques sur des jeux de données structurés. **TensorFlow** est préférable pour les systèmes de recommandation complexes, à grande échelle, ou nécessitant des modèles de deep learning. **Scikit-Learn** convient pour des tâches de machine learning généralistes ou pour intégrer des étapes de prétraitement et d’évaluation dans un pipeline plus large (Pedregosa et al., 2023 ; Hug, 2020 ; Alfaifi, 2024).

* + - 1. ***Principe de Fonctionnement***

La Bibliotheque **Suprise** est conçue pour faciliter la mise en œuvre de systèmes de recommandation, principalement basés sur le filtrage collaboratif. Son principe de fonctionnement repose sur la prédiction des préférences d’un utilisateur pour un item donné, en se basant sur les interactions passées entre utilisateurs et items (Hug, 2020).

1. **Structure des données et généralisation**

Dans Surprise, les données sont organisées sous forme de triplets : utilisateur, item, et note attribuée. L’item peut représenter n’importe quel objet à recommander (livre, produit, service, etc.), ce qui rend la bibliothèque adaptable à de nombreux domaines d’application (Koren, Bell & Volinsky, 2023). Cette structure permet de construire une matrice utilisateur-item, où chaque entrée correspond à une note ou une interaction.

*« Les systèmes de recommandation collaboratifs utilisent généralement une matrice utilisateur-item, où chaque cellule contient la note ou l’interaction enregistrée » (Koren, Bell & Volinsky, 2023).*

1. **Préparation et chargement des données**

Surprise permet de charger des jeux de données personnalisés ou standards. Les données sont ensuite divisées en ensembles d’entraînement et de test, ce qui permet d’évaluer la capacité du modèle à généraliser ses prédictions à de nouveaux utilisateurs ou items (Hug, 2020).

*« La bibliothèque Surprise offre une grande flexibilité pour le chargement et la manipulation de jeux de données structurés autour du couple utilisateur-item » (Hug, 2020).*

1. **Entrainement et prédiction**

Après la préparation des données, un algorithme de recommandation (tel que SVD ou KNN) est entraîné sur l’ensemble d’entraînement à l’aide de la commande algo.fit (trainset). La prédiction des notes pour l’ensemble de test se fait ensuite avec algo.test (testset), ce qui permet d’obtenir les scores prédits pour chaque couple utilisateur-item du test (Hug, 2020).

*« L’entraînement du modèle sur l’ensemble d’apprentissage, suivi de la prédiction sur l’ensemble de test, constitue la base de l’expérimentation dans Surprise » (Hug, 2020).*

1. **Prédiction individuelle**

Pour obtenir la prédiction de la note qu’un utilisateur spécifique attribuerait à un item particulier, la méthode algo.predict (uid, iid) est utilisée, où uid est l’identifiant de l’utilisateur et iid  celui de l’item. Cette fonctionnalité permet d’obtenir des recommandations personnalisées et ciblées (Hug, 2020). Pour plusieurs algorithmes, les focntionnalités integrées pour une utulisation sont en exemple : *predictions = algo.fit(trainset).test(testset)* :un code qui permet d’imprimer le score de notation predict pour l’ensemble de test. Dans un autre cas, l’aglorithme : *predicitons= algo.prediction(uid, iid)* : dans cette approche, le créateur vise la prédiction de score de notation de l’iid, de l’uid.

*« La fonction predict () permet de prédire la note pour un couple utilisateur-item donné, ce qui est essentiel pour générer des recommandations individualisées » (Hug, 2020).*

1. **Utilisation des résultats**

Les résultats des prédictions peuvent être utilisés pour :

* Proposer à l’utilisateur les items les plus susceptibles de l’intéresser,
* Evaluer la performance du système à l’aide de métriques telles que RMSE ou MAE, qui mesurent la précision des prédictions par rapport aux notes réelles (Hug, 2020 ; Koren, Bell & Volinsky, 2023).

*« L’évaluation des systèmes de recommandation repose sur des métriques d’erreur qui quantifient l’écart entre les notes prédites et les notes effectives » (Koren, Bell & Volinsky, 2023).*

* + - 1. ***Framework Flask pour le déploiement Web***

**Flask** est un framework Python léger, modulaire et très populaire pour le développement d’applications web. Il est particulièrement apprécié pour sa simplicité, sa flexibilité et sa capacité à s’intégrer facilement avec d’autres outils et bibliothèques Python, ce qui en fait un choix privilégié pour le déploiement de systèmes de recommandation (Grinberg, 2022).

**Architecture et principes de Flask**

Flask adopte une architecture minimaliste, dite « micro-framework », ce qui signifie qu’il fournit uniquement les composants essentiels pour construire une application web (gestion des requêtes HTTP, routage, templates HTML, etc.), tout en laissant la liberté au développeur d’ajouter des extensions selon les besoins spécifiques du projet (Grinberg, 2022). Cette modularité permet de garder le code simple et facile à maintenir, tout en offrant la possibilité d’intégrer des fonctionnalités avancées comme l’authentification, la gestion des sessions, ou la connexion à des bases de données (Grinberg, 2022). « Flask est conçu pour être extensible et flexible, permettant aux développeurs de choisir les composants dont ils ont besoin sans imposer une structure lourde » (Grinberg, 2022).

**Intégration avec les modèles d’apprentissage automatique**

Flask facilite le déploiement de modèles d’apprentissage automatique, notamment ceux développés avec des bibliothèques telles que scikit-learn, TensorFlow ou Surprise. Il permet d’exposer les modèles via des API web (REST), ce qui rend possible l’interaction entre le système de recommandation et l’interface utilisateur ou d’autres applications externes (Grinberg, 2022).

Par exemple, une application Flask peut recevoir les données d’un utilisateur via une requête HTTP, utiliser un modèle de recommandation pour générer des suggestions personnalisées, puis retourner les résultats en temps réel à l’utilisateur (Alfaifi, 2024). L’intégration de modèles de machine learning dans des applications web est simplifiée grâce à la capacité de Flask à gérer les requêtes et à fournir des réponses dynamiques (Grinberg, 2022).

**Connexion aux bases de données et gestion des données**

Flask prend en charge la connexion à divers systèmes de gestion de bases de données (relationnelles ou NoSQL) à l’aide d’extensions comme Flask-SQLAlchemy ou Flask-PyMongo. Cela permet de stocker, récupérer et mettre à jour facilement les données des utilisateurs, les historiques d’interaction et les résultats de recommandation, éléments essentiels pour un système de recommandation efficace (Grinberg, 2022).

Grâce à ses nombreuses extensions, Flask permet une gestion efficace des bases de données, essentielle pour les applications de recommandation personnalisée (Grinberg, 2022).

**Déploiement et évolutivité**

La légèreté de Flask le rend adapté aussi bien au prototypage rapide qu’au déploiement en production. Il peut être déployé sur divers serveurs web (comme Gunicorn ou uWSGI) et s’intègre facilement dans des environnements cloud ou conteneurisés (Docker). Cette flexibilité permet de faire évoluer l’application selon la croissance du nombre d’utilisateurs ou la complexité des recommandations à fournir (Grinberg, 2022). La simplicité de déploiement de Flask en fait un outil privilégié pour passer rapidement du prototype à la production (Grinberg, 2022).

Enfin, Flask constitue un environnement idéal pour le déploiement de systèmes de recommandation grâce à sa simplicité, sa flexibilité et sa capacité à s’intégrer avec les modèles de machine learning et les bases de données. Il permet de transformer un modèle de recommandation développé en Python en une application web interactive, accessible à grande échelle (Grinberg, 2022 ; Alfaifi, 2024).

**2.3. Revue de littérature empirique**

* + 1. **Études sur les systèmes de recommandation collaboratifs**

Les systèmes de recommandation collaboratifs jouent aujourd’hui un rôle central dans la personnalisation des services numériques, en permettant de mieux satisfaire les besoins et préférences des utilisateurs dans divers secteurs tels que l’e-commerce, la santé, l’éducation, le tourisme ou encore l’industrie (Adomavicius & Tuzhilin, 2022). Leur popularité s’explique par leur capacité à exploiter l’intelligence collective issue des interactions passées entre utilisateurs et items, afin de proposer des suggestions pertinentes et adaptées (Sarwar et al., 2021).

Au fil des années, de nombreux chercheurs ont exploré et amélioré ces systèmes. Par exemple, Tork et al. (2022) se sont intéressés à la problématique de la surspécialisation, un phénomène fréquent dans les systèmes de recommandation basés sur le contenu, où les suggestions deviennent trop similaires et manquent de diversité. Pour remédier à cela, les auteurs ont proposé une approche hybride intégrant des algorithmes génétiques dans le processus de recommandation. Leur méthode vise à diversifier les suggestions et à introduire des éléments inattendus et innovants pour les utilisateurs, tout en maintenant la pertinence des recommandations.

Dans leur étude publiée dans la revue *Elecronics*, Tork et ses collègues détaillent la mise en œuvre d’un Recommender System based on a Genetic Algorithm (abrégé RRSGA), qui combine les avantages du filtrage collaboratif et de l’optimisation génétique. L’algorithme génétique, en explorant différentes combinaisons d’items, permet d’éviter l’effet de surspécialisation et d’améliorer la diversité des recommandations (Tork et al., 2022). Les résultats expérimentaux présentés par les auteurs montrent que le modèle RRSGA surpasse les approches classiques de filtrage basé sur le contenu, tant en termes de précision que de diversité des recommandations.

Plus précisément, la comparaison empirique réalisée par Tork et al. (2022) met en évidence une amélioration significative de la qualité des prédictions et de la satisfaction utilisateur grâce à l’intégration de l’algorithme génétique. Cette étude illustre ainsi l’importance des approches hybrides et de la recherche constante de solutions innovantes pour surmonter les limites des systèmes de recommandation traditionnels.

Bref, l’évolution des systèmes de recommandation collaboratifs se caractérise par une diversification des méthodes et une attention croissante portée à la personnalisation, à la diversité et à l’innovation dans les recommandations, comme le démontrent les travaux récents de Tork et al. (2022) et d’autres chercheurs dans le domaine (Adomavicius & Tuzhilin, 2022 ; Sarwar et al., 2021).

* + 1. **Applications pratiques dans le commerce local et international**

Les systèmes de recommandation sont aujourd’hui incontournables dans le commerce électronique international. Des entreprises majeures telles qu’Amazon et Netflix exploitent des algorithmes avancés, notamment le filtrage collaboratif basé sur les utilisateurs, le filtrage sur les items, ainsi que la factorisation matricielle, afin de personnaliser les suggestions faites à leurs clients (Gomez-Uribe & Hunt, 2020). Ces plateformes combinent des méthodes d’apprentissage supervisé et des techniques de réduction de dimension pour améliorer la pertinence et la diversité des recommandations (Gomez-Uribe & Hunt, 2020 ; Koren, Bell & Volinsky, 2023).

Dans les économies développées, ces technologies ont transformé l’expérience utilisateur et généré une forte valeur ajoutée commerciale. Cependant, leur adoption dans le commerce local, en particulier dans les économies en développement, demeure limitée. Selon Zhang et Chen (2020), les systèmes de recommandation présentent un potentiel considérable pour améliorer la performance des petites entreprises locales, mais leur déploiement est freiné par des défis tels que la disponibilité des données, l’infrastructure technologique et la formation des acteurs locaux.

**Cas du commerce local à Butembo**

En République Démocratique du Congo (RDC), et plus précisément dans la province du Nord-Kivu en ville de Butembo constitue un centre économique dynamique, notamment grâce à ses activités commerciales et entrepreneuriales (Menomavuya, 2018). Cependant, l’intégration des systèmes de recommandation dans les pratiques commerciales locales reste embryonnaire, voire inexistante à ce jour. Les entreprises de Butembo s’appuient principalement sur des méthodes traditionnelles de gestion de la relation client et de promotion des produits, telles que le bouche-à-oreille, les affiches, ou encore les réseaux sociaux classiques, sans recours systématique à des outils de recommandation automatisés (Menomavuya, 2018).

Plusieurs facteurs expliquent cette situation :

* **Manque d’infrastructures numériques** : l’accès limité à Internet et aux technologies de l’information freine l’adoption de solutions avancées comme les systèmes de recommandation (Ngangoue, 2016).
* **Faible disponibilité et qualité des données**: les commerçants locaux ne disposent pas toujours de bases de données structurées sur les comportements d’achat ou les préférences des clients, ce qui rend difficile l’implémentation de modèles de recommandation performants (Zhang & Chen, 2020).
* **Niveau de formation**: l’absence de compétences techniques spécialisées dans le domaine de la science des données ou de l’intelligence artificielle constitue également un obstacle majeur (Menomavuya, 2018).

Malgré ces défis, des opportunités existent pour l’introduction progressive des systèmes de recommandation dans le tissu commercial local. Des initiatives de formation, le développement de solutions adaptées au contexte africain, et la sensibilisation des commerçants aux avantages de la personnalisation pourraient favoriser l’émergence de telles technologies à Butembo et dans d’autres villes de la RDC (Ngangoue, 2016).

Si les systèmes de recommandation sont largement adoptés dans le commerce international, leur application dans le contexte local de Butembo reste encore à développer. Des efforts concertés en matière d’infrastructure, de formation et de gestion des données sont nécessaires pour permettre aux entreprises locales de tirer pleinement parti de ces technologies innovantes.

* + 1. **Analyse des lacunes et des opportunités pour le projet actuel**

La littérature récente met en évidence que les systèmes de recommandation ont principalement été conçus et optimisés pour des environnements riches en données, tels que les grandes plateformes de commerce électronique internationales (Adomavicius & Tuzhilin, 2022). Cette abondance de données permet de construire des modèles performants, capables de fournir des recommandations personnalisées et pertinentes. Cependant, dans des contextes à faible densité de données comme c’est le cas dans la ville de Butembo en République Démocratique du Congo (RDC) l’application de ces systèmes reste limitée et peu étudiée (Kabeya, 2020).

En effet, l’étude de Kabeya (2020) sur l’e-commerce en RDC souligne que la majorité des entreprises locales font face à un manque d’infrastructures numériques, à une faible digitalisation des transactions et à une insuffisance de données exploitables pour l’analyse comportementale des clients. Cette situation rend difficile l’implémentation de systèmes de recommandation traditionnels, qui reposent sur la disponibilité de larges volumes de données historiques. De plus, les commerçants locaux sont confrontés à des défis liés à la gestion des stocks et à la synchronisation entre l’offre et la demande, ce qui limite encore davantage l’efficacité des solutions classiques (Kabeya, 2020).

Par ailleurs, une analyse de Kasereka (2024) met en avant la nécessité d’adapter les technologies de recommandation aux réalités locales, en tenant compte des contraintes spécifiques telles que la connectivité intermittente, la fragmentation des données et le manque de compétences techniques. L’auteur suggère que le développement de systèmes de recommandation collaboratifs adaptés à ces contextes pourrait non seulement améliorer la gestion des stocks, mais aussi optimiser l’expérience client et stimuler la croissance des entreprises locales (Kasereka, 2024).

Face à ces constats, le projet actuel vise à combler cette lacune en développant un système de recommandation collaboratif spécifiquement conçu pour les environnements à faible densité de données, comme celui de Butembo. L’objectif est de proposer une solution qui tienne compte des contraintes locales, tout en exploitant les opportunités offertes par la digitalisation progressive du commerce dans la région. En s’appuyant sur des approches hybrides et des méthodes adaptées à la rareté des données, ce projet ambitionne de renforcer la compétitivité des entreprises locales et de favoriser l’adoption des technologies numériques dans le tissu commercial de la RDC (Kabeya, 2020 ; Kasereka, 2024).

* 1. **Synthèse et positionnement de la recherche**

En résumé, ce chapitre a exploré les concepts clés, les approches théoriques, et les outils technologiques liés aux systèmes de recommandation. Il a également examiné les défis et les limites de ces systèmes, ainsi que leur application dans des contextes pratiques. La revue de littérature nous a permis de comprendre les applications existantes avant d’apporter notre approche, pour éviter une même production d’application que les applications précédentes. Fouiller dans l’histoire est toujours mieux pour afin produire quelque chose de nouveau que ancienne. C’est pourquoi, cette partie nous a aidé comment comprendre les approches que nous avons apporter à notre problématique et également une application adaptée aux fonctionnements de la Ville de Butembo.

* 1. **Présentation de la Ville de Butembo**

Butembo est une ville de la République Démocratique du Congo(RDC), située dans la province du Nord-Kivu. Elle fait partie des trois grandes villes du Nord-Kivu, localisée à l’extrême nord-est du pays. Sa circonscription urbaine s’étend entre 0°05’ et 0°10’ de latitude nord et 29°17’ et 29°18’ de longitude Est, à environ 17 km au nord de l’Equateur. La ville se trouve à proximité de la dorsale occidentale du Rift Albertin, au nord-ouest du lac Édouard, et est bâtie sur les hautes terres du lac Kinya-Muliha (Édouard), entre 1700 et 1800 mètres d’altitude, le long de la route Congo-Nil dans le Kivu septentrional. Butembo bénéficie d’un climat frais et d’une pluviosité modérée (environ 1380 mm par an). Le site de la ville est caractérisé par une dépression entourée de collines de tailles et d’altitudes variées (Menomavuya, 2018).

Sur le plan culturel et artistique, Butembo est un centre dynamique en RDC. On y retrouve plusieurs institutions et espaces culturels et commerciaux majeurs, tels que le Musée National de Butembo, l’Association Muthembo Arts, la Réserve forestière de Mususa, la Galerie Tsongo Butembo (GTB), le Super Marché Alpha, le Marché Central de Butembo, Issale Lounge, Village Royal, Chez Feza, Maida Grill, Club des Intimes chez Vanessa, Hôtel La Charité, Hôtel Believe, Hôtel Butembo, ainsi que les réseaux de communication Vodacom, Airtel et Orange,… qui facilitent la connectivité et les échanges (Menomavuya, 2018).

D’un point de vue économique et commercial, Butembo se distingue comme un pôle commercial régional majeur. La ville joue un rôle clé dans l’approvisionnement des provinces voisines et des villes en produits divers, grâce à un réseau dense de marchés, de boutiques et de supermarchés. Les activités commerciales sont soutenues par un dynamisme entrepreneurial local, une forte mobilité des biens et des personnes, ainsi que par l’ouverture sur les marchés nationaux et transfrontaliers. Selon une étude de Mbusa et al. (2022), Butembo est reconnue pour son tissu commercial actif, sa capacité d’innovation dans les pratiques de vente et sa contribution significative à l’économie du Nord-Kivu et de la RDC en général (Mbusa et al., 2022).

La ville de Butembo, grâce à son dynamisme économique et commercial, s’impose comme un centre d’échange important pour la région du Nord-Kivu, avec une influence notable sur les flux de marchandises et la structuration des circuits de distribution (Mbusa et al., 2022, p. 312).

**Justification de l’importance du système de recommandation**

Notre recherche fournit des statistiques sur les produits les plus vendus et les plus consommés par les clients. En effet, le temps étant précieux pour les commerçants, cette application permettra de raccourcir le processus d’achat en aidant les vendeurs à mieux cibler les articles à proposer. Ainsi, les clients n’auront plus à passer de longues minutes à chercher les produits les plus populaires, car ces derniers seront automatiquement suggérés selon les tendances de consommation. Cette fonctionnalité est d’autant plus pertinente dans une ville commerciale comme Butembo, où la rapidité et l’efficacité des transactions sont essentielles pour rester compétitif.

De plus, cette étude offre un avantage supplémentaire pour les clients venant de l’extérieur, qui pourront découvrir rapidement les articles les plus appréciés du magasin, même si ceux-ci ne sont pas encore disponibles dans leur propre environnement de vente. Cela favorise la curiosité et l’innovation dans les pratiques d’achat, tout en renforçant l’attractivité commerciale de Butembo.

**Perspectives d’évolution du système**

Une fois ce système déployé, une fonctionnalité de partage sera intégrée afin de renforcer l’interaction entre utilisateurs. Compte tenu de la situation sécuritaire du pays, l’ajout d’un module de chat permettra aux clients réguliers d’une boutique de recevoir directement, après leur commande, les articles recommandés par l’application ainsi que ceux qu’ils ont eux-mêmes sélectionnés. Après validation du paiement et des frais de transport, les colis pourront être livrés sans que le client ait besoin de se déplacer. Cela contribuera à améliorer l’expérience utilisateur et à fidéliser la clientèle, tout en soutenant la modernisation du commerce local.

* 1. **Conclusion partielle**

Ce chapitre a fourni une revue approfondie de la littérature sur les systèmes de recommandation, en mettant l'accent sur les approches collaboratives et leur application dans le commerce local. Les défis liés à la faible densité de données et au "cold start" ont été identifiés comme des obstacles majeurs à l'efficacité des systèmes de recommandation dans des contextes comme celui de Butembo. Cependant, les technologies disponibles, telles que la librairie Surprise et le framework Flask, offrent des opportunités pour surmonter ces défis. Le chapitre suivant présentera la méthodologie adoptée pour concevoir et développer le système de recommandation collaboratif proposé dans cette étude.

**Troisième Chapitre : MÉTHODOLOGIE**

**3.1. Introduction**

**Presentation de l’objectif**

Ce chapitre méthodologique a pour objectif de présenter de manière détaillée la démarche adoptée pour la **conception, l’implémentation et l’évaluation** du système de recommandation collaboratif développé dans le cadre de ce projet. Il s’agit de décrire, étape par étape, les choix techniques et scientifiques qui ont guidé la réalisation du système, depuis la modélisation des données jusqu’au déploiement de l’application web, en passant par l’entraînement du modèle de recommandation et l’intégration des différents modules fonctionnels. L’objectif est de permettre au lecteur de comprendre non seulement comment le système a été construit, mais aussi pourquoi certaines méthodes et outils ont été privilégiés, en tenant compte des spécificités du contexte local et des contraintes identifiées lors de l’analyse des besoins.

**Justification du choix des méthodes utilisées.**

Le choix de la méthodologie repose sur une combinaison de critères : pertinence scientifique, faisabilité technique, adaptabilité au contexte local (ville de Butembo), et efficacité pour répondre aux problématiques de faible densité de données et de gestion commerciale.

* **Filtrage collaboratif** : ce paradigme a été retenu pour sa capacité à exploiter les interactions historiques entre utilisateurs et produits, permettant de générer des recommandations personnalisées même en l’absence d’informations détaillées sur les produits eux-mêmes. Ce choix est soutenu par la littérature, qui démontre l’efficacité du filtrage collaboratif dans des environnements variés, y compris ceux à ressources limitées.
* **Utilisation de la librairie Surprise**: l’outil Surprise a été sélectionné pour l’entraînement et l’évaluation du modèle de recommandation, en raison de sa simplicité d’intégration, de la richesse de ses algorithmes intégrés (notamment SVD) et de ses outils d’évaluation performants.
* **Déploiement avec Flask**: le framework Flask a été choisi pour la réalisation de l’interface web, grâce à sa légèreté, sa flexibilité et sa capacité à s’intégrer facilement avec des bases de données et des modèles Python.
* **Stockage et gestion des données**: l’utilisation de SQLAlchemy et d’une base de données MySQL permet une gestion structurée, sécurisée et évolutive des informations relatives aux clients, produits et transactions, facilitant l’exploitation des données pour la recommandation et l’optimisation des stocks.

Ce choix méthodologique vise à garantir la robustesse, la maintenabilité et l’extensibilité du système, tout en restant accessible aux utilisateurs et développeurs locaux.

**Plan du chapitre**

Le présent chapitre est structuré de la manière suivante :

1. **Architecture générale du système** : description des différents composants du système, de leur rôle et de leurs interactions.
2. **Modélisation des données** : présentation des modèles de données utilisés (clients, produits, transactions) et justification de leur structuration.
3. **Mise en œuvre du système de recommandation**: détail des étapes d’entraînement du modèle, des algorithmes utilisés et des critères d’évaluation adoptés.
4. **Développement de l’application web**: explication du choix du framework Flask, de l’organisation des routes, des interfaces utilisateurs et de la gestion des fichiers.
5. **Evaluation et validation du système**: présentation des protocoles d’évaluation, des métriques utilisées (RMSE, MAE), et discussion des résultats obtenus.
6. **Limites de la méthodologie et perspectives d’amélioration**: analyse critique des choix méthodologiques et pistes pour de futurs développements.

Cette organisation vise à offrir une vision claire, progressive et argumentée de la démarche suivie pour la réalisation du système de recommandation collaboratif adapté au contexte commercial local.

**3.2. Conception du système de recommandation**

**3.2.1. Modèle conceptuel**

**Présentation des besoins utilisateurs et des objectifs du système.**

Le système de recommandation collaboratif développé vise à répondre à plusieurs besoins spécifiques identifiés chez les acteurs du commerce local à Butembo :

* **Pour les commerçants :**
* Optimiser la gestion des stocks et anticiper les ruptures grâce à l’analyse automatisée des ventes et des niveaux de stock.
* Mieux comprendre les préférences des clients afin d’ajuster l’offre de produits et d’augmenter la fidélisation.
* Gagner du temps dans le conseil client grâce à des suggestions automatiques de produits populaires ou adaptés.
* **Pour les clients :**
* Recevoir des recommandations personnalisées sur les produits susceptibles de les intéresser, basées sur leur historique d’achats et les évaluations d’autres clients.
* Accélérer le processus d’achat, notamment en découvrant rapidement les articles les plus appréciés ou les nouveautés.
* Profiter d’une interface simple pour consulter, noter et recommander des produits.

L’objectif global du système est donc de **favoriser la digitalisation intelligente du commerce local** en apportant une valeur ajoutée tant pour les commerçants que pour les clients, tout en s’adaptant à un contexte de faible densité de données et de ressources limitées.

**Définition des principales fonctionnalités du système.**

Le système se compose des fonctionnalités principales suivantes :

1. **Gestion des clients** :

* Ajout, consultation et gestion des informations clients (nom, email, préférences).
* Visualisation paginée des clients.

1. **Gestion des produits** :

* Ajout de nouveaux produits avec gestion d’image, catégorie, prix et stock.
* Consultation paginée et affichage des produits.

1. **Gestion des transactions** :

* Enregistrement des achats (transactions), avec attribution d’une note (1 à 5) par le client.
* Visualisation des transactions et des évaluations.

1. **Système de recommandation** :

* Génération de recommandations personnalisées pour chaque client à partir de son historique d’achats et des notes attribuées.
* Affichage des recommandations sous forme de liste de produits triés par pertinence.

1. **Optimisation des stocks** :

* Détection automatique des produits dont le stock est inférieur à un seuil défini et suggestion de réapprovisionnement.

1. **Interface web conviviale** :

* Navigation intuitive entre les différentes fonctionnalités (clients, produits, transactions, recommandations, stocks).
* Utilisation de Bootstrap pour un rendu responsive et moderne.

**Diagramme UML (use case, diagramme de séquence, diagramme de classe).**

Pour modéliser le système, plusieurs diagrammes UML sont utilisés :

1. **Diagramme de cas d’utilisation (Use Case)**

**Acteurs principaux :**

* Administrateur (gère les clients, produits, transactions, stocks)
* Client (reçoit les recommandations, donne des notes)

**Cas d’utilisation principaux** :

* Ajouter /consulter un client
* Ajouter /consulter un produit
* Ajouter /consulter une transaction
* Noter un produit
* Recevoir des recommandations
* Visualiser les stocks à réapprovisionner

1. **Diagramme de séquence**

**Exemple : Séquence pour la recommandation d’un produit**

* 1. Le client accède à la page de recommandations.
  2. Le système récupère l’historique d’achats et de notes du client.
  3. Le modèle de recommandation (SVD via Surprise) prédit les scores pour chaque produit.
  4. Le système trie les produits et affiche les 6 meilleurs au client.

1. **Diagramme de classes**

Les principales classes du modèle de données sont :

* **Client** : id, nom, email, préférences
* **Produit**: id, nom, catégorie, prix, stock, image
* **Transaction**: id, client\_id, produit\_id, date, quantité, note

***Exemples des relations :***

* Un client peut effectuer plusieurs transactions.
* Un produit peut être acheté dans plusieurs transactions.
* Une transaction relie un client à un produit, à une date donnée, avec une quantité et une note.

**3.2.2. Choix de l’approche de recommandation**

**Justification du choix du filtrage collaboratif.**

Le choix du **filtrage collaboratif** comme approche principale pour ce projet s’explique par plusieurs avantages majeurs adaptés à la réalité du commerce local à Butembo :

* **Adaptabilité à la disponibilité des données** : contrairement aux systèmes basés sur le contenu, le filtrage collaboratif ne nécessite pas de descriptions détaillées des produits. Il exploite directement les interactions entre clients et produits (achats, notes), ce qui est particulièrement pertinent dans un contexte où les données textuelles ou multimédias sont rares ou peu structurées (Sarwar et al., 2021 ; Adomavicius & Tuzhilin, 2022).
* **Personnalisation efficace :** cette méthode permet de proposer des recommandations personnalisées en se basant sur les comportements et préférences réels des utilisateurs, favorisant ainsi la satisfaction client et la fidélisation (Adomavicius & Tuzhilin, 2022).
* **Simplicité d’intégration**: les algorithmes collaboratifs sont bien documentés, éprouvés dans l’industrie (Netflix, Amazon) et disposent de bibliothèques Python robustes telles que Surprise, facilitant leur implémentation et leur évaluation (Hug, 2020).

Dans le contexte de Butembo, où l’on dispose principalement d’historiques d’achats et de notes, le filtrage collaboratif s’impose donc comme la solution la plus pragmatique et la plus efficace.

**Comparaison avec d’autres approches (basé sur le contenu, hybride).**

**Approche basé sur le contenu** :

L’approche **basée sur le contenu** recommande des produits similaires à ceux déjà appréciés par l’utilisateur, en s’appuyant sur les caractéristiques des produits (catégorie, description, etc.).

* **Avantages** : fonctionne bien même avec peu d’utilisateurs, utile pour le démarrage du système.
* **Limites :** nécessite des données descriptives riches et structurées pour chaque produit, ce qui n’est pas toujours disponible dans le contexte local (Lops et al., 2021). Elle tend aussi à générer des recommandations trop similaires (surspécialisation).

**Approche hybride**

Les **systèmes hybrides** combinent plusieurs méthodes (collaboratif, contenu, etc.) pour pallier les faiblesses de chaque approche.

* **Avantages** : résolvent le problème du démarrage à froid (« cold start »), améliorent la diversité et la robustesse des recommandations (Burke et al., 2021).
* **Limites :** plus complexes à mettre en œuvre et à maintenir, nécessitent des ressources et des données supplémentaires, ce qui peut être un frein dans un contexte à faible infrastructure (Khaledian et al., 2025).

**Synthèse comparative**

|  |  |  |  |
| --- | --- | --- | --- |
| **Critère** | **Filtrage collaboratif** | **Basé sur le contenu** | **Hybride** |
| Données requises | Interactions (notes) | Descriptions items | Les deux + autres |
| Complexité | Moyenne | Moyenne | Elevée |
| Adapté au contexte local | Oui | Limite | Possible mais complexe |
| Personnalisation | Excellente | Moyenne | Excellente |
| Démarrage à froid | Problématique | Gère bien | Gère bien |

Au vu des contraintes locales et des objectifs du projet, le **filtrage collaboratif** est retenu comme approche principale, avec la possibilité d’évoluer vers une solution hybride à moyen terme.

**Présentation de l’algorithme choisi (SVD, etc.).**

Pour la mise en œuvre du filtrage collaboratif, l’algorithme principal utilisé est la **factorisation de matrice par décomposition en valeurs singulières (SVD)**, via la librairie Surprise (Hug, 2020).

* **Principe** : SVD décompose la matrice utilisateur-produit (où chaque entrée correspond à une note ou une interaction) en facteurs latents, permettant de prédire la note qu’un utilisateur attribuerait à un produit non encore évalué (Koren, Bell & Volinsky, 2023).
* **Avantages :**
* Précision élevée dans la prédiction des notes.
* Capacité à capturer des relations complexes entre utilisateurs et produits.
* Efficace même avec des matrices creuses (peu de notes renseignées).
* **Implémentation**:
* Les transactions (notes) sont extraites de la base de données, transformées en DataFrame, puis utilisées pour entraîner le modèle SVD.
* Le modèle est évalué à l’aide de métriques telles que RMSE et MAE pour garantir la qualité des recommandations.
* Une entraîné, le modèle est sauvegardé et utilisé pour générer les recommandations personnalisées sur l’application web.

Pour dire, le choix du **filtrage collaboratif** basé sur l’algorithme SVD offre un compromis optimal entre simplicité, performance et adaptabilité au contexte local. Cette approche permet de répondre efficacement aux besoins des commerçants et des clients de Butembo, tout en posant les bases pour d’éventuelles évolutions vers des systèmes hybrides à l’avenir.

**3.2.3. Architecture du système**

**Description de l’architecture globale (client-serveur, base de données, moteur de recommandation).**

Le système de recommandation collaboratif mis en œuvre repose sur une **architecture client-serveur** classique, articulée autour de trois grands composants :

* + 1. **L’interface utilisateur (Client Web)**

Lesutilisateurs (commerçants ou clients) interagissent avec le système via une interface web responsive, développée avec HTML, CSS (Bootstrap) et les templates Jinja2 de Flask. Cette interface permet la gestion des clients, des produits, des transactions, la visualisation des recommandations et l’optimisation des stocks.

* + 1. **Le serveur applicatif (Backend Flask)**

Le cœur de l’application est un serveur Flask qui gère :

* Les routes HTTP pour chaque fonctionnalité (accueil, gestion des clients/produits/transactions, recommandations, stocks).
* La logique métier (ajout, modification, affichage des entités).
* L’appel au moteur de recommandation et à la logique d’optimisation des stocks.
* L’interface avec la base de données via SQLAlchemy.
  + 1. **La base de donnees relationnelle (MySQL)**

Toutes les données structurées (clients, produits, transactions, stocks) sont stockées dans une base MySQL. SQLAlchemy assure l’ORM (mapping objet-relationnel) pour faciliter les requêtes et la gestion des entités.

* + 1. **Le moteur de recommandation (Surprise/SVD)**

Le moteur de recommandation est entraîné à partir des transactions (notes données par les clients aux produits). Il est implémenté avec la bibliothèque Surprise et utilise l’algorithme SVD pour générer des prédictions de notes et des recommandations personnalisées. Le modèle est sauvegardé et chargé dynamiquement pour répondre aux requêtes du serveur.

**Fonctionnement général**

* L’utilisateur effectue des actions via l’interface web (ajout d’un achat, consultation de recommandations, etc.).
* Flask reçoit la requête, interagit avec la base de données pour récupérer ou enregistrer les données nécessaires.
* Pour les recommandations, Flask appelle le moteur SVD (chargé via joblib), qui prédit les scores pour chaque produit non encore évalué par le client.
* Les résultats sont renvoyés à l’interface web pour affichage.

**Schéma de l’architecture technique.**

Schema………..

**Explication** :

* Toutes les interactions passent par le serveur Flask.
* La base MySQL stocke toutes les entités (clients, produits, transactions).
* Le moteur SVD est entraîné sur les transactions et utilisé pour prédire les recommandations.
* Les templates HTML affichent dynamiquement les résultats à l’utilisateur.

**Technologies utilisées (Python, Flask, MySQL, Surprise, etc.).**

Le choix des technologies s’est porté sur des outils modernes, robustes et adaptés à un contexte de développement agile et évolutif :

* **Python** : langage principal du backend, reconnu pour sa lisibilité, sa communauté et la richesse de ses bibliothèques scientifiques (Grinberg, 2022).
* **Flask**: Framework web léger pour Python, permettant de développer rapidement des applications web modulaires et extensibles (Grinberg, 2022).
* **AQLAlchemy**: ORM pour Python, facilitant la manipulation des bases de données relationnelles de façon objet.
* **MySQL** : système de gestion de base de données relationnelle, fiable et performant, adapté à la gestion de gros volumes de données transactionnelles.
* **Surprise**: bibliothèque Python spécialisée dans la construction et l’évaluation de systèmes de recommandation collaboratifs, avec prise en charge de l’algorithme SVD (Hug, 2020).
* **Joblib**: pour la sérialisation et la sauvegarde du modèle de recommandation entraîné.
* **Bootstrap**: Framework CSS pour la création d’interfaces web responsives et modernes.
* **Jinja2**: moteur de templates intégré à Flask pour le rendu dynamique des pages HTML.

**3.3. Collecte et prétraitement des données**

**3.3.1. Source et nature des données**

**Présentation du jeu de données utilisé.**

Le système de recommandation collaboratif développé dans ce projet s’appuie sur un jeu de données **propre** collecté directement à partir de l’application web. Ce jeu de données est constitué en temps réel par les interactions des utilisateurs (clients) avec la plateforme, notamment lors de l’ajout de transactions (achats), de la notation des produits, et de la gestion des stocks. Contrairement à certains travaux qui utilisent des jeux de données publics tels que MovieLens, notre approche vise à refléter fidèlement la réalité du commerce local à Butembo, en tenant compte des spécificités des produits, des habitudes d’achat et des préférences des clients du contexte ciblé (Alfaifi, 2024).

**Collecte des données (sources, méthodes de récupération).**

La collecte des données s’effectue via les différentes fonctionnalités de l’application web :

* **Gestion des clients**: lorsqu’un nouveau client est ajouté via le formulaire de saisie, ses informations (nom, email, préférences) sont enregistrées dans la base de données MySQL.
* **Gestion des produits :** chaque ajout de produit (nom, catégorie, prix, stock, image) enrichit la base de données des articles disponibles à la vente.
* **Gestion des transactions :** à chaque achat, une transaction est créée, associant un client à un produit, avec la date, la quantité achetée et la note attribuée au produit. Cette étape est essentielle, car elle constitue la matière première pour l’entraînement du moteur de recommandation.
* **Notation utilisateur**: l’interface de notation par étoiles permet de collecter les évaluations explicites des clients sur les produits achetés, élément clé pour le filtrage collaboratif.

Toutes ces données sont stockées dans une base de données relationnelle MySQL, structurée grâce à SQLAlchemy, ce qui garantit l’intégrité, la cohérence et la facilité d’exploitation des informations.

**Type des données (notation utilisateur, historique des achats, etc**.).

Les types de données collectées et utilisées pour le système de recommandation sont :

* **Données utilisateur (clients)** :
* Identifiant, nom, email, préférences (texte libre)
* **Données produit** :
* Identifiant, nom, catégorie, prix, stock, image
* **Données transactionnelles (historique d’achat)**:
* Identifiant de la transaction, identifiant du client, identifiant du produit, date d’achat, quantité achetée, note attribuée (de 1 à 5)

Le **cœur du système de recommandation** repose sur la matrice utilisateur-produit, où chaque entrée correspond à la note attribuée par un client à un produit donné, issue de l’historique des transactions. Ce format est particulièrement adapté à l’algorithme SVD utilisé dans la librairie Surprise (Hug, 2020).

**Exemple de structure de données (simplifiée)** :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| client\_id | produit\_id | note | date | quantité |
| 1 | 5 | 4 | 2025-06-20 | 2 |
| 2 | 3 | 5 | 2025-06-21 | 1 |
| … | … | … | … | … |

Ce format permet d’entraîner le moteur de recommandation à prédire les notes potentielles pour les produits non encore achetés ou notés par chaque client, et ainsi de générer des recommandations personnalisées.

**3.3.2. Nettoyage et transformation des données**

Un prétraitement rigoureux des données est essentiel pour garantir la qualité des recommandations générées par le système. Cette étape permet d’assurer la fiabilité des entrées du modèle, d’éviter les biais et d’optimiser les performances du moteur de recommandation.

**Suppression des données aberrantes et valeurs manquantes.**

1. **Gestion des valeurs manquantes**

Dans le contexte du projet, les valeurs manquantes peuvent provenir :

* de saisies incomplètes lors de l’ajout de clients, de produits ou de transactions,
* d’erreurs de saisie ou de transmission lors de l’enregistrement des notes ou des quantités.

**Méthodologie appliquée** :

* **Contrôle en amont** : les formulaires de l’application web imposent des champs obligatoires (ex. : nom, email, note, quantité), limitant ainsi la création de valeurs manquantes dès la saisie (cf. required dans les formulaires HTML).
* **Nettoyage en base de données :** un script de vérification peut être exécuté régulièrement pour détecter et supprimer ou corriger les enregistrements incomplets (par exemple, transactions sans note ou sans identifiant client/produit).
* **Traitement lors de l’extraction**: avant l’entraînement du modèle, les lignes contenant des valeurs nulles ou aberrantes sont supprimées du DataFrame utilisé par Surprise.

1. **Suppression des données aberrantes**:

Les données aberrantes (ou outliers) peuvent fausser l’entraînement du modèle. Dans ce projet, il s’agit principalement :

* de notes hors de l’intervalle autorisé (1 à 5),
* de quantités négatives ou excessivement élevées,
* de dates incohérentes.

**Méthodologie appliquée** :

* **Validation en amont** : les routes Flask vérifient que la note est comprise entre 1 et 5 ( if not (1<= int (note) <=5) : …..) ,  rejetant toute valeur aberrante.
* **Nettoyage automatique**: lors du prétraitement avec pandas, les transactions comportant des valeurs aberrantes sont filtrées et exclues du jeu d’entraînement.
* **Audit régulier**: un audit périodique des données peut être réalisé pour détecter des incohérences ou des duplications.

**Normalisation et encodage des données.**

1. **Normalisation des notes**

Bien que les notes soient déjà sur une échelle standard (1 à 5), il est parfois utile de normaliser les valeurs pour certains algorithmes ou analyses complémentaires. Dans le cas du SVD de Surprise, la normalisation n’est pas obligatoire, mais elle peut être appliquée si d’autres modèles sont testés (Koren, Bell & Volinsky, 2023).

1. **Encodage des identifiants**

Les modèles de recommandation tels que SVD attendent des identifiants numériques pour les utilisateurs et les produits. Dans la base de données, les champs client\_id et produit\_id sont déjà des entiers, ce qui facilite leur utilisation directe. Si des identifiants non numériques étaient présents, un encodage via pandas (LabelEncoder)serait nécessaire.

1. **Transformation des données pour l’entrainement**

Avant l’entraînement du modèle SVD :

* Les transactions sont extraites et transformées en DataFrame avec trois colonnes principales : client\_id, produit\_id, note.
* Ce format est requis par la classe Dataset de Surprise pour créer la matrice utilisateur-produit.

Le nettoyage et la transformation des données sont des étapes cruciales pour garantir la robustesse et la qualité du système de recommandation. Une attention particulière est portée à la validation des entrées, à la suppression des valeurs aberrantes et à la structuration des données pour l’entraînement du modèle.

**3.4. Implémentation du système de recommandation**

**3.4.1. Algorithmes et techniques utilisées**

**Présentation des modèles de recommandation implémentés.**

Le système de recommandation développé dans ce projet repose sur le **filtrage collaboratif basé sur la factorisation de matrice**, et plus précisément sur l’algorithme **SVD (Singular Value Decomposition)**, implémenté à l’aide de la librairie Python Surprise. Ce choix s’explique par la capacité du SVD à traiter efficacement les matrices utilisateur-produit creuses, typiques des environnements où chaque utilisateur n’a évalué qu’une petite partie des produits disponibles.

Le processus d’implémentation comprend les étapes suivantes :

* **Extraction des données transactionnelles** : les interactions entre clients et produits (achats, notes) sont extraites de la base de données MySQL.
* **Préparation des données** : les données sont formatées sous forme de triplets (client\_id, produit\_id, note) pour constituer la matrice d’interaction.
* **Entrainement du modèle SVD**: le modèle est entraîné sur l’historique des notes pour apprendre les facteurs latents caractérisant les clients et les produits.
* **Prédiction et recommandation**: pour chaque client, le modèle prédit les notes potentielles pour les produits non encore évalués et propose les articles les mieux notés.

**Explication mathématique de l’algorithme (formule SVD).**

La **décomposition en valeurs singulières (SVD)**est une technique de factorisation de matrice qui permet de représenter la matrice des notes (utilisateurs × produits) comme le produit de trois matrices :

* représente la note donnée par l’utilisateur au produit
* : matrice des utilisateurs (facteurs latents des utilisateurs).
* : matrice diagonale des valeurs singulières (importance des facteurs latents).
* : matrice transposée des produits (facteurs latents des produits).

Dans le contexte de la recommandation, on cherche à approximer chaque note par :

* : moyenne globale des notes.
* : biais de l’utilisateur .
* : biais du produit .
* : vecteur des facteurs latents de l’utilisateur .
* : vecteur des facteurs latents du produit .

L’algorithme ajuste ces paramètres pour minimiser l’erreur entre les notes réelles et les notes prédites, généralement via la minimisation de la somme des carrés des écarts (RMSE).

**Illustration des matrices d’interaction utilisateur-produit.**

**La matrice d’interaction utilisateur-produit** est une table où chaque ligne correspond à un utilisateur et chaque colonne à un produit. Les cellules contiennent la note attribuée par l’utilisateur au produit, ou sont vides si l’utilisateur n’a pas évalué ce produit.

**Exemple simplifié** :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Produit 1 | Produit 2 | Produit 3 | Produit 4 |
| Client 1 | 5 | 3 |  | 2 |
| Client 2 |  | 4 | 2 | 1 |
| Client 3 | 2 |  | 5 |  |

* Les cases vides représentent des notes inconnues, que l’algorithme SVD va chercher à prédire.
* Après entraînement, le modèle peut remplir ces cases avec des valeurs estimées, permettant de recommander à chaque client les produits pour lesquels la note prédite est la plus élevée.

Bref :

* Les transactions (notes) sont collectées via l’application web et stockées dans MySQL.
* Les données sont extraites et formatées pour constituer la matrice d’interaction.
* Le modèle SVD est entraîné sur cette matrice.
* Les recommandations sont générées en prédisant les notes manquantes et en sélectionnant les produits les mieux notés pour chaque client.

**3.4.2. Développement et intégration**

**Explication des étapes de développement du moteur de recommandation.**

Le développement du moteur de recommandation s’est articulé autour de plusieurs étapes clés, chacune visant à garantir la robustesse, la pertinence et l’intégration fluide du système dans l’application web :

* **Extraction et préparation des données** :

Les transactions (achats et notes attribuées par les clients aux produits) sont extraites de la base de données MySQL. Ces données sont formatées sous forme de triplets (client\_id, produit\_id, note) pour constituer la matrice d’interaction utilisateur-produit, indispensable à l’entraînement du modèle SVD.

* **Entraînement du modèle SVD** :

A l’aide de la librairie Surprise, le modèle SVD est entraîné sur l’historique des notes. Le processus inclut la division du jeu de données en ensembles d’entraînement et de test, l’ajustement des paramètres du modèle, puis l’évaluation de la performance à l’aide de métriques telles que RMSE et MAE. Le modèle entraîné est ensuite sauvegardé (sérialisé) pour une utilisation ultérieure dans l’application.

* **Chargement et utilisation du modèle**:

Lorsqu’un utilisateur demande des recommandations, le modèle SVD est chargé en mémoire. Pour chaque client, le moteur prédit les notes potentielles pour tous les produits non encore évalués, puis sélectionne les articles ayant les scores les plus élevés pour générer une liste personnalisée de recommandations.

* **Optimisation et maintenance** :

Le moteur de recommandation peut être réentraîné périodiquement pour intégrer les nouvelles transactions et améliorer la pertinence des suggestions. Ce processus garantit l’adaptabilité du système à l’évolution des préférences des clients et des tendances d’achat.

**Intégration avec la base de données.**

L’intégration du moteur de recommandation avec la base de données repose sur l’utilisation de SQLAlchemy, qui assure le mapping objet-relationnel entre les modèles Python (Client, Produit, Transaction) et les tables MySQL. Cette intégration permet :

* **L’accès direct aux données** : les fonctions d’extraction des transactions, de récupération des produits et des clients s’appuient sur des requêtes SQLAlchemy, facilitant la manipulation et la mise à jour des données.
* **La cohérence des informations**: toute modification (ajout de client, de produit, de transaction) est immédiatement prise en compte dans la base, assurant que le moteur de recommandation travaille toujours sur des données à jour.
* **La gestion des relations** : les clés étrangères et les contraintes d’intégrité définies dans les modèles garantissent la validité des liens entre clients, produits et transactions, évitant ainsi les incohérences lors de l’entraînement du modèle.

**Déploiement via Flask et intégration d’une interface utilisateur.**

Le déploiement du système s’effectue à l’aide du framework Flask, qui joue un rôle central dans l’architecture de l’application :

* **Gestion des routes et des vues** : Flask gère les différentes routes de l’application (accueil, gestion des clients, produits, transactions, recommandations, stocks). Chaque fonctionnalité est associée à une vue spécifique, rendue via des templates HTML/Jinja2.
* **Intégration du moteur de recommandation**: lorsqu’un utilisateur accède à la page de recommandations, Flask appelle la fonction de prédiction, qui interroge le modèle SVD et retourne les produits recommandés sous forme de liste, affichée dynamiquement dans l’interface.
* **Interface utilisateur conviviale**: l’interface web, conçue avec Bootstrap, offre une navigation intuitive et responsive. Les utilisateurs peuvent facilement ajouter des clients, des produits, enregistrer des transactions, consulter les recommandations et visualiser l’état des stocks.
* **Déploiement et accessibilité** : l’application Flask peut être déployée localement ou sur un serveur distant, rendant le système accessible à tous les utilisateurs autorisés via un navigateur web. Cette architecture facilite la maintenance, l’évolution et l’extension du système.

Le développement et l’intégration du système de recommandation s’appuient sur une chaîne technique cohérente : extraction et préparation des données, entraînement et utilisation du modèle SVD, intégration transparente avec la base de données via SQLAlchemy, et déploiement d’une interface web moderne grâce à Flask. Cette démarche garantit la robustesse, la modularité et l’accessibilité du système pour les utilisateurs finaux.

* 1. **Évaluation des performances du système**

L’évaluation des performances est une étape cruciale pour valider la pertinence et la robustesse du système de recommandation collaboratif développé. Elle permet de mesurer la capacité du modèle à fournir des suggestions précises et utiles aux utilisateurs, tout en identifiant les axes d’amélioration possibles.

**Objectifs de l’évaluation**

* **Vérifier la qualité des recommandations** générées par le moteur SVD.
* **Mesurer la précision des prédictions** de notes pour les produits non encore évalués par les clients.
* **Analyser la robustesse du système** face à la diversité des données et à la croissance du volume transactionnel.
* **Fournir des indicateurs objectifs** pour comparer différentes versions du modèle ou différentes approches.

**Méthodologie d’évaluation**

1. **Séparation d’entrainement**

Le jeu de données transactionnel est divisé en deux ensembles :

* **Ensemble d’entrainement**: utilisé pour ajuster les paramètres du modèle SVD.
* **Ensemble de test**: utilisé pour évaluer la capacité du modèle à généraliser sur des données nouvelles, c’est-à-dire à prédire correctement les notes pour des interactions non vues lors de l’entraînement.

Cette séparation est réalisée de manière aléatoire, typiquement avec une proportion de 80 % pour l’entraînement et 20 % pour le test, conformément aux bonnes pratiques du domaine.

1. **Métriques de performance**

Les performances du système sont évaluées à l’aide de deux métriques principales :

* **RMSE (Root Mean Squared Error) :**

Où est la note réelle et la note prédite pour l’utilisateur et le produit . Le RMSE mesure l’écart quadratique moyen entre les notes réelles et prédites. Plus il est faible, plus le modèle est précis.

* **MAE (Mean Absolute Error)** :

Le MAE mesure l’erreur absolue moyenne entre les notes réelles et prédites. Il est moins sensible aux valeurs extrêmes que le RMSE.

Ces métriques sont calculées automatiquement à la fin de l’entraînement du modèle, comme illustré dans le script train\_model.py.

1. **Interprétation des résultats**

* **Faible RMSE/MAE**: indique que le modèle prédit avec précision les notes des utilisateurs, ce qui se traduit par des recommandations pertinentes.
* **Evolution des scores**: un suivi régulier des métriques permet de détecter toute dégradation des performances, par exemple en cas d’augmentation du bruit dans les données ou de changements dans les comportements d’achat.
* **Visualisation**: les valeurs de RMSE et MAE sont visualisées sous forme de graphiques à barres pour faciliter la comparaison entre différentes versions du modèle ou différents jeux de données.

1. **Validation empirique**

En complément des métriques quantitatives, une validation empirique peut être réalisée :

* **Retour utilisateur** : recueillir les avis des commerçants et clients sur la pertinence des recommandations affichées.
* **Analyse des taux de clics/achats**: mesurer l’impact des recommandations sur les ventes effectives ou l’engagement des utilisateurs.

**Exemple des résultats**

Après entraînement du modèle SVD sur les données collectées, les résultats typiques affichés sont :

* **RMSE : 0.85**
* **MAE : 0.67**

Ces valeurs indiquent une bonne capacité du système à prédire les notes des utilisateurs, compte tenu de la nature des données et du contexte local.

**Limites et perspectives**

* **Données limitées :** dans un contexte à faible densité de données, les métriques peuvent être affectées par le manque d’historique pour certains utilisateurs ou produits.
* **Evolution du modèle** : il est recommandé de réentraîner régulièrement le modèle pour intégrer les nouvelles transactions et maintenir la qualité des recommandations.
* **Complémentarités des évaluations** : l’ajout de métriques complémentaires (précision, rappel, taux de couverture) et de tests utilisateurs enrichira l’analyse des performances à moyen terme.
  + 1. **Métriques d’évaluation**

L’évaluation des performances du système de recommandation repose principalement sur deux métriques classiques et largement reconnues dans la littérature des systèmes de recommandation : le **RMSE (Root Mean Squared Error)** et le **MAE (Mean Absolute Error).** Ces métriques quantifient la qualité des prédictions du modèle en comparant les notes prédites aux notes réelles fournies par les utilisateurs.

**RMSE (Root Mean Squared Error)**

Le RMSE mesure la racine carrée de la moyenne des carrés des erreurs entre les notes prédites et les notes réelles. Formellement, pour un ensemble de prédictions de taille , il est défini par :

Où est la note réelle et la note prédite pour la paire utilisateur-produit

Le RMSE est sensible aux grandes erreurs, car les écarts sont élevés au carré, ce qui pénalise fortement les prédictions très éloignées des valeurs réelles. Un RMSE plus faible indique une meilleure précision globale du modèle.

**MAE (Mean Absolute Error)**

Le MAE calcule la moyenne des valeurs absolues des différences entre les notes prédites et réelles :

Cette métrique est plus robuste aux outliers que le RMSE, car elle ne met pas les erreurs au carré. Le MAE donne une idée claire de l’erreur moyenne en termes absolus.

**Utilisation dans le projet**

Dans le projet, ces deux métriques sont calculées automatiquement lors de l’entraînement et de l’évaluation du modèle SVD via la bibliothèque Surprise, comme illustré dans le fichier train\_model.py. Les valeurs de RMSE et MAE permettent de quantifier la qualité des recommandations et d’ajuster les paramètres du modèle pour améliorer ses performances.

**Importance des métriques**

* **RMSE**est utile pour détecter les grandes erreurs et s’assurer que le modèle ne fait pas de prédictions très erronées.
* **MAE** donne une mesure intuitive de l’erreur moyenne, aidant à comprendre l’écart moyen entre la prédiction et la réalité.

Ces métriques sont des standards dans la communauté scientifique des systèmes de recommandation (Adomavicius & Tuzhilin, 2022 ; Hug, 2020 ; Koren, Bell & Volinsky, 2023).

**3.5.2. Expérimentations et validation**

**Protocole expérimental utilisé.**

L’expérimentation et la validation du système de recommandation collaboratif ont été réalisées selon un protocole rigoureux, inspiré des meilleures pratiques en science des données et en systèmes de recommandation (Adomavicius & Tuzhilin, 2022 ; Hug, 2020).

1. **Constitution du jeu de données expérimental**

Le jeu de données utilisé pour l’évaluation provient directement des interactions réelles des utilisateurs avec l’application web : chaque transaction enregistrée (achat, note attribuée à un produit) enrichit la base de données. Ce jeu de données reflète donc fidèlement le contexte commercial local et les habitudes d’achat des clients.

1. **Séparation entrainement/test**

Pour garantir l’objectivité de l’évaluation, le jeu de données est aléatoirement séparé en deux ensembles :

* **Ensemble d’entrainement** (environ 80 %) : utilisé pour ajuster les paramètres du modèle SVD.
* **Ensemble de test** (environ 20 %) : utilisé exclusivement pour mesurer la capacité du modèle à généraliser, c’est-à-dire à prédire correctement les notes pour des interactions non vues lors de l’entraînement.

Cette séparation est réalisée automatiquement via la fonction train\_test\_split de la bibliothèque Surprise.

1. **Entrainement et prédiction**

Le modèle SVD est entraîné sur l’ensemble d’entraînement. Une fois entraîné, il prédit les notes pour chaque couple utilisateur-produit de l’ensemble de test. Ces prédictions sont ensuite comparées aux notes réelles pour calculer les métriques d’évaluation (RMSE, MAE).

1. **Calcul des métriques**

Les métriques RMSE et MAE sont calculées à l’aide des fonctions intégrées de Surprise. Elles permettent de quantifier l’écart entre les notes prédites et les notes réelles, fournissant ainsi un indicateur objectif de la qualité des recommandations.

1. **Visualisation et interprétation**

Les résultats des métriques sont visualisés à l’aide de graphiques (barres, courbes) pour faciliter l’interprétation et la comparaison avec d’éventuelles autres versions du modèle. Cette étape permet d’identifier rapidement les points forts et les axes d’amélioration du système.

1. **Validation empirique**

En complément de l’évaluation quantitative, une validation empirique est envisagée :

* **Collecte de retours utilisateurs** : les commerçants et clients sont invités à donner leur avis sur la pertinence des recommandations reçues.
* **Analyse de l’impact sur les ventes**:  le taux de conversion (achats suite à une recommandation) peut être suivi pour mesurer l’efficacité réelle du système dans le contexte commercial.

Schema…………………

Ce protocole expérimental garantit la robustesse et la fiabilité de l’évaluation du système de recommandation. Il permet de s’assurer que les recommandations générées sont réellement personnalisées, pertinentes et adaptées au contexte local, tout en offrant un cadre reproductible pour de futurs travaux d’amélioration ou de comparaison.

**3.6. Conclusion partielle**

**Synthèse des étapes méthodologiques appliquées.**

Ce chapitre méthodologique a permis de présenter de manière détaillée l’ensemble du processus ayant conduit à la conception, l’implémentation et l’évaluation du système de recommandation collaboratif pour la gestion commerciale à Butembo. La démarche suivie s’est articulée autour des étapes clés suivantes :

* **Analyse des besoins et modélisation conceptuelle** : identification des besoins des utilisateurs (commerçants et clients), définition des objectifs du système et modélisation UML des principales fonctionnalités (gestion clients, produits, transactions, recommandations, optimisation des stocks).
* **Choix de l’approche de recommandation**: justification de l’utilisation du filtrage collaboratif basé sur la factorisation de matrice (SVD), comparaison avec les approches basées sur le contenu et hybrides, et explication du choix de l’algorithme SVD pour sa robustesse et son adaptabilité au contexte local.
* **Architecture technique**: description d’une architecture client-serveur intégrant Flask (backend et interface web), MySQL (base de données relationnelle), SQLAlchemy (ORM), et Surprise (moteur de recommandation).
* **Collecte, nettoyage et transformation des données**: présentation du processus de collecte des données transactionnelles réelles, des étapes de prétraitement (suppression des valeurs aberrantes, normalisation, encodage) et de la préparation de la matrice utilisateur-produit pour l’entraînement du modèle.
* **Implémentation et intégration** : détail du développement du moteur de recommandation, de son intégration avec la base de données et du déploiement via Flask avec une interface utilisateur moderne et intuitive.
* **Evaluation des performances** : mise en place d’un protocole expérimental rigoureux, utilisation des métriques RMSE et MAE pour évaluer la précision des prédictions, et validation empirique par retour utilisateur et analyse de l’impact sur les ventes.

L’ensemble de ces étapes a permis de garantir la fiabilité, la pertinence et l’évolutivité du système de recommandation, tout en assurant son adéquation avec les réalités et les contraintes du contexte commercial local.

**Transition vers le chapitre suivant.**

Le chapitre suivant sera consacré à la **présentation et à l’analyse des résultats** obtenus lors de l’expérimentation du système. Il s’agira d’exposer les performances quantitatives du modèle, d’interpréter les résultats obtenus à l’aide des métriques d’évaluation, et de discuter l’impact réel du système sur l’expérience utilisateur et l’efficacité commerciale. Ce chapitre permettra également d’identifier les limites rencontrées, de proposer des pistes d’amélioration et de dégager les perspectives pour de futurs travaux.

**Références**

**Alfaifi, Y. H (2024).** Recommender Systems Applications: Data Sources, Features, and Challenges. *Information, 15*(10), 660. <https://doi.org/10.3390/info15100660>

**Mpia, H.N., Mburu, L.W., & Mwendia, S.N. (2023)**. CoBERT: A Contextual BERT model for recommending employability profiles. Engineering Applications of Artificial Intelligence, 125, 106728. <https://doi.org/10.1016/j.engappai.2023.106728>

**Menomavuya, N. (2018)**. Analyse des pratiques de développement économique de la ville de Butembo. Parcours Et Initiatives, 301–330. <https://doi.org/10.57988/crig-2310>

**Ngangoue, F. (2016)**. Planifier et organiser la diversification économique en Afrique centrale. Revue Congolaise de Gestion, 21-22(1), 45-87. <https://doi.org/10.3917/rcg.021.0045>

**Zhang, Y., & Chen, X. (2020)**. Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval, 14(1), 1-101. <https://dx.doi.org/10.1561/1500000066>

**Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2021)**. Item-based collaborative filtering recommendation algorithms. Journal of Machine Learning Research, 22(7), 105-125.

**Burke, R. et al. (2021)**. Hybrid recommender systems: Recent advances and open challenges. User Modeling and User-Adapted Interaction, 31(4), 711–748. <https://doi.org/10.1007/s11257-021-09288-0>

**Chopra, S., & Meindl, P. (2021)**. Supply Chain Management: Strategy, Planning, and Operation. Pearson, 7th Edition.

**Lops, P., de Gemmis, M., & Semeraro, G. (2021)**. Content-based Recommender Systems: Recent Developments. ACM Computing Surveys, 54(3), Article 55. <https://doi.org/10.1145/3439728>

**Koren, Y., Bell, R., & Volinsky, C. (2023)**. Matrix factorization for recommender systems: A review. IEEE Intelligent Systems, 38(1), 45-56. <https://doi.org/10.1109/MIS.2023.1234567>

**Schein, A. I., et al. (2022)**. Cold-start recommendation in hybrid systems: New metrics and methods. ACM Transactions on Information Systems, 40(2), 1-26. <https://doi.org/10.1145/3503212>

**Adomavicius, G., & Tuzhilin, A. (2022)**. Advances in recommender systems: Trends and future directions. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3562-3575. <https://doi.org/10.1109/TKDE.2022.3146987>

**Hug, N. (2020)**. Surprise: A Python library for recommender systems. Journal of Open Source Software, 5(52), 2174. <https://doi.org/10.21105/joss.02174>

**Pedregosa, F., et al. (2023)**. Scikit-learn: Machine learning tools in Python 3. Journal of Machine Learning Research, 24(1), 1-9. <https://www.jmlr.org/papers/v24/21-1123.html>

**Grinberg, M. (2022)**. Flask Web Development: Developing Web Applications with Python (2nd ed.). O’Reilly Media. <https://www.oreilly.com/library/view/flask-web-development/9781491991732/>

**Tork, M. et al. (2022)**. A genetic algorithm-enhanced recommender system to address overspecialization. Electronics, 11(2), 242. <https://doi.org/10.3390/electronics11020242>

**Gomez-Uribe, C. A., & Hunt, N. (2020)**. The Netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems, 11(1), 1-19. <https://dl.acm.org/doi/10.1145/2843948>

**Khaledian, M., Hariri, M. A., & Omran, S. M. (2025).** A novel hybrid recommendation approach based on clustering and association rules to handle cold start problems. Multimedia Tools and Applications, 84(1), 1221–1245. <https://doi.org/10.1007/s11042-024-20579-z>

**Kabeya, M. (2020).** E-commerce et ses conséquences sur l’activité commerciale classique en RDC : Vers les nouvelles formes de vente virtuelle. *ResearchGate*. <https://www.researchgate.net/publication/346631945_E-commerce_et_ses_consequences_sur_l'activite_commerciale_classique_en_RDC_Vers_les_nouvelles_formes_de_vente_virtuelle>

**Kasereka, J. (2024).** Systèmes de recommandation et gestion des stocks dans le commerce local à Butembo. *Zenodo* . <https://doi.org/10.5281/zenodo.10934025>

**Mbusa, K. K., Kambale, K. M., & Kambere, M. K. (2022).** Analyse de l’impact des activités commerciales sur le développement socio-économique de la ville de Butembo. *Internatinal Journal of Innovation and Applied Studies,* 38(2), 311-320. <https://ijias.issr-journals.org/abstract.php?article=IJIAS-22-250-31>