



*Université Ibn Tofail*

*Department d'infomatiques*

**MASTER BIG DATA & CLOUD COMPUTING**

---

*Data mining project*

---

# Movie Recommendation System using collaborative filtering in R

Superviser par : MOUMEN Idriss

Presenter par: ZIYAD Mouhssine & NAJAR Saad

**ANNEE UNIVERSITAIRE: 2021-2023**

## Remerciement

Tout d'abord nous tenons à remercier dans un premier temps, toute l'équipe pédagogique de notre université et les intervenants professionnels responsables de la formation technique de ses étudiants, et spécialement du master Big Data et cloud computing responsable de notre formation intéressante.

Cela étant, Nous présentons nos remerciements les plus considérables à notre professeur : Moumen Idriss pour sa motivation, ses conseils, ses qualités techniques et professionnelles et pour tous les efforts fournis favorisant l'aboutissement du de ce mini-projet que nous menons.

# Table de matière

Introduction.....	6
I. System de Recommandation.....	7
1. C'est quoi un system de recommandation ? .....	7
2. Comment fonctionnent les systèmes de recommandation ?.....	7
3. Techniques d'un system de recommandation .....	8
i. Filtrage collaboratif .....	8
ii. Filtrage collaboratif utilisateur .....	8
iii. Filtrage collaboratif objets .....	9
iv. Système basé sur le contenu.....	9
v. Système basé sur la connaissance.....	9
vi. Le problème de démarrage à froid.....	9
vii. Les systèmes hybrides.....	10
4. Exemples réels du système de recommandation.....	11
II. Data Mining.....	12
1. Définition.....	12
2. Fonctionnement du DataMining .....	12
3. Domaine d'application du DataMining.....	13
III. Etude de cas « Movie Recommendation System » .....	14
1. Environnement et outils de travaux.....	14
2. Importation du data .....	15
3. Sélection du 3 films favoris.....	16
4. Filtrage du rating .....	17
5. Filtrage spécifique.....	17
6. Semi-Recommandation List .....	18
7. Création du genre favoris.....	19
8. Recommandation cible .....	19
Conclusion .....	20
Bibliographie.....	21

## Table de figure

Figure 1 : Filtrage collaboratif.....	8
Figure 2 : Importation du data .....	15
Figure 3 : user's favorite movie .....	16
Figure 4 : filtrage du rating .....	17
Figure 5 : Filtrage spécifique .....	18
Figure 6 : semi-Recommendation list.....	18
Figure 7 : genres favoris.....	19
Figure 8 : Recommendation cible .....	19

## Résumé

À mesure que la quantité d'informations augmente, il devient difficile de savoir où trouver quelles données. Il existe des technologies informatiques qui peuvent faciliter cette enquête et extraire des informations pertinentes.

L'un d'eux est le processus de recommandation. Il guide l'exploration de l'utilisateur et trouve des informations qui peuvent être pertinentes pour l'utilisateur. Il s'agit d'une forme spéciale de filtrage visant à afficher des informations susceptibles d'intéresser l'utilisateur (films, musique, livres, images, web, etc.).

Les systèmes de recommandation introduisent des concepts uniques pour les recommandations basées sur la récupération d'informations, le filtrage, l'apprentissage automatique, les approches collaboratives.

## Introduction

Le développement du monde informatique, en particulier d'Internet, a considérablement changé nos vies à bien des égards. Nous faisons des achats sur Internet, recherchons des informations sur Internet, lisons des actualités sur Internet, regardons des films sur Internet et écoutons de la musique sur Internet. Disponible sur Internet. Avec cet engouement, Internet s'est révélé être une place de marché économique très rentable et un moyen de communication très accessible, attirant tous les acteurs de l'économie, de la politique et de la société. Tous les fournisseurs de services en ligne souhaitent s'implanter sur le web et attirer le plus d'utilisateurs possible. Cependant, la pléthore de services électroniques, de ressources et d'informations sur le Web rend cette tâche difficile, en particulier pour les utilisateurs.

Nous nous noyons dans cette mer d'informations et donnons accès aux ressources pertinentes travail fastidieux. Il devient alors impératif d'assister l'utilisateur et de faciliter son accès aux ressources susceptibles de l'intéresser et adaptées à ses besoins personnels. Parmi les solutions envisagées, on trouve l'adaptation des interfaces pour faciliter l'exploration et la recherche au sein d'un service, les systèmes à base de navigation sociale, la recherche personnalisée sur le web, l'aide à la navigation personnalisée, les outils statistiques aidant les utilisateurs à mieux définir leurs recherches par mots-clefs, la présentation personnalisée du contenu pour le web etc. Les systèmes de recommandation ont émergé comme un domaine de recherche à part entière au milieu des années 1990 à la suite de la publication des premiers articles sur le filtrage collaboratif. Bien que le domaine de recherche des systèmes de recommandation soit plus récent et moins répandu que le filtrage d'information et la recherche documentaire et en particulier les moteurs de recherche, son intérêt dans l'industrie des e-services n'en est pas moins important.

Les systèmes de recommandation sont pour beaucoup d'entre nous des entités mystérieuses qui semblent deviner nos pensées. Pensons simplement à Netflix, qui nous suggère des films, ou encore à Amazon, qui croit connaître les produits que nous devrions acheter. Depuis leurs débuts, les outils de recommandation ont été améliorés et raffinés et ils ne cessent d'enrichir l'expérience utilisateur. Bien que plusieurs d'entre eux soient des ensembles très complexes, leur fondement et leurs racines demeurent très simples. Je me suis penché sur la question afin de démystifier pour vous ces étranges systèmes.

# I. System de Recommendation

## 1. C'est quoi un system de recommandation ?

Les systèmes de recommandation sont une forme spécifique de filtrage de l'information (SI) visant à présenter les éléments d'information (films, musique, livres, news, images, pages Web) qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence, et cherche à prédire l'« avis » que donnerait un utilisateur. Ces caractéristiques peuvent provenir de :

- L'objet lui-même, on parle « d'approche basée sur le contenu » ou content-based approach ;
- L'utilisateur ;
- L'environnement social, on parle d'approche de filtrage collaboratif ou collaborative filtering.

## 2. Comment fonctionnent les systèmes de recommandation ?

Les systèmes de recommandation ont pour premier rôle d'identifier le sous-groupe d'utilisateurs auquel appartient un utilisateur afin de lui proposer des résultats susceptibles de l'intéresser. L'identification de sous-groupes d'utilisateurs auquel appartient un utilisateur se fait généralement en fonction de l'historique d'utilisation du service par cet utilisateur. Le système de recommandation peut toutefois s'appuyer des caractéristiques connues sur l'utilisateur (L'âge, la catégorie socio-professionnelle, le sexe, le secteur professionnel.) ou sur une combinaison de ces caractéristiques et de son historique. Il ne reste alors au système de recommandation qu'à trouver les autres utilisateurs partageant le plus de points communs avec cet utilisateur, analyser les items les plus commandés, partagés ou plébiscités par ces utilisateurs afin de pouvoir proposer une sélection personnalisée d'items recommandés. Les systèmes de recommandation reposent généralement sur un algorithme de clustering afin de distinguer les différents sous-groupes d'utilisateurs, pouvant être entraîné en apprentissage non supervisé.

### 3. Techniques d'un system de recommandation

#### i. Filtrage collaboratif

L'une des premières techniques utilisées et qui demeure encore aujourd'hui parmi les plus simples et les plus efficaces est le filtrage collaboratif. Ce procédé en trois étapes commence par la collecte d'information sur les utilisateurs. Puis, on forme une matrice afin de calculer des associations. Finalement, nous sommes en mesure de faire une recommandation avec un niveau de confiance assez élevé. Cette recommandation se divise en deux grandes catégories : l'une se base sur les utilisateurs et l'autre sur les items qui composent l'environnement.

#### ii. Filtrage collaboratif utilisateur

L'idée derrière le filtrage collaboratif utilisateur consiste à trouver des utilisateurs qui ont des goûts similaires à l'utilisateur cible. Plus concrètement, si Jean-Pierre et Jason ont évalué plusieurs films d'une manière similaire dans le passé, nous pouvons considérer ceux-ci comme des utilisateurs semblables et ainsi utiliser les cotes de Jean-Pierre pour prédire les cotes de Jason. Par exemple, si Jean-Pierre a beaucoup apprécié les films Return Of The Jedi et The Empire Strikes Back et que Jason a également aimé le film Return Of The Jedi, alors il serait judicieux de suggérer à Jason The Empire Strikes Back. Généralement, un nombre restreint d'utilisateurs semblables à Jason est suffisant pour prédire les évaluations de celui-ci.

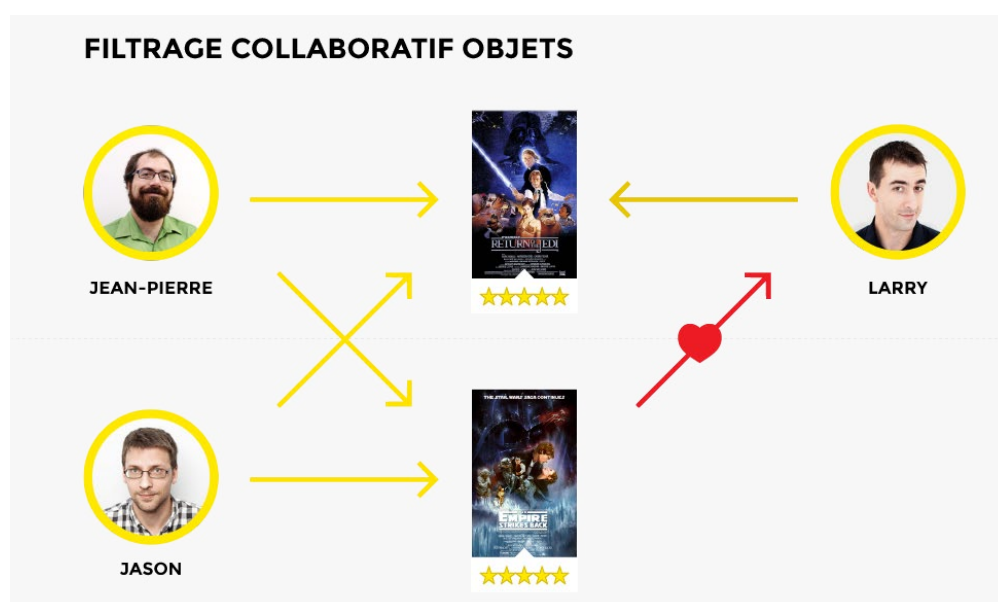


Figure 1 : Filtrage collaborative



### iii. Filtrage collaboratif objets

Le processus est simple. On calcule la ressemblance de deux éléments en se basant sur les notes données par les utilisateurs. Le filtrage collaboratif objets (ou items) est souvent favorisé puisque celui-ci ne comporte aucun des inconvénients du filtrage utilisateur. D'abord, les items du système (ici des films) ne changent pas, donc les suggestions seront plus pertinentes. De plus, il y a souvent moins d'éléments que d'utilisateurs, ce qui diminue le temps de traitement. Finalement, il est beaucoup plus complexe de tricher avec ces systèmes.

### iv. Système basé sur le contenu

Dans les systèmes de recommandation basés sur le contenu, ce sont les attributs descriptifs des éléments qui sont utilisés pour formuler des recommandations. Le terme « contenu » fait référence à ces descriptions. Par exemple, en regardant l'historique d'écoute de Sophie, le système remarque que celle-ci semble apprécier le genre country. Il sera donc en mesure d'ajouter plus de titres du même genre à ses recommandations.

### v. Système basé sur la connaissance

Les systèmes de recommandation fondés sur la connaissance sont particulièrement utiles dans le cas d'articles qui sont rarement achetés. On parle ici de maisons, d'automobiles, de services financiers et même d'objets de luxe. Dans de tels cas, le processus de recommandation souffre souvent d'un manque d'évaluation des produits. Il n'utilise alors pas les évaluations des utilisateurs, mais sera plutôt effectué sur la base des similitudes entre les exigences des clients et des descriptions d'articles ou grâce à l'utilisation de contraintes spécifiant les exigences des utilisateurs. En ce qui concerne les contraintes, si applicables, elles sont le plus souvent implémentées par les experts du domaine. Par exemple, lorsque l'utilisateur spécifie clairement que le prix de la maison recherchée ne doit pas dépasser X, le système doit prendre en considération cette spécification.

### vi. Le problème de démarrage à froid

L'un des problèmes majeurs des systèmes de recommandation est que le nombre d'évaluations initialement disponibles est relativement faible. Que faire lorsqu'un nouvel utilisateur n'a pas encore noté de films, ou lorsqu'un nouveau film est ajouté au système ? Dans de tels cas, il devient plus difficile d'appliquer les modèles traditionnels de filtrage collaboratif. Bien que les méthodes qui s'appuient sur le contenu et sur les connaissances soient plus robustes que les modèles collaboratifs, lorsqu'il s'agit de démarrage à froid, ces contenus ou connaissances pourraient ne

pas toujours être disponibles. Par conséquent, un certain nombre de procédés spécifiques comme les systèmes hybrides ont été conçus. Ils visent à remédier au problème du démarrage à froid dans le contexte des systèmes de recommandation.

#### vii. Les systèmes hybrides

Notons que les différents types de systèmes présentés ont tous des forces et des faiblesses et se basent sur différentes données pour générer des suggestions. Certains systèmes de recommandation, comme ceux qui s'appuient sur la connaissance, sont plus efficaces dans les environnements de démarrage à froid, où la quantité de données est limitée. D'autres systèmes de recommandation, comme les méthodes collaboratives, sont plus efficaces lorsque plusieurs données sont disponibles. Dans de nombreux cas, alors que les données sont très variées, nous avons la flexibilité d'utiliser différentes méthodes pour la même tâche. Nous pouvons donc combiner plusieurs techniques afin d'améliorer la qualité du système dans son ensemble. De nombreuses techniques de combinaisons ont été explorées, parmi lesquelles :

- **Pondéré** : un poids différent donné aux recommandations de chaque technique utilisée afin de favoriser certaines d'entre elles.
- **Mixte** : un seul ensemble de recommandations, sans préférences.
- **Augmenté** : les suggestions d'un système sont utilisées comme entrée pour le suivant, et ainsi de suite jusqu'au dernier.
- **Commutation** : choix d'une méthode au hasard.

#### 4. Exemples réels du système de recommandation.

The Netflix logo, consisting of the word "NETFLIX" in a bold, red, sans-serif font.

Le système de recommandation de **NETFLIX** est l'un des meilleurs du marché. Bien qu'il soit difficile à croire, avant qu'il ne devienne grand, Netflix était un service de location de DVD qui a débuté en 1997. Le système de recommandation qu'ils utilisaient initialement s'appelait « **Cinematch** ». Depuis lors, ils se sont constamment améliorés et c'est maintenant l'un des meilleurs systèmes de recommandation.

The Amazon logo, featuring the word "amazon" in a black, lowercase, sans-serif font, with a curved orange arrow underneath it.

**Amazon** a lancé le filtrage collaboratif basé sur les éléments en 1998, et il n'y a pas eu de retour en arrière ! Amazon a un magasin hautement personnalisé pour chacun de ses utilisateurs. Tout le monde voit des produits différents lorsqu'il entre dans l'interface utilisateur d'Amazon. Et cela a sûrement fait des merveilles pour l'entreprise, car elle est maintenant devenue la plus grande entreprise de commerce électronique au monde, surpassant d'autres géants du commerce électronique comme Alibaba. Et, le système de recommandation très puissant contribue à **35%** de **son chiffre d'affaires**.

The YouTube logo, featuring the word "You" in black and "Tube" in white inside a red rounded rectangle.

**YouTube** a commencé en 2005 en tant que plate-forme de partage de vidéos et est maintenant devenue l'une des applications les plus utilisées dans le monde, créant de nombreux nouveaux emplois et étant un pionnier dans la création de contenu.

Le système de recommandation YouTube est assez complexe et utilise les réseaux de neurones profonds (NN) pour fournir un contenu frais et pertinent à chaque utilisateur individuel. L'entreprise a même été confrontée à un défi pour développer son système de recommandation car elle devait gérer le plus grand ensemble de données sur les comportements des utilisateurs ainsi que la quantité de contenu qu'il contenait.

## II. Data Mining

### 1. Définition

Le Data mining est la pratique consistant à rechercher automatiquement de grandes quantités de données afin de découvrir des tendances et des modèles qui vont au-delà de la simple analyse. Il est souvent couplé au Deep Learning et au machine learning. Data Mining utilise des algorithmes mathématiques sophistiqués pour segmenter les données et évaluer la probabilité d'événements futurs. Data Mining est également connue sous le nom de découverte de connaissances dans les données.

Les propriétés clés de Data Mining sont :

- Découverte automatique des motifs
- Prévion des résultats probables
- Création d'informations exploitables

Mettre l'accent sur les grands ensembles de données et les grandes bases de données

Data Mining peut répondre à des questions qui ne peuvent être résolues par de simples techniques d'interrogation et de rapport.

### 2. Fonctionnement du DataMining

- **Traitement des données**

Cet outil d'analyse statistique (analyse de données) est parfaitement adapté au traitement d'une grande quantité de données, surtout à l'ère du Big Data. Le Data Mining utilise des supports informatiques dont la capacité de stockage est importante. Ainsi, il est possible de trier et de mettre en ordre les informations. De façon concrète, les données recueillies peuvent être relatives aux caractéristiques des produits ou services, à la gestion de la production ou encore aux comportements des clients.

- **Les techniques utilisées**

Dans le domaine du Data Mining, de nombreuses techniques peuvent être utilisées selon le volume des données, leur nature et le type d'analyse à réaliser. Celles-ci se

servent de méthodes supervisées comme la classification ou la régression et de méthodes non supervisées comme la segmentation ou la réduction de dimension.

- **L'utilisation des algorithmes**

Par ailleurs, le Data Mining permet d'avoir des données intelligibles par l'utilisation d'algorithmes provenant de diverses disciplines, dont l'informatique, l'intelligence artificielle et les statistiques. Le processus assure, en ce sens, l'identification des corrélations entre les ensembles de données et l'extraction des connaissances utiles à l'entreprise.

### 3. Domaine d'application du DataMining

- **Le scoring**

Les logiciels de Data Mining sont employés dans plusieurs domaines. C'est le cas du scoring qui consiste en marketing à attribuer une note à un client ou à un prospect. Cela permet de déterminer le profil client idéal. Les opérations marketing seront ainsi mieux ciblées.

- **Prévenir les risques de crimes**

À partir du Data Mining, on peut aussi prévenir de potentiels crimes. De nombreuses expériences ont même été menées dans ce domaine. Aux États unis par exemple, le Data Mining a permis d'identifier les associations criminelles et de connaître des plages horaires où les crimes pouvaient se produire.

- **La détection des fraudes**

Le Data Mining est également une solution efficace pour détecter les fraudes. Il utilise la classification sur les données, un mécanisme assurant l'identification des données qui sortent de l'ordinaire, dont l'empreinte diffère des comportements normaux. Le Data Mining décèle les cas suspects à surveiller.

### III. Etude de cas « Movie Recommendation System »

#### 1. Environnement et outils de travaux



scientifique), qui vise à simplifier la gestion des paquets et de déploiement. Les versions de paquetages sont gérées par le système de gestion de paquets conda.

**Anaconda** est une distribution libre et open source des langages de programmation Python et R appliqué au développement d'applications dédiées à la science des données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul



**Visual Studio Code** est un éditeur de code source léger mais puissant qui s'exécute sur votre bureau. Il est livré avec un support intégré pour JavaScript, TypeScript et Node.js et dispose d'un riche écosystème d'extensions pour d'autres langages (Python, PHP, Java...) et moteurs d'exécution (tels que .NET et Unity).



ouverts et des services pour l'informatique interactive. Jupyter est une évolution du projet IPython. Jupyter permet de réaliser des calepins ou notebooks, c'est-à-dire des programmes contenant à la fois du texte en markdown et du code. Ces calepins sont utilisés en science des données pour explorer et analyser des données.

## 2. Importation du data

On a choisi de travailler avec deux datasets, movies.csv et rating.csv

Les figures ci-dessus montre l'importation du dataset « movie\_data » et « rating\_data », ainsi que l'affichage du 6 premières lignes.

```
▷ 'data.frame': 10329 obs. of 3 variables:
  movie_data <- read.csv("movies.csv", stringsAsFactors = FALSE)
  rating_data <- read.csv("ratings.csv")
[138] ✓ 0.2s
```

```
▷ #affichage du 6 premiere lignes du movie_data
  head(movie_data)
[139] ✓ 0.8s
```

... A data.frame: 6 × 3

	movieId	title	genres
	<int>	<chr>	<chr>
1	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
2	2	Jumanji (1995)	Adventure Children Fantasy
3	3	Grumpier Old Men (1995)	Comedy Romance
4	4	Waiting to Exhale (1995)	Comedy Drama Romance
5	5	Father of the Bride Part II (1995)	Comedy
6	6	Heat (1995)	Action Crime Thriller

```
#affichage du 6 premiere lignes du movie_data
  head(rating_data)
[140] ✓ 0.8s
```

... A data.frame: 6 × 4

	userId	movieId	rating	timestamp
	<int>	<int>	<dbl>	<int>
1	1	16	4.0	1217897793
2	1	24	1.5	1217895807
3	1	32	4.0	1217896246
4	1	47	4.0	1217896556
5	1	50	4.0	1217896523
6	1	110	4.0	1217896150

Figure 2 : Importation du data

### 3. Sélection du 3 films favoris

Dans cette étape, on a choisi de sélectionner 3 films favoris afin de filtrer les utilisateurs d'après ces derniers. Dont le but de d'extraire les communes significations favorites entre eux.

```
#creation du 3 films favoris de la recommandation cible
fav1 <- ("Toy Story (1995)")
fav2 <- ("Jumanji (1995)")
fav3 <- ("Heat (1995)")

[143] ✓ 0.5s

#filtrage d'utilisateur qui ont regarder les 3 films favoris
user_fav_mov <- movie_data[movie_data$title %in% c(fav1,fav2,fav3),1]
user_fav_mov
typeof(x)

[144] ✓ 0.6s

... 1 · 2 · 6

</> 'integer'
```

Figure 3 : user's favorite movie



## 4. Filtrage du rating

Afin de rassembler les utilisateurs qui vont une influence sur la recommandation cible, on a filtré les utilisateurs qui on regarde les 3 films favoris dont l'évaluation est supérieure de 4.5 (4.5~5 stars)

```
#user_rated_mov représente l'utilisateur qui ont favoris les 3 films favoris a partir d'un 4.5 etoile est plus
user_rated_mov <- rating_data[rating_data$movieId %in% x & rating_data$rating >= 4.5 , 1]
y1 <- rating_data[rating_data$movieId %in% x[1] & rating_data$rating >= 4.5 , 1]
y2 <- rating_data[rating_data$movieId %in% x[2] & rating_data$rating >= 4.5 , 1]
y3 <- rating_data[rating_data$movieId %in% x[3] & rating_data$rating >= 4.5 , 1]
user_rated_mov

# eliminant la redondance des rates
unique_user_rated_mov=unique(y)
unique_user_rated_mov

y1
y2
y3
typeof(y1)
p=table(y)
length(which(y==662))

[145] ✓ 0.1s R
... 2-8-17-30-35-38-62-69-71-72-87-88-88-96-103-108-109-109-116-117-119-122-122-128-134-143-147-151-156-158-158-164-165-171-177-184-187-198-213-224-224-261-271-278-282-286-287-289-292-294-299-303-307-328-335-339-347-350-380-381-387-393-398-399-402-405-409-428-432-432-439-440-454-455-456-458-460-460-462-471-473-475-484-499-511-524-531-541-552-552-555-556-560-561-569-571-572-575-578-580-583-589-591-597-611-615-627-632-633-637-648-650-662-662-662-668

2-8-17-30-35-38-62-69-71-72-87-88-96-103-108-109-116-117-119-122-128-134-143-147-151-156-158-164-165-171-177-184-187-198-213-224-261-271-278-282-286-287-289-292-294-299-303-307-328-335-339-347-350-380-381-387-393-398-399-402-405-409-428-432-439-440-454-455-456-458-460-462-471-473-475-484-499-511-524-531-541-552-555-556-560-561-569-571-572-575-578-580-583-589-591-597-611-615-627-632-633-637-648-650-662-668

2-8-17-30-38-71-72-88-96-108-109-116-122-147-151-156-158-165-171-187-198-213-224-278-282-286-289-299-303-328-335-339-347-350-387-393-399-405-409-432-439-440-454-455-460-462-471-484-511-531-552-555-560-561-571-572-575-580-589-597-627-632-637-648-650-662

88-103-143-184-292-460-552-662

35-62-69-87-88-109-117-119-122-128-134-158-164-177-224-261-271-287-294-307-380-381-398-402-428-432-456-458-473-475-499-524-541-556-569-578-583-591-611-615-633-662-668

"integer"

3
```

Figure 4 : filtrage du rating

## 5. Filtrage spécifique

Dans cette phase, on a essayé d'extraire respectivement les utilisateurs qui on regarde exactement les 3 films favoris dont leur évaluation est supérieure à 4.5, dont le but d'améliorer la qualité de la recommandation cible.

```

#filtrage des utilisateurs qui ont regarde exactement les 3 films favoris parmi tous les utilisateurs
tree_user_mov=list()
for (k in unique_userRated_mov){
  if(length(which(userRated_mov==k)) == 3){
    print(k)
    tree_user_mov<-append(tree_user_mov,k)
  }
}
tree_user_mov
length(tree_user_mov)

```

[148] ✓ 0.7s

```

... [1] 88
      [1] 662

```

</>

```

      1.88
      2.662

```

</> 2

Figure 5 : Filtrage spécifique

## 6. Semi-Recommandation List

Semi-Recommandation : est une liste qui contient tous les films qui ont regarder par les utilisateurs traite par avant, ainsi qu'on élimine la redondance des films choisi par ces derniers.

```

#tous les films qui ont regarde par les utilisateurs qui ont regarde exactement les 3 films favoris
semi_recomm <- rating_data[rating_data$userId %in% tree_user_mov & rating_data$rating > 4,2]
length(semi_recomm)

```

[166] ✓ 0.1s

```

... 380

```

```

#eliminant les films redondants
semi_recomm = unique(semi_recomm)
semi_recomm
length(semi_recomm)

```

[150] ✓ 0.6s

```

... 1 · 2 · 6 · 21 · 25 · 32 · 36 · 47 · 50 · 69 · 150 · 163 · 180 · 223 · 231 · 260 · 290 · 293 · 296 · 318 · 337 · 353 · 356 · 386 · 440 · 441 · 457 · 480 · 508 · 527 · 541 · 553 · 555 · 589 · 590 · 592 · 593 · 608 · 750 · 78
1036 · 1073 · 1079 · 1089 · 1090 · 1097 · 1136 · 1148 · 1179 · 1193 · 1196 · 1197 · 1198 · 1201 · 1206 · 1207 · 1208 · 1210 · 1213 · 1214 · 1215 · 1219 · 1221 · 1222 · 1226 · 1234 · 1240 · 1242 · 1246 · 124
1287 · 1297 · 1302 · 1304 · 1358 · 1387 · 1393 · 1394 · 1408 · 1466 · 1500 · 1517 · 1527 · 1610 · 1617 · 1641 · 1653 · 1663 · 1682 · 1704 · 1732 · 1784 · 1799 · 1845 · 1945 · 1954 · 1958 · 1961 · 1982 · 199
2231 · 2245 · 2248 · 2261 · 2268 · 2288 · 2301 · 2321 · 2352 · 2406 · 2407 · 2416 · 2463 · 2502 · 2571 · 2599 · 2640 · 2700 · 2706 · 2716 · 2728 · 2746 · 2762 · 2763 · 2779 · 2791 · 2795 · 2797 · 2804 · 285
2087 · 2098 · 2100 · 2104 · 2107 · 2117 · 2153 · 2210 · 2271 · 2275 · 2261 · 2263 · 2262 · 2121 · 2148 · 2171 · 2536 · 2535 · 2553 · 2617 · 2625 · 2630 · 2671 · 2703 · 2702 · 2722 · 2740 · 2702 · 2800 · 285

```

Figure 6 : semi-Recommendation list

## 7. Création du genre favoris

Afin de construire la recommandation cible, on avait choisi spécifier l'extraction à partir des genres des films choisis dans la liste de recommandation.

```
#création du genre favoris de la recommandation cible
g1<-"Adventure"
g2<-"Fantasy"

[156] ✓ 0.7s
```

Figure 7 : genres favoris

## 8. Recommandation cible

Ci-dessus, l'étape finale du traitement, notre recommandation cible est bien extraite.

Les 20 premières lignes représentent les films recommander dont l'évaluation est supérieure strictement à 4.5 à partir des 3 films favoris.

```
#on choisit la liste des films favoris recommander à partir des genres favoris pour la recommandation cible
recommended_df=movie_data[grepl(g1, movie_data$genres, fixed = TRUE) & grepl(g2, movie_data$genres, fixed = TRUE) & movie_data$movieId %in% semi_recomm,]

#affichage en format dataframe de la recommandation cible
recommended_df

[159] ✓ 0.1s

...
A data.frame: 20 × 3
  movieId title genres
  <int> <chr> <chr>
1 1 Toy Story (1995) Adventure|Animation|Children|Comedy|Fantasy
2 2 Jumanji (1995) Adventure|Children|Fantasy
741 919 Wizard of Oz The (1939) Adventure|Children|Fantasy|Musical
923 1136 Monty Python and the Holy Grail (1975) Adventure|Comedy|Fantasy
960 1197 Princess Bride, The (1987) Action|Adventure|Comedy|Fantasy|Romance
977 1215 Army of Darkness (1993) Action|Adventure|Comedy|Fantasy|Horror
1035 1275 Highlander (1986) Action|Adventure|Fantasy
1569 2005 Goonies, The (1985) Action|Adventure|Children|Comedy|Fantasy
1697 2140 Dark Crystal, The (1982) Adventure|Fantasy
1717 2161 NeverEnding Story, The (1984) Adventure|Children|Fantasy
2389 2987 Who Framed Roger Rabbit? (1988) Adventure|Animation|Children|Comedy|Crime|Fantasy|Mystery
2958 3740 Big Trouble in Little China (1986) Action|Adventure|Comedy|Fantasy
```

Figure 8 : Recommendation cible

## Conclusion

Les systèmes de recommandation sont devenus particulièrement importants sur le web ces dernières années. Grâce à eux, les entreprises de commerce électronique ont pu se différencier de leurs concurrents, faciliter la vie des clients existants et toucher des clients potentiels. Le fonctionnement du système de recommandation est très simple, mais la mise en œuvre est complexe. Ces difficultés résident dans des aspects spécifiques tels que la collecte et la sélection de données pertinentes, la taille et la qualité des données, la rareté des données, la construction de profils d'utilisateurs et la prévision de nouveaux profils d'utilisateurs ou de nouveaux produits.

## Bibliographie

[Aciar et al., 2007] Silvana Aciar, Debbie Zhang, Simeon Simoff, and John Debenham, (2007). Informed recommender: Basing recommendations on consumer product reviews. *Intelligent Systems, IEEE*, 22(3), pp.39–47.

[Adomavicius and Tuzhilin, 2005] Gediminas Adomavicius and Alexander Tuzhilin, (2005). Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 17(6), pp.734–749.

[Aggarwal et al., 1999] Charu C. Aggarwal, Joel L. Wolf, Kun-Lung Wu, and Philip S. Yu. Horting hatches an egg : A new graph-theoretic approach to collaborative filtering. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '99*, pages 201–212, New York, NY, USA, 1999. ACM.

[Ahn et al., 2007] Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. Open user profiles for adaptive news systems : Help or harm ? In *Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 11–20, New York, NY, USA, 2007. ACM.

Aggarwal, C. C. (2016b). Context-sensitive recommender systems. In *Recommender Systems*, pages 255–281. Springer.

Originally published at [blog.arcbees.com](http://blog.arcbees.com) on February 8, 2017.

Mining of Massive Datasets (Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman)