

# Recommendation System

Agata FILIANA

May 28, 2014

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Introduction

Apa recommendation system (sistem rekomendasi)?  
Contoh?

# Introduction

## Apa recommendation system (sistem rekomendasi)? Contoh?

Livres

Page 1 sur 4



Les Confidents : Et ...  
Philippe Claudel  
★★★★☆ (1)  
EUR 3,80 EUR 3,61  
Pourquoi est-ce recommandé ?



L'immoraliste  
André Gide  
★★★★☆ (2)  
EUR 5,60 EUR 5,32  
Pourquoi est-ce recommandé ?



Les nourritures ...  
André Gide  
★★★★☆ (13)  
EUR 5,90 EUR 5,89  
Pourquoi est-ce recommandé ?



The Importance of ...  
Oscar Wilde  
★★★★☆ (7)  
EUR 2,00 EUR 1,90  
Pourquoi est-ce recommandé ?



I Heart New York  
Lindsey Kelk  
★★★★☆ (4)  
EUR 9,60 EUR 9,12  
Pourquoi est-ce recommandé ?



Verbes anglais ...  
Jean-Michel Daube  
EUR 18,50 EUR 17,58  
Pourquoi est-ce recommandé ?



La couleur des ...  
Kathryn Stockett  
★★★★☆ (297)  
EUR 9,70 EUR 9,22  
Pourquoi est-ce recommandé ?

> Voir toutes les recommandations dans Livres

# Introduction



Format Kindle


[Voir une image plus grande \(ouvrez un zoom\)](#)
[Partagez vos propres images client](#)

## The Fault in Our Stars [Format Kindle]

John Green (Auteur)

★★★★★ (41 commentaires client)

Prix éditeur - format imprimé: ~~EUR 7,99~~Prix Kindle: **EUR 5,49** TTC & envoi gratuit via réseau sans fil par **Amazon Whispernet**Économisez: **EUR 2,11 (28%)**

- Longueur : 337 pages (Contient les vrais numéros de page)
- Langue : Anglais
- Vous n'avez pas encore de Kindle ? [Achetez-le ici](#) Ou commencez à lire dès maintenant avec l'une de nos [applications de lecture Kindle](#) gratuites.

Formats	Prix Amazon	Nouveau à partir de	Occasion à partir de
Format Kindle	EUR 5,49	---	---
Relié	EUR 11,99	EUR 8,85	EUR 7,87
Broché	EUR 6,99	EUR 4,22	EUR 5,53
MP3 CD, Livre audio	EUR 7,70	EUR 4,64	---

[Essayez gratuitement pendant 30 jours Amazon Premium](#) et bénéficiez de la livraison en 1 jour ouvré gratuite et illimitée sur des millions d'articles, et d'autres avantages.

 Fête des Mères : [découvrez toutes nos idées cadeaux, promotions et sélections](#)

## Les clients ayant acheté cet article ont également acheté



Looking for Alaska

John Green

★★★★★ (112)

Format Kindle

EUR 4,92



An Abundance of Katherines

John Green

★★★★★ (13)

Format Kindle

EUR 5,49



The Rosie Project

Graeme Simsion

★★★★★ (15)

Format Kindle

EUR 3,99



Insurgent (Divergent Trilogy, Book 2)

Veronica Roth

★★★★★ (20)

Format Kindle

EUR 6,50



The Husband's Secret

Liane Moriarty

★★★★★ (8)

Format Kindle

EUR 3,99



Paper Towns

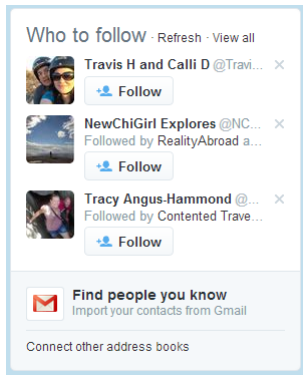
John Green

★★★★★ (14)

Format Kindle

EUR 4,84

# Introduction



# Recommender system

Memberikan rekomendasi, biasanya barang, kepada orang (user)  
Misalnya :



# Recommender system

Memberikan rekomendasi, biasanya barang, kepada orang (user)

Misalnya :

- ▶ Kamera digital mana yang harus saya beli?

# Recommender system

Memberikan rekomendasi, biasanya barang, kepada orang (user)  
Misalnya :

- ▶ Kamera digital mana yang harus saya beli?
- ▶ Akun twitter mana yang akan harus saya follow?

# Recommender system

Memberikan rekomendasi, biasanya barang, kepada orang (user)  
Misalnya :

- ▶ Kamera digital mana yang harus saya beli?
- ▶ Akun twitter mana yang akan harus saya follow?
- ▶ Universitas mana yang tepat bagi saya?

# Recommender system

Kegunaan recommender system :

- ▶ Mempersempit information overload

# Recommender system

Kegunaan recommender system :

- ▶ Mempersempit information overload
- ▶ Kegunaan bagi user : mendapatkan hal yang menarik, mempersempit pilihan, menemukan hal yang baru

# Recommender system

Kegunaan recommender system :

- ▶ Mempersempit information overload
- ▶ Kegunaan bagi user : mendapatkan hal yang menarik, mempersempit pilihan, menemukan hal yang baru
- ▶ Kegunaan bagi provider : memberikan rekomendasi yang lebih personal kepada user-nya, meningkatkan loyalitas user, meningkatkan pembelian, peluang untuk promosi, mendapatkan pengetahuan tentang user-nya

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Recommenders vs Search Engines

- ▶ Search engines bukan sebuah sistem rekomendasi



# Recommenders vs Search Engines











- ▶ Search engines bukan sebuah sistem rekomendasi
- ▶ Query untuk mencari rekomendasi pada search engine menghasilkan kumpulan sistem rekomendasi

# Recommenders vs Search Engines

Google best movies of 2014

Web News Videos Images Shopping More Search tools

Movies frequently mentioned on the web

									
X-Men: Days of Future Past 2014	Captain America: The Winter Sol... 2014	The Amazing Spider-Man 2 2014	Transformers: Age of Extinction 2014	The Lego Movie 2014	A Million Ways to Die in the West 2014	Guardians of the Galaxy 2014	Dawn of the Planet of the A... 2014	The Monuments Men 2014	Divergent 2014

## IMDb: Most Popular Feature Films Released in 2014

[www.imdb.com/title/?..2014.2014..movie&..](http://www.imdb.com/title/?..2014.2014..movie&..) Internet Movie Database  
6: A Million Ways to Die in the West (2014) · A Million ... 27: The Lego Movie (2014) ... 28: Alexander and the Terrible, Horrible, No Good, Very Bad Day (2014)

Most Popular Feature Films ... - User Rating - US Release Date - The Hunger Games

## Top 100 Movies of 2014 - Rotten Tomatoes

[www.rottentomatoes.com/top/bestofrt/?year=2014](http://www.rottentomatoes.com/top/bestofrt/?year=2014) Rotten Tomatoes  
Rotten Tomatoes lists the best 100 Movies based on their Adjusted Score!  
Child's Pose - 12 O'clock Boys - Like Father, Like Son - Teenage

## The top 25 must-see movies of 2014 | Den of Geek

[www.denofgeek.com](http://www.denofgeek.com) Movies · Odd List

Introduction

Recommenders vs Search Engines

**Kategori recommendation system**

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Kategori Recommendation System

Content based filtering: *"Rekomendasikan buku yang sesuai dengan tipe buku yang saya suka"*. Biasanya menggunakan fitur barang.

# Kategori Recommendation System

**Content based filtering:** *"Rekomendasikan buku yang sesuai dengan tipe buku yang saya suka"*. Biasanya menggunakan fitur barang.

**Collaborative filtering:** *"Rekomendasikan buku yang disukai oleh teman-teman saya"*. Menggunakan preferensi komunitas.

# Kategori Recommendation System

**Content based filtering:** *"Rekomendasikan buku yang sesuai dengan tipe buku yang saya suka"*. Biasanya menggunakan fitur barang.

**Collaborative filtering:** *"Rekomendasikan buku yang disukai oleh teman-teman saya"*. Menggunakan preferensi komunitas.

**Hybrid:** Kombinasi dari CF dan content-based

Introduction

Recommenders vs Search Engines

Kategori recommendation system

**Input recommendation system**

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Input Recommendation System

(Vozalis & Margaritis, 2003) menyatakan input sistem rekomendasi ada tiga:



# Input Recommendation System

(Vozalis & Margaritis, 2003) menyatakan input sistem rekomendasi ada tiga:

- ▶ Demographic data : umur, jenis kelamin, dll

# Input Recommendation System

(Vozalis & Margaritis, 2003) menyatakan input sistem rekomendasi ada tiga:

- ▶ Demographic data : umur, jenis kelamin, dll
- ▶ Content data : analisis tekstual dari barang-barang yg pernah dibeli oleh user

# Input Recommendation System

(Vozalis & Margaritis, 2003) menyatakan input sistem rekomendasi ada tiga:

- ▶ Demographic data : umur, jenis kelamin, dll
- ▶ Content data : analisis tekstual dari barang-barang yg pernah dibeli oleh user
- ▶ Ratings : scalar, binary, unary. Selain itu ada juga rating implisit dan eksplisit.

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Collaborative Filtering

*"wisdom of crowd"*

# CF : Keuntungan

(Melville et al., 2002) menyebutkan dua keuntungan CF:

## CF : Keuntungan

(Melville et al., 2002) menyebutkan dua keuntungan CF:

- ▶ bisa digunakan pada domains dimana di dalamnya terdapat content yang tidak berhubungan dengan items

## CF : Keuntungan

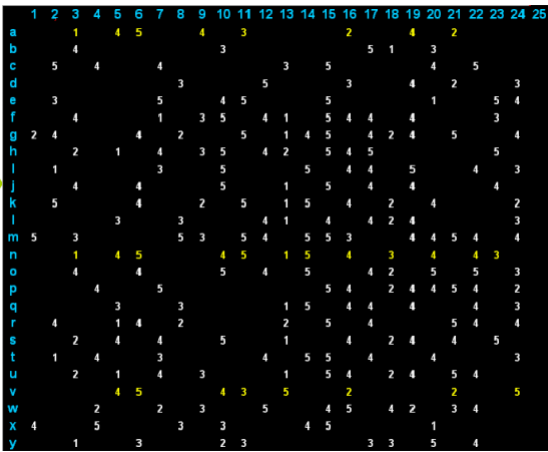
(Melville et al., 2002) menyebutkan dua keuntungan CF:

- ▶ bisa digunakan pada domains dimana di dalamnya terdapat content yang tidak berhubungan dengan items
- ▶ serendipitous recommendations



## CF : Input

Input dari CF : matriks user-item ratings



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
a			1		4	5			4		3					2			4		2				
b			4							3							5	1		3					
c	5			4			4					3		5						4		5			
d							3					5				3			4		2			3	
e	3					5				4	5				5					1			5	4	
f		4					1		3	5		4	1		5	4	4		4				3		
g	2	4				4		2			5		1	4	5		4	2	4		5			4	
h			2		1		4		3	5		4	2		5	4	5					5			
i		1					3			5				5		4	4		5			4		3	
j		4				4				5			1		5		4		4				4		
k	5					4			2		5		1	5		4		2		4				2	
l				3				3				4	1		4		4	2	4				3		
m	5		3					5	3		5	4		5	5	3			4	4	5	4		4	
n			1		4	5				4	5		1	5		4		3		4		4	3		
o			4			4				5		4		5			4	2		5		5		3	
p				4			5							5	4			2	4	4	5	4		2	
q					3			3					1	5		4	4		4				4		3
r		4			1	4		2					2		5		4				5	4		4	
s		2		4		4				5			1			4		2	4		4		5		
t		1		4			3					4		5	5		4			4				3	
u		2			1		4		3				1		5	4		2	4		5	4			
v				4	5					4	3		5			2				2				5	
w				2			2		3			5			4	5		4	2		3	4			
x	4			5				3		3				4	5					1					
y		1				3				2	3						3	3		5		4			

# CF : Output

Output dari CF : top-N list, prediksi rating score

# CF : Metode

## Nearest Neighbour

- ▶ user-to-user

# CF : Metode

## Nearest Neighbour

- ▶ user-to-user
- ▶ item-to-item

## CF : User-to-user

Terdapat sebuah matriks berisi rating. Kira-kira berapa rating Alice untuk item5?

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

## CF User-to-user : Pearson correlation coefficient

Bagaimana caranya mengukur kesamaan antar user?

Solusi:

Pearson correlation coefficient

$\text{sim}(a, b) =$

$$\frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

### Variables

- ▶  $a, b$ : users
- ▶  $P$ : items, yang sudah dirating oleh  $a$  dan  $b$
- ▶  $r_{a,p}$ : rating dari user  $a$  untuk item  $p$

## CF User-to-user : Menentukan prediksi

Weighted normalized adjusted average

$pred(a, b) =$

$$\bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

## CF User-to-user : Menentukan prediksi

Lainnya :


- ▶ Simple average
- ▶ Weighted average



## CF User-to-user : Menentukan prediksi

Jika kita ambil 2-Nearest-Neighbour

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1



$sim = 0.85$   
 $sim = 0.00$   
 $sim = 0.70$   
 $sim = -0.79$

$$P_{Alice, item5} = \frac{3+4}{2} = 3.5$$

## CF : Item-to-item

Memberikan prediksi berdasarkan kesamaan antar barang (items)

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

## CF Item-to-item : Cosine similarity

Kesamaan dihitung dengan cosine similarity (yang paling sering digunakan), dimana ratings dianggap sebagai vektor

### Cosine similarity

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

## CF Item-to-item : Cosine similarity

Opsi lain adalah adjusted cosine similarity yang menggunakan rata-rata ratings dari user untuk mengubah rating awal

### Adjusted Cosine similarity

$$\text{sim}(\vec{a}, \vec{b}) =$$

$$\frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

### Variables

- $U$ : satu set users yang sudah melakukan rating terhadap items  $a$  dan  $b$

## CF Item-to-item : Menentukan prediksi

Prediction function

$pred(u, p) =$

$$\frac{\sum_{i \in ratedItems(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ratedItem(u)} sim(i, p)}$$

## CF Item-to-item : Menentukan prediksi

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

e.g.  $P(\text{Alice}, \text{item5}) = (5+4)/2 = 4.5$

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

**Content-based**

Evaluasi recommendation system

Masalah dalam recommendation system

# Content-based

- ▶ Membutuhkan informasi barang, tidak perlu informasi komunitas



# Content-based

- ▶ Membutuhkan informasi barang, tidak perlu informasi komunitas
- ▶ Dibutuhkan : informasi tentang konten barang informasi tentang apa yg disukai oleh user

# Content-based

- ▶ Membutuhkan informasi barang, tidak perlu informasi komunitas
- ▶ Dibutuhkan : informasi tentang konten barang informasi tentang apa yg disukai oleh user
- ▶ Biasanya dipakai untuk text-based, misalnya berita && text classification

# Content-based : Keuntungan

Keuntungan dari content-based:

# Content-based : Keuntungan

Keuntungan dari content-based:

- ▶ komunitas tidak diperlukan

# Content-based : Keuntungan

Keuntungan dari content-based:

- ▶ komunitas tidak diperlukan
- ▶ lebih mudah daripada CF?

## Content-based : tf-idf

Karena merupakan text classification maka digunakan tf-idf dan classifiers klasik seperti Naive Bayes dan SVM

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Evaluasi Recommendation System

Statistical accuracy metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)



# Evaluasi Recommendation System

**Statistical accuracy metrics:** Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

**Decision support accuracy:** biasanya untuk data binary, menunjukkan kualitas dari barang yang direkomendasi : Precision, Recall, ROC

Introduction

Recommenders vs Search Engines

Kategori recommendation system

Input recommendation system

Collaborative Filtering

Content-based

Evaluasi recommendation system

Masalah dalam recommendation system

# Masalah dalam Recommendation System

Data sparsity Solusi? SVD (Singular Value Decomposition)

# Masalah dalam Recommendation System

Data sparsity Solusi? SVD (Singular Value Decomposition)

Cold start First rater problem

# Masalah dalam Recommendation System

Data sparsity Solusi? SVD (Singular Value Decomposition)

Cold start First rater problem

Shilling attacks Biased ratings