

R5B10  
MODÉLISATION MATHÉMATIQUES

# **SYSTEME DE RECOMMANDATION**

## **RAPPORT**

Rives Olivia  
Heraudeau Lucas

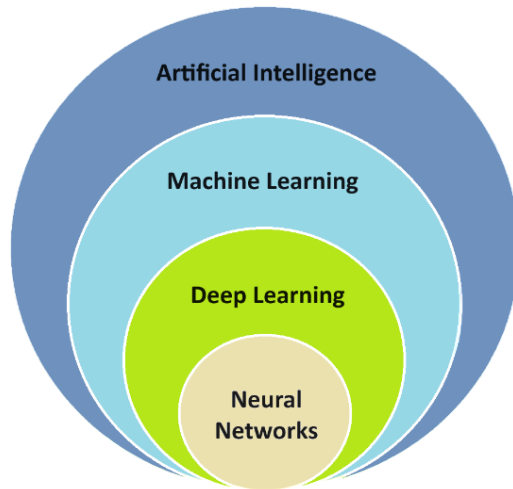
# Table des matières

Introduction.....	2
<b>L'algorithme de recommandation de Netflix.....</b>	<b>3</b>
Modèles Mathématiques dans les Systèmes de Recommandation.....	4
Explication des données utilisées : utilisateurs, items et notations.....	4
Filtrage basé sur l'utilisateur.....	4
Filtrage basé sur l'item.....	4
Mesures de Similarité.....	5
Similarité entre utilisateurs.....	5
Similarité entre items.....	5
Travail Réalisé.....	6
Analyse des résultats obtenus.....	7
<b>Conclusion.....</b>	<b>7</b>
Ressources Consultées.....	9

## Introduction

---

Le Machine Learning, une branche vitale de l'intelligence artificielle, se distingue comme un domaine en plein essor au sein de la science des données. Notre travail se concentre sur les systèmes de recommandation, qui permettent aux machines d'apprendre et de faire des prédictions en analysant de grandes quantités de données grâce à des modèles mathématiques.



Ces systèmes améliorent la manière dont nous naviguons sur Internet en recommandant des articles, des produits et des services qui correspondent à ce que nous aimons.

Ils deviennent de plus en plus précis en apprenant de ce que nous avons choisi par le passé pour offrir des suggestions de plus en plus personnalisées.

Devenus des éléments centraux des plateformes numériques, les systèmes de recommandations sont intégrés dans des services largement utilisés tels que Netflix, YouTube, Instagram et Facebook, touchant ainsi divers aspects de nos vies quotidiennes. Ces systèmes utilisent les données des utilisateurs, comme les avis, l'historique des achats et les interactions, pour créer des recommandations personnalisées, rendant ainsi le grand nombre d'options disponibles plus pertinent et adapté à chaque utilisateur.

Le machine learning joue un rôle crucial dans le développement de ces systèmes de recommandation. Il permet de traiter et d'analyser de grands volumes de données pour identifier des motifs, des tendances et des associations qui échapperaient à une analyse humaine traditionnelle. Par exemple, des techniques telles que le clustering ou la régression peuvent être utilisées pour prédire la note qu'un utilisateur pourrait donner à un item non encore évalué, tandis que la classification peut aider à déterminer si un utilisateur aimera ou non un item.

En outre, le machine learning améliore la capacité du système à s'adapter et à apprendre de nouveaux comportements d'utilisateurs au fil du temps, raffinant ainsi continuellement la qualité des recommandations. En utilisant des algorithmes d'apprentissage supervisé et non supervisé, les systèmes peuvent évoluer et s'ajuster dynamiquement pour répondre aux changements de goûts et de préférences des utilisateurs, ce qui est essentiel pour maintenir leur pertinence et leur précision.

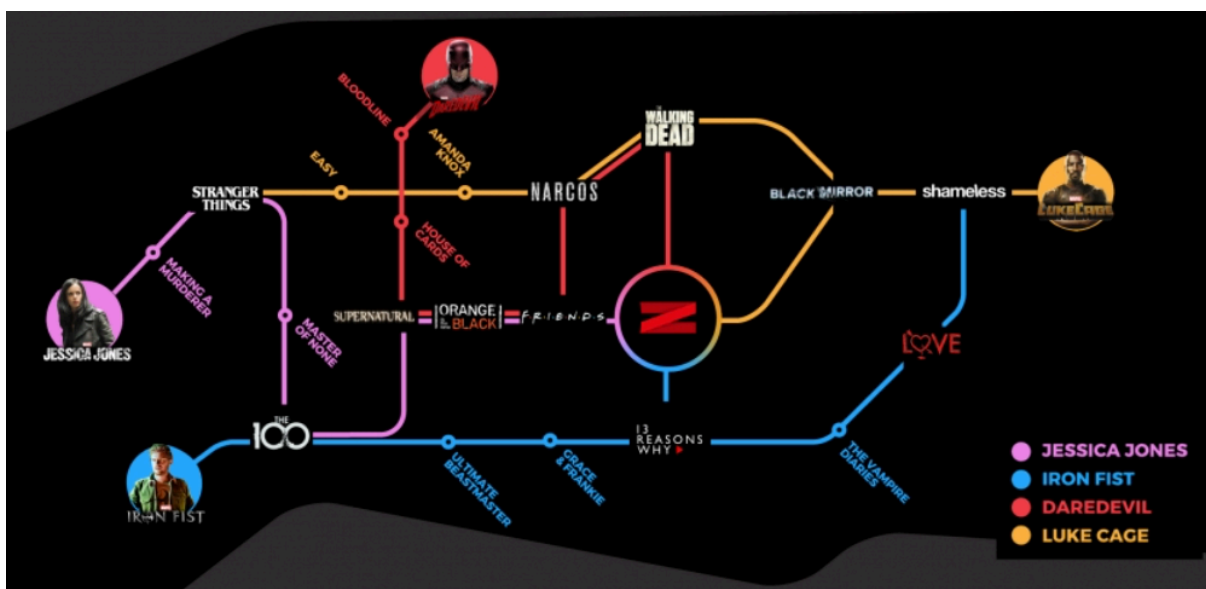
## L'algorithme de recommandation de Netflix

Nous allons vous présenter l'exemple de Netflix. Netflix utilise un système de recommandation sophistiqué pour personnaliser les expériences de ses utilisateurs en suggérant des films et des séries télévisées susceptibles de leur plaire. L'image ci-dessous illustre un exemple de la manière dont Netflix peut interconnecter diverses séries en fonction des goûts et des comportements de visionnage des utilisateurs.

Comme le montre le schéma, les séries telles que "Stranger Things", "Narcos" et "Orange is the New Black" sont reliées par des lignes colorées, qui suggère un chemin de recommandation basé sur des facteurs tels que le genre, les acteurs, la popularité ou des caractéristiques de contenu similaires. Cela permet à Netflix de créer un "chemin de visionnage" personnalisé pour chaque utilisateur, en augmentant la probabilité que les suggestions soient bien reçues.

Les algorithmes de Netflix vont au-delà des simples évaluations basiques et examinent un ensemble complexe de données pour recommander du contenu. Cela inclut non seulement les choix précédents des utilisateurs, mais aussi leur comportement de visionnage, le moment de la journée où ils regardent, l'appareil utilisé, et même la durée pendant laquelle ils regardent certaines séries ou films.

L'objectif de Netflix est de maximiser le temps passé sur sa plateforme en s'assurant que les utilisateurs ont toujours quelque chose d'intéressant à regarder, ce qui est essentiel dans un marché de streaming de plus en plus compétitif.



## Modèles Mathématiques dans les Systèmes de Recommandation

---

Explication des données utilisées : utilisateurs, items et notations.

Les systèmes de recommandation utilisent des jeux de données composés d'utilisateurs, d'items et de notations. Chaque utilisateur fournit des notes à certains des items disponibles. Ces données servent de base pour entraîner les modèles de recommandation. Dans notre cas, on a utilisé un jeu de données composé de notes de 0 à 5. La structure des données est une matrice dans laquelle chaque ligne représente un utilisateur et chaque colonne représente une évaluation ou un score pour un élément particulier.

*Les filtres collaboratifs sont l'une des approches les plus courantes pour les systèmes de recommandation. Ils se divisent en deux catégories principales :*

### Filtrage basé sur l'utilisateur

Ce type de filtrage recommande des items en fonction des préférences des utilisateurs similaires. Pour cela, il est nécessaire de définir une mesure de similarité entre utilisateurs et d'agréger les notes des utilisateurs similaires pour formuler des recommandations pertinentes.

### Filtrage basé sur l'item

Ici, le système identifie la similarité entre les items et recommande des items similaires à ceux déjà appréciés par l'utilisateur. Il implique de mesurer la similarité entre les items pour proposer des recommandations pertinentes.

L'objectif principal des systèmes de recommandation est de prédire la note qu'un utilisateur donnerait à un item particulier. En utilisant les données disponibles sur les notes attribuées par les utilisateurs, les algorithmes tentent de prédire les notes manquantes ou de suggérer de nouveaux items susceptibles d'intéresser l'utilisateur en se basant sur ses préférences passées.

Les modèles mathématiques sont essentiels dans les systèmes de recommandation, car ils permettent de formaliser les relations entre utilisateurs, items et leurs interactions (notations). Des approches telles que la régression, les méthodes matricielles, les techniques de factorisation de matrices et diverses mesures de similarité sont utilisées pour créer des modèles prédictifs.

## Mesures de Similarité

### Similarité entre utilisateurs

La similarité de Pearson mesure la corrélation linéaire entre deux ensembles de données. Elle évalue la similarité des préférences entre utilisateurs en prenant en compte la moyenne de leurs notes et la corrélation de leurs notations par rapport à cette moyenne. Cette mesure varie de -1 (corrélation négative totale) à 1 (corrélation positive totale), avec 0 n'indiquant aucune corrélation linéaire.

La similarité cosinus évalue la similarité directionnelle entre deux vecteurs (dans ce cas, les vecteurs représentent les préférences des utilisateurs). Elle mesure l'angle entre ces vecteurs dans un espace multidimensionnel et une valeur de 1 indique une similarité parfaite tandis que 0 indique une absence de similarité.

### Similarité entre items

La similarité entre items est cruciale pour le filtrage basé sur l'item.

Pour calculer cette similarité, on peut utiliser des mesures comme :

- Coefficient de corrélation de Pearson entre les notations des items.
- Similarité Cosinus entre les vecteurs représentant les notations des items.

Ces mesures permettent de quantifier la proximité entre les éléments du système de recommandation. Elles sont fondamentales pour déterminer les relations entre utilisateurs ou items. En utilisant ces mesures, les systèmes de recommandation peuvent identifier les utilisateurs similaires ou les items semblables, ce qui est crucial pour générer des recommandations précises et pertinentes.

Le choix de la mesure de similarité dépend souvent de la nature des données et des caractéristiques des utilisateurs ou des items. Certaines mesures peuvent être plus adaptées à certaines situations ou à des types de données spécifiques.

## Travail Réalisé

---

Au lancement du projet, nous avons consacré du temps à établir une stratégie de travail structurée, en créant un tableau sur Trello pour une meilleure organisation. Tout d'abord, nous avons réservé du temps pour comprendre le concept de la théorie des systèmes de recommandations.

Nous avons débuté par une étude approfondie des algorithmes de recommandation, en particulier le filtrage collaboratif. Notre attention s'est portée sur deux mesures de similarité clés : la similarité de Pearson et la similarité cosinus. Ces mesures ont été appliquées pour évaluer les relations entre les utilisateurs dans nos jeux de données (complets et incomplets).

Au niveau de la démarche algorithmique, on a importé les bibliothèques nécessaires, notamment Pandas pour la manipulation de données et NumPy pour les calculs numériques. Les données provenaient de deux jeux de données distincts : l'un incomplet et l'autre complet.

Pour plus de précision sur les jeux de données sur lesquels nous travaillons, nous avons rajouté une fonction afin de calculer le nombre d'utilisateurs dans chaque fichier (complet et incomplet).

Après le chargement, on a entamé la gestion des valeurs manquantes dans le jeu de données incomplet. On a remplacé les valeurs "-1" par des NaN pour symboliser les données manquantes, facilitant ainsi leur manipulation ultérieure.

Pour l'application de la similarité cosinus, on a remplacé ces NaN par des zéros. Cette étape est cruciale pour calculer la similarité entre les utilisateurs, en considérant les notes manquantes comme des absences de préférence.

Ensuite, on a utilisé la similarité de Pearson et la similarité cosinus pour évaluer la corrélation entre les utilisateurs. Ces calculs ont fourni des mesures quantitatives de la ressemblance entre les préférences des utilisateurs, un élément clé pour comprendre les relations entre eux.

Pour faciliter l'analyse, on a affiché ces similarités pour le premier utilisateur dans les données, offrant ainsi une perspective pratique de ces valeurs.

Par la suite, on a mis en place des fonctions de filtrage basées sur les utilisateurs et les items. Ces fonctions étaient essentielles pour évaluer la prédiction des préférences manquantes d'un utilisateur pour des items non notés précédemment.

En appliquant ces filtres aux données incomplètes, on a pu évaluer la précision de ces prédictions en utilisant le RMSE qui signifie "Root Mean Squared Error" et en français "erreur quadratique moyenne". Cette mesure a permis de quantifier la différence entre les

prédictions et les vraies notes, fournissant ainsi un indicateur de la performance des filtres de recommandation.

Enfin, on a affiché les résultats du RMSE pour les filtres utilisateur et item basés sur ces données incomplètes. Ces résultats ont fourni une évaluation quantitative de la précision de nos modèles de recommandation pour prédire les préférences manquantes des utilisateurs.

## Analyse des résultats obtenus

...

## Conclusion

L'objectif des algorithmes de recommandation, comme celui utilisé par Netflix, est de garder les spectateurs impliqués et intéressés. L'algorithme de Netflix est particulièrement avancé, capable d'adapter les suggestions de films et de séries aux goûts individuels de chaque utilisateur.

Ces systèmes intelligents utilisent beaucoup de données pour prédire ce qui pourrait plaire aux utilisateurs. Ils aident non seulement à garder les spectateurs accrochés en leur montrant des choses qu'ils aiment, mais ils les aident aussi à découvrir de nouvelles choses qu'ils n'auraient peut-être pas trouvées par eux-mêmes.

Ce travail a apporté des contributions significatives à notre compréhension des systèmes de recommandation et d'apprentissage automatique, en offrant une immersion pratique dans ces domaines. Notre exploration a permis une concrétisation des concepts théoriques, mettant en œuvre des méthodes mathématiques pour la création de modèles de recommandation et leur évaluation.

L'application des connaissances acquises à des ensembles de données réels a été d'une importance capitale. Nous avons manipulé ces données pour calculer les similarités entre utilisateurs et items, construire des filtres utilisateur et item, puis évaluer la performance de ces modèles grâce à des mesures comme le RMSE et la comparaison des similarités. Ces analyses ont offert une perspective sur la qualité des recommandations produites et ont permis d'identifier les forces et les lacunes des approches utilisées.

Ce travail a également souligné l'importance de la préparation des données dans la construction de systèmes de recommandation robustes. La transformation des valeurs manquantes, la gestion des données aberrantes et la compréhension profonde des jeux de données ont été des étapes cruciales pour assurer la fiabilité et la précision des modèles.

De manière plus significative, cette expérience a éclairé l'influence des systèmes de recommandation sur les décisions et les comportements des utilisateurs. Comprendre les mécanismes sous-jacents des algorithmes de recommandation et leur impact sur les choix des utilisateurs constitue un aspect fondamental dans un paysage numérique en constante évolution.



En conclusion, cette étude a élargi nos compétences pratiques dans la construction et l'évaluation de systèmes de recommandation. Elle a ainsi mis en lumière la complexité et le potentiel de ces systèmes dans des contextes réels, ouvrant la voie à des explorations plus approfondies et à des améliorations continues dans ce domaine en constante évolution.

## Ressources Consultées

---

Pour comprendre correctement le fonctionnement de l'algorithme de recommandation :

<https://www.mediego.com/fr/blog/principaux-algorithmes-de-recommandation/>

<https://www.youtube.com/watch?v=0io52iBxoxY>

<https://www.youtube.com/watch?v=n3RKsY2H-NE>

<https://www.youtube.com/watch?v=IFcpp70vg2s>

<https://datascientest.com/quest-ce-qui-se-cache-derriere-lalgorithme-de-recommandation-di-nstagram-reponse-ici>

Pour la corrélation de Pearson

[http://www.biostat.ulg.ac.be/pages/Site\\_r/corr\\_pearson.html](http://www.biostat.ulg.ac.be/pages/Site_r/corr_pearson.html)

<https://support.minitab.com/fr-fr/minitab/21/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/all-statistics-and-graphs/>

Afin de bien comprendre l'erreur quadratique moyenne, nous avons effectué plusieurs recherches, voici la liste des sites que nous avons consulté :

<https://datascientest.com/erreur-quadratique-moyenne>

<https://www.youtube.com/watch?v=pZuxM18jA-A>