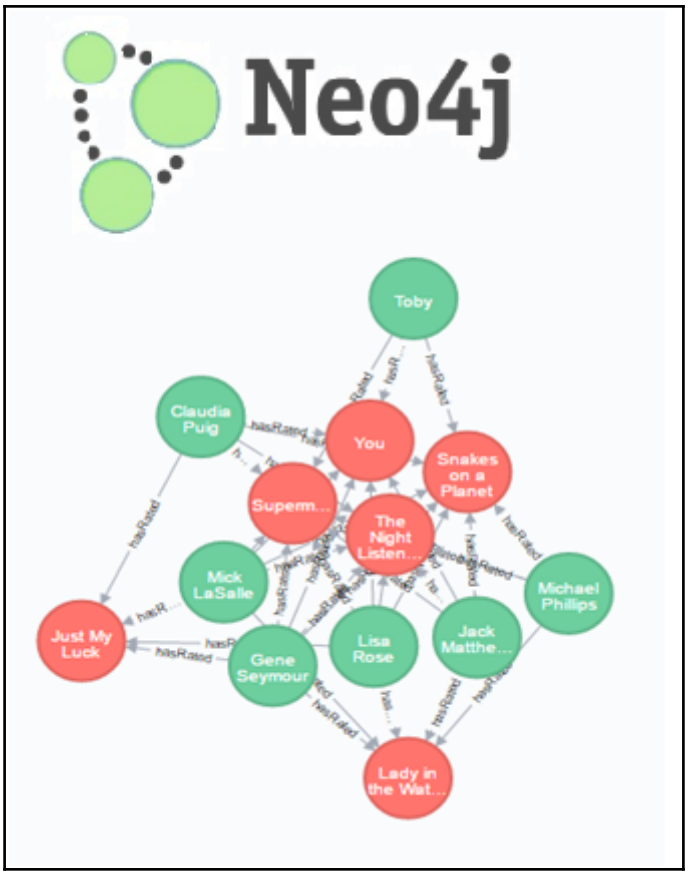
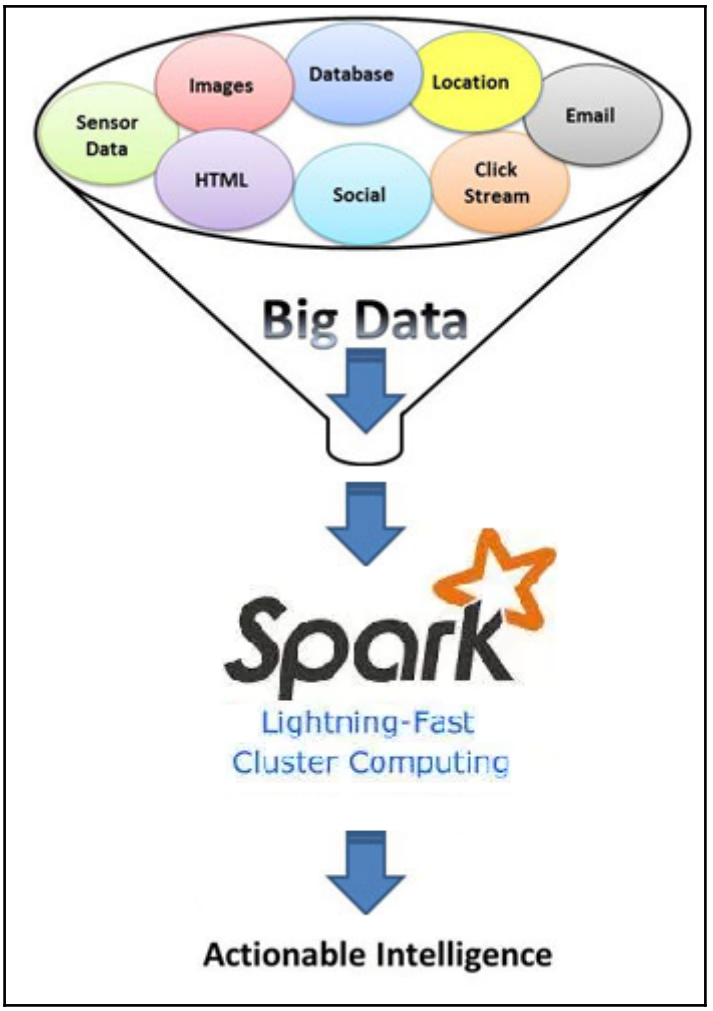
# Graphics Bundle

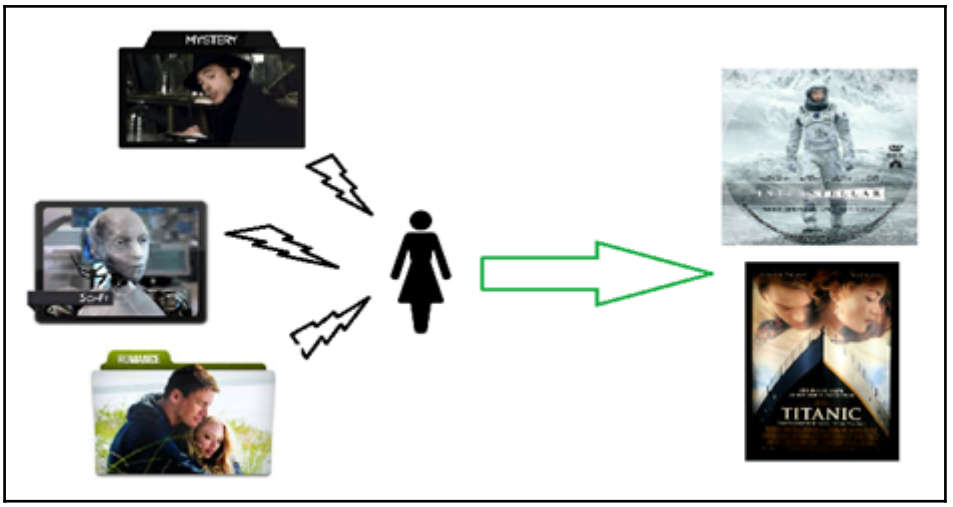
## Chapter 1: Introduction to Recommendation Engines

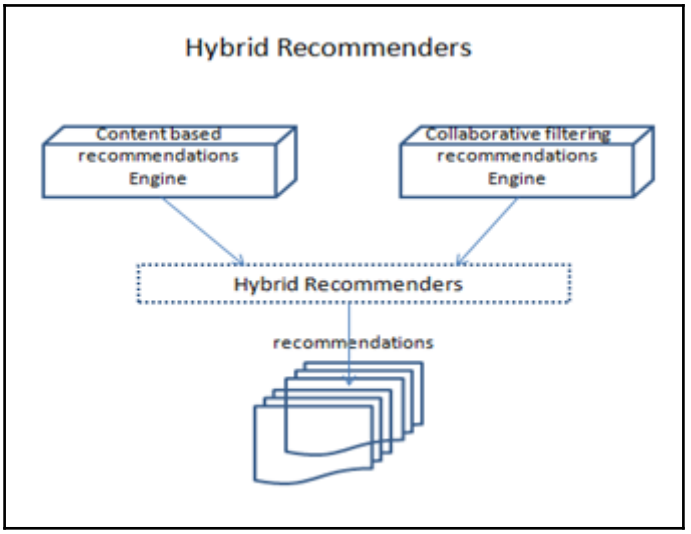
$$\mathbf{u} \in U, I'_u = argmax_u u(u, i)$$

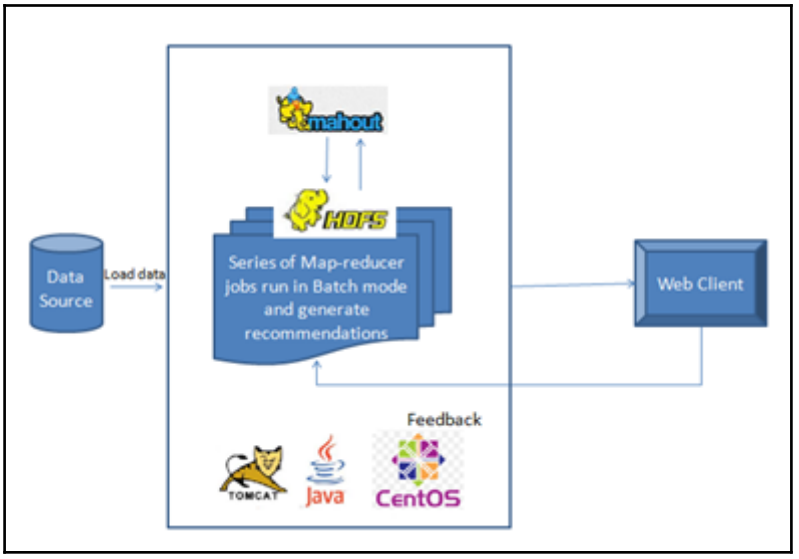

$$F{:}U\,X\,I \rightarrow R$$

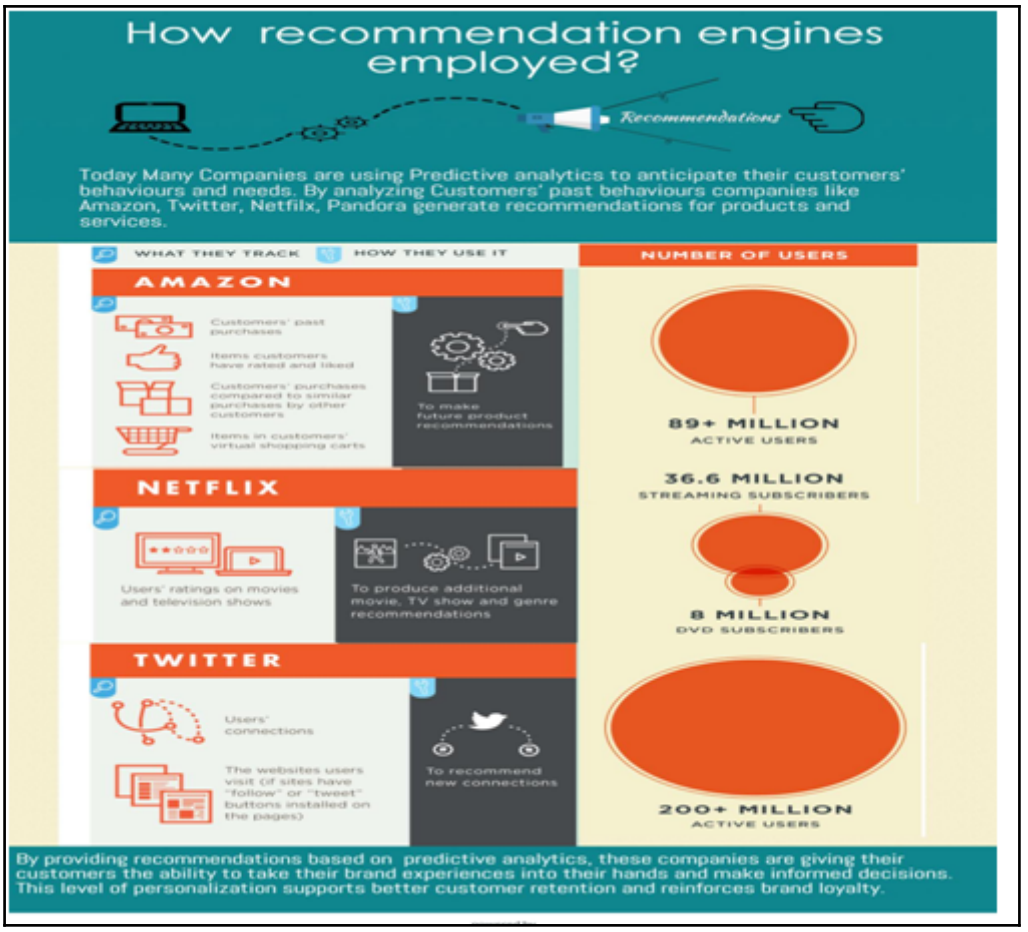

Neo4j
Toby
Claudia Puig
You
Snakes on a Planet
Superm...
The Night Listen...
Mick LaSalle
Michael Phillips
Just My Luck
Gene Seymour
Lisa Rose
Jack Matthe...
Lady in the Wat...

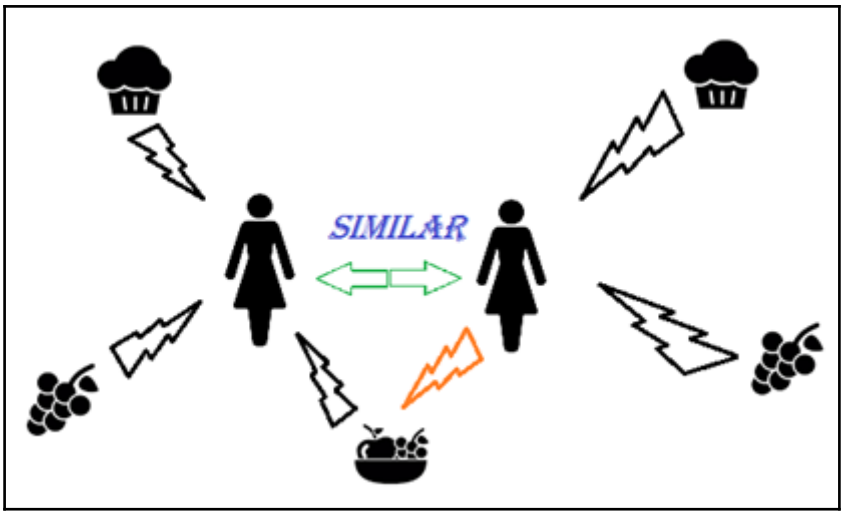

Images
Database
Location
Email
Sensor Data
HTML
Social
Click Stream
Big Data
Spark
Lightning-Fast
Cluster Computing
Actionable Intelligence


MYSTERY
SCI-FI
ROMANCE
TITANIC

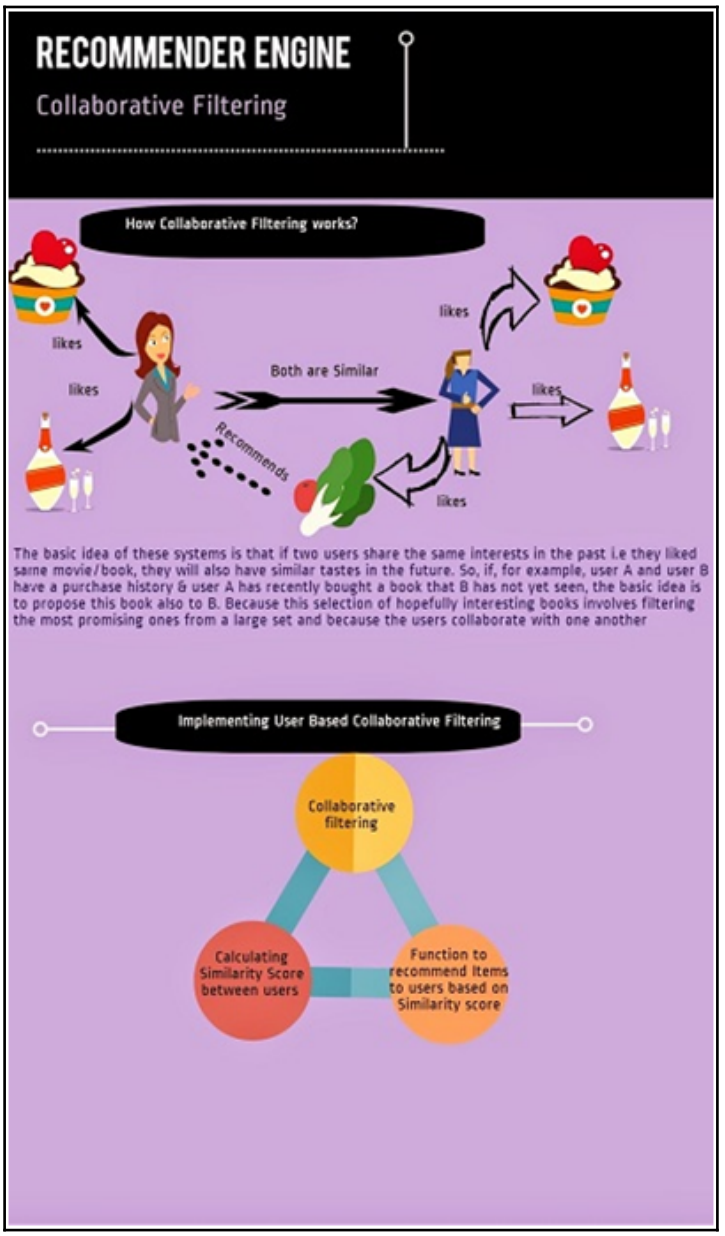

Hybrid Recommenders
Content based recommendations Engine
Collaborative filtering recommendations Engine
Hybrid Recommenders
recommendations

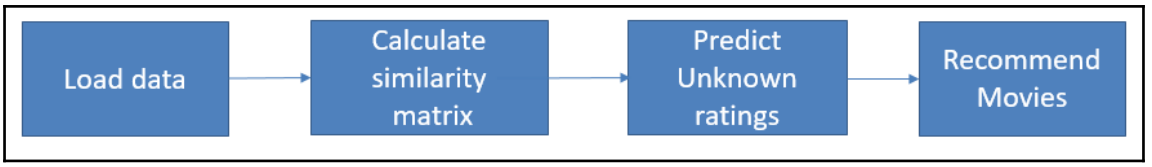

mahout
HDFS
Data Source
Load data
Series of Map-reducer jobs run in Batch mode and generate recommendations
Web Client
Feedback
TOMCAT
Java
CentOS


How recommendation engines employed?
Recommendations
Today Many Companies are using Predictive analytics to anticipate their customers' behaviours and needs. By analyzing Customers' past behaviours companies like Amazon, Twitter, Netfilx, Pandora generate recommendations for products and services.
WHAT THEY TRACK
HOW THEY USE IT
NUMBER OF USERS
AMAZON
Customers' past purchases
Items customers have rated and liked
Customers' purchases compared to similar purchases by other customers
Items in customers' virtual shopping carts
To make future product recommendations
89+ MILLION
ACTIVE USERS
NETFLIX
Users' ratings on movies and television shows
To produce additional movie, TV show and genre recommendations
36.6 MILLION
STREAMING SUBSCRIBERS
8 MILLION
DVD SUBSCRIBERS
TWITTER
Users' connections
The websites users visit (if sites have "follow" or "tweet" buttons installed on the pages)
To recommend new connections
200+ MILLION
ACTIVE USERS
By providing recommendations based on predictive analytics, these companies are giving their customers the ability to take their brand experiences into their hands and make informed decisions. This level of personalization supports better customer retention and reinforces brand loyalty.


SIMILAR

# Chapter 2: Build Your First Recommendation Engine






| | critic | title | rating |
|---|---|---|---|
| 1 | Jack Matthews | Lady in the Water | 3.0 |
| 2 | Jack Matthews | Snakes on a Plane | 4.0 |
| 3 | Jack Matthews | You Me and Dupree | 3.5 |
| 4 | Jack Matthews | Superman Returns | 5.0 |
| 5 | Jack Matthews | The Night Listener | 3.0 |
| 6 | Mick LaSalle | Lady in the Water | 3.0 |
| 7 | Mick LaSalle | Snakes on a Plane | 4.0 |
| 8 | Mick LaSalle | Just My Luck | 2.0 |
| 9 | Mick LaSalle | Superman Returns | 3.0 |
| 10 | Mick LaSalle | You Me and Dupree | 2.0 |
| 11 | Mick LaSalle | The Night Listener | 3.0 |
| 12 | Claudia Puig | Snakes on a Plane | 3.5 |
| 13 | Claudia Puig | Just My Luck | 3.0 |
| 14 | Claudia Puig | You Me and Dupree | 2.5 |
| 15 | Claudia Puig | Superman Returns | 4.0 |
| 16 | Claudia Puig | The Night Listener | 4.5 |
| 17 | Lisa Rose | Lady in the Water | 2.5 |
| 18 | Lisa Rose | Snakes on a Plane | 3.5 |
| 19 | Lisa Rose | Just My Luck | 3.0 |
| 20 | Lisa Rose | Superman Returns | 3.5 |

```
> head(ratings)
          critic               title rating
1 Jack Matthews  Lady in the Water     3.0
2 Jack Matthews  Snakes on a Plane     4.0
3 Jack Matthews  You Me and Dupree     3.5
4 Jack Matthews   Superman Returns     5.0
5 Jack Matthews The Night Listener     3.0
6  Mick LaSalle  Lady in the Water     3.0
```

```
> str(ratings)
'data.frame':   31 obs. of  3 variables:
 $ critic: Factor w/ 6 levels "Claudia Puig",..: 3 3 3 3 3 5 5 5 5 5 ...
 $ title : Factor w/ 6 levels "Just My Luck",..: 2 3 6 4 5 2 3 1 4 6 ...
 $ rating: num  3 4 3.5 5 3 3 4 2 3 2 ...
```

| | Claudia Puig | Gene Seymour | Jack Matthews | Lisa Rose | Mick LaSalle | Toby |
|---|---|---|---|---|---|---|
| Just My Luck | 3.0 | 1.5 | *NA* | 3.0 | 2 | *NA* |
| Lady in the Water | *NA* | 3.0 | 3.0 | 2.5 | 3 | *NA* |
| Snakes on a Plane | 3.5 | 3.5 | 4.0 | 3.5 | 4 | 4.5 |
| Superman Returns | 4.0 | 5.0 | 5.0 | 3.5 | 3 | 4.0 |
| The Night Listener | 4.5 | 3.0 | 3.0 | 3.0 | 3 | *NA* |
| You Me and Dupree | 2.5 | 3.5 | 3.5 | 2.5 | 2 | 1.0 |

| | Claudia Puig | Gene Seymour | Jack Matthews | Lisa Rose | Mick LaSalle | Toby |
|---|---|---|---|---|---|---|
| Claudia Puig | 1.0000000 | 0.7559289 | 0.9285714 | 0.9449112 | 0.6546537 | 0.8934051 |
| Gene Seymour | 0.7559289 | 1.0000000 | 0.9449112 | 0.5000000 | 0.0000000 | 0.3812464 |
| Jack Matthews | 0.9285714 | 0.9449112 | 1.0000000 | 0.7559289 | 0.3273268 | 0.6628490 |
| Lisa Rose | 0.9449112 | 0.5000000 | 0.7559289 | 1.0000000 | 0.8660254 | 0.9912407 |
| Mick LaSalle | 0.6546537 | 0.0000000 | 0.3273268 | 0.8660254 | 1.0000000 | 0.9244735 |
| Toby | 0.8934051 | 0.3812464 | 0.6628490 | 0.9912407 | 0.9244735 | 1.0000000 |

Filter

| | title | rating |
|---|---|---|
| 1 | Just My Luck | *NA* |
| 2 | Lady in the Water | *NA* |
| 3 | Snakes on a Plane | 4.5 |
| 4 | Superman Returns | 4.0 |
| 5 | The Night Listener | *NA* |
| 6 | You Me and Dupree | 1.0 |

```
> titles_na_critic
[1] "Just My Luck"       "Lady in the Water"  "The Night Listener"
```

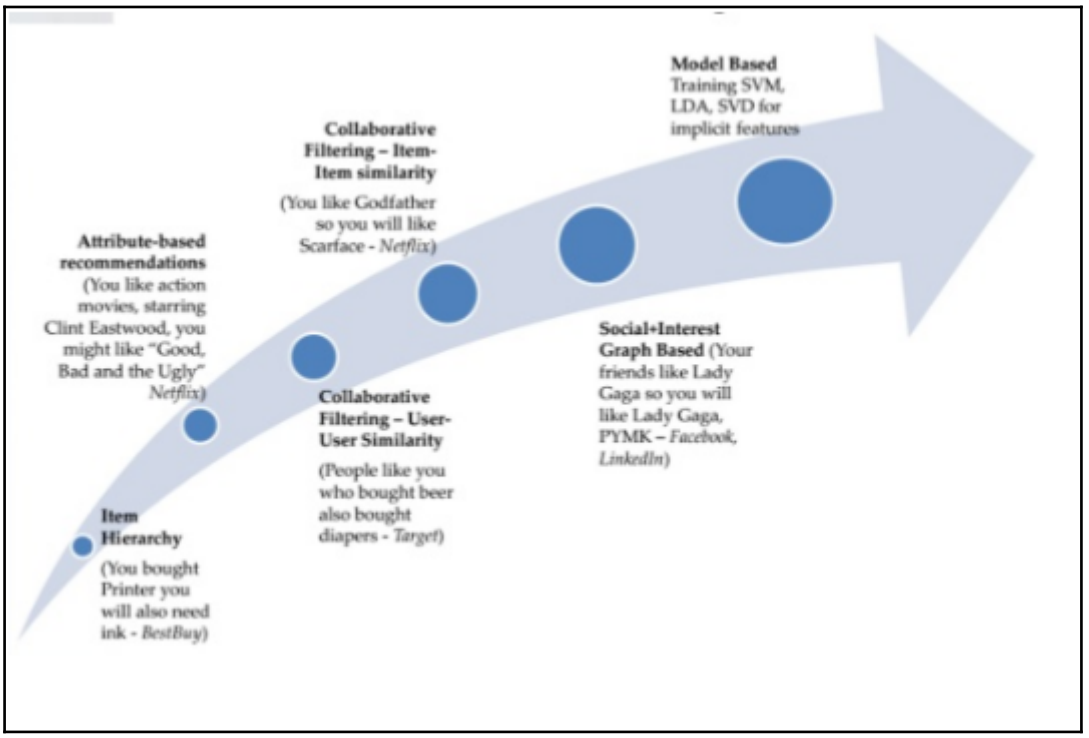
| critic | title | rating |
|---|---|---|
| Jack Matthews | Lady in the Water | 3.0 |
| Jack Matthews | The Night Listener | 3.0 |
| Mick LaSalle | Lady in the Water | 3.0 |
| Mick LaSalle | Just My Luck | 2.0 |
| Mick LaSalle | The Night Listener | 3.0 |
| Claudia Puig | Just My Luck | 3.0 |
| Claudia Puig | The Night Listener | 4.5 |
| Lisa Rose | Lady in the Water | 2.5 |
| Lisa Rose | Just My Luck | 3.0 |
| Lisa Rose | The Night Listener | 3.0 |
| Gene Seymour | Lady in the Water | 3.0 |
| Gene Seymour | Just My Luck | 1.5 |
| Gene Seymour | The Night Listener | 3.0 |

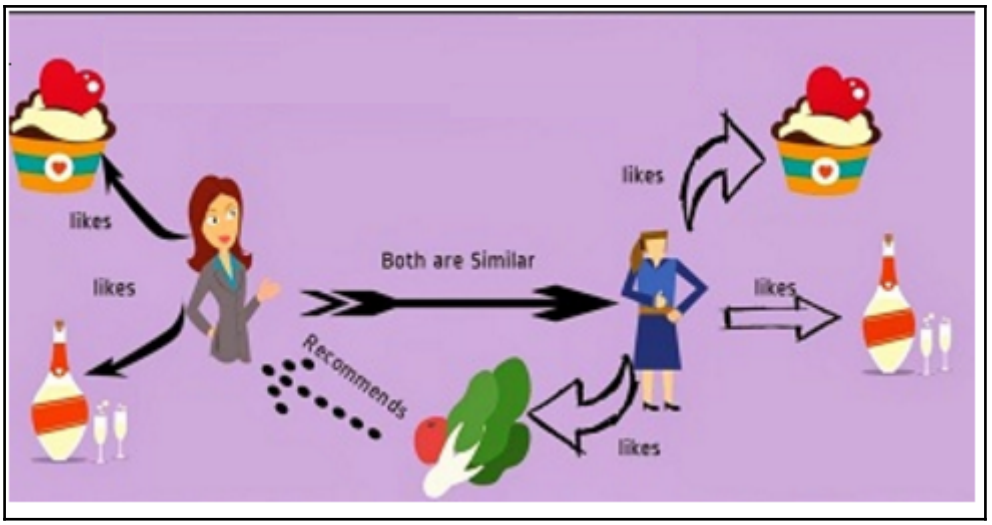
| critic | title | rating | similarity |
|---|---|---|---|
| Claudia Puig | Just My Luck | 3.0 | 0.8934051 |
| Claudia Puig | The Night Listener | 4.5 | 0.8934051 |
| Gene Seymour | Lady in the Water | 3.0 | 0.3812464 |
| Gene Seymour | Just My Luck | 1.5 | 0.3812464 |
| Gene Seymour | The Night Listener | 3.0 | 0.3812464 |
| Jack Matthews | Lady in the Water | 3.0 | 0.6628490 |
| Jack Matthews | The Night Listener | 3.0 | 0.6628490 |
| Lisa Rose | Lady in the Water | 2.5 | 0.9912407 |
| Lisa Rose | Just My Luck | 3.0 | 0.9912407 |
| Lisa Rose | The Night Listener | 3.0 | 0.9912407 |
| Mick LaSalle | Lady in the Water | 3.0 | 0.9244735 |
| Mick LaSalle | Just My Luck | 2.0 | 0.9244735 |
| Mick LaSalle | The Night Listener | 3.0 | 0.9244735 |

| critic | title | rating | similarity | sim_rating |
|---|---|---|---|---|
| Claudia Puig | Just My Luck | 3.0 | 0.8934051 | 2.6802154 |
| Claudia Puig | The Night Listener | 4.5 | 0.8934051 | 4.0203232 |
| Gene Seymour | Lady in the Water | 3.0 | 0.3812464 | 1.1437393 |
| Gene Seymour | Just My Luck | 1.5 | 0.3812464 | 0.5718696 |
| Gene Seymour | The Night Listener | 3.0 | 0.3812464 | 1.1437393 |
| Jack Matthews | Lady in the Water | 3.0 | 0.6628490 | 1.9885469 |
| Jack Matthews | The Night Listener | 3.0 | 0.6628490 | 1.9885469 |
| Lisa Rose | Lady in the Water | 2.5 | 0.9912407 | 2.4781018 |
| Lisa Rose | Just My Luck | 3.0 | 0.9912407 | 2.9737221 |
| Lisa Rose | The Night Listener | 3.0 | 0.9912407 | 2.9737221 |
| Mick LaSalle | Lady in the Water | 3.0 | 0.9244735 | 2.7734204 |
| Mick LaSalle | Just My Luck | 2.0 | 0.9244735 | 1.8489469 |
| Mick LaSalle | The Night Listener | 3.0 | 0.9244735 | 2.7734204 |

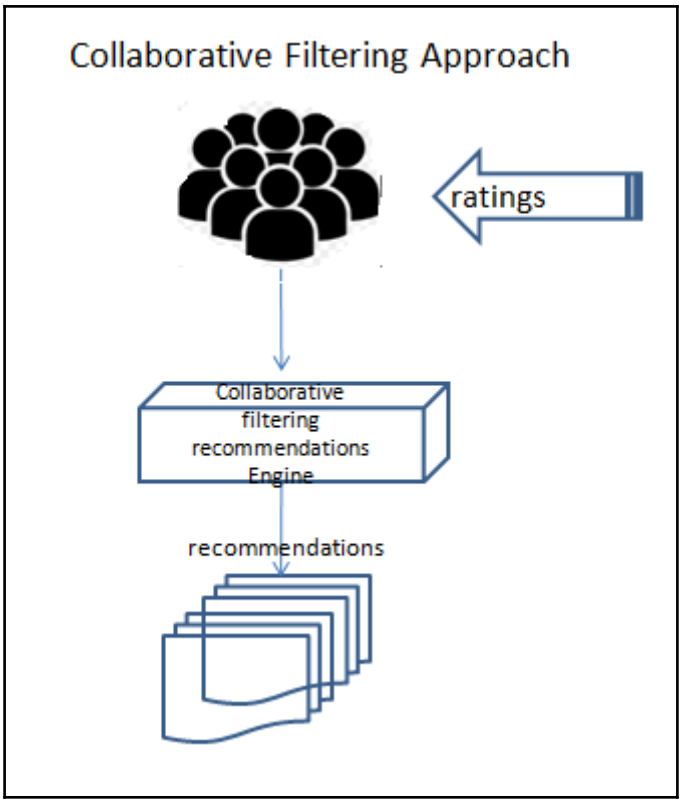
```
> generateRecommendations(1)
Source: local data frame [1 x 2]

              title sum(sim_rating)/sum(similarity)
             (fctr)                           (dbl)
1 Lady in the Water                        2.856137
> generateRecommendations(2)
Source: local data frame [0 x 2]

Variables not shown: title (fctr), sum(sim_rating)/sum(similarity) (lgl)
> generateRecommendations(3)
Source: local data frame [1 x 2]

         title sum(sim_rating)/sum(similarity)
        (fctr)                           (dbl)
1 Just My Luck                        2.409926
> generateRecommendations(4)
Source: local data frame [0 x 2]

Variables not shown: title (fctr), sum(sim_rating)/sum(similarity) (lgl)
> generateRecommendations(5)
Source: local data frame [0 x 2]

Variables not shown: title (fctr), sum(sim_rating)/sum(similarity) (lgl)
> generateRecommendations(6)
Source: local data frame [3 x 2]

               title sum(sim_rating)/sum(similarity)
              (fctr)                           (dbl)
1       Just My Luck                        2.530981
2  Lady in the Water                        2.832550
3 The Night Listener                        3.347790
>
```

# Chapter 3: Recommendation Engines Explained




likes
likes
likes
Both are Similar
likes
likes
Recommends
likes

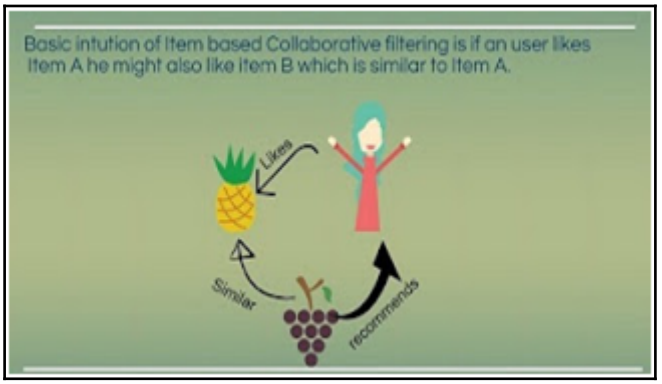

Customers Who Bought This Item Also Bought
A Fistful of Evil: An Urban Fantasy Novel ...
Rebecca Chastain
(1)
Kindle Edition
$2.99
Trace of Magic: 1 (The Diamond City Magic Novels)
Diana Pharaoh Francis
Kindle Edition
$7.07
Murder of Crows (The Twenty-Sided Sorceress Book 2)
Annie Bellet
Kindle Edition
$2.99




| Movie/User | Claudia Puig | Gene Seymour | Jack Matthews | Lisa Rose | Mick LaSalle | Toby |
|---|---|---|---|---|---|---|
| Just My Luck | 3 | 1.5 | | 3 | 2 | |
| Lady in the Water | | 3 | 3 | 2.5 | 3 | |
| Snakes on a Plane | 3.5 | 3.5 | 4 | 3.5 | 4 | 4.5 |
| Superman Returns | 4 | 5 | 5 | 3.5 | 3 | 4 |
| The Night Listener | 4.5 | 3 | 3 | 3 | 3 | |
| You Me and Dupree | 2.5 | 3.5 | 3.5 | 2.5 | 2 | 1 |

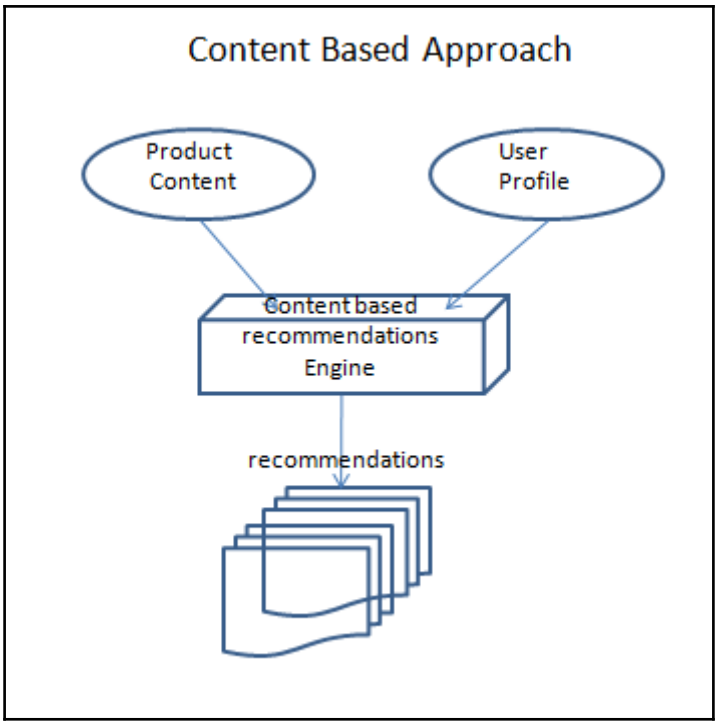
$$\text{Euclidean Distance}(x,y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

| | Claudia Puig | Gene Seymour | Jack Matthews | Lisa Rose | Mick LaSalle | Toby |
|---|---|---|---|---|---|---|
| Claudia Puig | 1 | 0.7559289 | 0.9285714 | 0.9449112 | 0.6546537 | 0.8934051 |
| Gene Seymour | 0.7559289 | 1 | 0.9449112 | 0.5 | 0 | 0.3812464 |
| Jack Matthews | 0.9285714 | 0.9449112 | 1 | 0.7559289 | 0.3273268 | 0.662849 |
| Lisa Rose | 0.9449112 | 0.5 | 0.7559289 | 1 | 0.8660254 | 0.9912407 |
| Mick LaSalle | 0.6546537 | 0 | 0.3273268 | 0.8660254 | 1 | 0.9244735 |
| Toby | 0.8934051 | 0.3812464 | 0.662849 | 0.9912407 | 0.9244735 | 1 |




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

| | Just My Luck | Lady in the Water | Snakes on a Plane | Superman Returns | The Night Listener | You Me and Dupree |
|---|---|---|---|---|---|---|
| Just My Luck | 1.0000000 | 0.6339001 | 0.7372414 | 0.7194516 | 0.8935046 | 0.7598559 |
| Lady in the Water | 0.6339001 | 1.0000000 | 0.7950515 | 0.8149529 | 0.7977412 | 0.8897565 |
| Snakes on a Plane | 0.7372414 | 0.7950515 | 1.0000000 | 0.9779829 | 0.8585983 | 0.9200319 |
| Superman Returns | 0.7194516 | 0.8149529 | 0.9779829 | 1.0000000 | 0.8857221 | 0.9680784 |
| The Night Listener | 0.8935046 | 0.7977412 | 0.8585983 | 0.8857221 | 1.0000000 | 0.9412504 |
| You Me and Dupree | 0.7598559 | 0.8897565 | 0.9200319 | 0.9680784 | 0.9412504 | 1.0000000 |




| Movies | Genre |
|---|---|
| Just My luck | Romance |
| Lady in the water | Thriller |
| snakes on a plane | Action |
| Superman Returns | ScienceFiction |
| The Night Listener | Mystery |
| You Me and Dupree | Comedy |

| | Romance | Thriller | Action | ScienceFiction | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|---|
| Just My luck | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Lady in the water | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| snakes on a plane | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Superman Returns | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| The Night Listener | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| You Me and Dupree | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

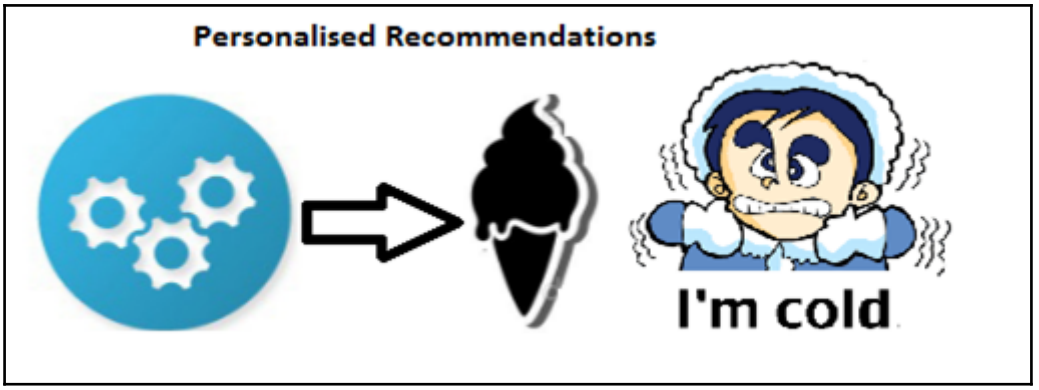
| Romance | Thriller | Action | ScienceFiction | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|
| 1.0986123 | 1.0986123 | 1.7917595 | 1.7917595 | 1.7917595 | 1.7917595 | 0.6931472 | 1.7917595 |

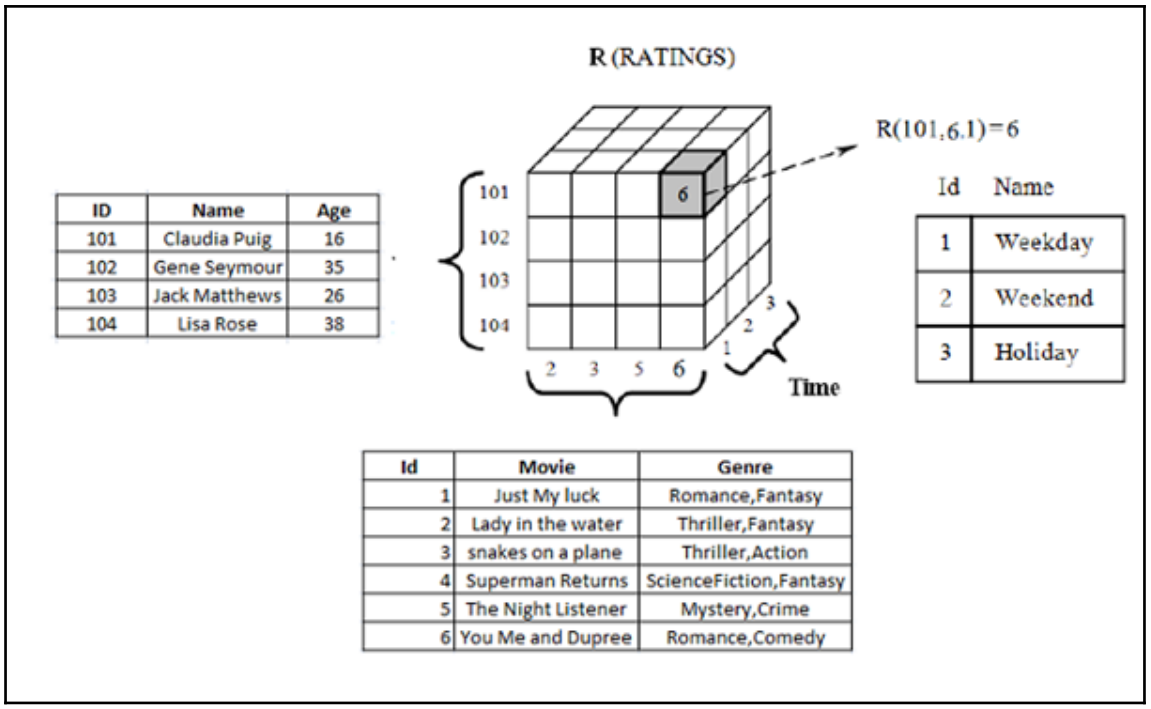
| | Romance | Thriller | Action | ScienceFiction | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|---|
| Just My luck | 1.098612 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.6931472 | 0.000000 |
| Lady in the water | 0.000000 | 1.098612 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.6931472 | 0.000000 |
| snakes on a plane | 0.000000 | 1.098612 | 1.791759 | 0.000000 | 0.000000 | 0.000000 | 0.0000000 | 0.000000 |
| Superman Returns | 0.000000 | 0.000000 | 0.000000 | 1.791759 | 0.000000 | 0.000000 | 0.6931472 | 0.000000 |
| The Night Listener | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 1.791759 | 0.000000 | 0.0000000 | 1.791759 |
| You Me and Dupree | 1.098612 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 1.791759 | 0.0000000 | 0.000000 |

| | Claudia.Puig | Gene.Seymour | Jack.Matthews | Lisa.Rose | Mick.LaSalle | Toby |
|---|---|---|---|---|---|---|
| Just My luck | 1 | 1 | *NA* | 1 | 1 | *NA* |
| Lady in the water | *NA* | 1 | 1 | 1 | 1 | *NA* |
| snakes on a plane | 1 | 1 | 1 | 1 | 1 | 1 |
| Superman Returns | 1 | 1 | 1 | 1 | 1 | 1 |
| The Night Listener | 1 | 1 | 1 | 1 | 1 | *NA* |
| You Me and Dupree | 1 | 1 | 1 | 1 | 1 | 1 |

| | Romance | Thriller | Action | ScienceFiction | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|---|
| Claudia.Puig | 6.042368 | 3.845143 | 6.271158 | 7.167038 | 8.062918 | 4.479399 | 4.852030 | 8.062918 |
| Gene.Seymour | 5.493061 | 7.140980 | 6.271158 | 8.958797 | 5.375278 | 6.271158 | 6.584898 | 5.375278 |
| Jack.Matthews | 3.845143 | 7.690286 | 7.167038 | 8.958797 | 5.375278 | 6.271158 | 5.545177 | 5.375278 |
| Lisa.Rose | 6.042368 | 6.591674 | 6.271158 | 6.271158 | 5.375278 | 4.479399 | 6.238325 | 5.375278 |
| Mick.LaSalle | 4.394449 | 7.690286 | 7.167038 | 5.375278 | 5.375278 | 3.583519 | 5.545177 | 5.375278 |
| Toby | 1.098612 | 4.943755 | 8.062918 | 7.167038 | 0.000000 | 1.791759 | 2.772589 | 0.000000 |

| | Just My luck | Lady in the water | snakes on a plane | Superman Returns | The Night Listener | You Me and Dupree |
|---|---|---|---|---|---|---|
| Claudia.Puig | 0.8919446 | 0.8826889 | 0.8057865 | 0.8173293 | 0.7461213 | 0.7964116 |
| Gene.Seymour | 0.9478958 | 0.9442628 | 0.8729500 | 0.8891199 | 0.8219716 | 0.8696591 |
| Jack.Matthews | 0.8879721 | 0.9029502 | 0.8005526 | 0.8502198 | 0.7538020 | 0.8210059 |
| Lisa.Rose | 0.9478958 | 0.9442628 | 0.8729500 | 0.8891199 | 0.8219716 | 0.8696591 |
| Mick.LaSalle | 0.9478958 | 0.9442628 | 0.8729500 | 0.8891199 | 0.8219716 | 0.8696591 |
| Toby | 0.6232739 | 0.6408335 | 0.5219196 | 0.5785800 | 0.4712180 | 0.5430630 |




| ID | Name | Age |
|---|---|---|
| 101 | Claudia Puig | 16 |
| 102 | Gene Seymour | 35 |
| 103 | Jack Matthews | 26 |
| 104 | Lisa Rose | 38 |

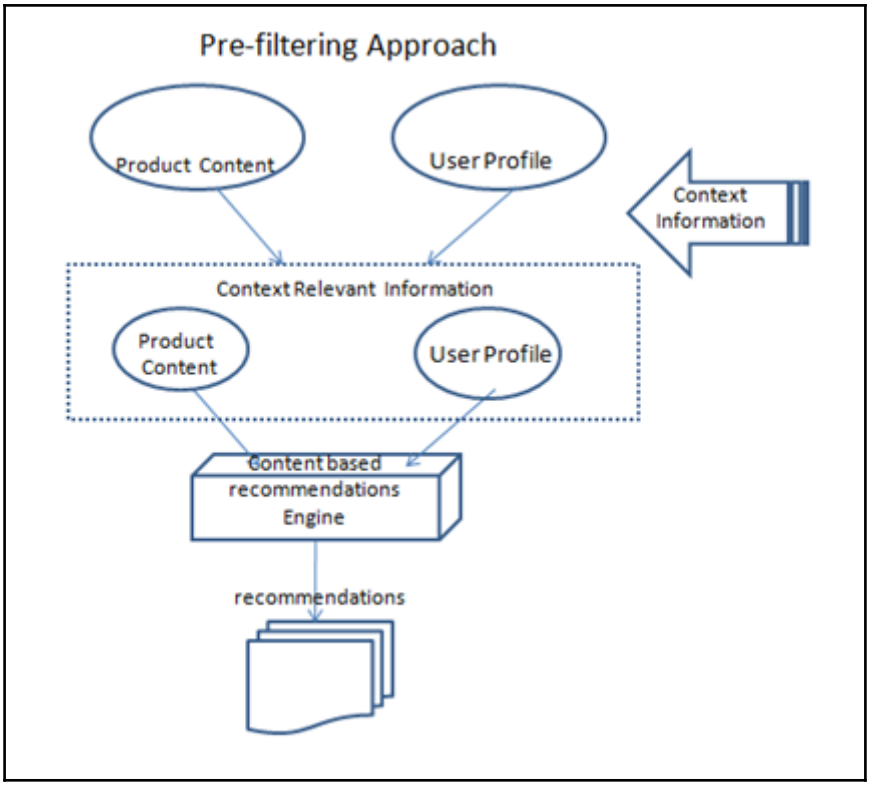
| Id | Movie | Genre |
|---|---|---|
| 1 | Just My luck | Romance,Fantasy |
| 2 | Lady in the water | Thriller,Fantasy |
| 3 | snakes on a plane | Thriller,Action |
| 4 | Superman Returns | ScienceFiction,Fantasy |
| 5 | The Night Listener | Mystery,Crime |
| 6 | You Me and Dupree | Romance,Comedy |

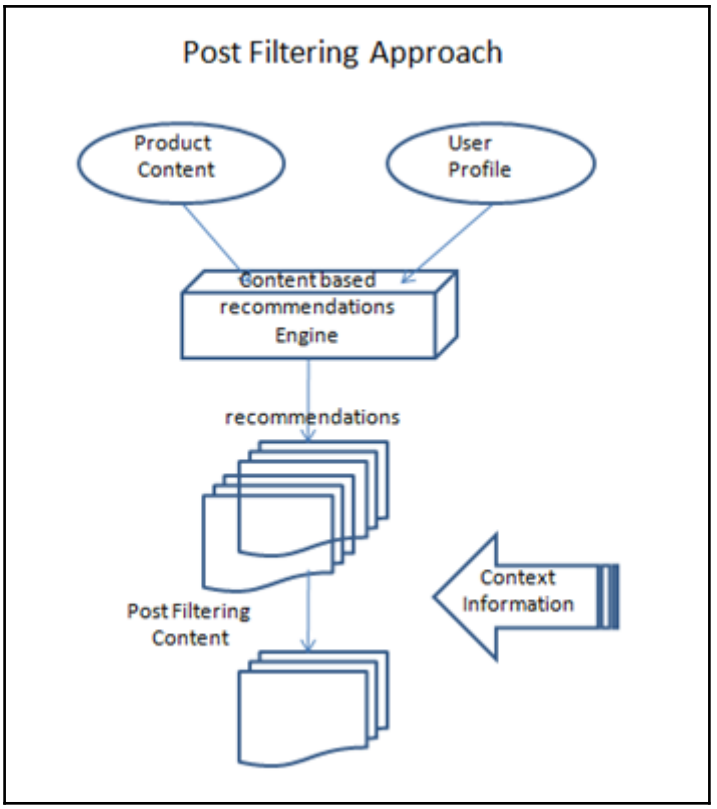
| | Romance | Thriller | Action | ScienceFi | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|---|
| weekday | 0.3 | 0 | 0.2 | 0 | 0 | 0.5 | 0 | 0 |
| weekend | 0.4 | 0 | 0 | 0.3 | 0 | 0 | 0.3 | 0 |
| Holiday | 0 | 0.5 | 0 | 0.4 | 0 | 0 | 0.1 | 0 |

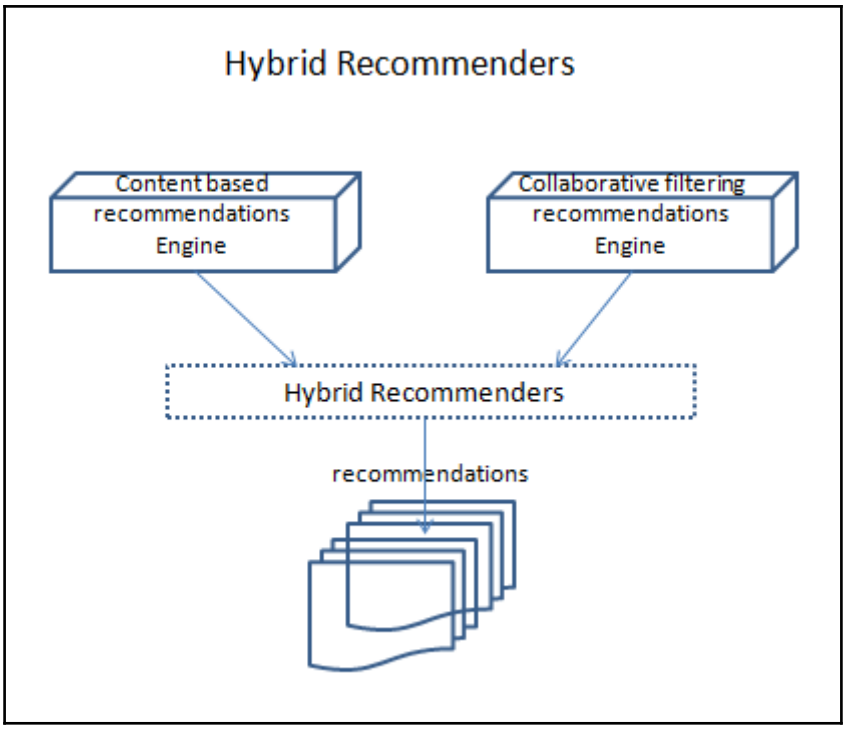
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Preferences of TOBY | 1.09 | 4.94 | 8.06 | 7.16 | 0 | 1.79 | 2.77 | 0 |

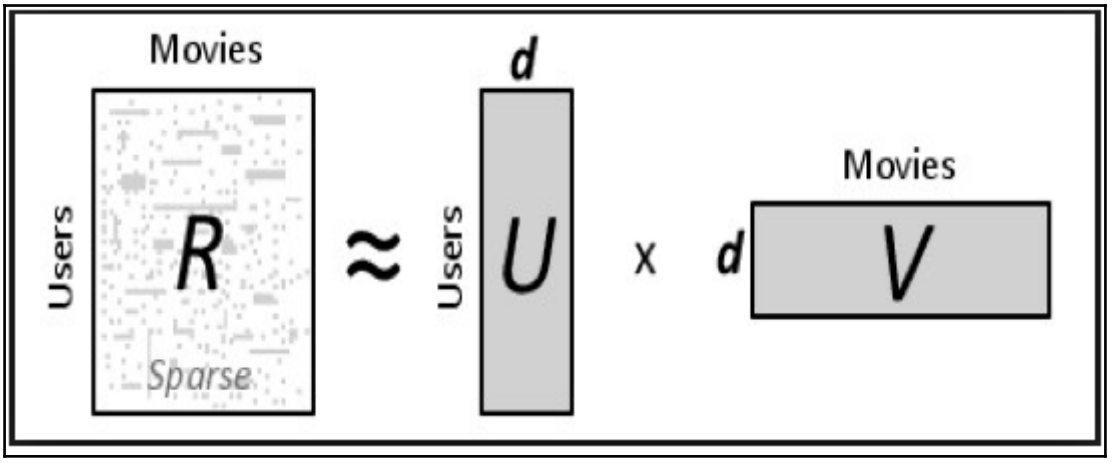
| | | Romance | Thriller | Action | ScienceFi | Mystery | Comedy | Fantasy | Crime |
|---|---|---|---|---|---|---|---|---|---|
| TOBY CONTEXT | weekday | 0.32958 | 0 | 0.35835 | 0 | 0 | 0.89588 | 0 | 0 |
| | weekend | 0.43944 | 0 | 0 | 0.53753 | 0 | 0 | 0.20794 | 0 |
| | Holiday | 0 | 0.54931 | 0 | 0.7167 | 0 | 0 | 0.06931 | 0 |

| | Just My luck | Lady in the water | snakes on a plane | Superman Returns | The Night Listener | You Me and Dupree |
|---|---|---|---|---|---|---|
| weekday | 0.27337511 | 0.000000 | 0.2996171 | 0.0000000 | 0 | 0.918004 |
| weekend | 0.66588719 | 0.153096 | 0.0000000 | 0.7952170 | 0 | 0.316934 |
| Holiday | 0.04083948 | 0.553805 | 0.3170418 | 0.7656785 | 0 | 0.000000 |

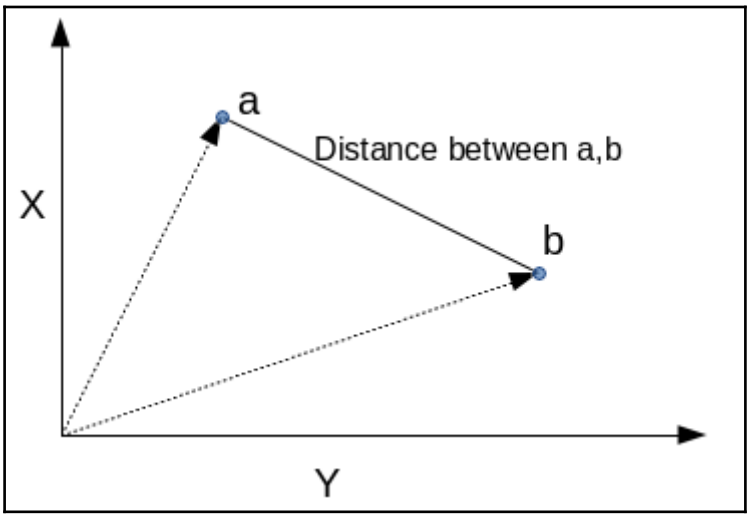



Post Filtering Approach
Product Content
User Profile
Content based recommendations Engine
recommendations
Post Filtering Content
Context Information


Hybrid Recommenders
Content based recommendations Engine
Collaborative filtering recommendations Engine
Hybrid Recommenders
recommendations


Movies
d
Users
R
Sparse
≈
Users
U
x
d
Movies
V

# Chapter 4: Data Mining Techniques Used in Recommendation Engines



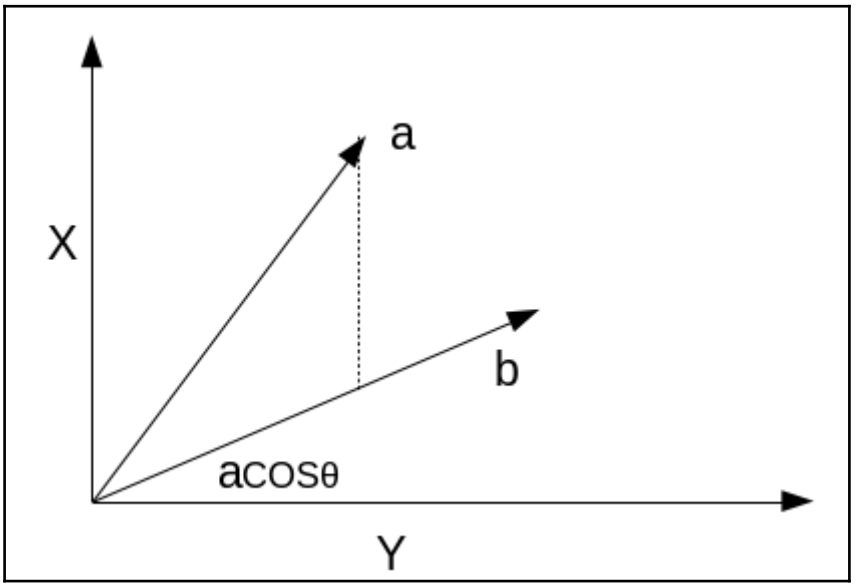

$$\text{Euclidean Distance}(x, y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

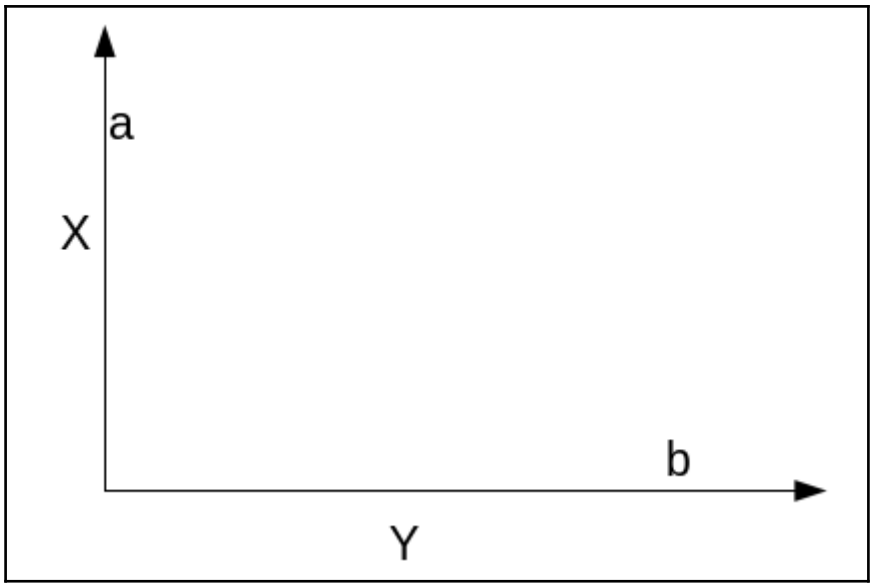
```
> x1
 [1]  0.7824548 -0.2623895  0.5276719  1.2552186 -0.9803315 -0.4561338  2.4051567
 [8]  0.6858002 -0.8711695 -0.2618928  0.2973917  0.8448787  0.2188954 -1.2323462
[15]  0.9133412 -0.4238214 -0.5814376 -0.2448999  1.1896259 -0.9937443  0.6576142
[22] -0.1357882 -0.5627333 -0.8575745  0.2385076  0.7217603 -1.7579127 -0.7489078
[29] -0.3605539 -0.7173789
> x2
 [1]  0.36918961 -0.85669259 -0.66356226  0.70927104 -0.24235742  0.68548041  0.97911641
 [8]  0.19732953 -0.83348519  0.38272366 -1.61543924  2.31314283  1.44765481 -0.77416639
[15]  1.20584033 -0.94992148 -0.73585753 -1.32329554  0.10810163 -0.62878243  1.22097185
[22]  0.33721922 -0.03807742 -0.55773028  0.68864984  1.26823921 -0.94928127  0.88784091
[29]  0.81162258  1.37679405
> Euc_dist = dist(rbind(x1,x2),method = "euclidean")
> Euc_dist
         x1
x2 5.259711
>
```

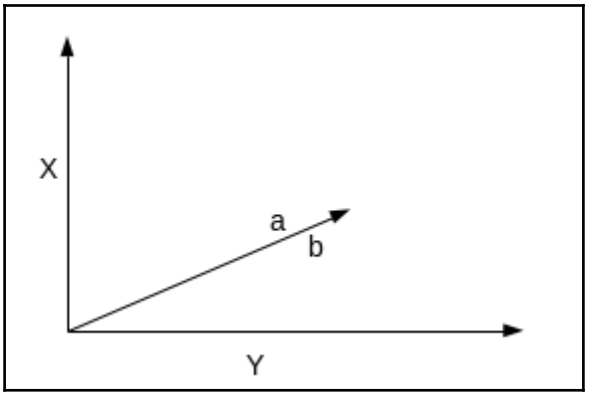
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

$$\vec{a} \cdot \vec{b} = \|\vec{a}\|\|\vec{b}\| \cos\theta$$

$$\vec{a} \cdot \vec{b} = \|\vec{b}\|\|\vec{a}\| \cos\theta$$





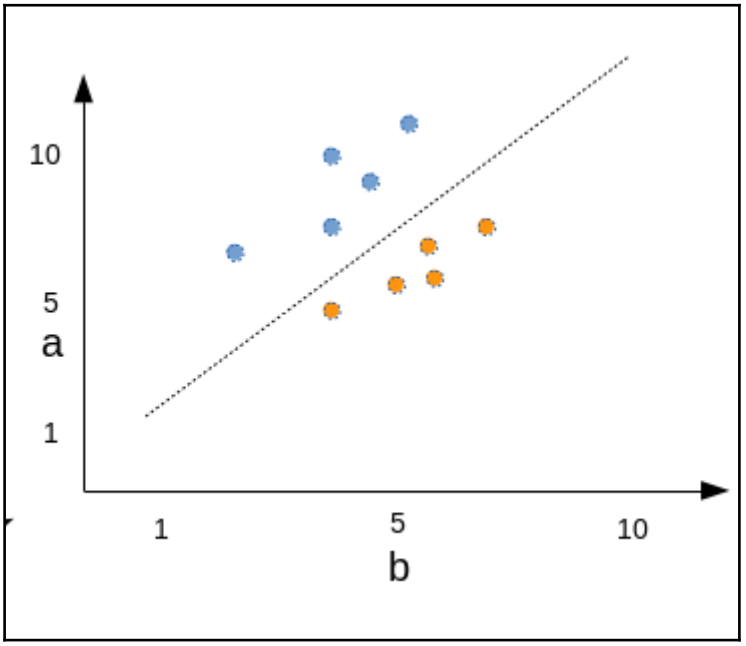

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\theta$$

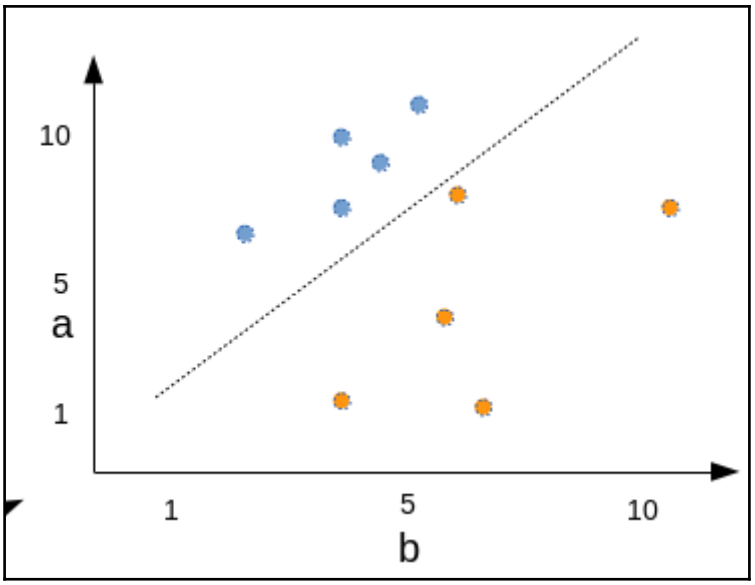
$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

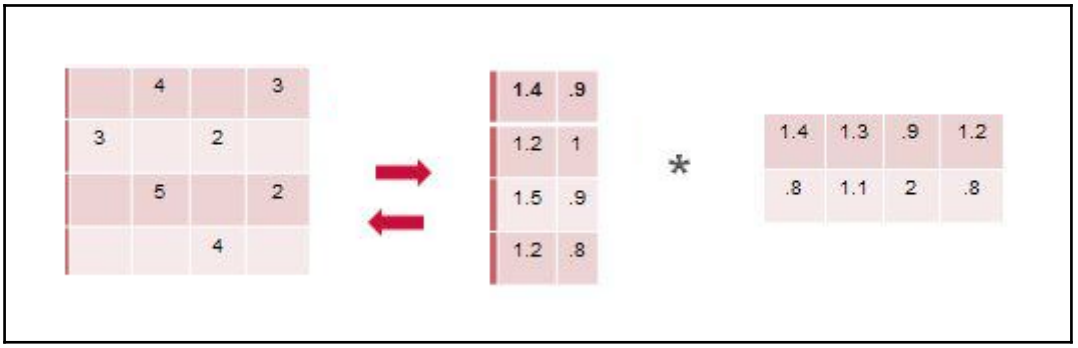
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

$$r = r_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)$$

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

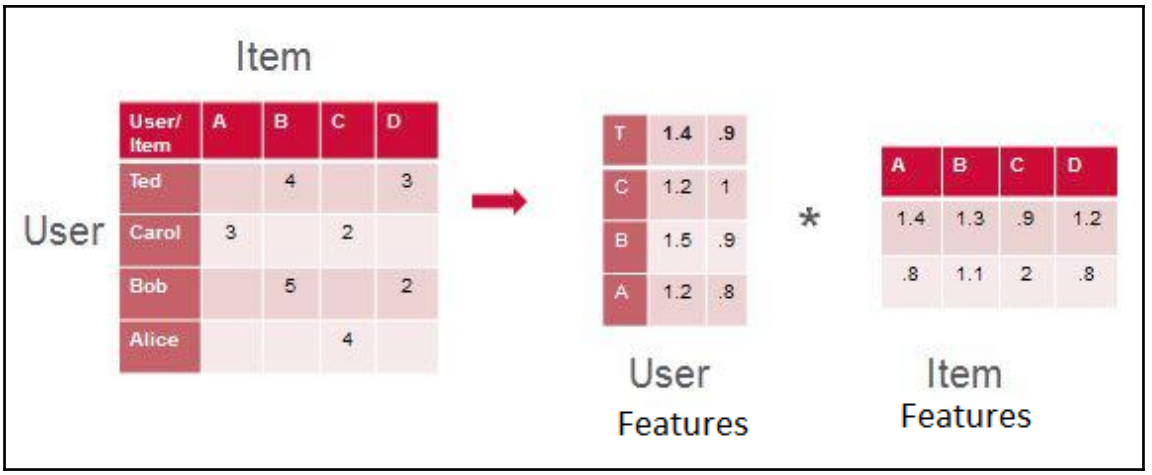





$$\hat{r}_{ij} = p_i^T q_j = \sum_{k=1}^{k} p_{ik} q_{kj}$$

$$e_{ij}^2 = (r_{ij} - \hat{r}_{ij})^2 = (r_{ij} - \sum_{k=1}^{K} p_{ik} q_{kj})^2$$




$$\min_{q^*,p^*} \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

$$\mathbf{A} = \mathbf{U} \times \mathbf{\Sigma} \times \mathbf{V}^T$$

(m x n) (m x r) (r x r) (r x n)

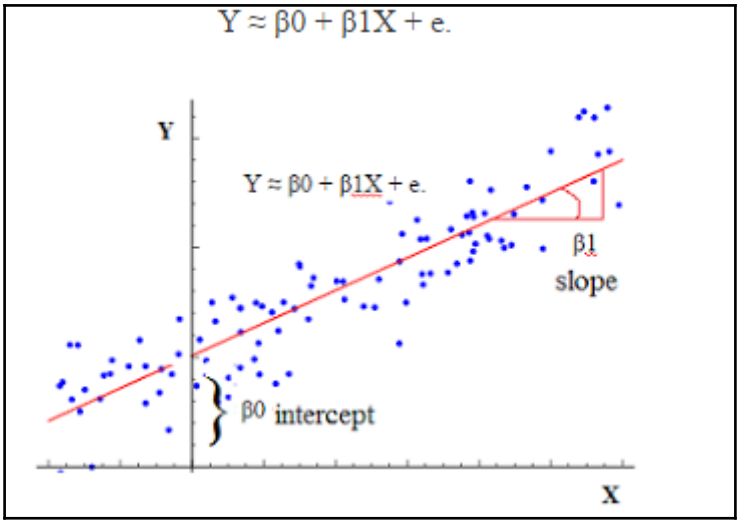
Where **U** is (m x r) matrix
V is (n x r) matrix
$\Sigma$ is (r x r ) matrix

**U** is (m x k) matrix
V is (n x k) matrix
$\Sigma$ is (k x k) matrix

Step 1: $U \otimes \Sigma^{1/2}$ (m x k)

Step 2: $\Sigma^{1/2} \otimes V^T$ (k x n)

Step 3: $\hat{A}$ (m x n) = $U \otimes \Sigma^{1/2}$ (m x k) $\otimes$ $\Sigma^{1/2} \otimes V^T$ (k x n)



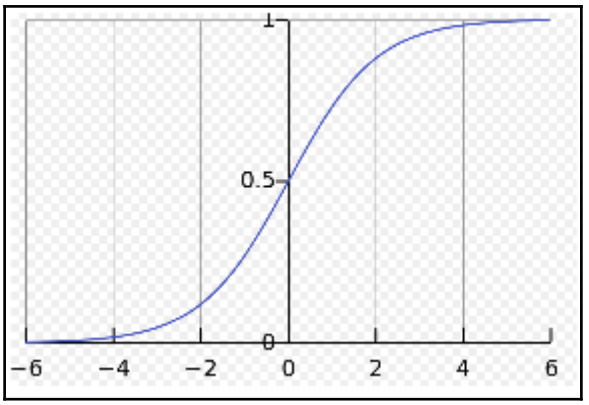

$$P_{u,i} = \frac{\sum_{\text{all similar items, N}}(s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, N}}(|s_{i,N}|)}$$

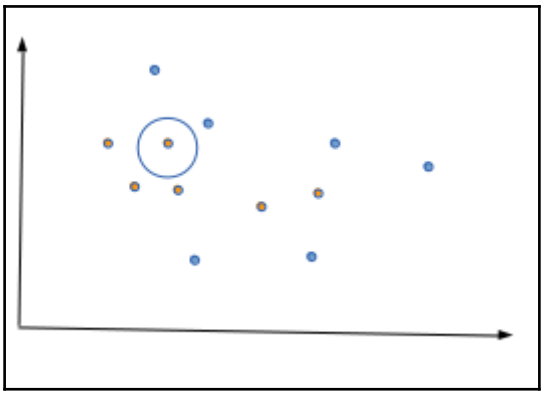
$$\bar{R_N'} = \alpha \bar{R_i} + \beta + \epsilon$$

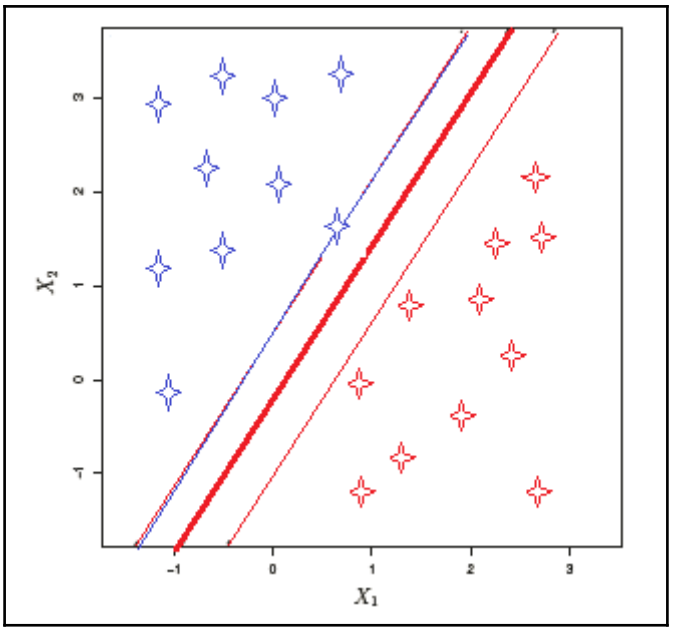
$$F(x) = \frac{1}{1 + e^{-x}}$$

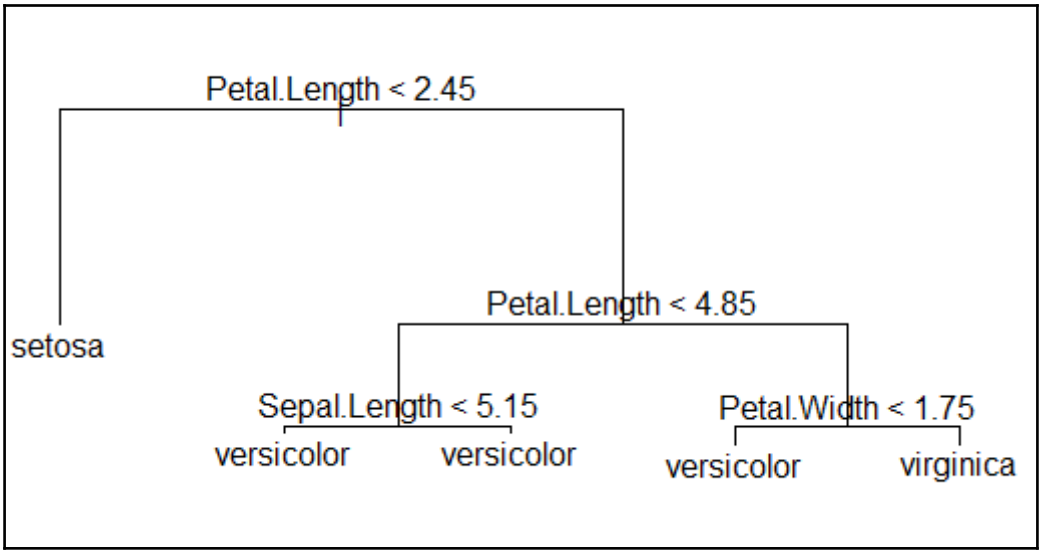
$$x = \beta_0 + \beta_1 x$$

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

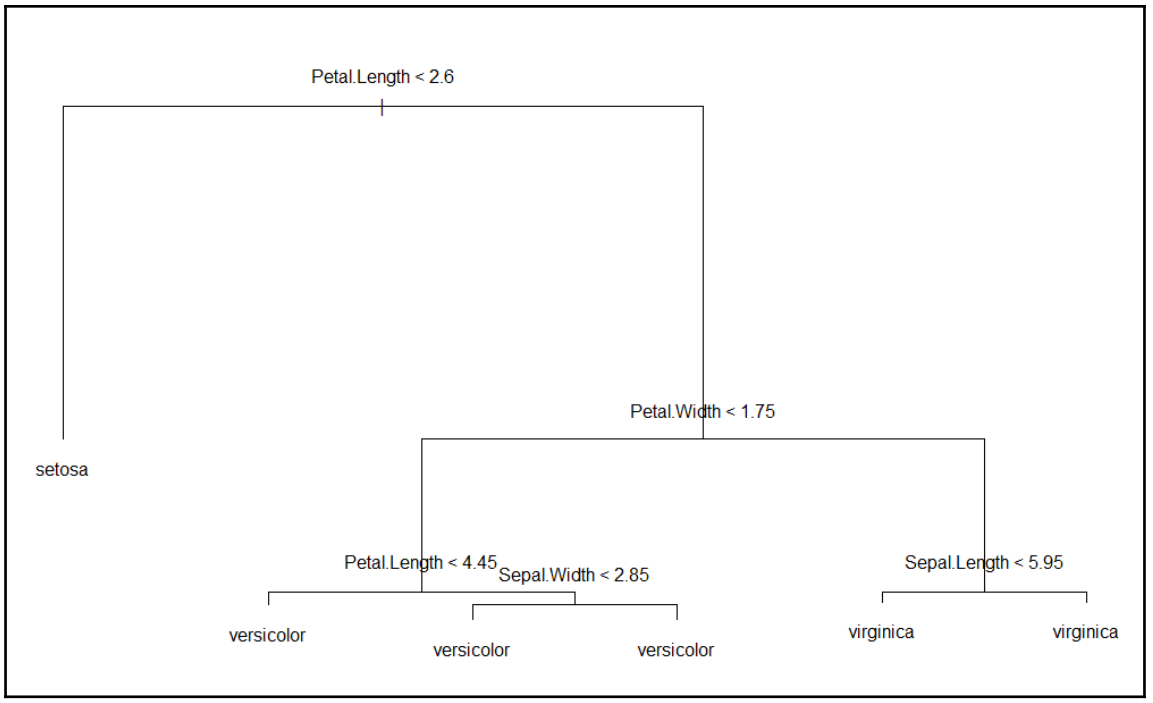



$X_2$
$X_1$
-1
0
1
2
3


Petal.Length < 2.45
setosa
Petal.Length < 4.85
Sepal.Length < 5.15
versicolor
versicolor
Petal.Width < 1.75
versicolor
virginica

```
Classification tree:
tree(formula = Species ~ ., data = train)
Number of terminal nodes:  6
Residual mean deviance:  0.1471 = 14.56 / 99
Misclassification error rate: 0.0381 = 4 / 105
```

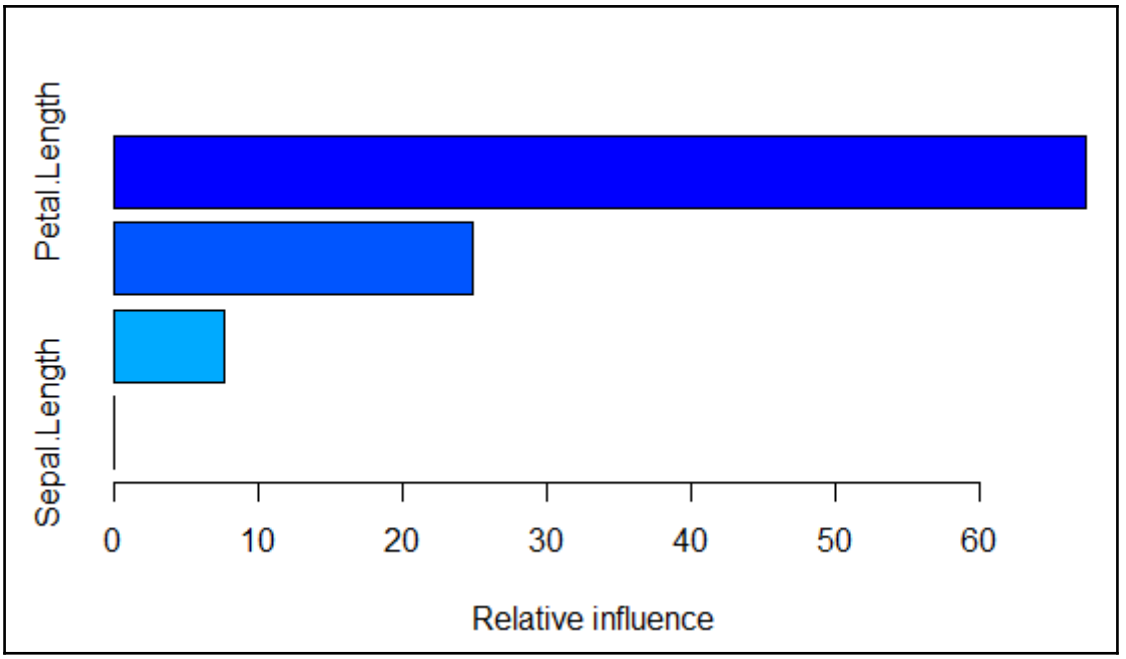



```
> pred
 [1] versicolor virginica  versicolor versicolor versicolor setosa     virginica
 [8] setosa     virginica  versicolor setosa     setosa     setosa     setosa
[15] versicolor versicolor versicolor setosa     versicolor versicolor virginica
[22] virginica  virginica  setosa     setosa     versicolor virginica  versicolor
[29] versicolor virginica  virginica  virginica  versicolor virginica  versicolor
[36] virginica  versicolor setosa     virginica  setosa     virginica  versicolor
[43] versicolor versicolor setosa
Levels: setosa versicolor virginica
```

```
> model

Call:
 randomForest(formula = Species ~ ., data = train, mtry = 2, importance = TRUE,      proximity = TRUE)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 2

        OOB estimate of  error rate: 6.67%
Confusion matrix:
           setosa versicolor virginica class.error
setosa         40          0         0  0.00000000
versicolor      0         28         3  0.09677419
virginica       0          4        30  0.11764706
```

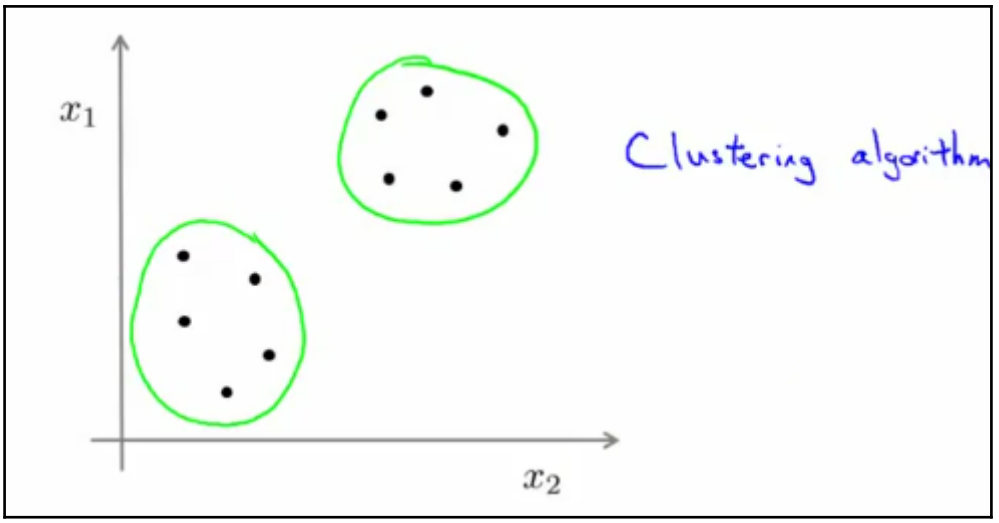
```
> pred
        96        105        138         99         39         37        149        106         29
versicolor  virginica  virginica versicolor     setosa     setosa  virginica  virginica     setosa
        66         61         70        140         83        126         77         53        102
versicolor versicolor versicolor  virginica versicolor  virginica versicolor versicolor  virginica
       135         82        103         52        146         58         67         19         87
 virginica versicolor  virginica versicolor  virginica versicolor versicolor     setosa versicolor
         5        124         57         42         68        100        145         32          6
    setosa  virginica versicolor     setosa versicolor versicolor  virginica     setosa     setosa
       139         21         86        148        130        108         47         98         73
 virginica     setosa versicolor  virginica  virginica  virginica     setosa versicolor  virginica
Levels: setosa versicolor virginica
```

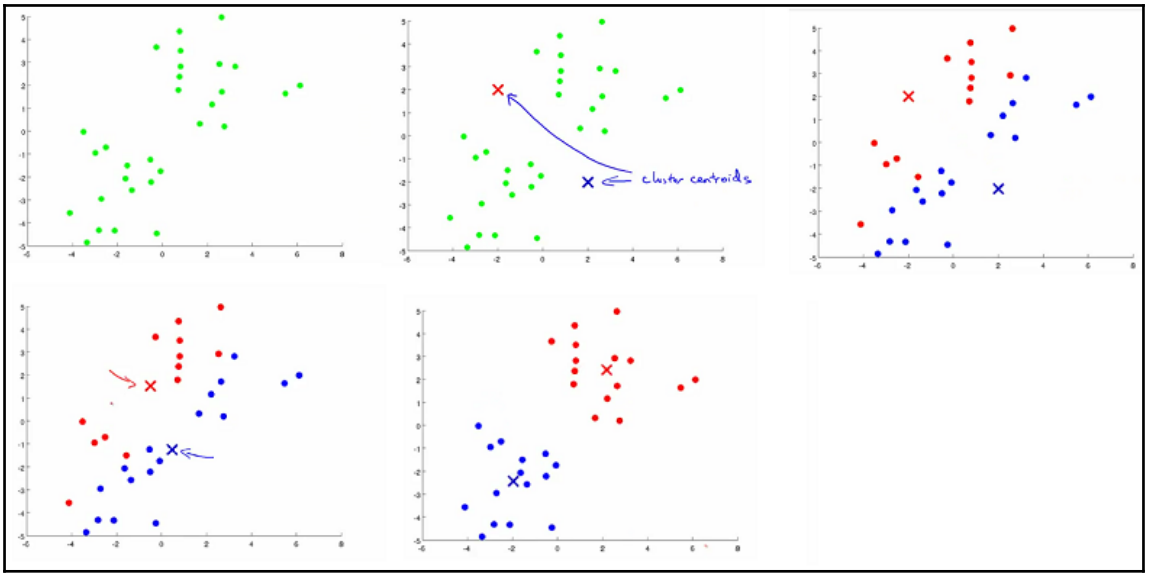
```
> summary(model)
                             var   rel.inf
Petal.Length Petal.Length 67.440852
Petal.Width   Petal.Width 24.942084
Sepal.Width   Sepal.Width  7.617065
Sepal.Length Sepal.Length  0.000000
```

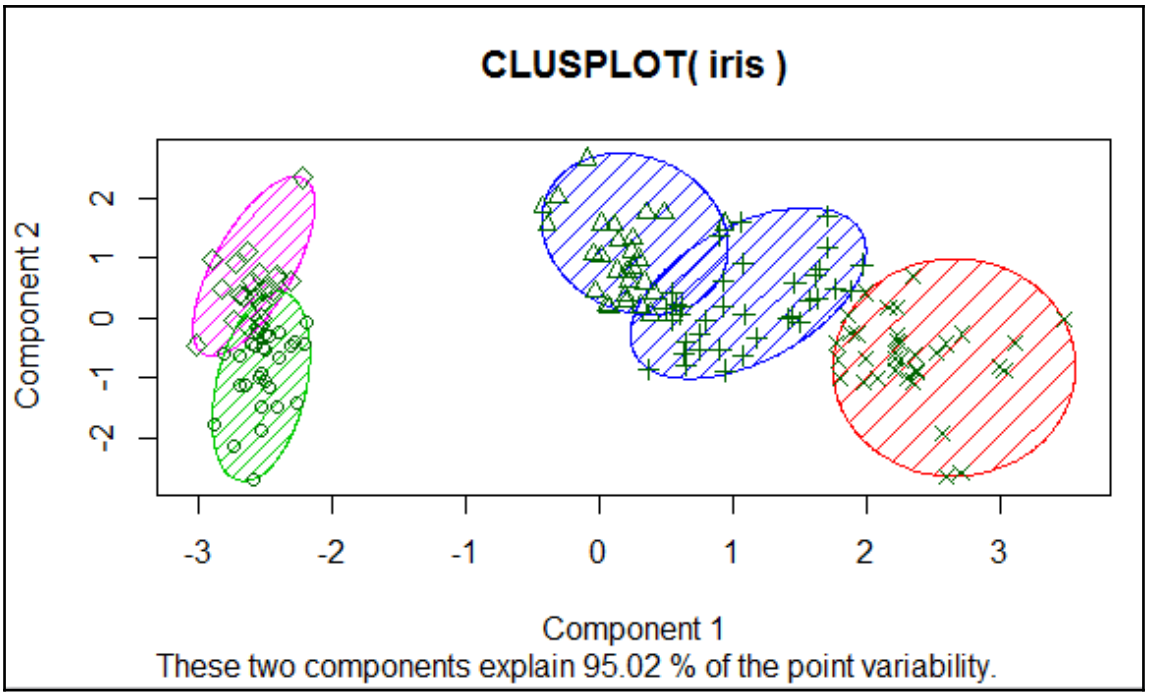


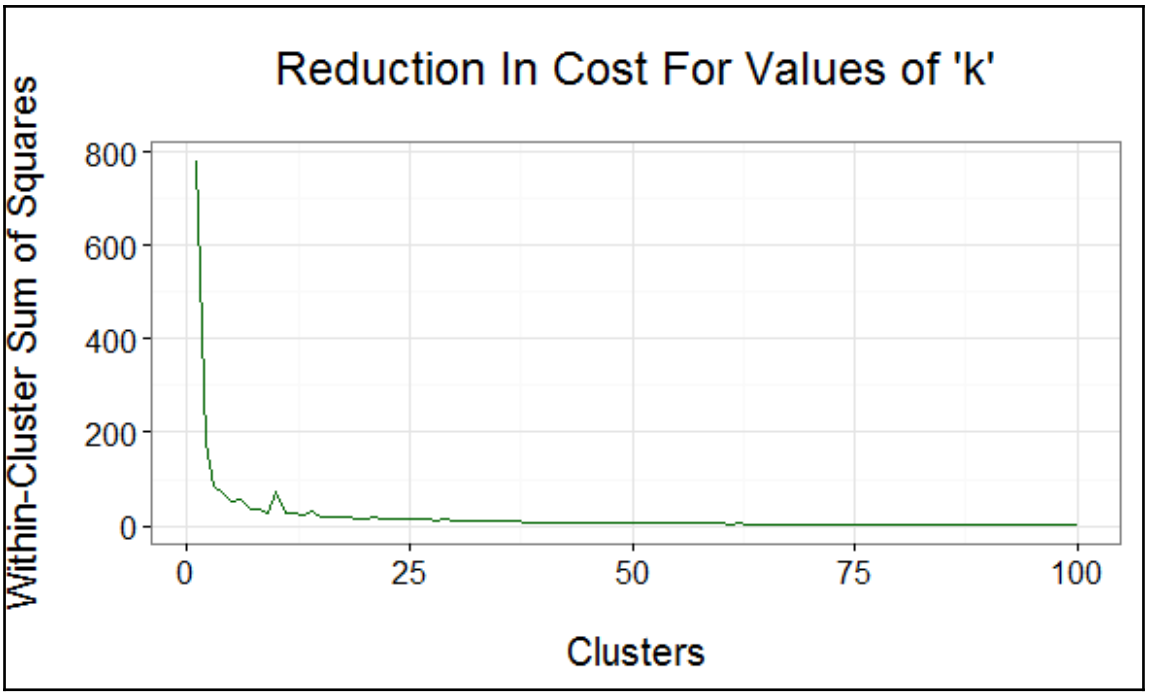

```
> pred[1:5,,]
        setosa versicolor virginica
[1,]  5.647443  -2.951628 -4.964130
[2,] -5.238890  -3.222812  4.295997
[3,] -5.289086   3.447595 -3.277463
[4,] -5.288114   2.690219 -2.389940
[5,] -5.245599  -1.588168  3.026419
>
```

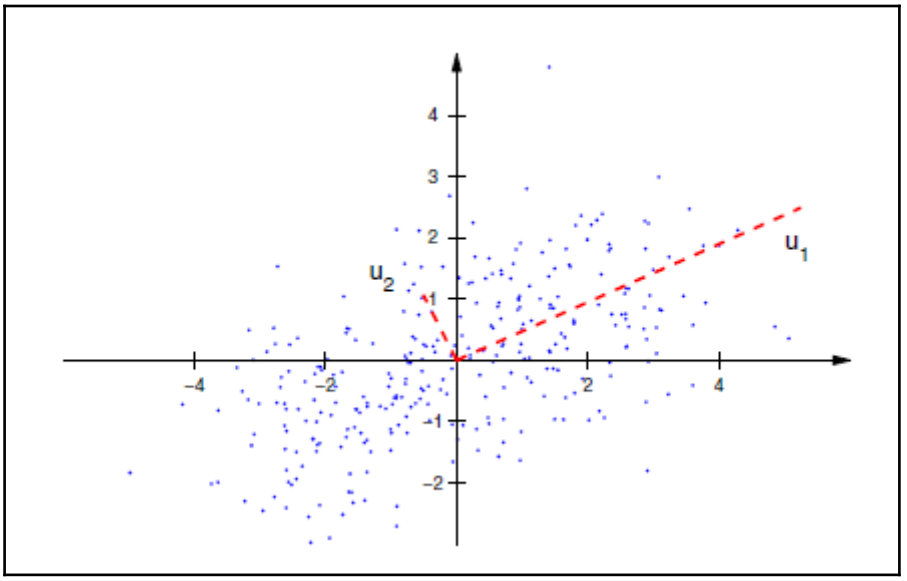
```
> p.pred
  [1] 1 3 2 2 3 1 1 1 1 3 1 3 2 2 2 3 2 1 2 1 3 2 3 1 1 2 3 1 2 1 2 2 1 3 3 1 2 1 1 3 2 2 3 2 2
>
```

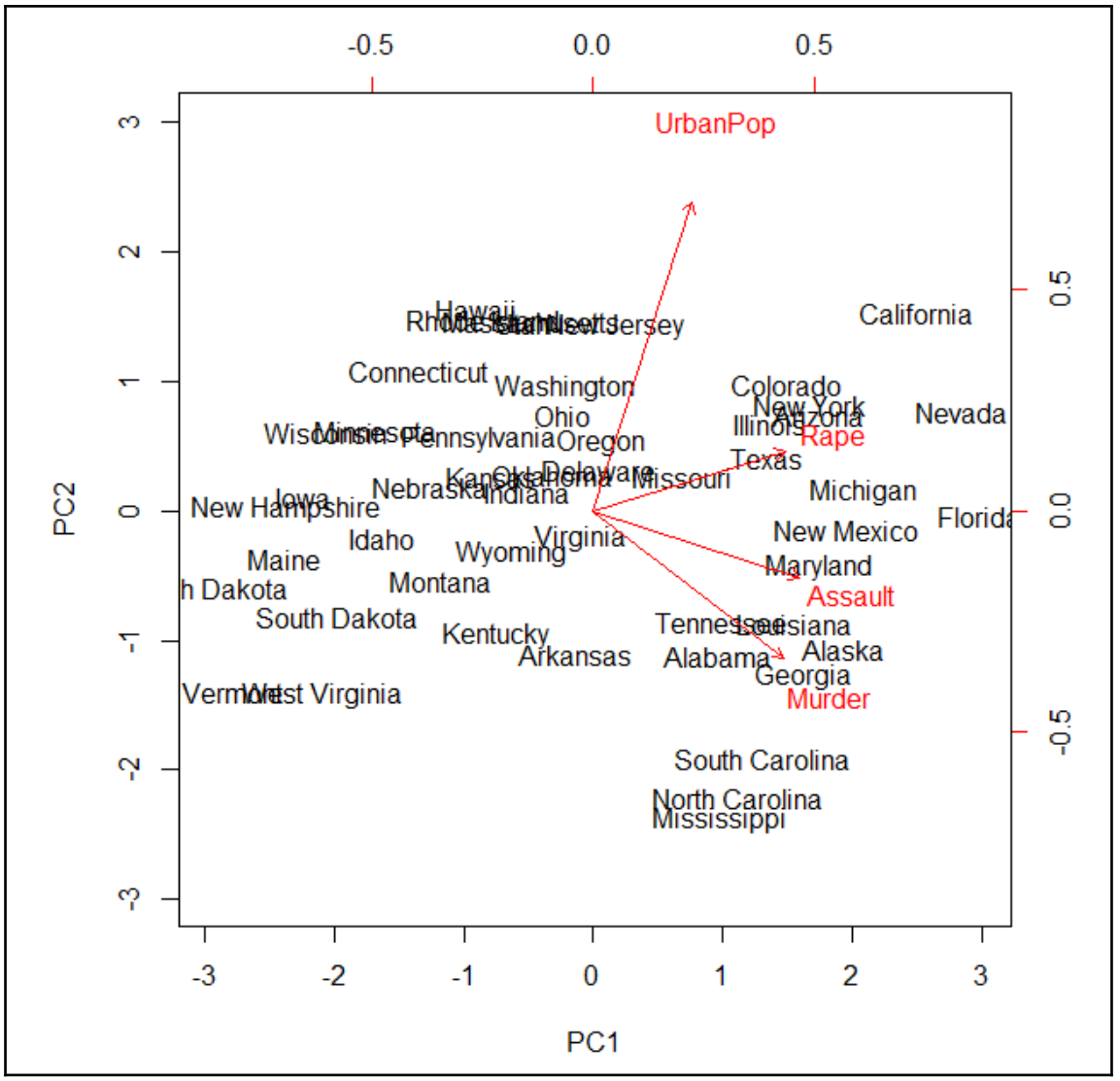

$x_1$
$x_2$
Clustering algorithm


cluster centroids

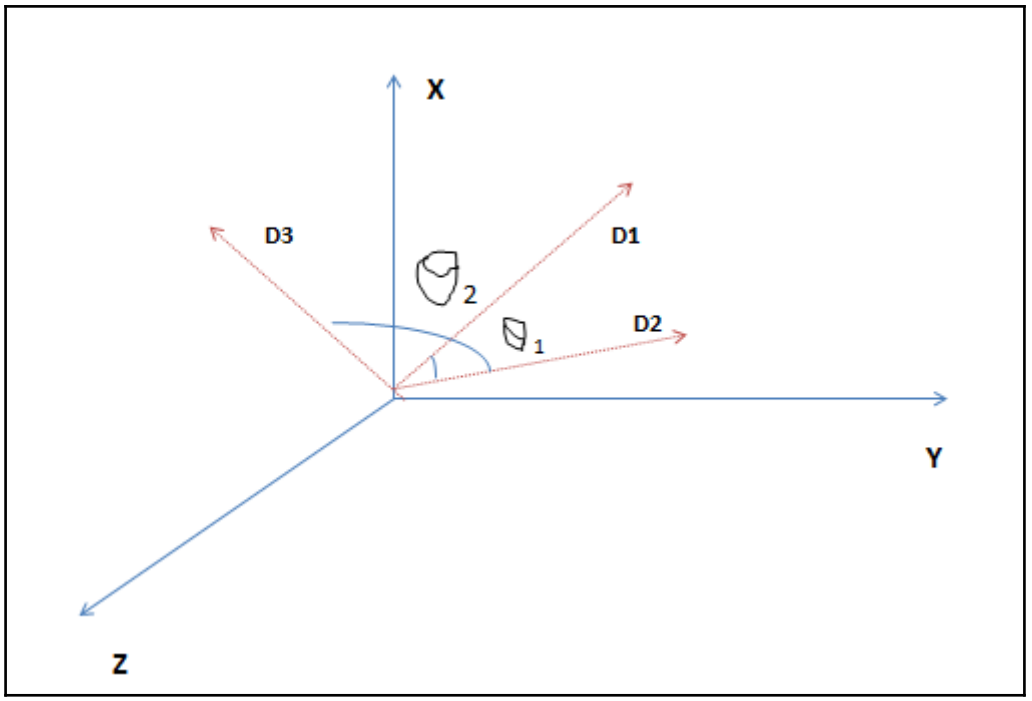

CLUSPLOT( iris )
Component 2
-2
-1
0
1
2
-3
-2
-1
0
1
2
3
Component 1
These two components explain 95.02 % of the point variability.


Reduction In Cost For Values of 'k'
Within-Cluster Sum of Squares
800
600
400
200
0
0
25
50
75
100
Clusters


4
3
2
1
−1
−2
−4
−2
2
4
u2
u1




| | THE | CAT | CHASES | RAT | DOG | MAN | WALKS | ON | MAT |
|---|---|---|---|---|---|---|---|---|---|
| D1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| D2 | 1 | 1 | 1 | | 1 | 0 | 0 | 0 | 0 |
| D3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |



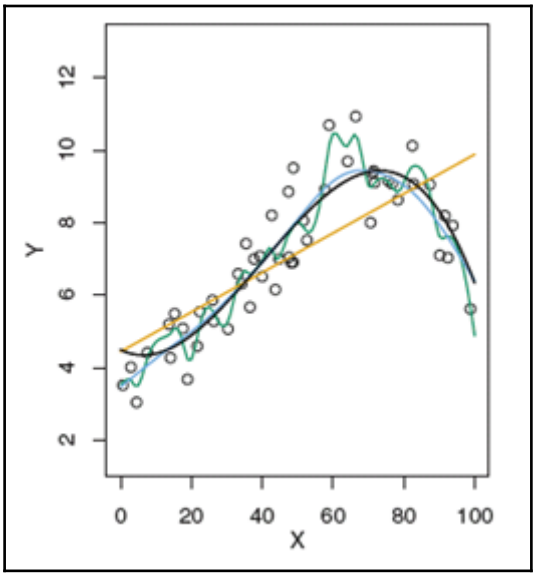

| | THE | CAT | CHASES | RAT | DOG | MAN | WALKS | ON | MAT |
|---|---|---|---|---|---|---|---|---|---|
| D1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| D2 | 1 | 1 | 1 | | 1 | 0 | 0 | 0 | 0 |
| D3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| DF | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| IDF | -0.12494 | 0 | 0 | 0.477121 | 0.477121 | 0.477121 | 0.477121 | 0.477121 | 0.477121 |

| | THE | CAT | CHASES | RAT | DOG | MAN | WALKS | ON | MAT |
|---|---|---|---|---|---|---|---|---|---|
| D1 | -0.12494 | 0 | 0 | 0.477121 | 0 | 0 | 0 | 0 | 0 |
| D2 | -0.12494 | 0 | 0 | 0 | 0.477121 | 0 | 0 | 0 | 0 |
| D3 | -0.12494 | 0 | 0 | 0 | 0 | 0.477121 | 0.477121 | 0.477121 | 0.477121 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| "for | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "growth | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "if | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "is | 0.000000000 | 0.000000000 | 0.015238202 | 0.000000000 | 0.000000000 | 0.046137890 | 0.000000000 |
| "may | 0.000000000 | 0.000000000 | 0.019825358 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "none | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.054457838 |
| "opec | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "opec's | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "our | 0.000000000 | 0.021082576 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "the | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "there | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "they | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |
| "this | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 | 0.000000000 |




$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{act} - X_{pred})^2}{n}}$$

Item

| User / Item | A | B | C | D |
|---|---|---|---|---|
| Ted | | 4 | | 3 |
| Carol | 3 | | 2 | |
| Bob | | 5 | | 2 |
| Alice | | | 4 | |

User

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|$$

| | | Preferred | |
|---|---|---|---|
| | | **TRUE** | **FALSE** |
| **Recommended** | **POSITIVE** | **25** | **10** |
| | **NEGATIVE** | **5** | **10** |

| | | ACTUAL | |
|---|---|---|---|
| | | **TRUE** | **FALSE** |
| **PREDICTED** | **POSITIVE** | *TRUE POSITIVE* | *FALSE POSITIVE* |
| | **NEGATIVE** | *FALSE NEGATIVE* | *TRUE NEGATIVE* |

$$\text{Precision} = \frac{\#\text{tp}}{\#\text{tp} + \#\text{fp}}$$

$$\text{Recall (True Positive Rate)} = \frac{\#\text{tp}}{\#\text{tp} + \#\text{fn}}$$

$$\sum$$

$$\hat{A}:$$

$$m = \sqrt{p}$$

$$p$$

# Chapter 5: Building Collaborative Filtering Recommendation Engines

**Lab for Developing and Testing Recommender Algorithms**

Documentation for package 'recommenderlab' version 0.2-0

- DESCRIPTION file.
- User guides, package vignettes and other documentation.

Help Pages

A B C D E F G H I J L M N P R S T misc

-- A --

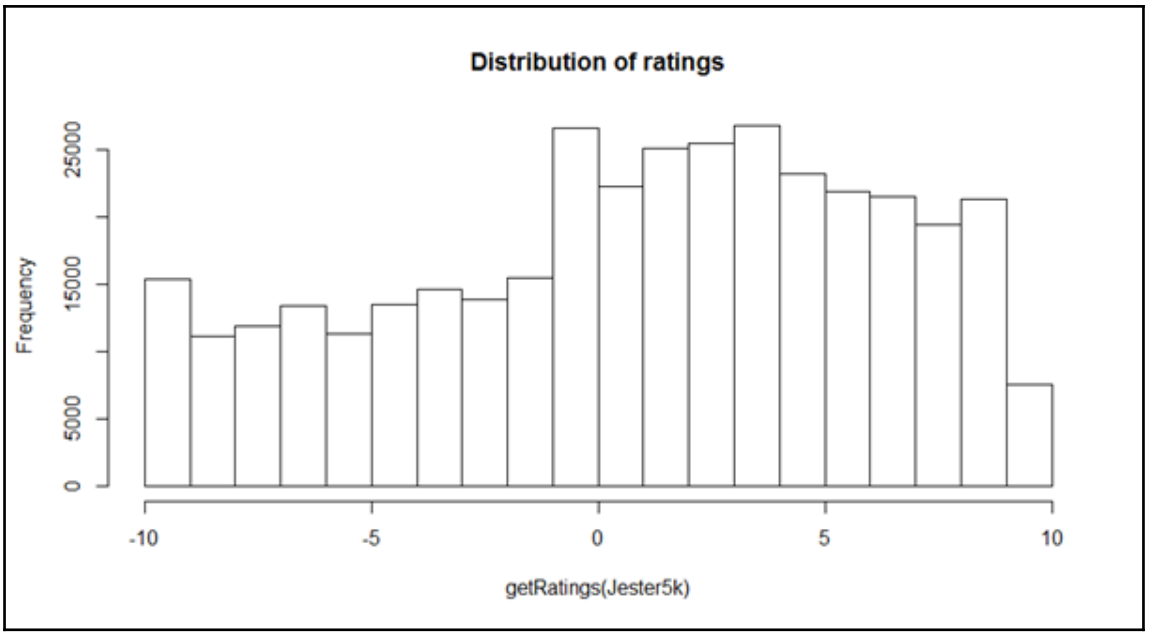
| | |
|---|---|
| avg | Class "evaluationResults": Results of the Evaluation of a Single Recommender Method |
| avg-method | Class "evaluationResultList": Results of the Evaluation of a Multiple Recommender Methods |
| avg-method | Class "evaluationResults": Results of the Evaluation of a Single Recommender Method |

-- B --

| Item | Title |
|---|---|
| Jester5k | Jester dataset (5k sample) |
| JesterJokes (Jester5k) | Jester dataset (5k sample) |
| MSWeb | Anonymous web data from www.microsoft.com |
| MovieLense | MovieLense Dataset (100k) |
| MovieLenseMeta (MovieLense) | MovieLense Dataset (100k) |

| | | |
|---|---|---|
| [ | dimnames<- | Recommender |
| binarize | dissimilarity | removeKnownRatings |
| calcPredictionAccuracy | evaluationScheme | rowCounts |
| calcPredictionAccuracy | getData.frame | rowMeans |
| colCounts | getList | rowSds |
| colMeans | getNormalize | rowSums |
| colSds | getRatings | sample |
| colSums | getTopNLists | show |
| denormalize | image | similarity |
| dim | normalize | |
| dimnames | nratings | |

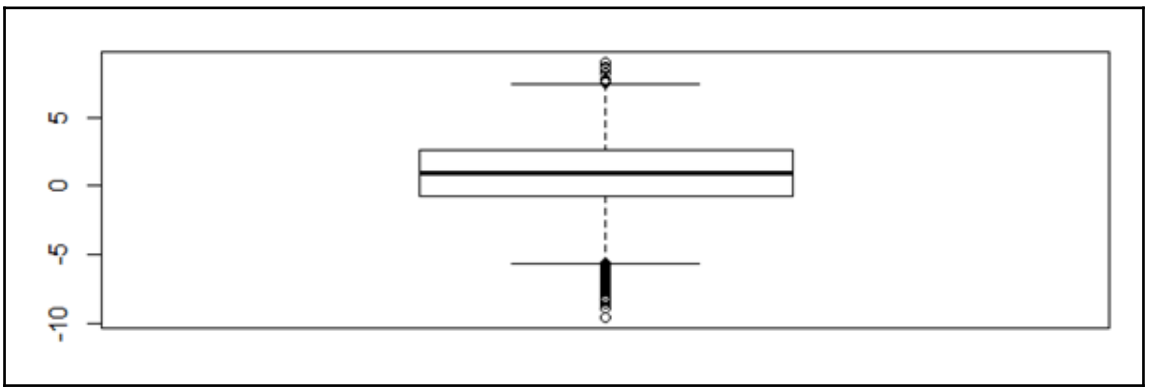
| **Models** |
|---|
| IBCF_realRatingMatrix |
| RERECOMMEND_realRatingMatrix |
| UBCF_realRatingMatrix |
| POPULAR_realRatingMatrix |
| RANDOM_realRatingMatrix |
| SVD_realRatingMatrix |
| SVDF_realRatingMatrix |

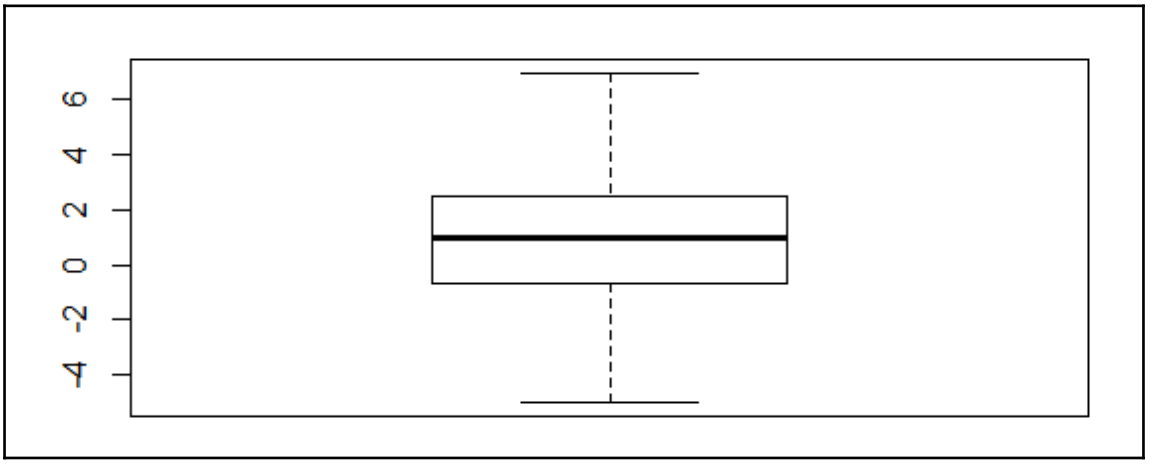


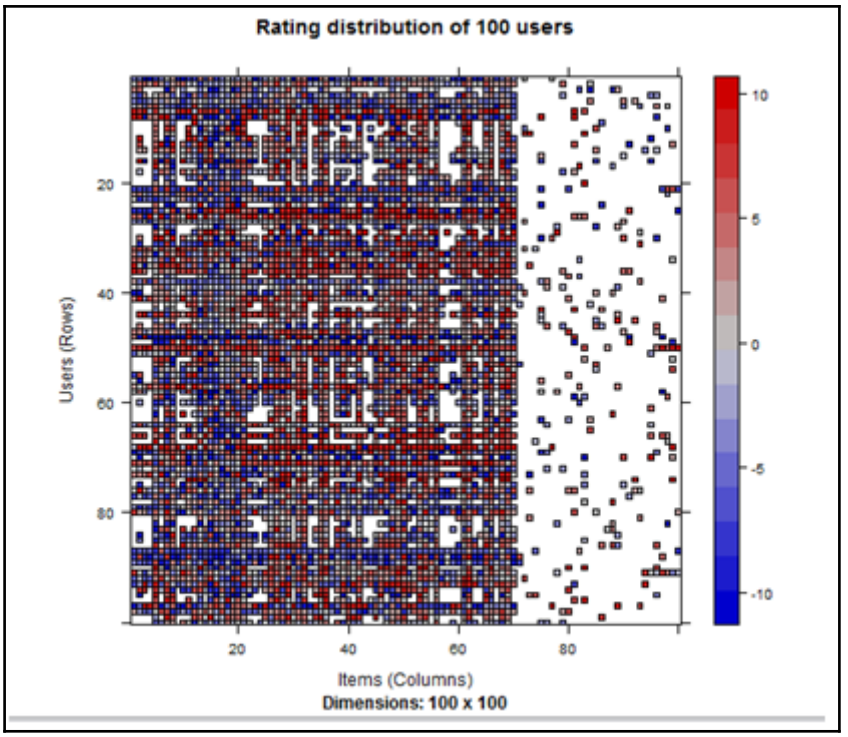

| | j1 | j2 | j3 | j4 | j5 | j6 | j7 | j8 | j9 | j10 |
|---|---|---|---|---|---|---|---|---|---|---|
| u2841 | 7.91 | 9.17 | 5.34 | 8.16 | -8.74 | 7.14 | 8.88 | -8.25 | 5.87 | 6.21 |
| u15547 | -3.20 | -3.50 | -9.56 | -8.74 | -6.36 | -3.30 | 0.78 | 2.18 | -8.40 | -8.79 |
| u15221 | -1.70 | 1.21 | 1.55 | 2.77 | 5.58 | 3.06 | 2.72 | -4.66 | 4.51 | -3.06 |
| u15573 | -7.38 | -8.93 | -3.88 | -7.23 | -4.90 | 4.13 | 2.57 | 3.83 | 4.37 | 3.16 |
| u21505 | 0.10 | 4.17 | 4.90 | 1.55 | 5.53 | 1.50 | -3.79 | 1.94 | 3.59 | 4.81 |
| u15994 | 0.83 | -4.90 | 0.68 | -7.18 | 0.34 | -4.32 | -6.17 | 6.12 | -5.58 | 5.44 |

| Users | Recommended Jokes |
|---|---|
| **u21505** | "j81" "j73" "j83" "j75" "j100" "j80" "j72" "j95" "j87" "j96" |
| | |
| **u5809** | "j97" "j93" "j76" "j78" "j77" "j85" "j89" "j98" "j91" "j80" |
| | |
| **u12519** | "j98" "j100" "j80" "j93" "j99" "j87" "j76" "j89" "j84" "j96" |
| | |
| **u12094** | "j89" "j96" "j78" "j94" "j88" "j86" "j87" "j93" "j91" "j99" |

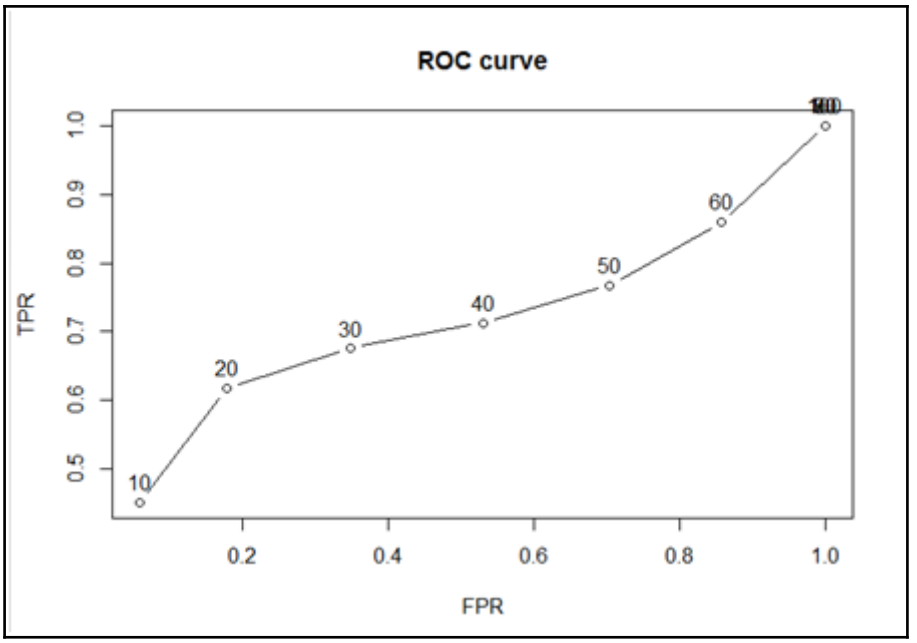
| 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 114 | 77 | 80 | 78 | 74 | 70 | 74 | 75 | 80 | 81 | 70 | 60 | 69 | 61 | 62 | 47 | 42 | 52 | 52 | 48 | 56 | 54 | 34 | 43 |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 |
| 48 | 41 | 42 | 42 | 41 | 53 | 51 | 39 | 39 | 29 | 87 | 221 | 364 | 312 | 131 | 48 | 36 | 19 | 33 | 32 | 38 | 20 | 19 | 18 |
| 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | | | | | | | |
| 15 | 23 | 21 | 12 | 12 | 13 | 14 | 16 | 10 | 12 | 12 | 11 | 16 | 13 | 8 | 16 | 1422 | | | | | | | |

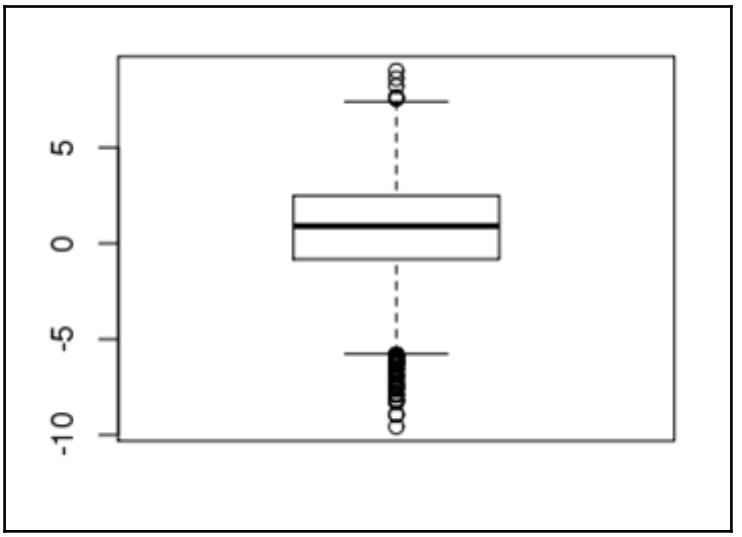


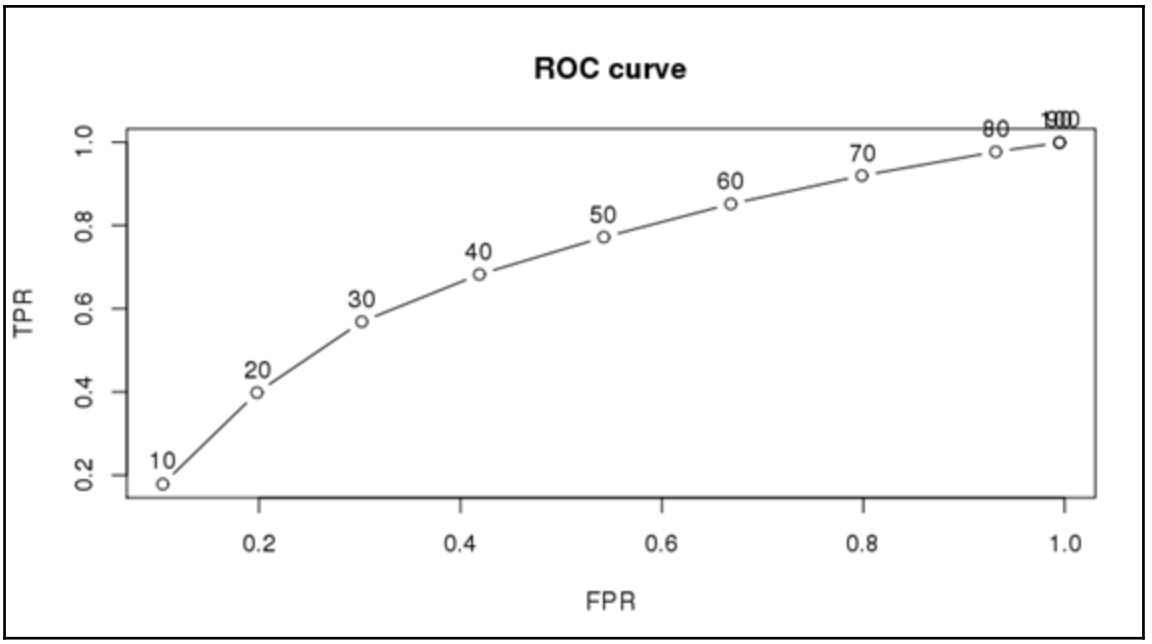



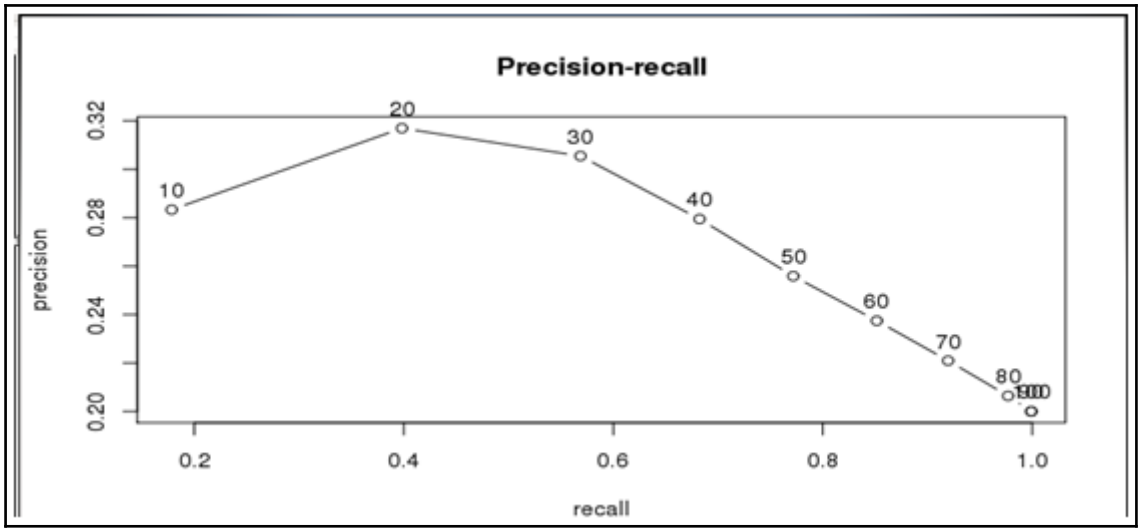



| | |
|---|---|
| data | data as ratingMatrix |
| method | Split/Cross-validation/bootstrap |
| k | number of nearest neighbours to be considered for similarity calculation |
| goodRating | Minimum value for considering as good rating |
| given | Minimum number of records each row should contain |

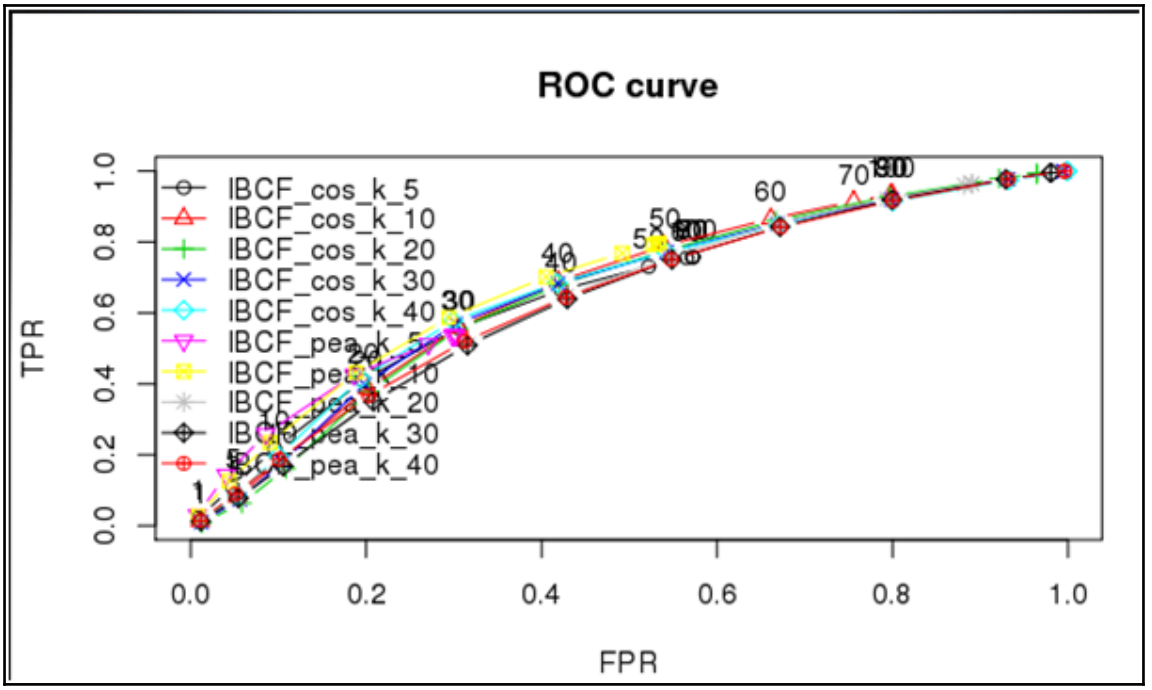

ROC curve
TPR
FPR
0.5
0.6
0.7
0.8
0.9
1.0
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
60

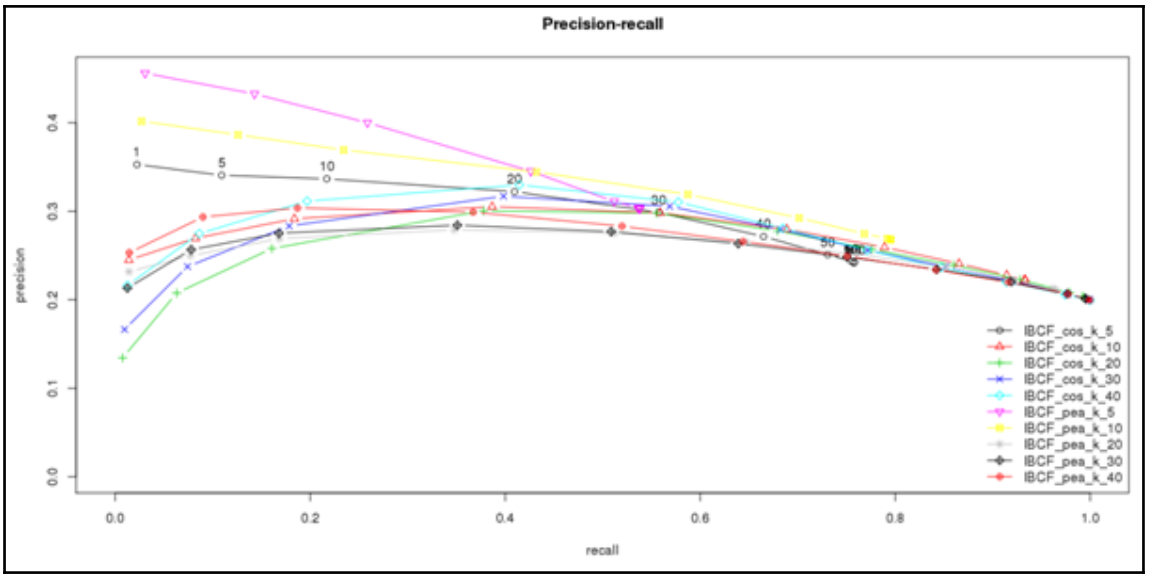

-10
-5
0
5


ROC curve
TPR
FPR
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
60
70
80

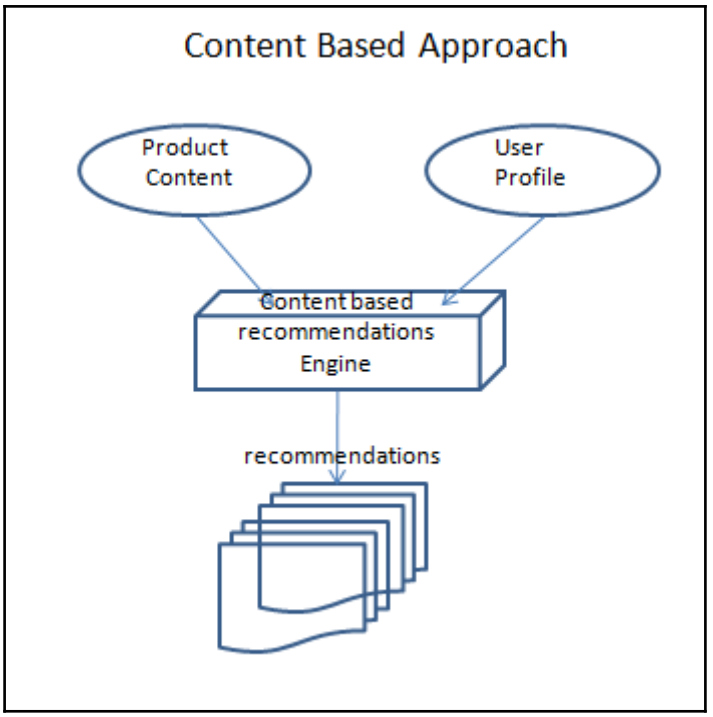

Precision-recall
precision
recall
0.20
0.24
0.28
0.32
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
60
70
80

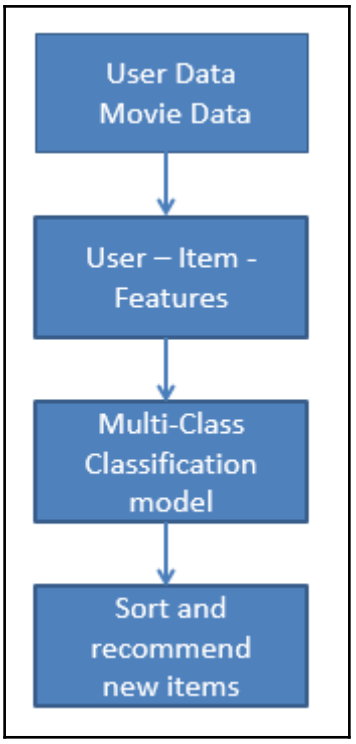
| | | | | |
|---|---|---|---|---|
| *$IBCF_cos_k_5* | *$IBCF_cos_k_30* | *$IBCF_pea_k_5* | *$IBCF_pea_k_20* | $IBCF_pea_k_40 |
| *$IBCF_cos_k_5$name* | *$IBCF_cos_k_30$name* | *$IBCF_pea_k_5$name* | *$IBCF_pea_k_20$name* | $IBCF_pea_k_40$name |
| *[1] "IBCF"* | *[1] "IBCF"* | *[1] "IBCF"* | *[1] "IBCF"* | [1] "IBCF" |
| | | | | |
| *$IBCF_cos_k_5$param* | *$IBCF_cos_k_30$param* | *$IBCF_pea_k_5$param* | $IBCF_pea_k_20$param | $IBCF_pea_k_40$param |
| *$IBCF_cos_k_5$param$method* | *$IBCF_cos_k_30$param$method* | *$IBCF_pea_k_5$param$method* | *$IBCF_pea_k_20$param$method* | $IBCF_pea_k_40$param$method |
| *[1] "cosine"* | *[1] "cosine"* | *[1] "pearson"* | *[1] "pearson"* | [1] "pearson" |
| | | | | |
| *$IBCF_cos_k_5$param$k* | $IBCF_cos_k_30$param$k | *$IBCF_pea_k_5$param$k* | *$IBCF_pea_k_20$param$k* | $IBCF_pea_k_40$param$k |
| *[1] 5* | *[1] 30* | *[1] 5* | *[1] 20* | [1] 40 |
| | | | | |
| *$IBCF_cos_k_10* | *$IBCF_cos_k_40* | *$IBCF_pea_k_10* | $IBCF_pea_k_30 | *$IBCF_cos_k_20* |
| *$IBCF_cos_k_10$name* | *$IBCF_cos_k_40$name* | *$IBCF_pea_k_10$name* | $IBCF_pea_k_30$name | *$IBCF_cos_k_20$name* |
| *[1] "IBCF"* | *[1] "IBCF"* | *[1] "IBCF"* | [1] "IBCF" | *[1] "IBCF"* |
| | | | | |
| *$IBCF_cos_k_10$param* | *$IBCF_cos_k_40$param* | *$IBCF_pea_k_10$param* | $IBCF_pea_k_30$param | *$IBCF_cos_k_20$param* |
| *$IBCF_cos_k_10$param$method* | *$IBCF_cos_k_40$param$method* | *$IBCF_pea_k_10$param$method* | $IBCF_pea_k_30$param$method | $IBCF_cos_k_20$param$method |
| *[1] "cosine"* | *[1] "cosine"* | *[1] "pearson"* | [1] "pearson" | [1] "cosine" |
| | | | | |
| *$IBCF_cos_k_10$param$k* | *$IBCF_cos_k_40$param$k* | *$IBCF_pea_k_10$param$k* | $IBCF_pea_k_30$param$k | $IBCF_cos_k_20$param$k |
| *[1] 10* | *[1] 40* | *[1] 10* | [1] 30 | [1] 20 |




Precision-recall
precision
recall
0.0
0.1
0.2
0.3
0.4
0.6
0.8
1.0
1
5
10
20
30
40
50
IBCF_cos_k_5
IBCF_cos_k_10
IBCF_cos_k_20
IBCF_cos_k_30
IBCF_cos_k_40
IBCF_pea_k_5
IBCF_pea_k_10
IBCF_pea_k_20
IBCF_pea_k_30
IBCF_pea_k_40

# Chapter 6: Building Personalized Recommendation Engines


Content Based Approach
Product Content
User Profile
Content based recommendations Engine
recommendations




| MovieId | UserId | Rating | Gender | Occupation | Unknown | Action | Adventure | Animation | Children | Comedy | Crime |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 5 | M | technician | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 117 | 4 | M | student | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 429 | 3 | M | student | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 919 | 4 | M | other | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 457 | 4 | F | salesman | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 468 | 5 | M | engineer | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 17 | 4 | M | programmer | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 892 | 5 | M | other | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 16 | 5 | M | entertainment | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 580 | 3 | M | student | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 268 | 3 | M | engineer | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 894 | 4 | M | educator | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 1 | 535 | 3 | F | educator | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

```
> head(ratings)
  UserId MovieId Rating
1    196     242      3
2    186     302      3
3     22     377      1
4    244      51      2
5    166     346      1
6    298     474      4
> names(ratings)
[1] "UserId"  "MovieId" "Rating"
> str(ratings)
'data.frame':   100000 obs. of  3 variables:
 $ UserId : int  196 186 22 244 166 298 115 253 305 6 ...
 $ MovieId: int  242 302 377 51 346 474 265 465 451 86 ...
 $ Rating : int  3 3 1 2 1 4 2 5 3 3 ...
>
```

| | MovieId | Unknown | Action | Adventure | Animation | Children | Comedy | Crime | Documentary | Drama | Fantasy | FilmNoir | Horror | Musical | Mystery | Romance | SciFi | Thriller |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 | 4 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 6 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 8 | 8 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 12 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 13 | 13 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 15 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 | 16 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 17 | 17 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |

```
> names(movies)
 [1] "MovieId"     "Unknown"     "Action"      "Adventure"   "Animation"   "Children"
 [7] "Comedy"      "Crime"       "Documentary" "Drama"       "Fantasy"     "FilmNoir"
[13] "Horror"      "Musical"     "Mystery"     "Romance"     "SciFi"       "Thriller"
[19] "War"         "Western"
> str(movies)
'data.frame':   1682 obs. of  20 variables:
 $ MovieId    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Unknown    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Action     : int  0 1 0 1 0 0 0 0 0 0 ...
 $ Adventure  : int  0 1 0 0 0 0 0 0 0 0 ...
 $ Animation  : int  1 0 0 0 0 0 0 0 0 0 ...
 $ Children   : int  1 0 0 0 0 0 0 1 0 0 ...
 $ Comedy     : int  1 0 0 1 0 0 0 1 0 0 ...
 $ Crime      : int  0 0 0 0 1 0 0 0 0 0 ...
 $ Documentary: int  0 0 0 0 0 0 0 0 0 0 ...
 $ Drama      : int  0 0 0 1 1 1 1 1 1 1 ...
 $ Fantasy    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ FilmNoir   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Horror     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Musical    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Mystery    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Romance    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ SciFi      : int  0 0 0 0 0 0 1 0 0 0 ...
 $ Thriller   : int  0 1 1 0 1 0 0 0 0 0 ...
 $ War        : int  0 0 0 0 0 0 0 0 0 1 ...
 $ Western    : int  0 0 0 0 0 0 0 0 0 0 ...
>
```

| | MovieId | UserId | Rating | Unknown | Action | Adventure | Animation | Children | Comedy | Crime | Documentary | Drama | Fantasy | FilmNoir | Horror | Musical | Mystery | Romance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 650 | 3 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 635 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1 | 514 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 1 | 250 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 1 | 210 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 1 | 5 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 1 | 72 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 1 | 77 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 252 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 1 | 120 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 1 | 45 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 1 | 265 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 1 | 263 | 5 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 | 1 | 881 | 4 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

```
> names(ratings)
 [1] "MovieId"     "UserId"      "Rating"      "Unknown"     "Action"      "Adventure"   "Animation"   "Children"    "Comedy"      "Crime"       "Documentary"
[12] "Drama"       "Fantasy"     "FilmNoir"    "Horror"      "Musical"     "Mystery"     "Romance"     "SciFi"       "Thriller"    "War"         "Western"
> |
```

```
> head(ratings)
  MovieId UserId Rating Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller
1       1    650      3       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
2       1    635      4       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
3       1      1      5       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
4       1    514      5       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
5       1    250      4       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
6       1    210      5       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
  War Western nrat
1   0       0    0
2   0       0    1
3   0       0    1
4   0       0    1
5   0       0    1
6   0       0    1
> |
```

```
> apply(ratings[,-c(1:3,23)],2,function(x)table(x))
  Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller   War Western
0   99990  74411     86247     96395    92818  70168 91945       99242 60105   98648    98267  94683   95046   94755   80539 87270    78128 90602   98146
1      10  25589     13753      3605     7182  29832  8055         758 39895    1352     1733   5317    4954    5245   19461 12730    21872  9398    1854
>
```

```
> head(scaled_ratings)
      Action  Adventure Animation Children  Comedy      Crime Documentary      Drama   Fantasy   FilmNoir    Horror    Musical    Mystery    Romance      SciFi
1 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
2 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
3 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
4 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
5 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
6 -0.5864161 -0.3993232  5.170975 3.594937 1.53365 -0.2959828 -0.08739462 -0.8147076 -0.117069 -0.1327985 -0.236971 -0.2283016 -0.2352716 -0.4915609 -0.3819263
    Thriller        War   Western MovieId UserId nrat
1 -0.5291012 -0.3220673 -0.137441       1    650    0
2 -0.5291012 -0.3220673 -0.137441       1    635    1
3 -0.5291012 -0.3220673 -0.137441       1      1    1
4 -0.5291012 -0.3220673 -0.137441       1    514    1
5 -0.5291012 -0.3220673 -0.137441       1    250    1
6 -0.5291012 -0.3220673 -0.137441       1    210    1
```

```
> dim(model_data_test)
[1] 19955    21
> dim(model_data_train)
[1] 80045    21
```

```
> fit

Call:
 randomForest(formula = as.factor(nrat) ~ ., data = model_data_train[,      -c(19, 20)])
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 4

        OOB estimate of  error rate: 39.98%
Confusion matrix:
      0     1 class.error
0 14390 21315   0.5969752
1 10686 33654   0.2410014
```

```
> summary(fit)
                Length Class  Mode
call                 3 -none- call
type                 1 -none- character
predicted        80045 factor numeric
err.rate          1500 -none- numeric
confusion            6 -none- numeric
votes           160090 matrix numeric
oob.times        80045 -none- numeric
classes              2 -none- character
importance          18 -none- numeric
importanceSD         0 -none- NULL
localImportance      0 -none- NULL
proximity            0 -none- NULL
ntree                1 -none- numeric
mtry                 1 -none- numeric
forest              14 -none- list
y                80045 factor numeric
test                 0 -none- NULL
inbag                0 -none- NULL
terms                3 terms  call
```

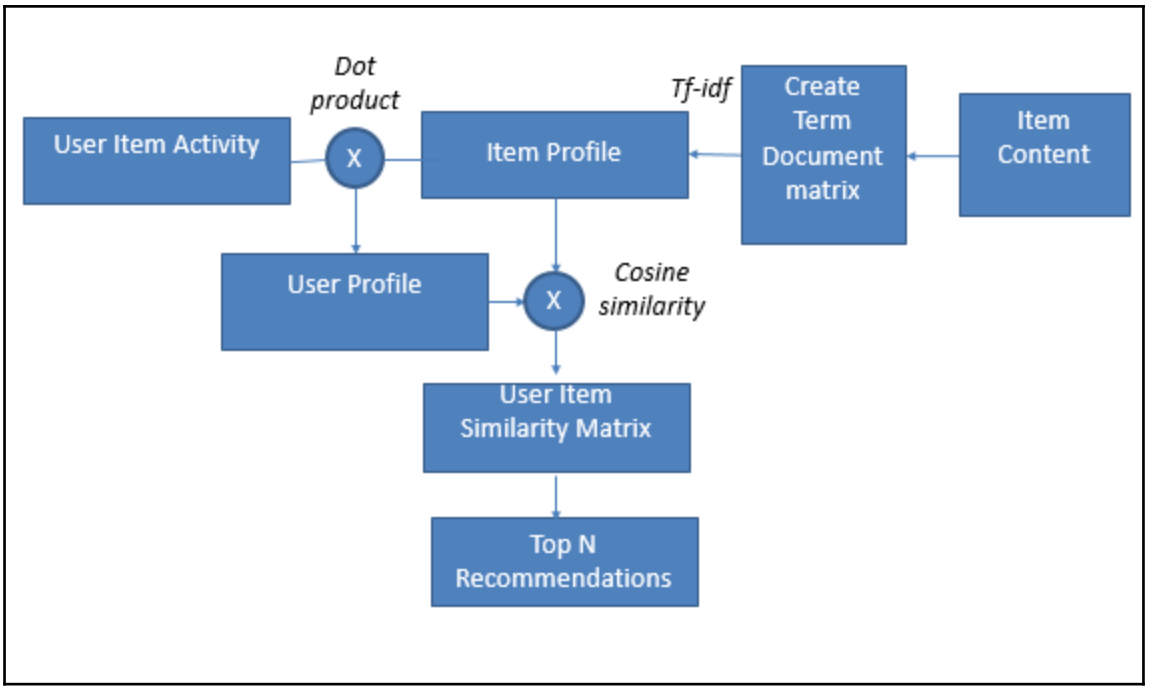
```
> predictions[0:20]
  1   8  19  23  24  25  29  36  39  45  48  49  50  73  80  81  82  93  99 107
  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
Levels: 0 1
```

```
> cm

predictions    0    1
          0 3541 2738
          1 5379 8297
> (accuracy <- sum(diag(cm)) / sum(cm))
[1] 0.5932348
> (precision <- diag(cm) / rowSums(cm))
        0         1
0.5639433 0.6066832
> recall <- (diag(cm) / colSums(cm))
> recall
        0         1
0.3969731 0.7518804
>
```

```
> head(totalMovieIds)
[1] 1 2 3 4 5 6
> tail(totalMovieIds)
[1] 1677 1678 1679 1680 1681 1682
>
```

```
> head(activeusernonratedmoviedf)
  UserId MovieId Rating
1    943       1      0
2    943       3      0
3    943       4      0
4    943       5      0
5    943       6      0
6    943       7      0
>
```

```
> head(activeuserratings)
  MovieId UserId Rating Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller
1       1    943      0       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0
2       3    943      0       0      0         0         0        0      0     0           0     0       0        0      0       0       0       0     0        1
3       4    943      0       0      1         0         0        0      1     0           0     1       0        0      0       0       0       0     0        0
4       5    943      0       0      0         0         0        0      0     1           0     1       0        0      0       0       0       0     0        1
5       6    943      0       0      0         0         0        0      0     0           0     1       0        0      0       0       0       0     0        0
6       7    943      0       0      0         0         0        0      0     0           0     1       0        0      0       0       0       0     1        0
  War Western
1   0       0
2   0       0
3   0       0
4   0       0
5   0       0
6   0       0
>
```


Dot product
Tf-idf
User Item Activity
X
Item Profile
Create Term Document matrix
Item Content
User Profile
X
Cosine similarity
User Item Similarity Matrix
Top N Recommendations

```
I,4,"www.microsoft.com","created by getlog.pl"
T,1,"VRoot",0,0,"VRoot"
N,0,"0"
N,1,"1"
T,2,"Hide1",0,0,"Hide"
N,0,"0"
N,1,"1"
A,1277,1,"NetShow for PowerPoint","/stream"
A,1253,1,"MS Word Development","/worddev"
A,1109,1,"TechNet (World Wide Web Edition)","/technet"
A,1038,1,"SiteBuilder Network Membership","/sbnmember"
A,1205,1,"Hardware Supprt","/hardwaresupport"
A,1076,1,"NT Workstation Support","/ntwkssupport"
A,1100,1,"MS in Education","/education"
A,1229,1,"Uruguay","/uruguay"
A,1172,1,"Belgium","/belgium"
A,1173,1,"Microsoft OnLine Institute","/moli"
A,1283,1,"Cinemainia","/cinemania"
A,1167,1,"Windows Hardware Testing","/hwtest"
A,1290,1,"Activate the Internet Conference","/devmovies"
A,1193,1,"Office Developer Support","/offdevsupport"
A,1153,1,"Venezuela","/venezuela"
A,1013,1,"Visual Basic Support","/vbasicsupport"
A,1241,1,"India","/india"
A,1169,1,"MS Project","/msproject"
A,1260,1,"Exchange Trial","/trial"
A,1063,1,"Intranet Strategy","/intranet"
A,1252,1,"Community Affairs","/giving"
```

```
In [6]: raw_data
Out[6]:
       0      1      2                                3                4
0      A   1277      1            NetShow for PowerPoint          /stream
1      A   1253      1               MS Word Development          /worddev
2      A   1109      1  TechNet (World Wide Web Edition)          /technet
3      A   1038      1    SiteBuilder Network Membership        /sbnmember
4      A   1205      1                  Hardware Supprt  /hardwaresupport
5      A   1076      1            NT Workstation Support     /ntwkssupport
6      A   1100      1                  MS in Education        /education
7      A   1229      1                         Uruguay          /uruguay
8      A   1172      1                         Belgium          /belgium
9      A   1173      1       Microsoft OnLine Institute             /moli
10     A   1283      1                      Cinemainia        /cinemania
11     A   1167      1         Windows Hardware Testing           /hwtest
12     A   1290      1  Activate the Internet Conference        /devmovies
...   ..    ...    ...                               ...               ...
20455  V   1004      1                               NaN               NaN
20456  C  14992  14992                               NaN               NaN
20457  V   1001      1                               NaN               NaN
20458  V   1034      1                               NaN               NaN
20459  V   1004      1                               NaN               NaN
20460  C  14993  14993                               NaN               NaN
20461  V   1010      1                               NaN               NaN
20462  V   1004      1                               NaN               NaN
20463  C  14994  14994                               NaN               NaN
```

```
In [85]: user_activity.head(15)
Out[85]:
    category  value
294        C  10001
295        V   1038
296        V   1026
297        V   1034
298        C  10002
299        V   1008
300        V   1056
301        V   1032
302        C  10003
303        V   1064
304        V   1065
305        V   1020
306        V   1007
307        V   1038
308        V   1026
```

```
In [102]: user_activity.head(30)
Out[102]:
    category  value   userid    webid
294        C  10001  10001.0  10001.0
295        V   1038  10001.0   1038.0
296        V   1026  10001.0   1026.0
297        V   1034  10001.0   1034.0
298        C  10002  10002.0  10002.0
299        V   1008  10002.0   1008.0
300        V   1056  10002.0   1056.0
301        V   1032  10002.0   1032.0
302        C  10003  10003.0  10003.0
303        V   1064  10003.0   1064.0
304        V   1065  10003.0   1065.0
305        V   1020  10003.0   1020.0
306        V   1007  10003.0   1007.0
307        V   1038  10003.0   1038.0
308        V   1026  10003.0   1026.0
309        V   1052  10003.0   1052.0
310        V   1041  10003.0   1041.0
311        V   1028  10003.0   1028.0
312        C  10004  10004.0  10004.0
313        V   1004  10004.0   1004.0
314        C  10005  10005.0  10005.0
315        V   1017  10005.0   1017.0
316        V   1156  10005.0   1156.0
317        V   1004  10005.0   1004.0
318        V   1018  10005.0   1018.0
319        V   1008  10005.0   1008.0
320        V   1027  10005.0   1027.0
321        V   1009  10005.0   1009.0
322        V   1046  10005.0   1046.0
323        V   1038  10005.0   1038.0
```

```
In [104]: user_activity.head()
Out[104]:
    category  value   userid   webid
295        V   1038  10001.0  1038.0
296        V   1026  10001.0  1026.0
297        V   1034  10001.0  1034.0
299        V   1008  10002.0  1008.0
300        V   1056  10002.0  1056.0
```

```
In [13]: items.head()
Out[13]:
   webid                              desc
0   1277            NetShow for PowerPoint
1   1253               MS Word Development
2   1109  TechNet (World Wide Web Edition)
3   1038    SiteBuilder Network Membership
4   1205                    Hardware Supprt
```

```
In [122]: items_sort.head(5)
Out[122]:
     webid                     desc
113   1000                   regwiz
40    1001          Support Desktop
278   1002  End User Produced View
102   1003           Knowledge Base
243   1004    Microsoft.com Search
```

```
In [128]: itemprof
Out[128]:
matrix([[ 1.        ,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.32213709,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.43709646,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        ...,
        [ 0.38159493,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.30073274,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.36402686,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ]])
```

```
In [130]: userprof
Out[130]:
matrix([[ 0.00062937,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.00089668,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.00144708,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        ...,
        [ 0.00046412,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.00067229,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ],
        [ 0.00079067,  0.        ,  0.        , ...,  0.        ,
          0.        ,  0.        ]])

In [131]: userprof.shape
Out[131]: (5000, 100)
```

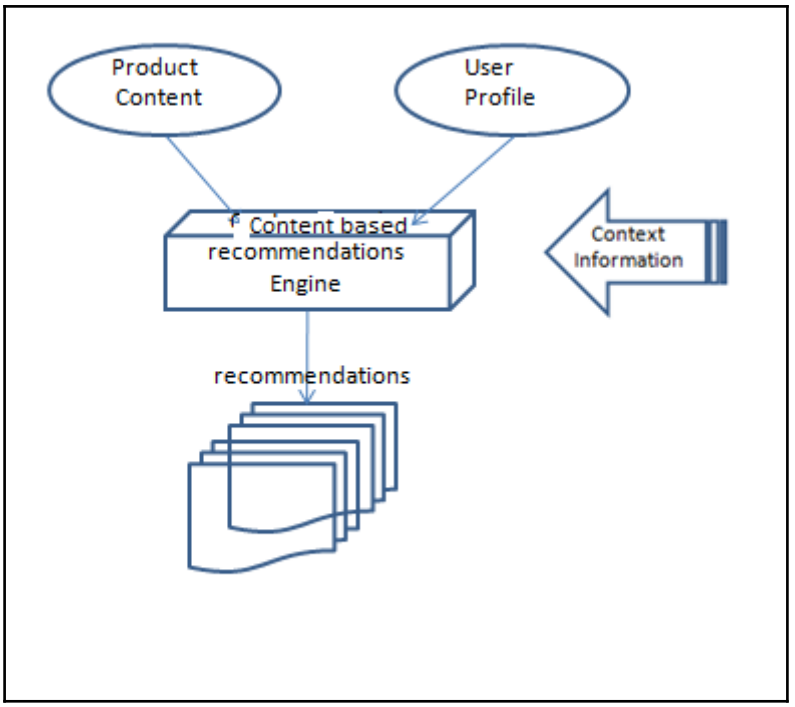


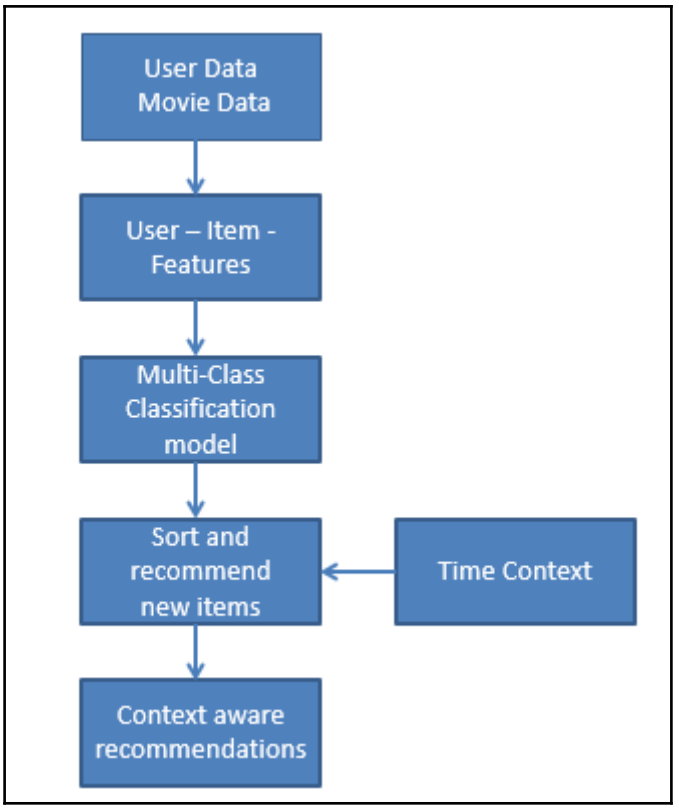
```
In [138]: similarityCalc
Out[138]:
array([[ 0.54168902,  0.17449812,  0.23677035, ...,  0.20670579,
         0.16290362,  0.19718935],
       [ 0.78844617,  0.25398775,  0.34462703, ...,  0.30086706,
         0.23711158,  0.28701558],
       [ 0.63172381,  0.20350167,  0.27612424, ...,  0.29413451,
         0.18998003,  0.22996444],
       ...,
       [ 0.56969503,  0.1835199 ,  0.24901168, ...,  0.21739274,
         0.17132595,  0.20738429],
       [ 0.49394733,  0.15911875,  0.21590263, ...,  0.1884878 ,
         0.14854613,  0.17981009],
       [ 0.86518334,  0.27870764,  0.37816858, ...,  0.33014958,
         0.26018896,  0.31494998]])

In [139]: similarityCalc.shape
Out[139]: (5000, 236)

In [140]:
```

```
In [141]: final_pred[1]

In [141]: Out[141]:
array([1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0,
       0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
       0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1,
       0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 0])

In [142]: final_pred[2]
Out[142]:
array([1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0,
       1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
       0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1,
       0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 0])

In [143]: final_pred[3]
Out[143]:
array([0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```



```
In [145]: indexes_of_user
Out[145]: (array([  9,  37,  68, 152], dtype=int64),)
```


Product
Content
User
Profile
Content based
recommendations
Engine
Context
Information
recommendations




```
> head(raw_data)
  UserId MovieId Rating TimeStamp
1    196     242      3 881250949
2    186     302      3 891717742
3     22     377      1 878887116
4    244      51      2 880606923
5    166     346      1 886397596
6    298     474      4 884182806
>
```

```
> head(movies)
  MovieId Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller War Western
1       1       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0        0   0       0
2       2       0      1         1         0        0      0     0           0     0       0        0      0       0       0       0     0        1   0       0
3       3       0      0         0         0        0      0     0           0     0       0        0      0       0       0       0     0        1   0       0
4       4       0      1         0         0        0      1     0           0     1       0        0      0       0       0       0     0        0   0       0
5       5       0      0         0         0        0      0     1           0     1       0        0      0       0       0       0     0        1   0       0
6       6       0      0         0         0        0      0     0           0     1       0        0      0       0       0       0     0        0   0       0
>
```

```
> head(ratings_ctx)
  MovieId UserId Rating TimeStamp Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi
1       1    650      3 891369759       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
2       1    635      4 878879283       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
3       1      1      5 874965758       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
4       1    514      5 875309276       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
5       1    250      4 883263374       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
6       1    210      5 887731052       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
  Thriller War Western
1        0   0       0
2        0   0       0
3        0   0       0
4        0   0       0
5        0   0       0
6        0   0       0
>
```

```
> ts = ratings_ctx$TimeStamp
> head(ts)
[1] 891369759 878879283 874965758 875309276 883263374 887731052
> hours <- as.POSIXlt(ts,origin="1960-10-01")$hour
> head(hours)
[1]  0 10  3  2  4 21
>
```

```
> head(ratings_ctx)
  MovieId UserId Rating TimeStamp Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi
1       1    650      3 891369759       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
2       1    635      4 878879283       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
3       1      1      5 874965758       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
4       1    514      5 875309276       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
5       1    250      4 883263374       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
6       1    210      5 887731052       0      0         0         1        1      1     0           0     0       0        0      0       0       0       0     0
  Thriller War Western hours
1        0   0       0     0
2        0   0       0    10
3        0   0       0     3
4        0   0       0     2
5        0   0       0     4
6        0   0       0    21
>
```

```
> head(UCP)
     MovieId Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller War Western hours
496        2      1         1         0        0      0     0           0     0       0        0      0       0       0       0     0        1   0       0     9
1676       9      0         0         0        0      0     0           0     1       0        0      0       0       0       0     0        0   0       0     8
2210      11      0         0         0        0      0     1           0     0       0        0      0       0       0       0     0        1   0       0     9
2356      12      0         0         0        0      0     1           0     0       0        0      0       0       0       0     0        1   0       0     9
3781      22      1         0         0        0      0     0           0     1       0        0      0       0       0       0     0        0   1       0     9
3944      23      0         0         0        0      0     0           0     1       0        0      0       0       0       0     0        1   0       0     9
>
```

```
> head(UCP_pref)
  hours Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller War Western
1     0      8         7         0        1      1     0           0     3       0        0      1       0       0       1     6        0   0       1
2     6      3         2         0        0      0     0           0     2       0        0      0       0       1     1        1   1       0
3     8     11         5         0        0     11     4           0    10       0        0      2       0       2     5        4  10   3       0
4     9     38        20         2        8     36    14           0    41       2        0     11       4       1    24        9  23  11       7
5    10      4         1         0        0      8     0           0     1       0        0      0       0       0     2        2   1   0       1
>
```

```
> head(UCP_pref_sc)
     context    Action Adventure  Animation Children    Comedy     Crime Documentary     Drama    Fantasy FilmNoir    Horror    Musical    Mystery   Romance
[1,]       0 1.0000000 0.8750000 0.00000000 0.125000 0.1250000 0.0000000           0 0.3750000 0.00000000        0 0.1250000 0.00000000 0.00000000 0.1250000
[2,]       6 1.0000000 0.6666667 0.00000000 0.000000 0.0000000 0.0000000           0 0.6666667 0.00000000        0 0.0000000 0.00000000 0.00000000 0.3333333
[3,]       8 1.0000000 0.4545455 0.00000000 0.000000 1.0000000 0.3636364           0 0.9090909 0.00000000        0 0.1818182 0.00000000 0.18181818 0.4545455
[4,]       9 0.9268293 0.4878049 0.04878049 0.195122 0.8780488 0.3414634           0 1.0000000 0.04878049        0 0.2682927 0.09756098 0.02439024 0.5853659
[5,]      10 0.5000000 0.1250000 0.00000000 0.000000 1.0000000 0.0000000           0 0.1250000 0.00000000        0 0.0000000 0.00000000 0.00000000 0.2500000
       SciFi  Thriller       War   Western
[1,] 0.7500000 0.0000000 0.0000000 0.1250000
[2,] 0.3333333 0.3333333 0.3333333 0.0000000
[3,] 0.3636364 0.9090909 0.2727273 0.0000000
[4,] 0.2195122 0.5609756 0.2682927 0.1707317
[5,] 0.2500000 0.1250000 0.0000000 0.1250000
>
```

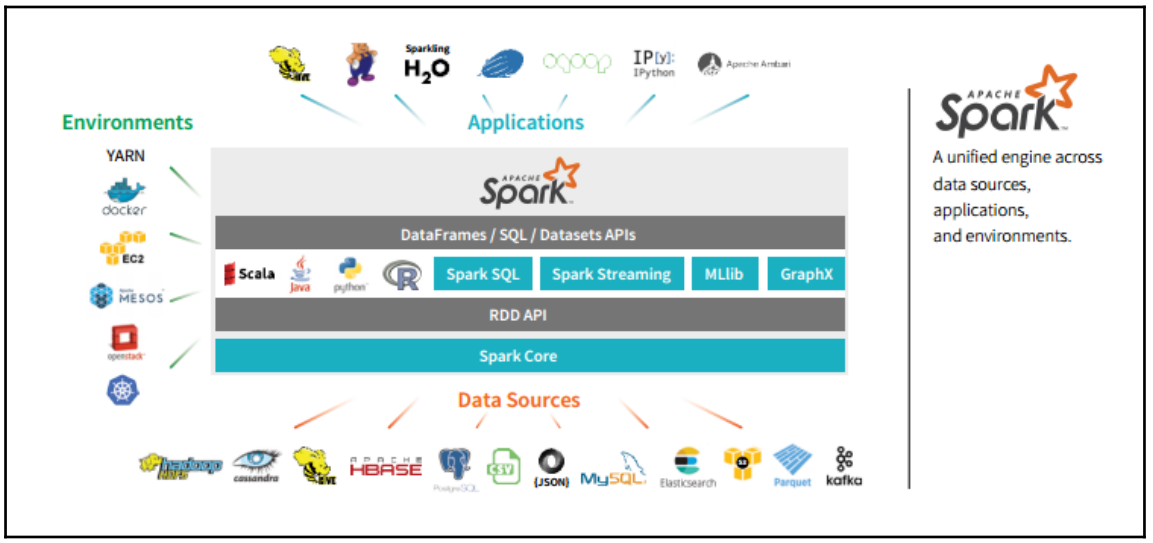
```
> recommend$MovieId
  [1]   3   4   5   6   7   8  10  14  15  16  18  19  20  30  32  34  35  36  37  43  44  45  46  47  48  49
 [32]  65  66  70  71  74  75  77  78  81  83  84  86  87  88  89  90  93  99 102 103 106 107 112 113 114 115
 [63] 131 133 134 135 136 137 140 141 142 143 146 149 150 155 156 157 159 160 162 165 166 170 175 178 179 180
 [94] 199 203 206 207 209 212 213 214 220 221 223 224 236 244 245 246 253 255 256 258 262 264 267 268 270 272
[125] 280 283 285 286 287 288 289 291 292 293 296 297 299 302 303 304 305 306 307 308 309 310 311 312 313 315
[156] 324 325 326 327 329 330 331 332 333 336 337 339 340 344 345 346 347 348 351 354 355 357 359 360 361 365
[187] 379 381 382 387 389 392 394 404 408 409 411 413 416 417 418 420 424 429 432 434 436 437 438 439 440 441
[218] 448 451 454 457 458 459 460 461 462 463 464 465 466 467 469 473 474 478 479 480 481 483 484 486 487 488
[249] 499 500 501 503 504 505 506 507 509 510 511 512 513 514 515 517 518 519 520 521 522 523 525 527 528 529
[280] 535 536 537 538 542 543 544 545 547 548 551 552 553 555 556 557 558 560 561 563 565 567 571 573 574 580
[311] 589 591 592 593 594 596 598 600 601 602 603 605 606 607 608 610 611 612 615 616 617 618 619 620 621 622
[342] 632 633 634 635 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 656 657 658 659 660
[373] 669 670 671 673 674 675 676 677 678 681 682 686 690 691 692 693 694 695 696 697 698 699 701 702 703 704
[404] 711 712 713 714 715 716 723 726 727 729 730 731 733 735 736 737 738 740 741 742 744 747 749 750 753 754
[435] 766 767 768 770 772 773 774 775 776 777 778 781 782 783 784 786 787 788 789 791 799 800 803 806 807 809
[466] 817 821 822 823 826 828 832 834 835 836 837 839 842 844 846 847 848 850 851 852 855 856 857 858 860 861
[497] 869 870 872 873 874 875 877 880 882 883 884 885 886 887 888 889 891 895 896 898 899 900 902 903 904 905
```

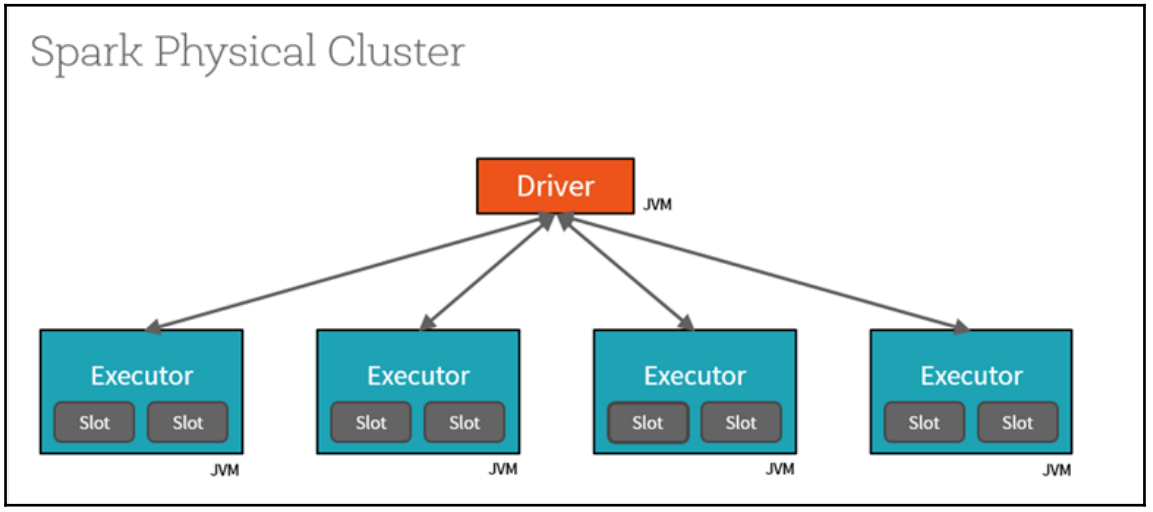
```
> head(UCP_pref_content)
  MovieId predictions Unknown Action Adventure Animation Children Comedy Crime Documentary Drama Fantasy FilmNoir Horror Musical Mystery Romance SciFi Thriller
1       3           1       0      0         0         0        0      0     0           0     0       0        0      0       0       0       0     0        1
2       4           1       0      1         0         0        0      1     0           0     1       0        0      0       0       0       0     0        0
3       5           1       0      0         0         0        0      0     1           0     1       0        0      0       0       0       0     0        1
4       6           1       0      0         0         0        0      0     0           0     1       0        0      0       0       0       0     0        0
5       7           1       0      0         0         0        0      0     0           0     1       0        0      0       0       0       0     1        0
6       8           1       0      0         0         0        1      1     0           0     1       0        0      0       0       0       0     0        0
  War Western
1   0       0
2   0       0
3   0       0
4   0       0
5   0       0
6   0       0
>
```

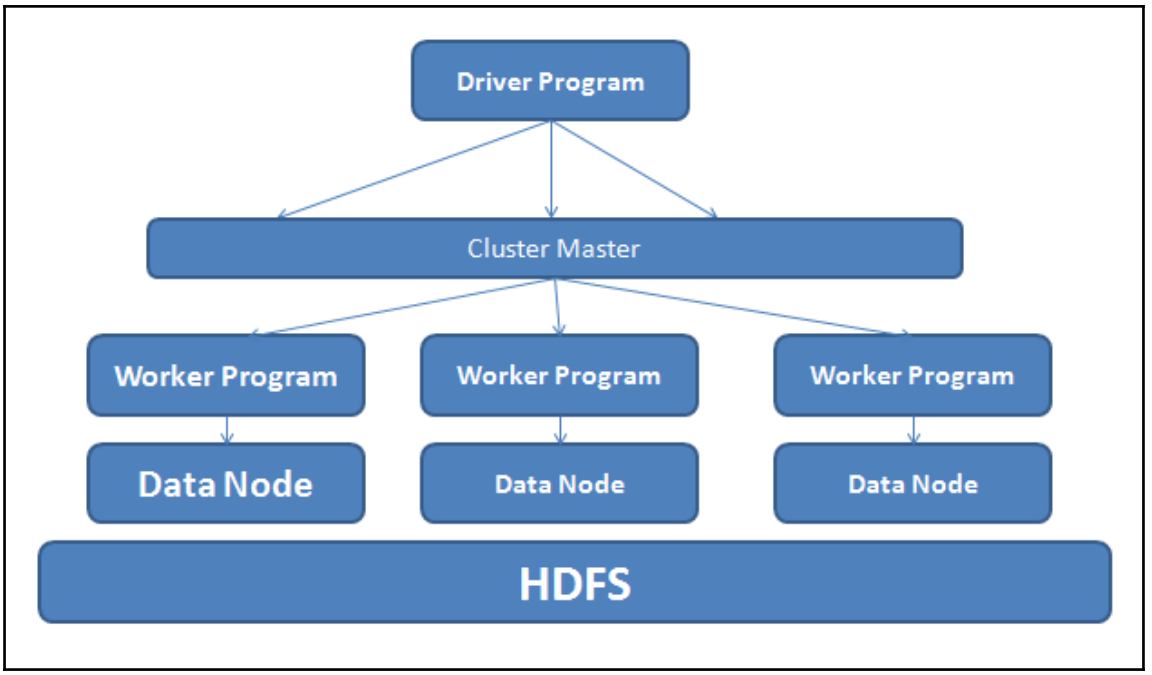
```
> head(active_user)
     [,1]      [,2]
[1,]    3 0.5609756
[2,]    4 2.8048780
[3,]    5 1.9024390
[4,]    6 1.0000000
[5,]    7 1.2195122
[6,]    8 2.0731707
>
```

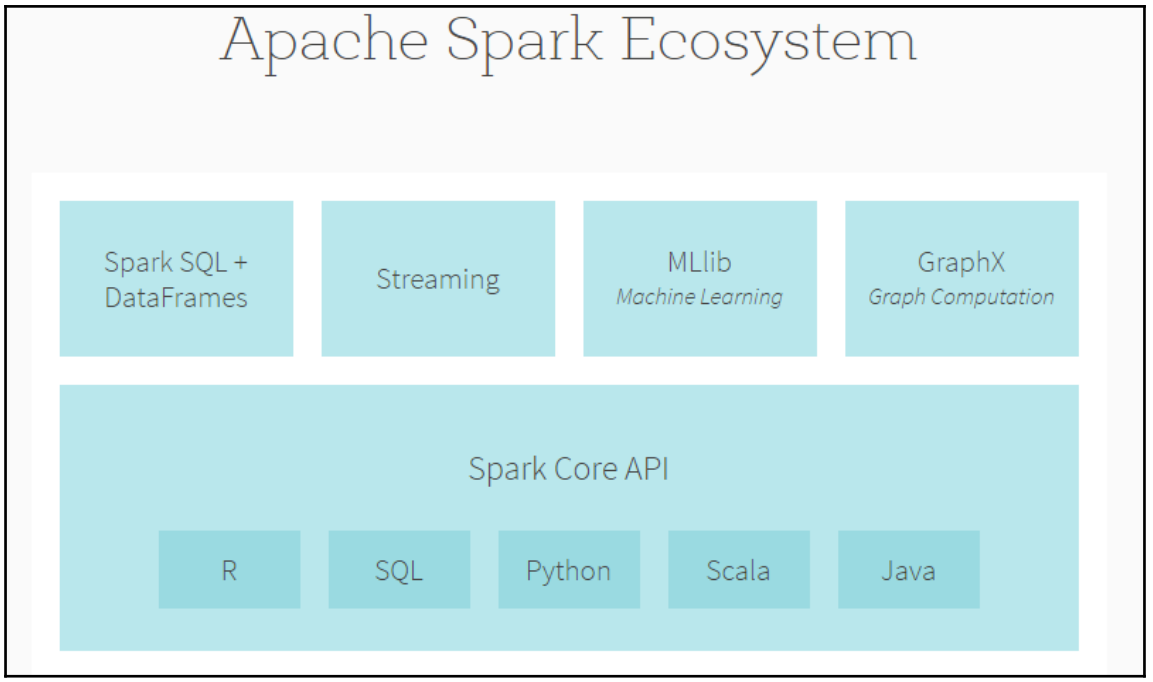

```
> head(active_user)
     [,1]      [,2]
[1,]    3 0.5609756
[2,]    4 2.8048780
[3,]    5 1.9024390
[4,]    6 1.0000000
[5,]    7 1.2195122
[6,]    8 2.0731707
>
```

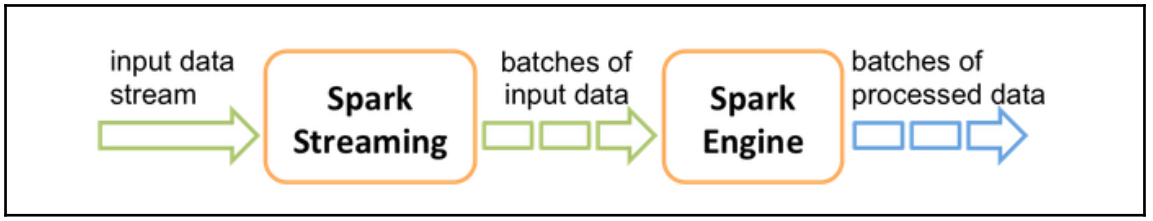
# Chapter 7: Building Real-Time Recommendation Engines with Spark

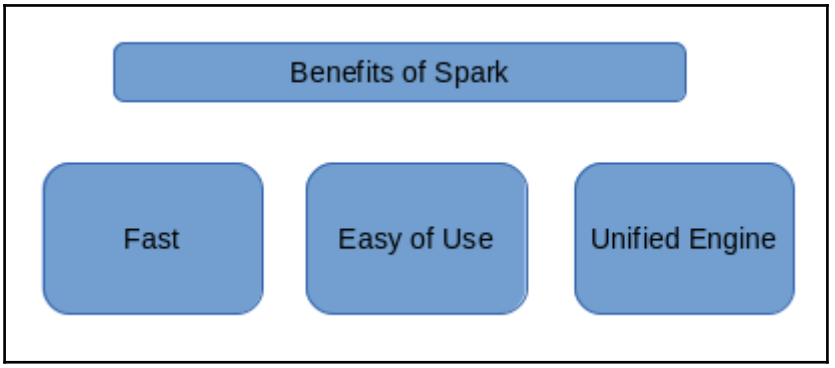


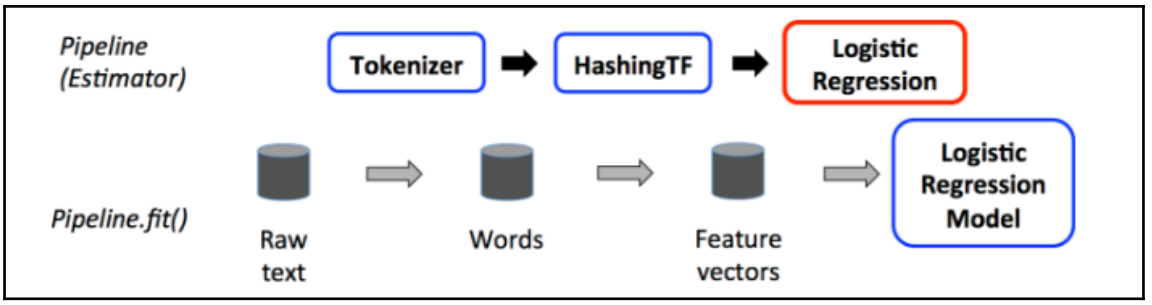



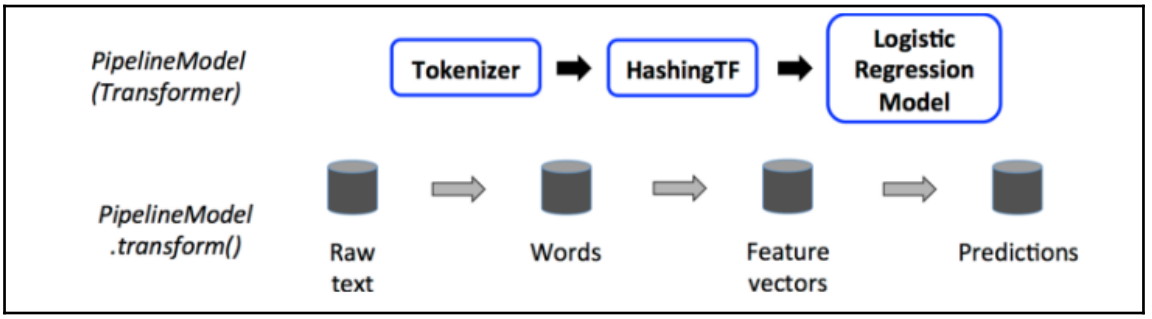

Driver Program
Cluster Master
Worker Program
Worker Program
Worker Program
Data Node
Data Node
Data Node
HDFS

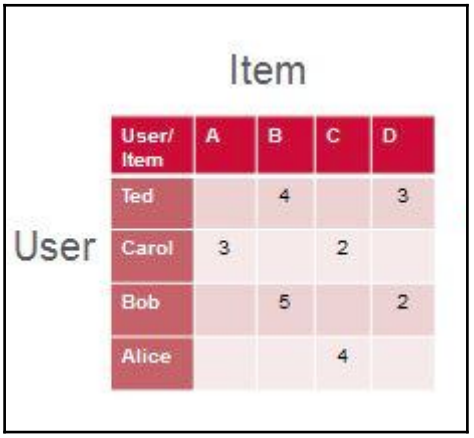

Apache Spark Ecosystem
Spark SQL + DataFrames
Streaming
MLlib
Machine Learning
GraphX
Graph Computation
Spark Core API
R
SQL
Python
Scala
Java

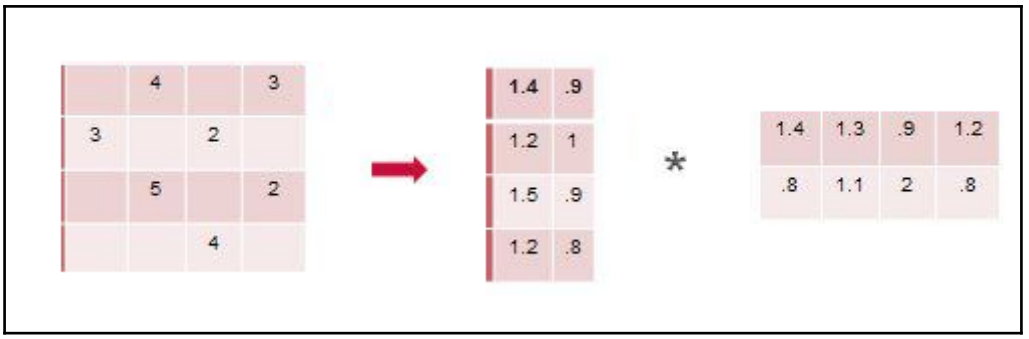

input data stream
Spark Streaming
batches of input data
Spark Engine
batches of processed data

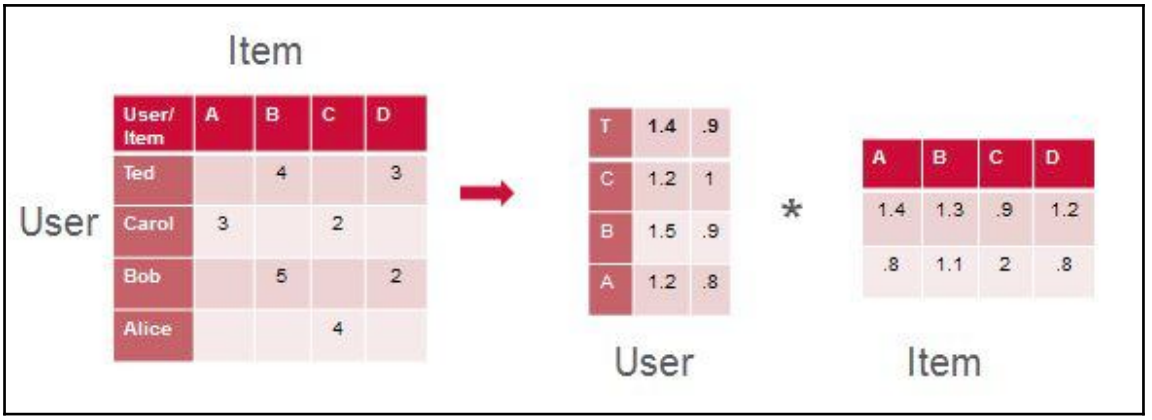

Benefits of Spark
Fast
Easy of Use
Unified Engine


Pipeline
(Estimator)
Tokenizer
HashingTF
Logistic Regression
Pipeline.fit()
Raw text
Words
Feature vectors
Logistic Regression Model

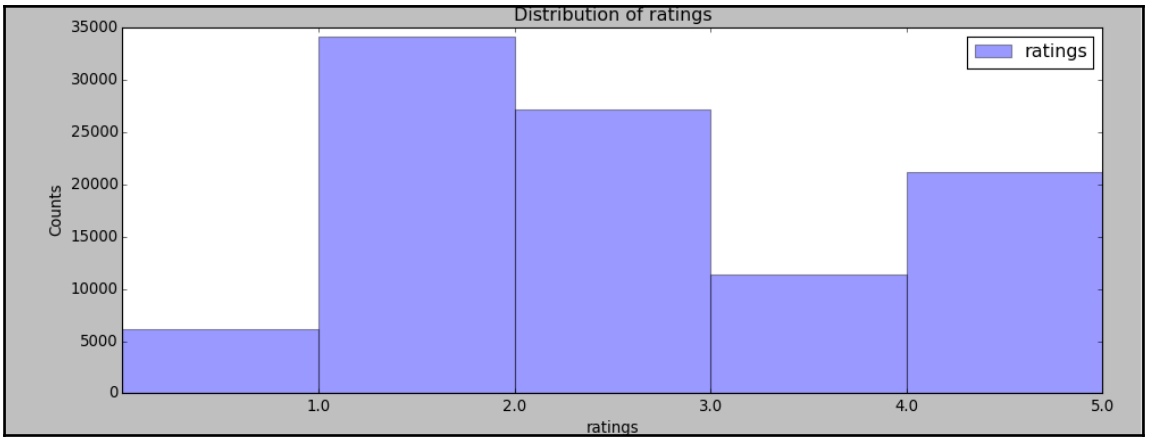

PipelineModel
(Transformer)
Tokenizer
HashingTF
Logistic Regression Model
PipelineModel
.transform()
Raw text
Words
Feature vectors
Predictions


Item
User
User/ Item
A
B
C
D
Ted
4
3
Carol
3
2
Bob
5
2
Alice
4

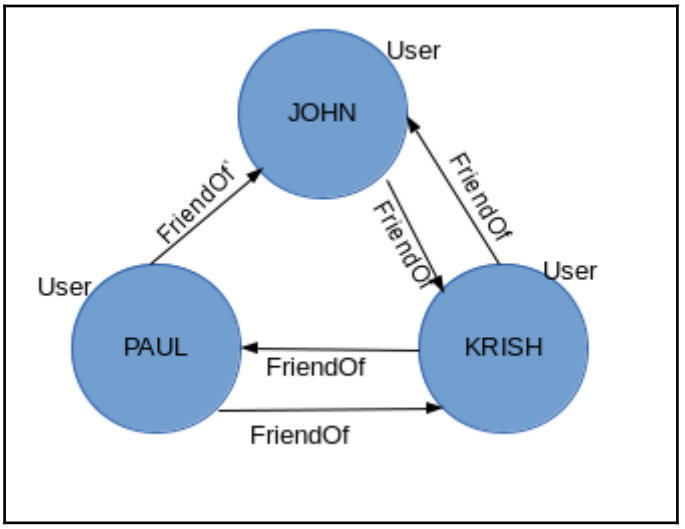


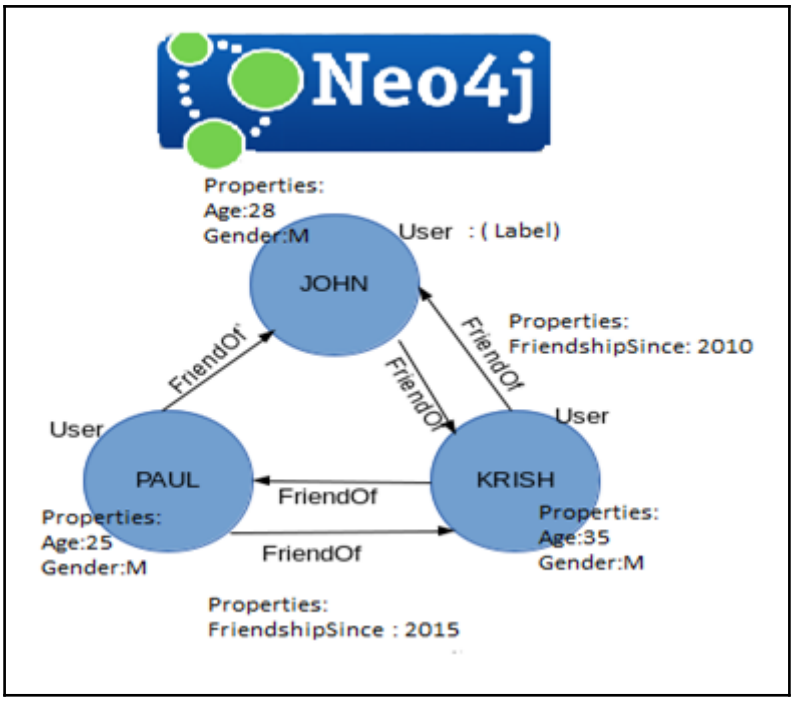




$$\min_{q^*,p^*} \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$


Distribution of ratings
ratings
Counts
35000
30000
25000
20000
15000
10000
5000
0
1.0
2.0
3.0
4.0
5.0
ratings

# Chapter 8: Building Real-Time Recommendation Engines with Neo4j




## Cypher Query

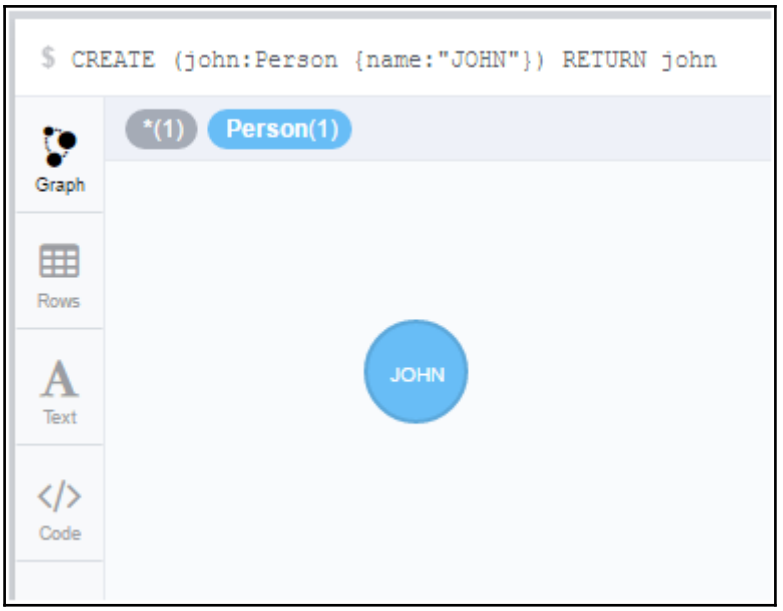
```
(u:User) -[f:friendof]-> (m:User)
```

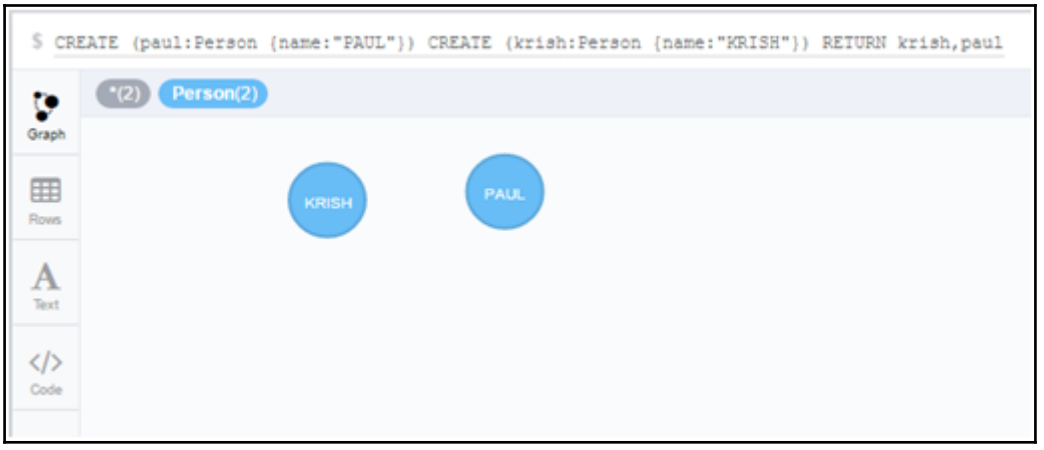
Above Cypher query pulls up all the friendship relations between pairs of users

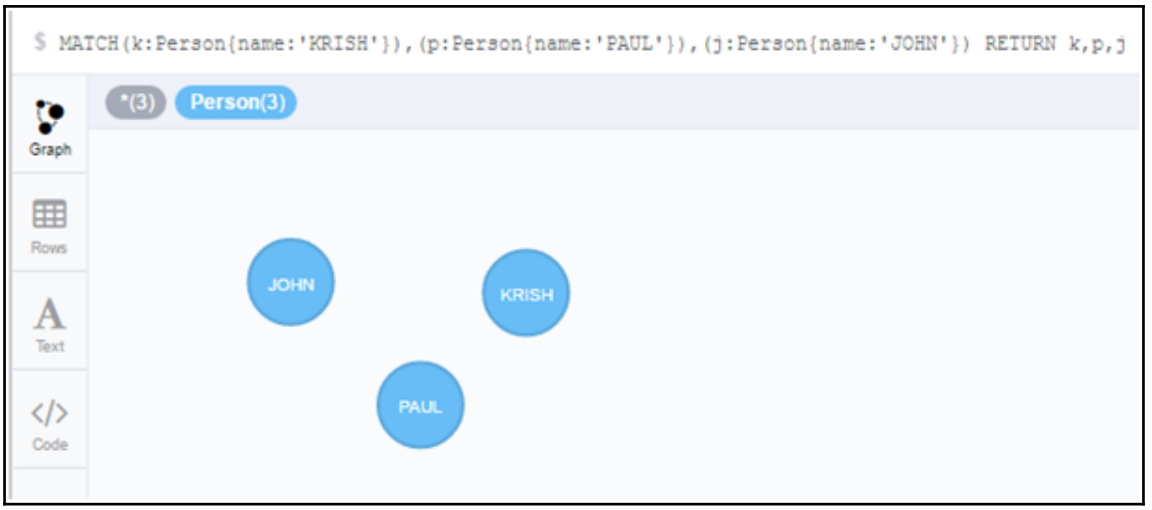
## Cypher Query

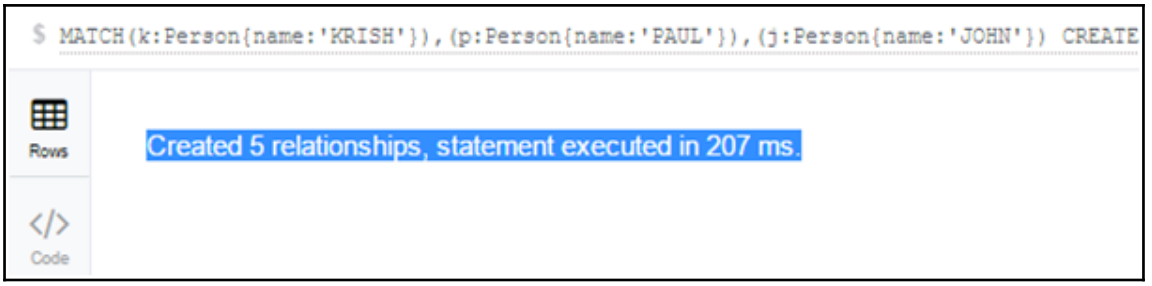
```
MATCH(u:User) -[f:friendof]-> (m:User) RETURN f
```

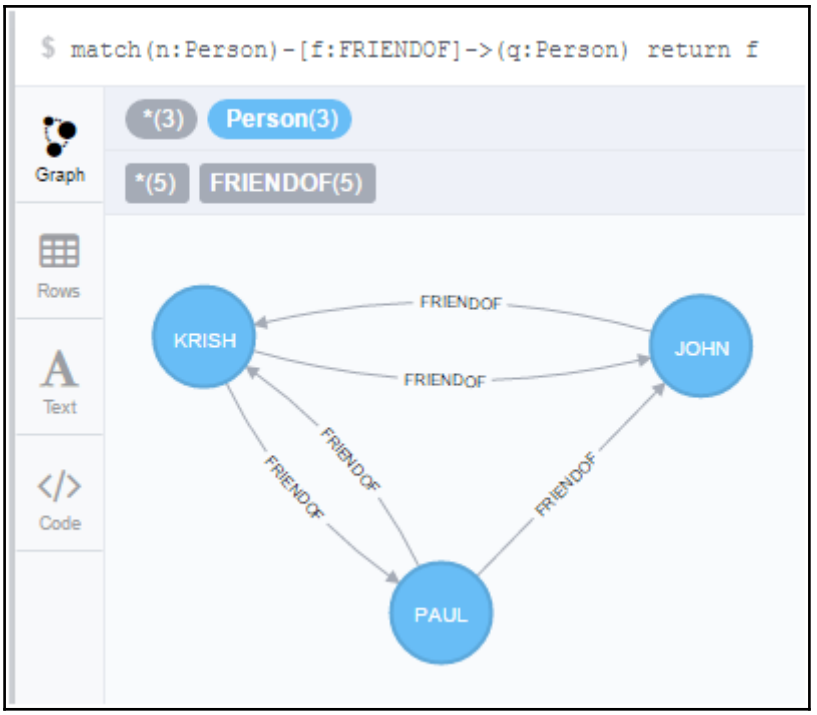
Above Cypher query pulls up all the friendship relations between pairs of users

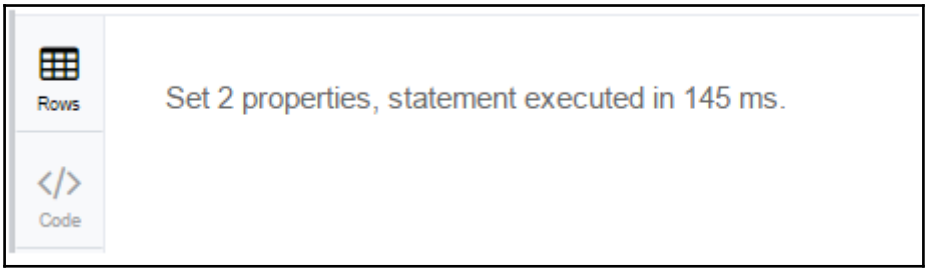

$ CREATE (john:Person {name:"JOHN"}) RETURN john
Graph
Rows
Text
Code
*(1)
Person(1)
JOHN

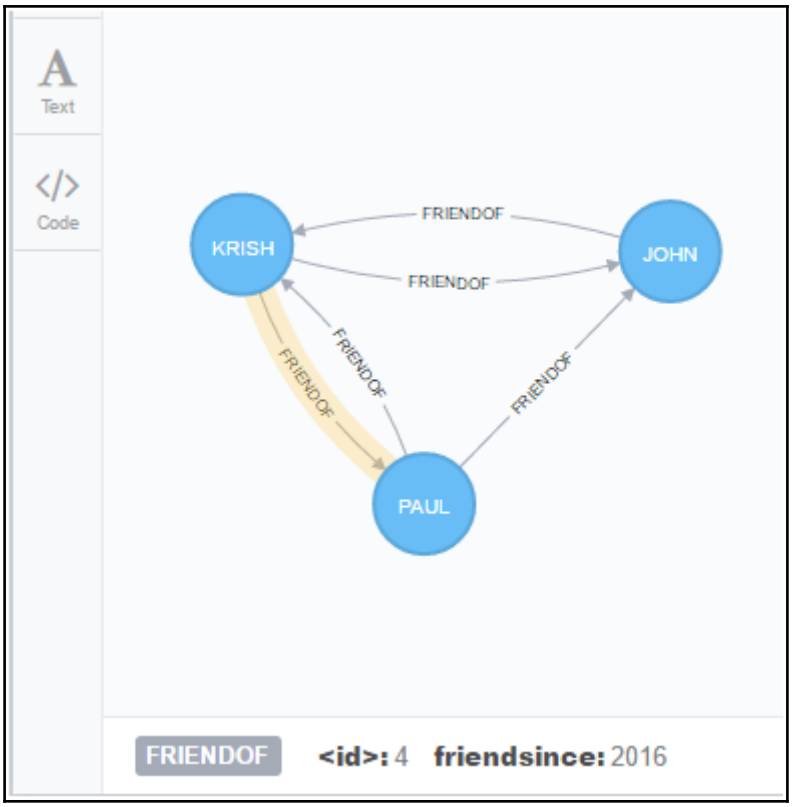

$ CREATE (paul:Person {name:"PAUL"}) CREATE (krish:Person {name:"KRISH"}) RETURN krish,paul
Graph
Rows
Text
Code
*(2)
Person(2)
KRISH
PAUL

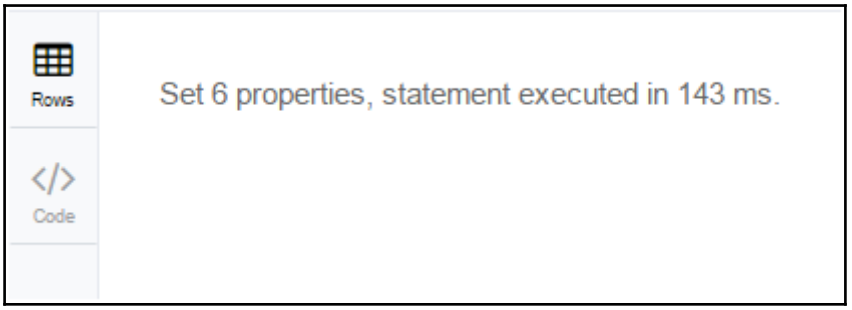

$ MATCH(k:Person{name:'KRISH'}),(p:Person{name:'PAUL'}),(j:Person{name:'JOHN'}) RETURN k,p,j
Graph
Rows
Text
Code
*(3)
Person(3)
JOHN
KRISH
PAUL

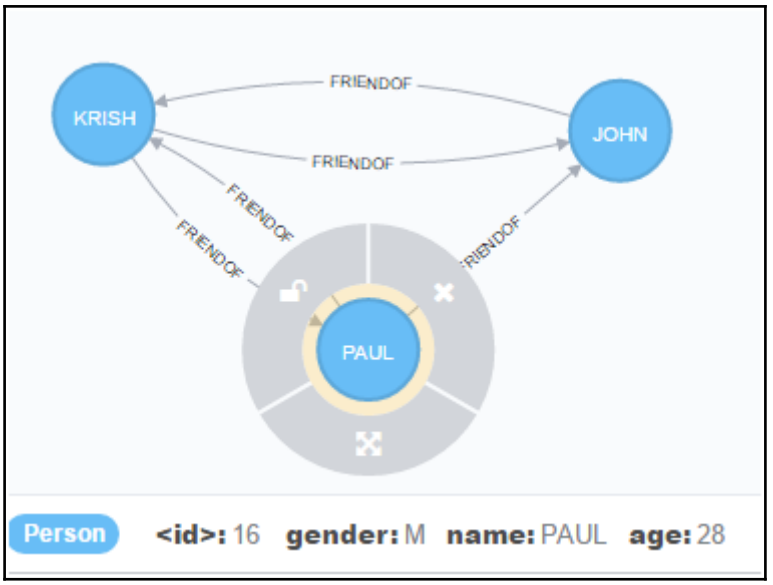

$ MATCH(k:Person{name:'KRISH'}),(p:Person{name:'PAUL'}),(j:Person{name:'JOHN'}) CREATE
Rows
Code
Created 5 relationships, statement executed in 207 ms.


$ match(n:Person)-[f:FRIENDOF]->(q:Person) return f
Graph
Rows
Text
Code
*(3)
Person(3)
*(5)
FRIENDOF(5)
KRISH
JOHN
PAUL
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF

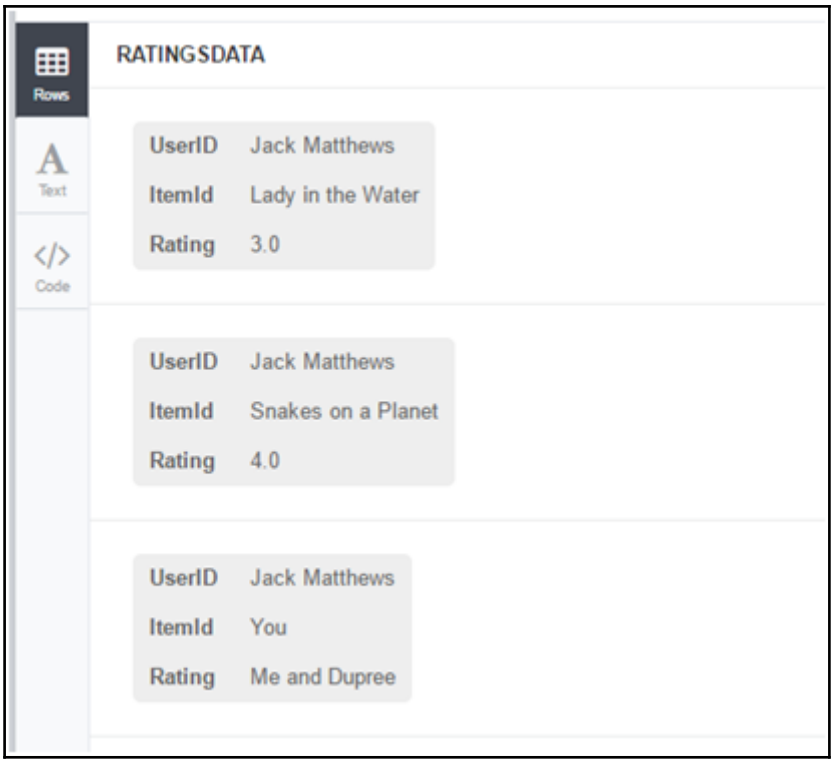

Rows
Code
Set 2 properties, statement executed in 145 ms.


Text
Code
KRISH
JOHN
PAUL
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
<id>: 4
friendsince: 2016

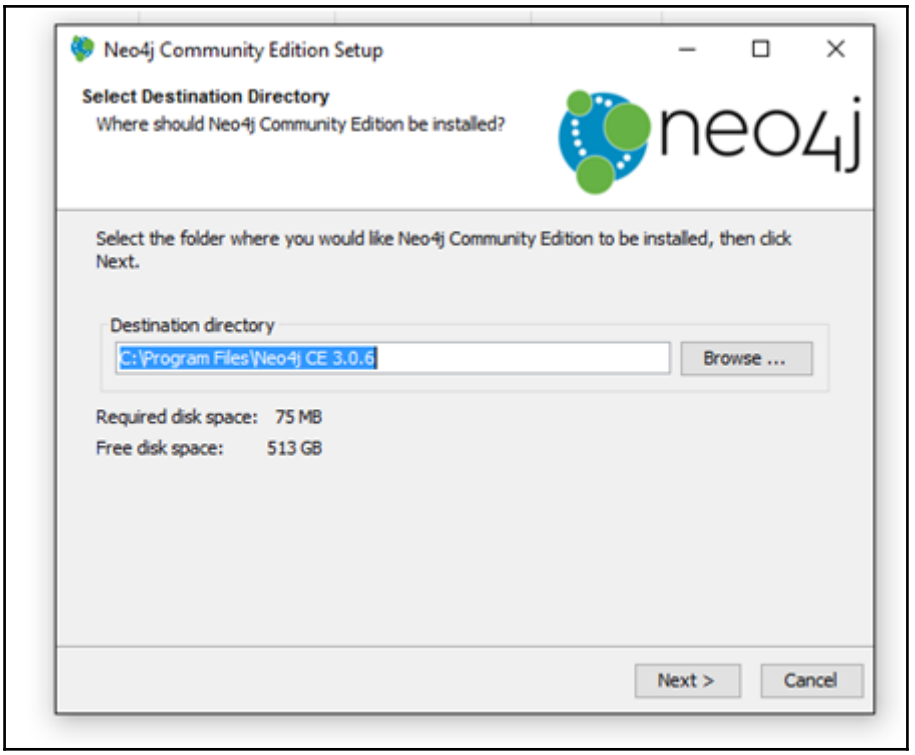

Rows
Code
Set 6 properties, statement executed in 143 ms.


KRISH
JOHN
PAUL
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
FRIENDOF
Person
<id>: 16
gender: M
name: PAUL
age: 28

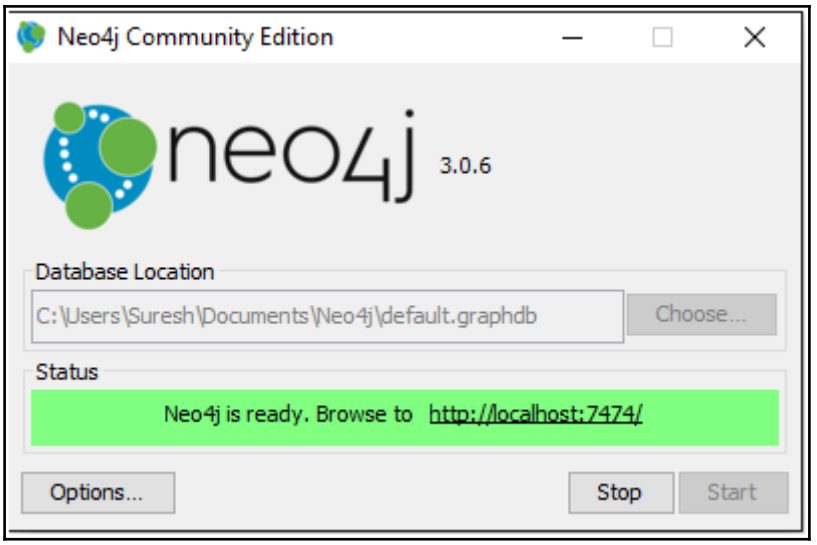

Rows
Text
Code
line
[UserID, ItemId, Rating]
[Jack Matthews, Lady in the Water, 3.0]
[Jack Matthews, Snakes on a Planet, 4.0]
[Jack Matthews, You, Me and Dupree, 3.5]
[Jack Matthews, Superman Returns, 5.0]
[Jack Matthews, The Night Listener, 3.0]
[Mick LaSalle, Lady in the Water, 3.0]
[Mick LaSalle, Snakes on a Planet, 4.0]
[Mick LaSalle, Just My Luck, 2.0]
[Mick LaSalle, Superman Returns, 3.0]
[Mick LaSalle, You, Me and Dupree, 2.0]
[Mick LaSalle, The Night Listener, 3.0]
[Claudia Puig, Snakes on a Planet, 3.5]

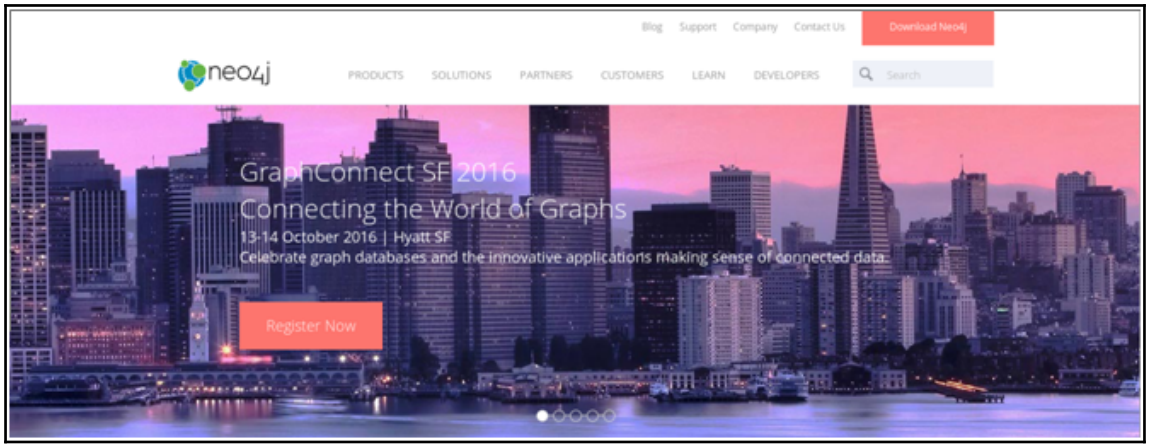

Rows
Text
Code
RATINGSDATA
UserID Jack Matthews
ItemId Lady in the Water
Rating 3.0
UserID Jack Matthews
ItemId Snakes on a Planet
Rating 4.0
UserID Jack Matthews
ItemId You
Rating Me and Dupree


Blog Support Company Contact Us Download Neo4j
neo4j
PRODUCTS SOLUTIONS PARTNERS CUSTOMERS LEARN DEVELOPERS
Search
Download Neo4j Community Edition
Neo4j 3.0.6
Release Notes | Other Releases
Ideal for learning and smaller do-it-yourself projects that do not require high levels of scaling.
Download Now - Windows 64 bit (exe)


Neo4j Community Edition Setup
Select Destination Directory
Where should Neo4j Community Edition be installed?
neo4j
Select the folder where you would like Neo4j Community Edition to be installed, then click Next.
Destination directory
C:\Program Files\Neo4j CE 3.0.6
Browse ...
Required disk space: 75 MB
Free disk space: 513 GB
Next >
Cancel

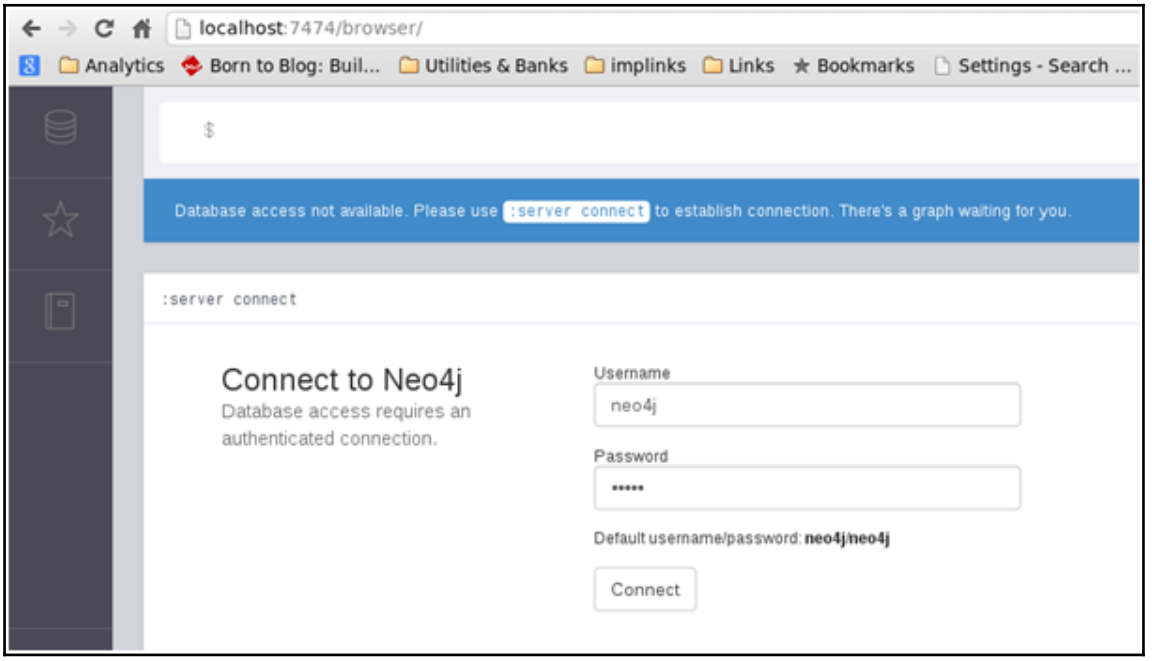

Neo4j Community Edition
neo4j 3.0.6
Database Location
[redacted]Neo4j\default.graphdb
Choose...
Status
Choose a graph database directory, then start the server
Options...
Stop
Start

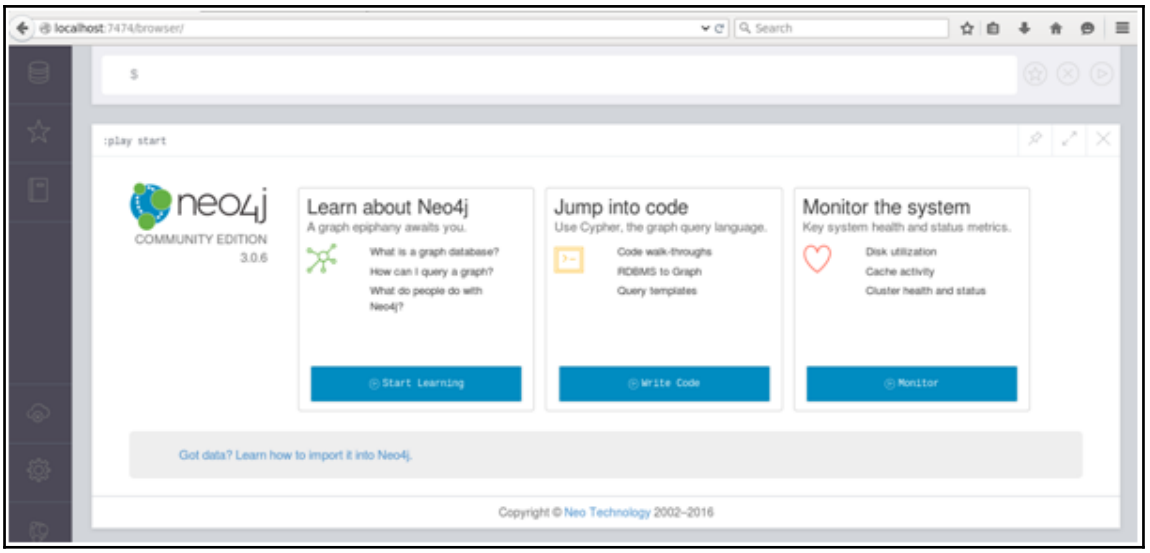

Neo4j Community Edition
neo4j
3.0.6
Database Location
C:\Users\Suresh\Documents\Neo4j\default.graphdb
Choose...
Status
Neo4j is ready. Browse to http://localhost:7474/
Options...
Stop
Start

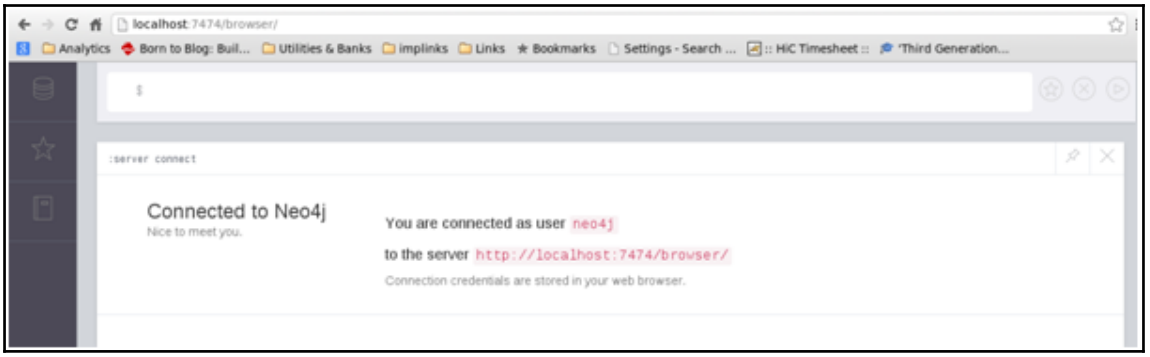

Blog
Support
Company
Contact Us
Download Neo4j
neo4j
PRODUCTS
SOLUTIONS
PARTNERS
CUSTOMERS
LEARN
DEVELOPERS
Search
GraphConnect SF 2016
Connecting the World of Graphs
13-14 October 2016 | Hyatt SF
Celebrate graph databases and the innovative applications making sense of connected data.
Register Now

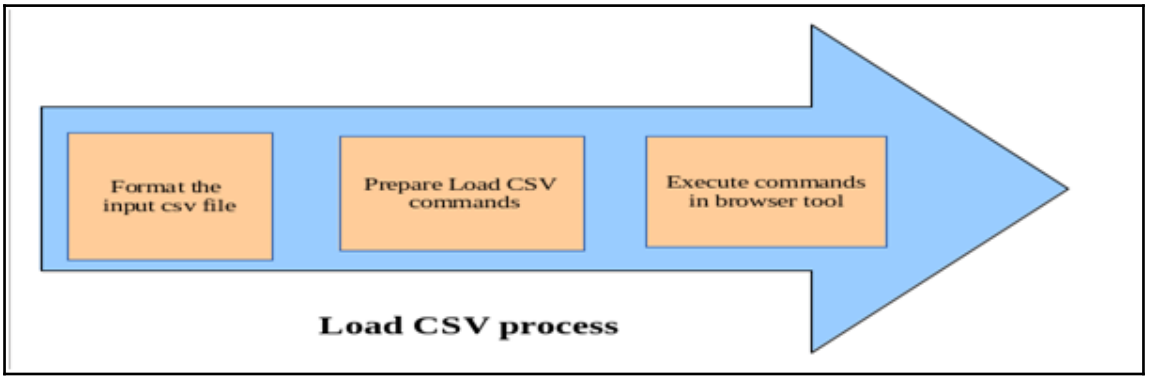

File Edit View Go Bookmarks Tabs Help
Back
Forward
Location:
100%
Icon View
neo4j-community-3.
0.6-unix.tar.gz


```
[1060929@01hw745020 home]$ neo4j start
Starting Neo4j.
WARNING: Max 1024 open files allowed, minimum of 40000 recommended. See the Neo4j manual.
Started neo4j (pid 1039). By default, it is available at http://localhost:7474/
There may be a short delay until the server is ready.
See /home/1060929/Softwares/neo4j/neo4j-community-3.0.6/logs/neo4j.log for current status.
[1060929@01hw745020 home]$ gedit ~/.bashrc
```




localhost:7474/browser/
Search
$
:play start
neo4j
COMMUNITY EDITION
3.0.6
Learn about Neo4j
A graph epiphany awaits you.
What is a graph database?
How can I query a graph?
What do people do with Neo4j?
Start Learning
Jump into code
Use Cypher, the graph query language.
Code walk-throughs
RDBMS to Graph
Query templates
Write Code
Monitor the system
Key system health and status metrics.
Disk utilization
Cache activity
Cluster health and status
Monitor
Got data? Learn how to import it into Neo4j.
Copyright © Neo Technology 2002–2016


localhost:7474/browser/
Analytics
Born to Blog: Buil...
Utilities & Banks
implinks
Links
Bookmarks
Settings - Search ...
:: HiC Timesheet ::
'Third Generation...
$
:server connect
Connected to Neo4j
Nice to meet you.
You are connected as user neo4j
to the server http://localhost:7474/browser/
Connection credentials are stored in your web browser.

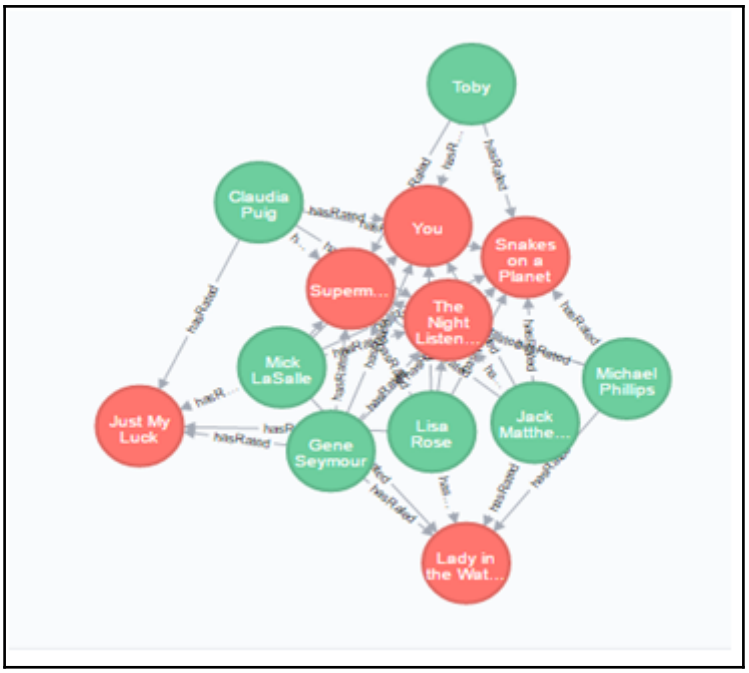

Format the input csv file
Prepare Load CSV commands
Execute commands in browser tool
Load CSV process

```
Jack Matthews,Lady in the Water,3.0
Jack Matthews,Snakes on a Planet,4.0
Jack Matthews,You, Me and Dupree,3.5
Jack Matthews,Superman Returns,5.0
Jack Matthews,The Night Listener,3.0
Mick LaSalle,Lady in the Water,3.0
Mick LaSalle,Snakes on a Planet,4.0
Mick LaSalle,Just My Luck,2.0
Mick LaSalle,Superman Returns,3.0
Mick LaSalle,You, Me and Dupree,2.0
Mick LaSalle,The Night Listener,3.0
Claudia Puig,Snakes on a Planet,3.5
Claudia Puig,Just My Luck,3.0
Claudia Puig,You, Me and Dupree,2.5
Claudia Puig,Superman Returns,4.0
Claudia Puig,The Night Listener,4.5
Lisa Rose,Lady in the Water,2.5
Lisa Rose,Snakes on a Planet,3.5
Lisa Rose,Just My Luck,3.0
Lisa Rose,Superman Returns,3.5
Lisa Rose,The Night Listener,3.0
Lisa Rose,You, Me and Dupree,2.5
Toby,Snakes on a Planet,4.5
Toby,Superman Returns,4.0
Toby,You, Me and Dupree,1.0
Gene Seymour,Lady in the Water,3.0
Gene Seymour,Snakes on a Planet,3.5
Gene Seymour,Just My Luck,1.5
Gene Seymour,Superman Returns,5.0
Gene Seymour,You, Me and Dupree,3.5
Gene Seymour,The Night Listener,3.0
Michael Phillips,Lady in the Water,2.5
Michael Phillips,Snakes on a Planet,3.0
Michael Phillips,Superman Returns,3.5
```

Rows

Code

Added 13 labels, created 13 nodes, set 48 properties, created 35 relationships, statement executed in 276 ms.




| Recommendations |
|---|
| Just My Luck |
| Lady in the Water |
| The Night Listener |

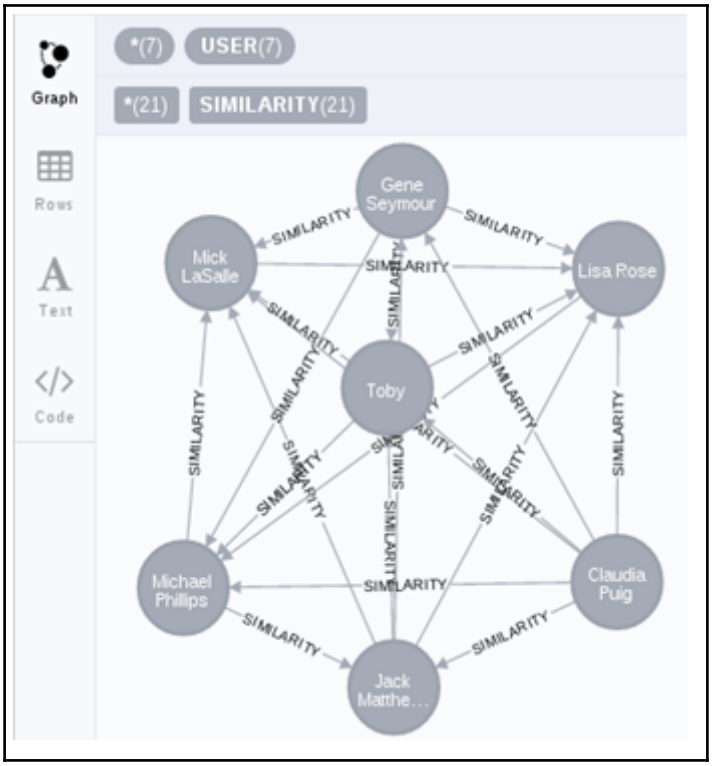
| MOVIE | Recommendation |
|---|---|
| The Night Listener | 3.3333333333333335 |
| Lady in the Water | 2.6 |
| Just My Luck | 2.25 |

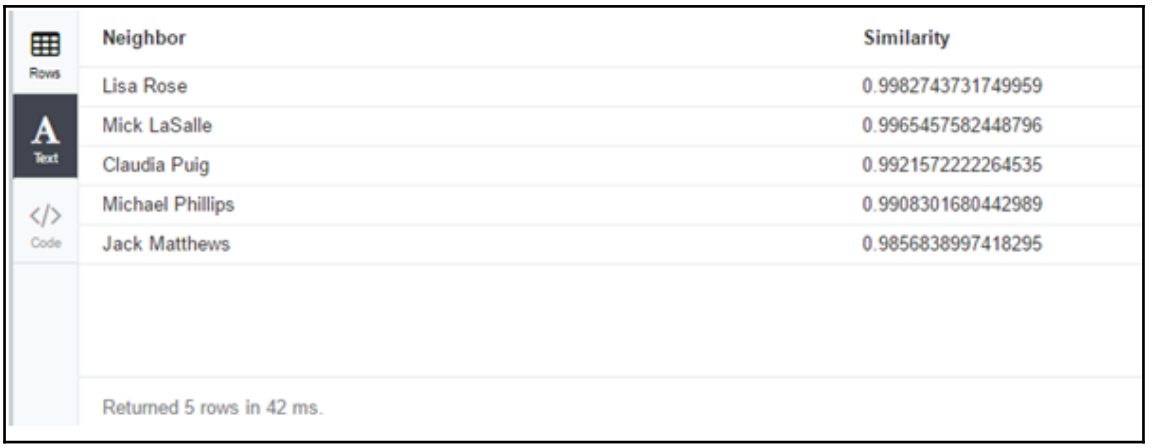
| u1 | u2 | CommonMovies | user1Rating | user2Rating |
|---|---|---|---|---|
| USERID Toby | USERID Jack Matthews | [You Me and Dupree, Superman Returns, Snakes on a Planet] | [1.0, 4.0, 4.5] | [3.5, 5.0, 4.0] |
| USERID Toby | USERID Michael Phillips | [Superman Returns, Snakes on a Planet] | [4.0, 4.5] | [3.5, 3.0] |
| USERID Toby | USERID Mick LaSalle | [You Me and Dupree, Superman Returns, Snakes on a Planet] | [1.0, 4.0, 4.5] | [2.0, 3.0, 4.0] |
| USERID Toby | USERID Gene Seymour | [You Me and Dupree, Superman Returns, Snakes on a Planet] | [1.0, 4.0, 4.5] | [3.5, 5.0, 3.5] |
| USERID Toby | USERID Claudia Puig | [You Me and Dupree, Superman Returns, Snakes on a Planet] | [1.0, 4.0, 4.5] | [2.5, 4.0, 3.5] |
| USERID Toby | USERID Lisa Rose | [You Me and Dupree, Superman Returns, Snakes on a Planet] | [1.0, 4.0, 4.5] | [2.5, 3.5, 3.5] |

| CoReviewer | similarity |
|---|---|
| Mick LaSalle | 0.7834936490538904 |
| Claudia Puig | 0.7397917500667335 |
| Lisa Rose | 0.7299691375663392 |
| Michael Phillips | 0.7204915028125263 |
| Jack Matthews | 0.6047152924789525 |
| Gene Seymour | 0.585421901205575 |

| MOVIE | Recommendation |
|---|---|
| The Night Listener | 3.3333333333333335 |
| Lady in the Water | 2.6 |
| Just My Luck | 2.25 |


Rows
Code
Set 42 properties, created 21 relationships, statement executed in 292 ms.

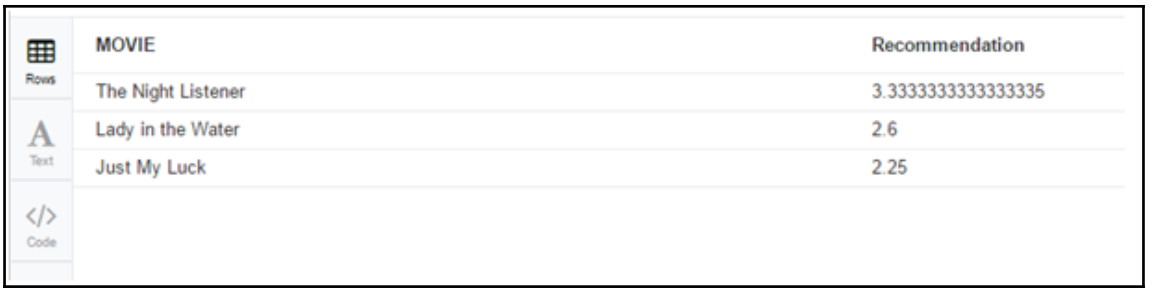

Graph
Rows
Text
Code
*(7)
USER(7)
*(21)
SIMILARITY(21)
Gene Seymour
Mick LaSalle
Lisa Rose
Toby
Michael Phillips
Claudia Puig
Jack Matthe…
SIMILARITY



Rows | Text | Code

| Neighbor | Similarity |
|---|---|
| Lisa Rose | 0.9982743731749959 |
| Mick LaSalle | 0.9965457582448796 |
| Claudia Puig | 0.9921572222264535 |
| Michael Phillips | 0.9908301680442989 |
| Jack Matthews | 0.9856838997418295 |

Returned 5 rows in 42 ms.

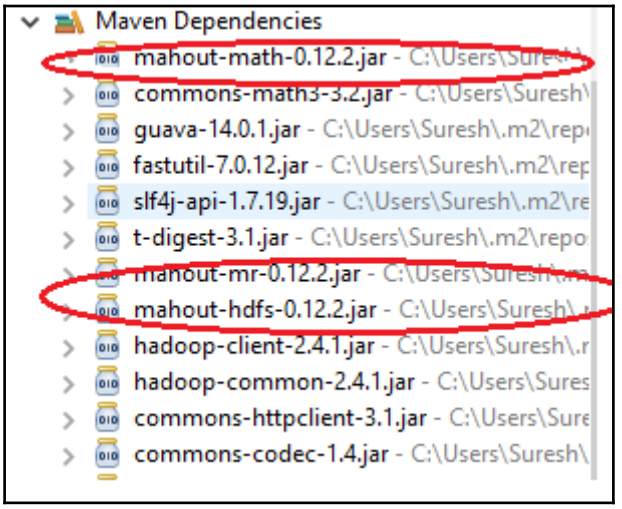


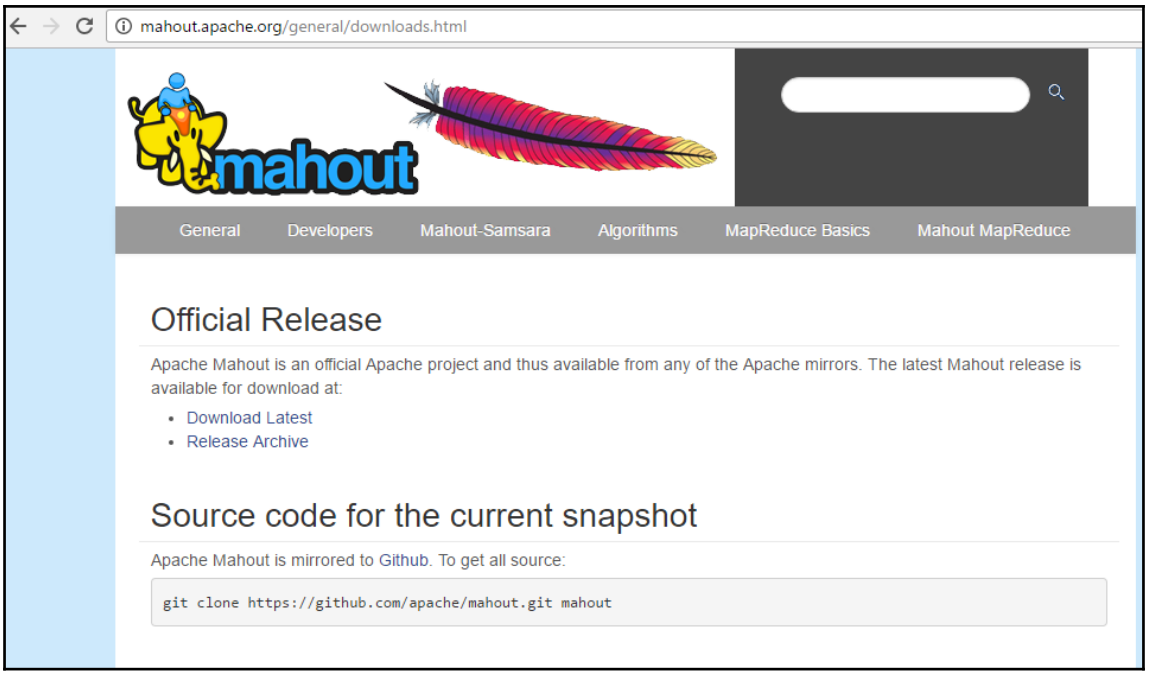
Rows | Text | Code

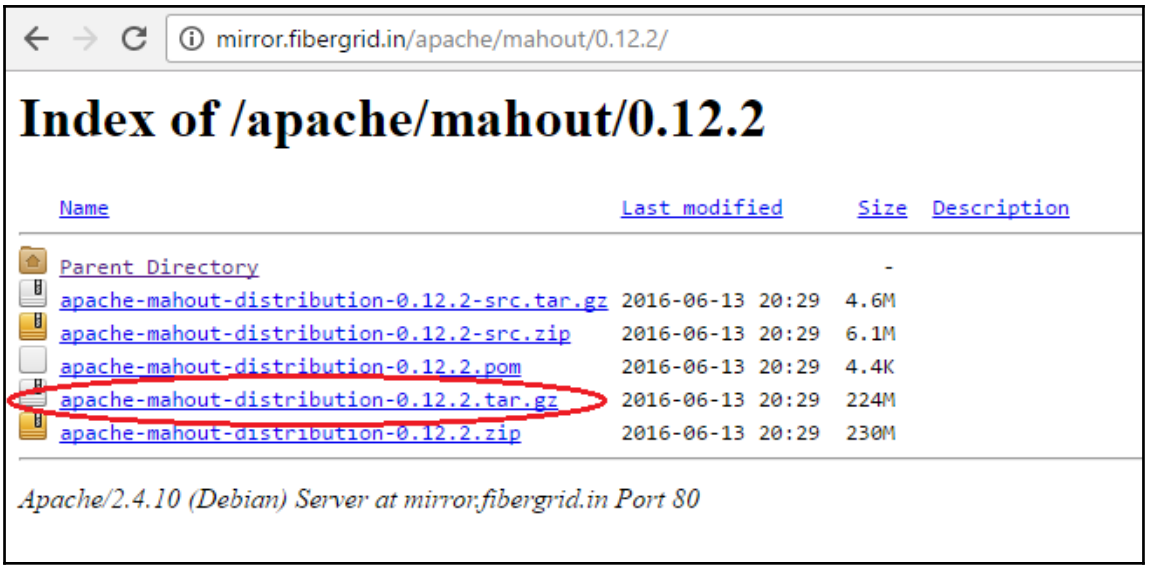
| MOVIE | Recommendation |
|---|---|
| The Night Listener | 3.3333333333333335 |
| Lady in the Water | 2.6 |
| Just My Luck | 2.25 |

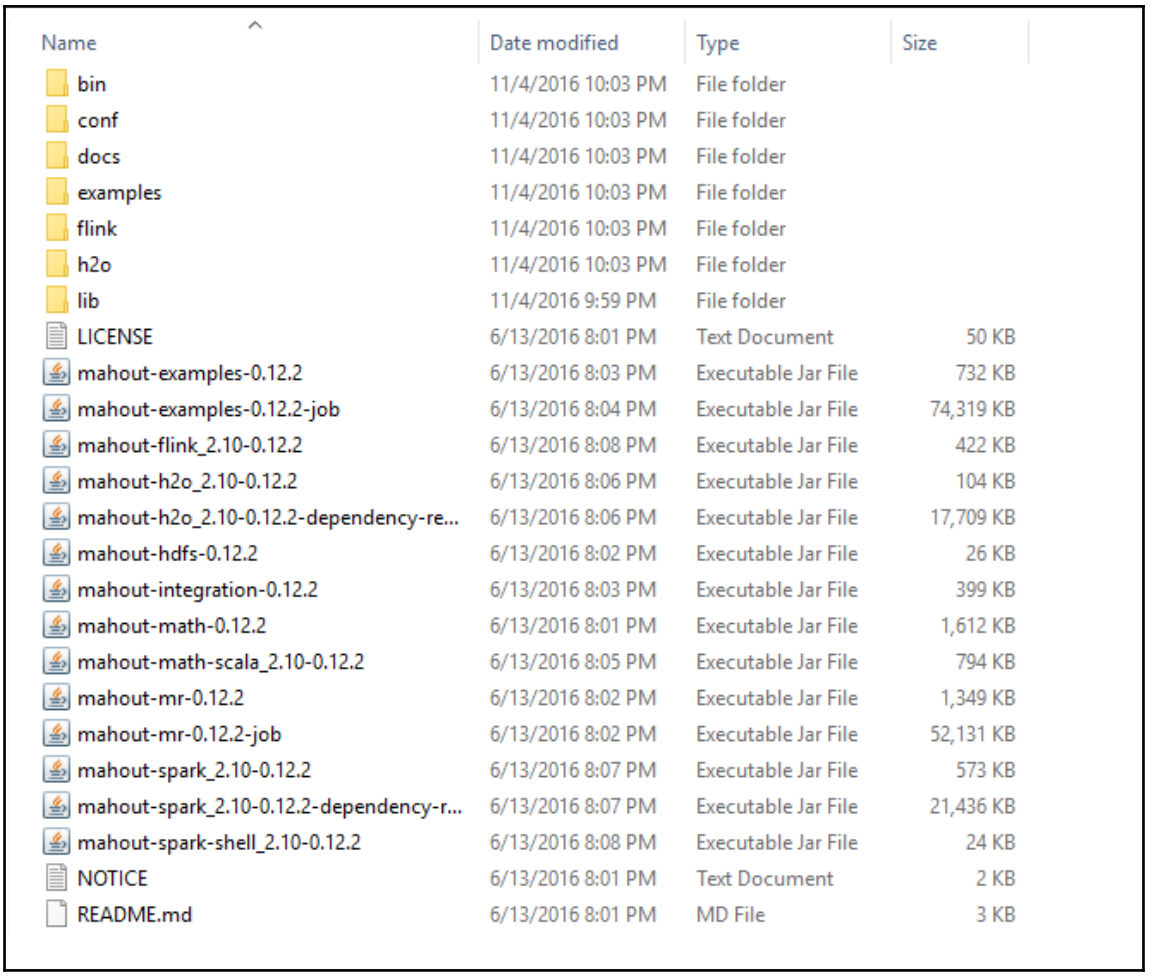
# Chapter 9: Building Scalable Recommendation Engines with Mahout


```
18   <dependency>
19     <groupId>org.apache.mahout</groupId>
20     <artifactId>mahout-math</artifactId>
21     <version>0.12.2</version>
22 </dependency>
23 <dependency>
24     <groupId>org.apache.mahout</groupId>
25     <artifactId>mahout-mr</artifactId>
26     <version>0.12.2</version>
27 </dependency>
```

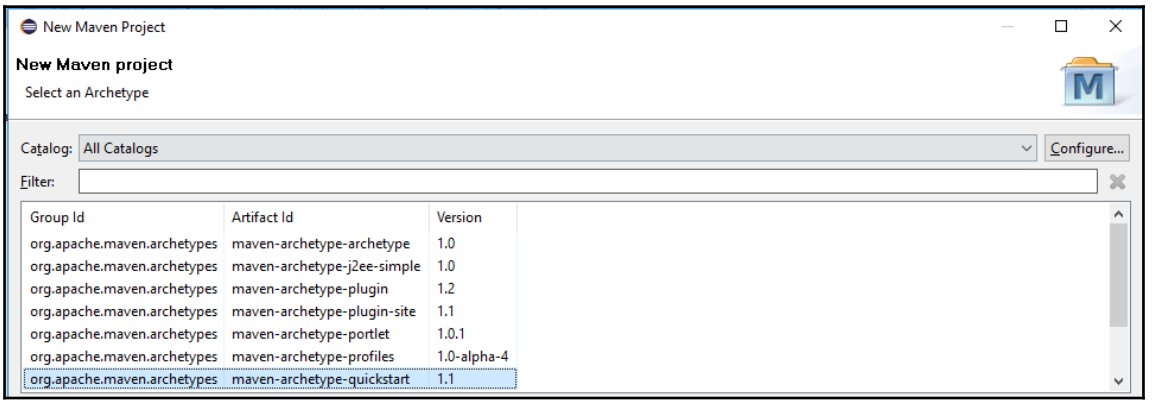


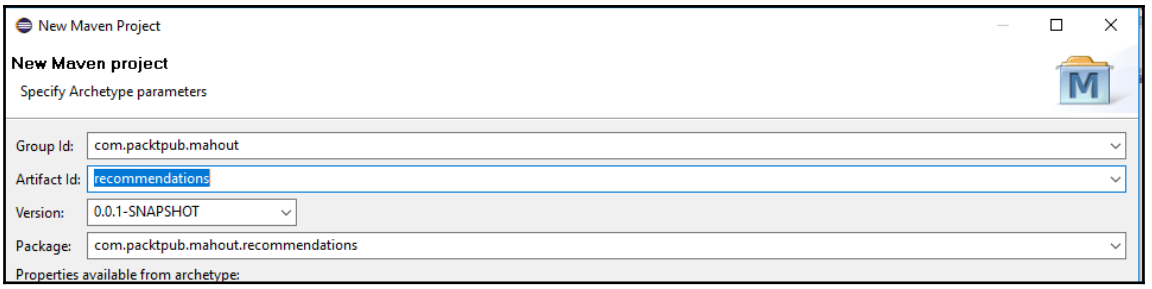

mahout.apache.org/general/downloads.html
mahout
General
Developers
Mahout-Samsara
Algorithms
MapReduce Basics
Mahout MapReduce
Official Release
Apache Mahout is an official Apache project and thus available from any of the Apache mirrors. The latest Mahout release is available for download at:
Download Latest
Release Archive
Source code for the current snapshot
Apache Mahout is mirrored to Github. To get all source:
git clone https://github.com/apache/mahout.git mahout

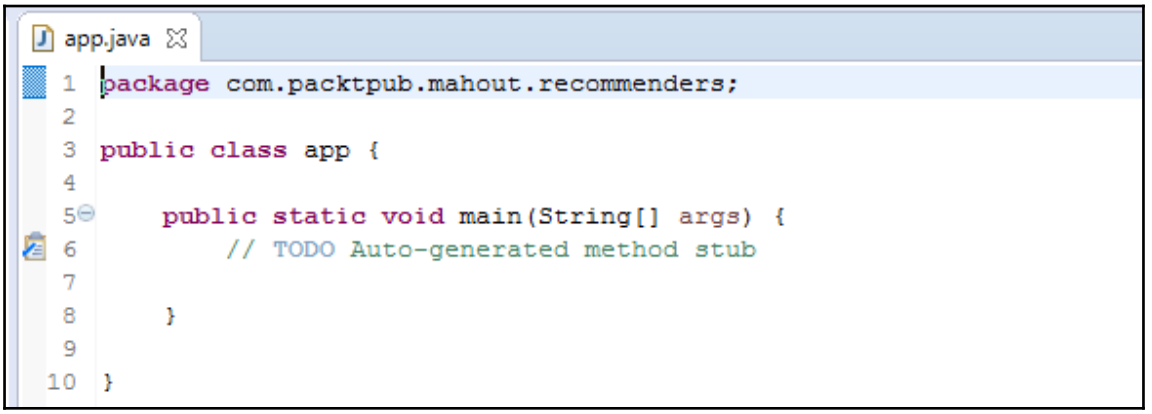

mirror.fibergrid.in/apache/mahout/0.12.2/
Index of /apache/mahout/0.12.2
Name
Last modified
Size
Description
Parent Directory -
apache-mahout-distribution-0.12.2-src.tar.gz 2016-06-13 20:29 4.6M
apache-mahout-distribution-0.12.2-src.zip 2016-06-13 20:29 6.1M
apache-mahout-distribution-0.12.2.pom 2016-06-13 20:29 4.4K
apache-mahout-distribution-0.12.2.tar.gz 2016-06-13 20:29 224M
apache-mahout-distribution-0.12.2.zip 2016-06-13 20:29 230M
Apache/2.4.10 (Debian) Server at mirror.fibergrid.in Port 80

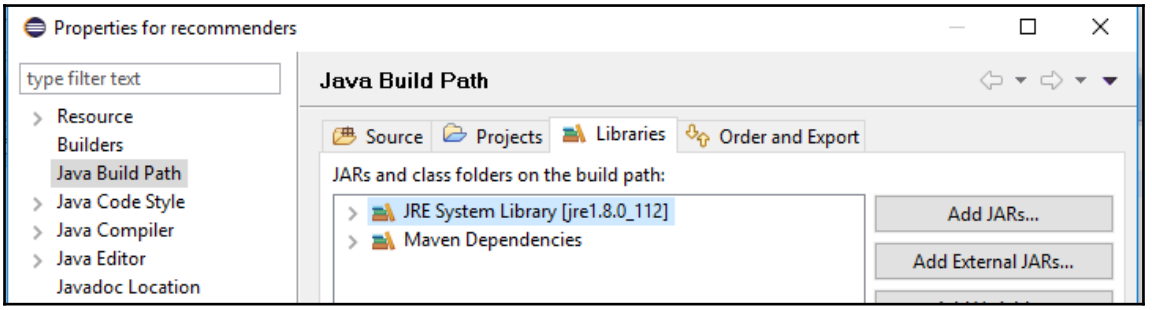


| Name | Date modified | Type | Size |
|---|---|---|---|
| bin | 11/4/2016 10:03 PM | File folder | |
| conf | 11/4/2016 10:03 PM | File folder | |
| docs | 11/4/2016 10:03 PM | File folder | |
| examples | 11/4/2016 10:03 PM | File folder | |
| flink | 11/4/2016 10:03 PM | File folder | |
| h2o | 11/4/2016 10:03 PM | File folder | |
| lib | 11/4/2016 9:59 PM | File folder | |
| LICENSE | 6/13/2016 8:01 PM | Text Document | 50 KB |
| mahout-examples-0.12.2 | 6/13/2016 8:03 PM | Executable Jar File | 732 KB |
| mahout-examples-0.12.2-job | 6/13/2016 8:04 PM | Executable Jar File | 74,319 KB |
| mahout-flink_2.10-0.12.2 | 6/13/2016 8:08 PM | Executable Jar File | 422 KB |
| mahout-h2o_2.10-0.12.2 | 6/13/2016 8:06 PM | Executable Jar File | 104 KB |
| mahout-h2o_2.10-0.12.2-dependency-re... | 6/13/2016 8:06 PM | Executable Jar File | 17,709 KB |
| mahout-hdfs-0.12.2 | 6/13/2016 8:02 PM | Executable Jar File | 26 KB |
| mahout-integration-0.12.2 | 6/13/2016 8:03 PM | Executable Jar File | 399 KB |
| mahout-math-0.12.2 | 6/13/2016 8:01 PM | Executable Jar File | 1,612 KB |
| mahout-math-scala_2.10-0.12.2 | 6/13/2016 8:05 PM | Executable Jar File | 794 KB |
| mahout-mr-0.12.2 | 6/13/2016 8:02 PM | Executable Jar File | 1,349 KB |
| mahout-mr-0.12.2-job | 6/13/2016 8:02 PM | Executable Jar File | 52,131 KB |
| mahout-spark_2.10-0.12.2 | 6/13/2016 8:07 PM | Executable Jar File | 573 KB |
| mahout-spark_2.10-0.12.2-dependency-r... | 6/13/2016 8:07 PM | Executable Jar File | 21,436 KB |
| mahout-spark-shell_2.10-0.12.2 | 6/13/2016 8:08 PM | Executable Jar File | 24 KB |
| NOTICE | 6/13/2016 8:01 PM | Text Document | 2 KB |
| README.md | 6/13/2016 8:01 PM | MD File | 3 KB |



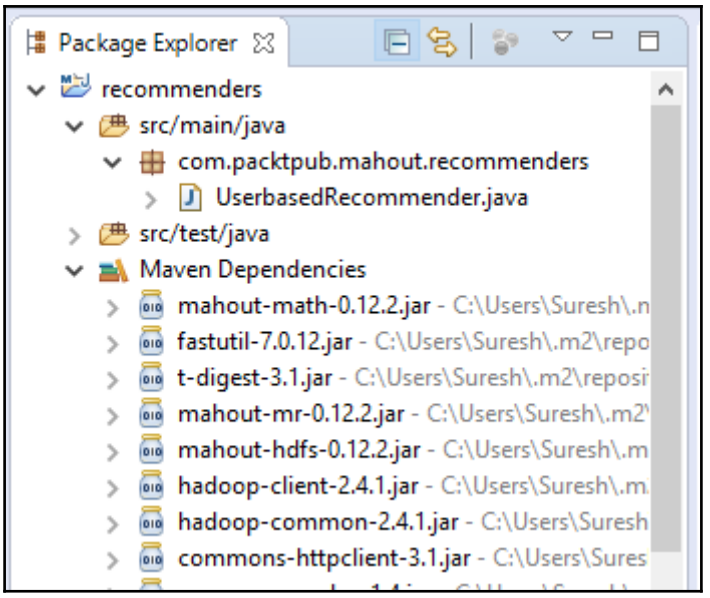

New Maven Project
New Maven project
Specify Archetype parameters
Group Id: com.packtpub.mahout
Artifact Id: recommendations
Version: 0.0.1-SNAPSHOT
Package: com.packtpub.mahout.recommendations
Properties available from archetype:

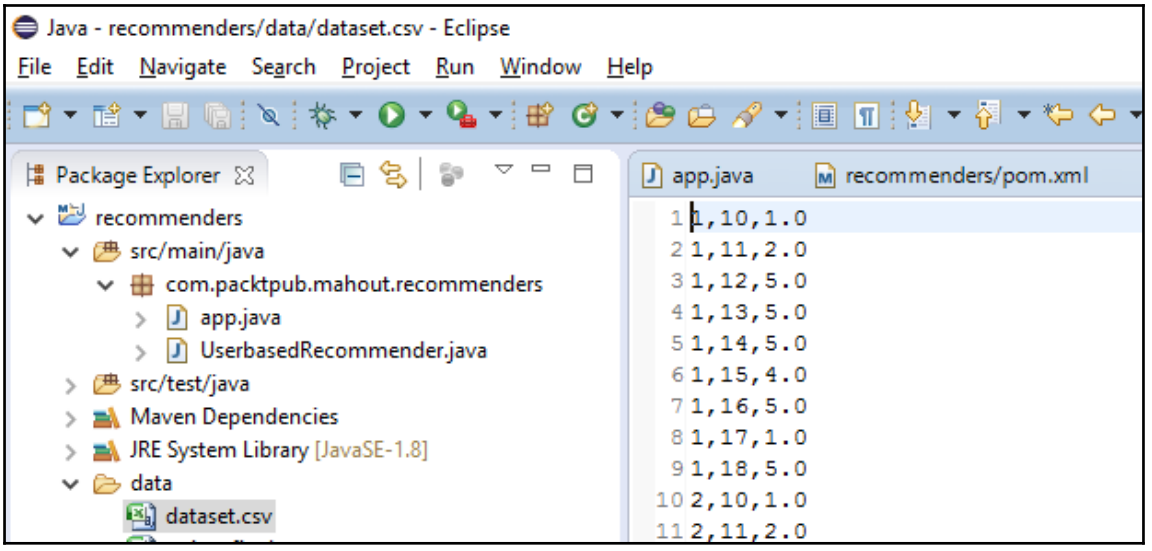

app.java
1 package com.packtpub.mahout.recommenders;
2
3 public class app {
4
5 public static void main(String[] args) {
6 // TODO Auto-generated method stub
7
8 }
9
10 }

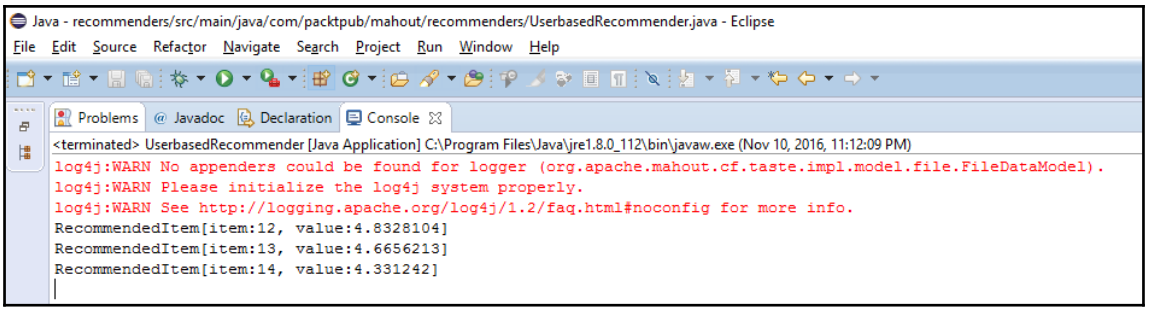

Properties for recommenders
type filter text
Resource
Builders
Java Build Path
Java Code Style
Java Compiler
Java Editor
Javadoc Location
Java Build Path
Source
Projects
Libraries
Order and Export
JARs and class folders on the build path:
JRE System Library [jre1.8.0_112]
Maven Dependencies
Add JARs...
Add External JARs...

app.java | recommenders/pom.xml

```
13    <properties>
14      <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
15    </properties>
16    <dependencies>
17    <dependency>
18      <groupId>org.apache.mahout</groupId>
19      <artifactId>mahout-math</artifactId>
20      <version>0.12.2</version>
21  </dependency>
22  <dependency>
23      <groupId>org.apache.mahout</groupId>
24      <artifactId>mahout-mr</artifactId>
25      <version>0.12.2</version>
26  </dependency>
27    <dependency>
28      <groupId>org.slf4j</groupId>
29      <artifactId>slf4j-api</artifactId>
30      <version>1.7.21</version>
31  </dependency>
32  <dependency>
33      <groupId>org.slf4j</groupId>
34      <artifactId>slf4j-log4j12</artifactId>
35      <version>1.7.21</version>
36  </dependency>
37  <dependency>
38      <groupId>com.google.guava</groupId>
39      <artifactId>guava</artifactId>
40      <version>19.0</version>
41  </dependency>
42  <dependency>
43      <groupId>org.apache.commons</groupId>
44      <artifactId>commons-math3</artifactId>
```


Package Explorer
recommenders
src/main/java
com.packtpub.mahout.recommenders
UserbasedRecommender.java
src/test/java
Maven Dependencies
mahout-math-0.12.2.jar - C:\Users\Suresh\.n
fastutil-7.0.12.jar - C:\Users\Suresh\.m2\repo
t-digest-3.1.jar - C:\Users\Suresh\.m2\reposi
mahout-mr-0.12.2.jar - C:\Users\Suresh\.m2\
mahout-hdfs-0.12.2.jar - C:\Users\Suresh\.m
hadoop-client-2.4.1.jar - C:\Users\Suresh\.m
hadoop-common-2.4.1.jar - C:\Users\Suresh
commons-httpclient-3.1.jar - C:\Users\Sures


Java - recommenders/data/dataset.csv - Eclipse
File Edit Navigate Search Project Run Window Help
Package Explorer
recommenders
src/main/java
com.packtpub.mahout.recommenders
app.java
UserbasedRecommender.java
src/test/java
Maven Dependencies
JRE System Library [JavaSE-1.8]
data
dataset.csv
app.java
recommenders/pom.xml
1,10,1.0
1,11,2.0
1,12,5.0
1,13,5.0
1,14,5.0
1,15,4.0
1,16,5.0
1,17,1.0
1,18,5.0
2,10,1.0
2,11,2.0




```
[cloudera@quickstart ~]$ hadoop fs -cat mahout/output1/part-r-
00000
3 [10:3.8597424]
4 [13:4.0]
```

```
1 userID,placeID,rating,food_rating,service_rating
2 U1077,135085,2,2,2
3 U1077,135038,2,2,1
4 U1077,132825,2,2,2
5 U1077,135060,1,2,2
6 U1068,135104,1,1,2
7 U1068,132740,0,0,0
8 U1068,132663,1,1,1
9 U1068,132732,0,0,0
10 U1068,132630,1,1,1
11 U1067,132584,2,2,2
12 U1067,132733,1,1,1
13 U1067,132732,1,2,2
14 U1067,132630,1,0,1
15 U1067,135104,0,0,0
16 U1067,132560,1,0,0
17 U1103,132584,1,2,1
18 U1103,132732,0,0,2
```

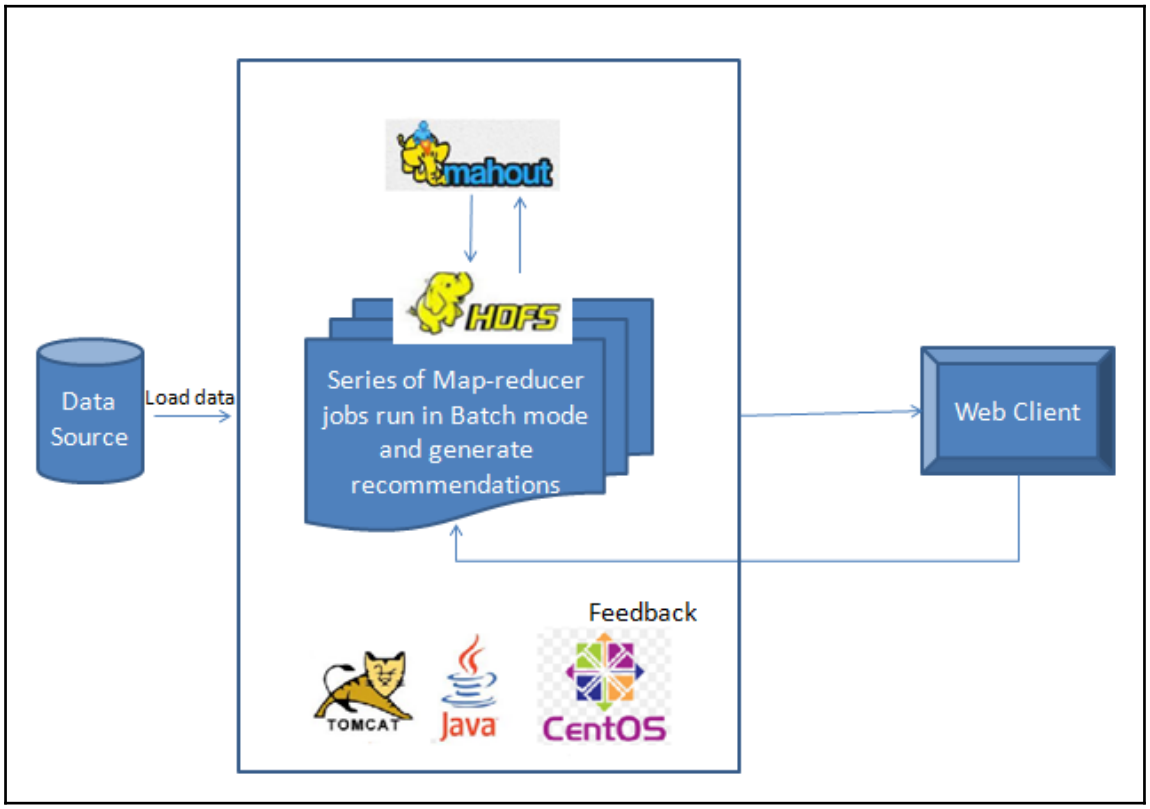
```
1077,135085,2
1077,135038,2
1077,132825,2
1077,135060,1
1068,135104,1
1068,132740,0
1068,132663,1
1068,132732,0
1068,132630,1
1067,132584,2
1067,132733,1
1067,132732,1
1067,132630,1
1067,135104,0
1067,132560,1
1103,132584,1
1103,132732,0
1103,132630,1
```

```
Problems  Javadoc  Declaration  Console
<terminated> UserbasedRecommender [Java Application] C:\Program Files\Java\jre1.8.0_112\bin\javaw.exe (Nov 11, 2016, 11:04:44 PM)
log4j:WARN No appenders could be found for logger (org.apache.mahout.cf.taste.impl.model.file.FileDataModel).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
RecommendedItem[item:132613, value:1.2205102]
RecommendedItem[item:132667, value:1.0]
RecommendedItem[item:132584, value:0.98069793]
```

```
Problems  Javadoc  Declaration  Console
<terminated> ItembasedRecommendations [Java Application] C:\Program Files\Java\jre1.8.0_112\bin\javaw.exe (Nov 12, 2016, 12:03:12 AM)
log4j:WARN No appenders could be found for logger (org.apache.mahout.cf.taste.impl.model.file.FileDataModel).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
RecommendedItem[item:132667, value:0.96383345]
RecommendedItem[item:132732, value:0.9602005]
RecommendedItem[item:132733, value:0.9543598]
```

**Class Summary**

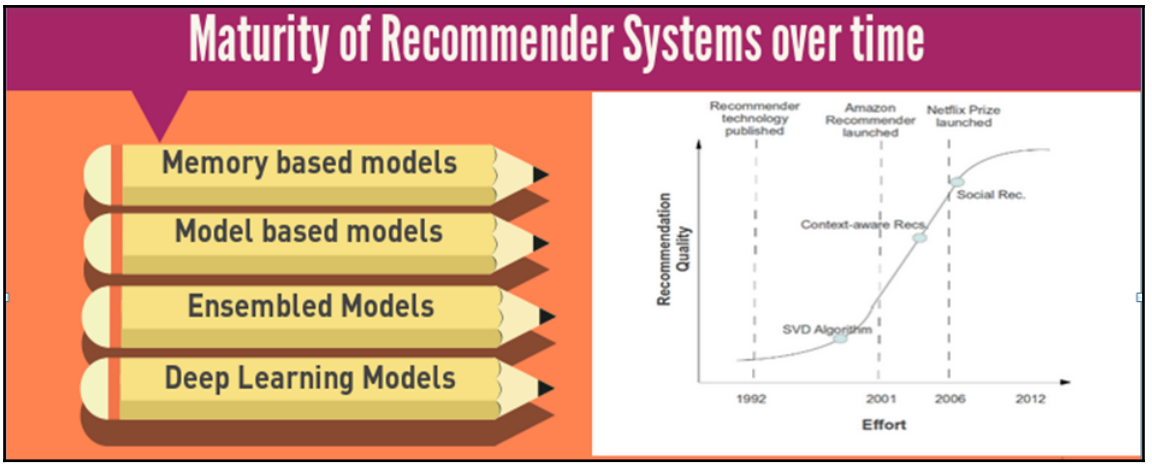
| Class | Description |
|---|---|
| AbstractDifferenceRecommenderEvaluator | Abstract superclass of a couple implementations, providing shared functionality. |
| AverageAbsoluteDifferenceRecommenderEvaluator | A `RecommenderEvaluator` which computes the average absolute difference between predicted and actual ratings for users. |
| GenericRecommenderIRStatsEvaluator | For each user, these implementation determine the top `n` preferences, then evaluate the IR statistics based on a `DataModel` that does not have these values. |
| GenericRelevantItemsDataSplitter | Picks relevant items to be those with the strongest preference, and includes the other users' preferences in full. |
| IRStatisticsImpl | |
| LoadEvaluator | Simple helper class for running load on a Recommender. |
| LoadStatistics | |
| OrderBasedRecommenderEvaluator | Evaluate recommender by comparing order of all raw prefs with order in recommender's output for that user. |
| RMSRecommenderEvaluator | A `RecommenderEvaluator` which computes the "root mean squared" difference between predicted and actual ratings for users. |

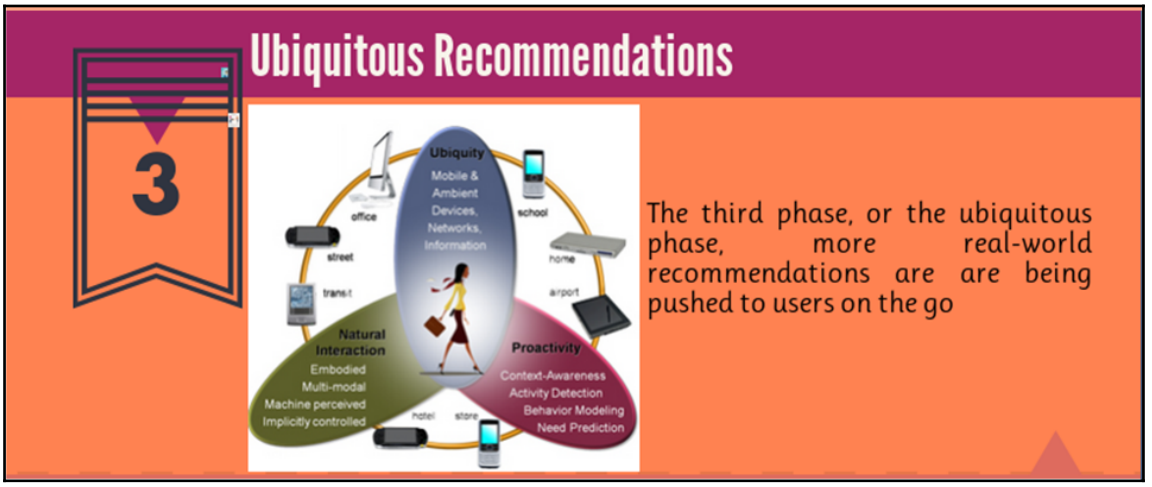


```
[cloudera@quickstart ~]$ hadoop fs -cat mahout/u.data |head
196 242 3 881250949
186 302 3 891717742
22 377 1 878887116
244 51 2 880606923
166 346 1 886397596
298 474 4 884182806
115 265 2 881171488
253 465 5 891628467
305 451 3 886324817
6 86 3 883603013
```

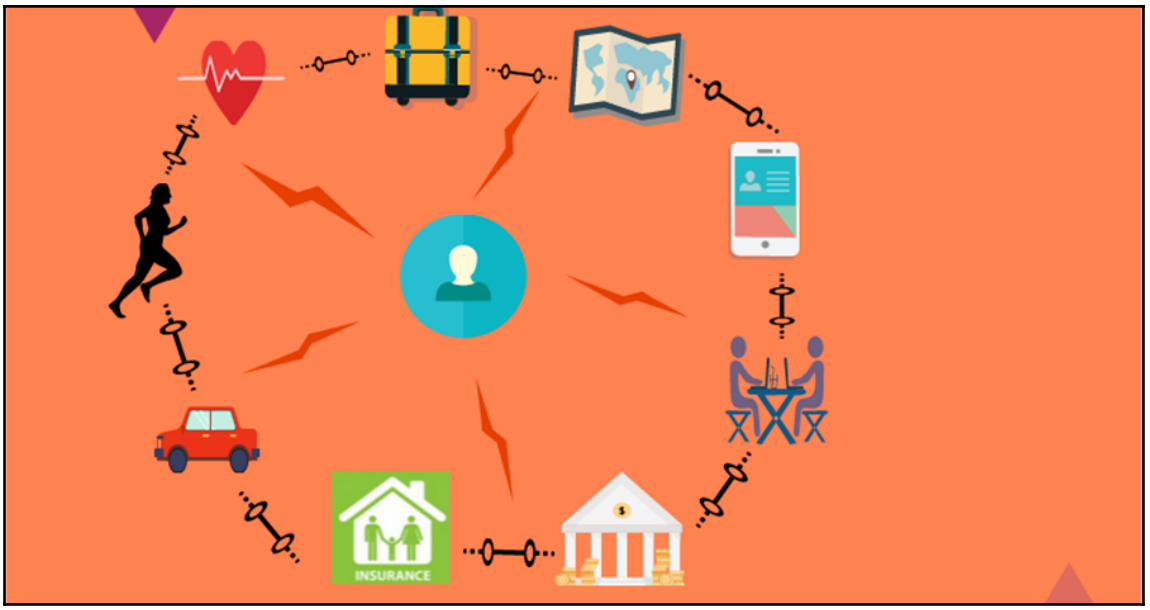
```
[cloudera@quickstart ~]$ hadoop fs -cat
recommendations/topNrecommendations/part-m-00000 |head

1
[1536:5.0,1467:4.831182,1449:4.80844,814:4.742634,1599:4.68286
9,1398:4.649307,1629:4.570285,1639:4.562079,408:4.536842,1367:
4.528492,483:4.4752526,318:4.4236937,1500:4.4102707,1201:4.408
1335,603:4.3991466]

2
[1536:5.0,814:4.78269,1449:4.7134724,1398:4.6964526,1599:4.563
0975,1467:4.551129,169:4.437805,408:4.38913,114:4.381019,64:4.
3791966,1367:4.3718753,1064:4.3644257,483:4.350657,851:4.32865
8,318:4.325402]

3
[1536:3.842033,1642:3.749693,1467:3.6595325,1449:3.6565793,150
0:3.652883,1398:3.5153584,814:3.5086145,169:3.4643984,1651:3.4
516013,1636:3.4516013,1645:3.4516013,1650:3.4516013,114:3.4097
04,1639:3.3940363,1524:3.367357]

4
[1642:5.0,1651:5.0,1636:5.0,1650:5.0,1645:5.0,1201:5.0,1639:5.
0,1536:5.0,1449:5.0,1367:5.0,1500:5.0,483:5.0,113:5.0,1122:5.0
,1398:5.0]

5
[1536:4.328833,1449:4.1486154,814:4.09392,1599:4.0884914,1467:
4.0109262,1398:4.0078206,1500:3.8654287,1639:3.821516,1629:3.8

013735,1122:3.8007212,1463:3.7989178,318:3.7716885,483:3.76697
33,114:3.7642615,1642:3.7634466]

6
[1536:4.669815,1467:4.5344944,814:4.49133,1449:4.4251847,1599:
4.367457,1639:4.355604,1398:4.228149,1500:4.21864,1642:4.19020
7,1367:4.1796536,1463:4.175493,1452:4.0896783,1458:4.0896783,1
629:4.063499,851:4.0583687]

7
[1500:4.8419685,1122:4.636387,1536:4.611435,1449:4.604633,1467
:4.470685,1651:4.439256,1650:4.439256,1636:4.439256,1645:4.439
256,1642:4.408753,1189:4.399697,1201:4.3634996,1398:4.3018093,
169:4.25464,1450:4.2321644]

8
[1536:5.0,1467:5.0,1642:5.0,1639:5.0,814:4.9976716,1449:4.9946
73,1398:4.813181,851:4.7969217,119:4.7551465,169:4.7425375,146
3:4.7282143,1367:4.6779137,483:4.661983,1201:4.658822,1458:4.6
515884]

9
[1500:5.0,1645:4.7686267,1636:4.7686267,1650:4.7686267,1651:4.
7686267,1431:4.6634703,1491:4.651774,1558:4.6267195,1122:4.591
5747,1449:4.568913,1201:4.5492425,1175:4.515396,1643:4.506893,
1512:4.4682612,1155:4.4282117]

10
[1536:4.821219,1500:4.7425957,1449:4.724121,1467:4.6232443,112
2:4.5975246,1642:4.548042,1398:4.512008,1599:4.461757,1650:4.4
549794,1645:4.4549794,1651:4.4549794,1636:4.4549794,1189:4.440
5603,169:4.3742394,814:4.372402]
```

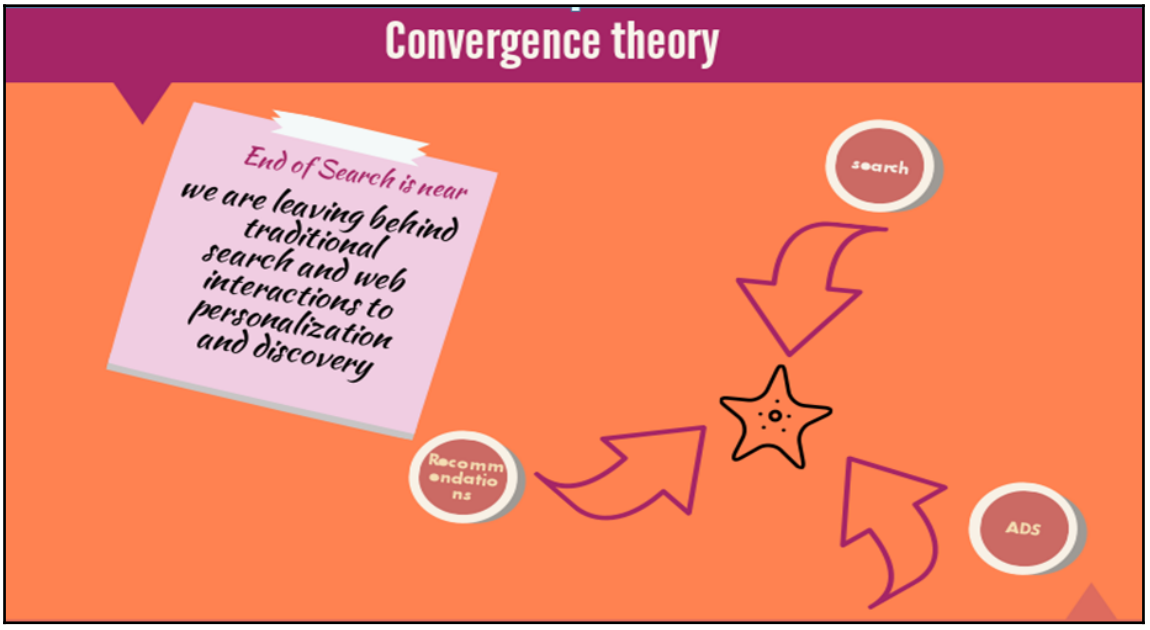
# Chapter 10: What Next?

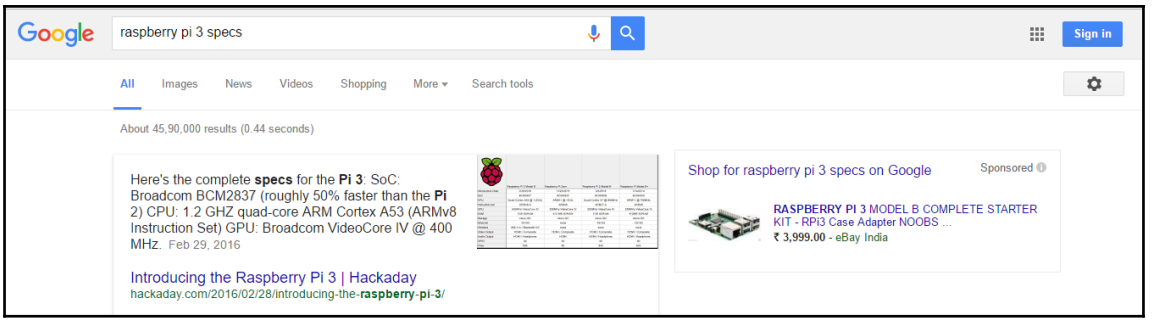


Maturity of Recommender Systems over time
Memory based models
Model based models
Ensembled Models
Deep Learning Models
Recommender technology published
Amazon Recommender launched
Netflix Prize launched
Recommendation Quality
SVD Algorithm
Context-aware Recs.
Social Rec.
1992
2001
2006
2012
Effort


3
Ubiquitous Recommendations
Ubiquity
Mobile & Ambient Devices, Networks, Information
office
school
street
home
transit
airport
Natural Interaction
Embodied
Multi-modal
Machine perceived
Implicitly controlled
Proactivity
Context-Awareness
Activity Detection
Behavior Modeling
Need Prediction
hotel
store
The third phase, or the ubiquitous phase, more real-world recommendations are are being pushed to users on the go


Google Allo


INSURANCE


Convergence theory
End of Search is near
we are leaving behind traditional search and web interactions to personalization and discovery
search
Recommendations
ADS


Google
raspberry pi 3 specs
Sign in
All
Images
News
Videos
Shopping
More
Search tools
About 45,90,000 results (0.44 seconds)
Here's the complete specs for the Pi 3: SoC: Broadcom BCM2837 (roughly 50% faster than the Pi 2) CPU: 1.2 GHZ quad-core ARM Cortex A53 (ARMv8 Instruction Set) GPU: Broadcom VideoCore IV @ 400 MHz. Feb 29, 2016
Introducing the Raspberry Pi 3 | Hackaday
hackaday.com/2016/02/28/introducing-the-raspberry-pi-3/
Shop for raspberry pi 3 specs on Google
Sponsored
RASPBERRY PI 3 MODEL B COMPLETE STARTER KIT - RPi3 Case Adapter NOOBS ...
₹ 3,999.00 - eBay India

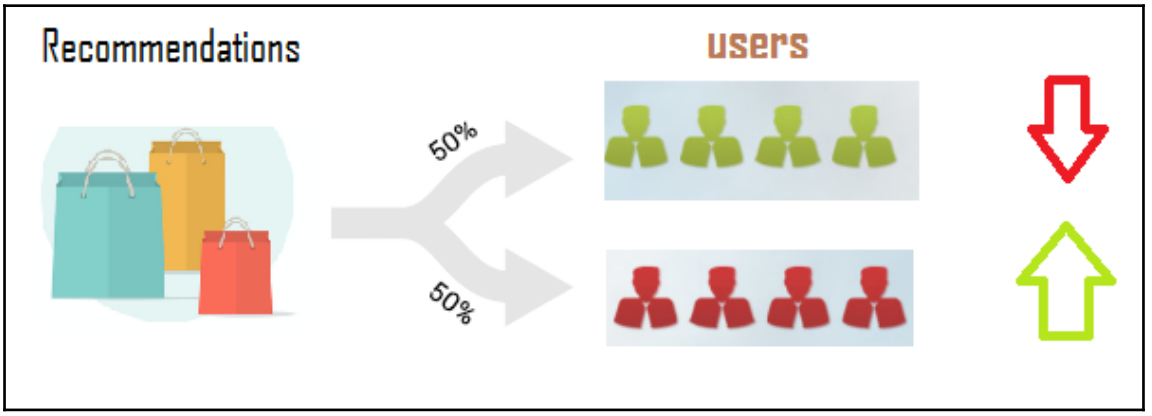

leaving behind the web
Systems that discover content on the move, i.e. ubiquitous ssytems
Including Social experience

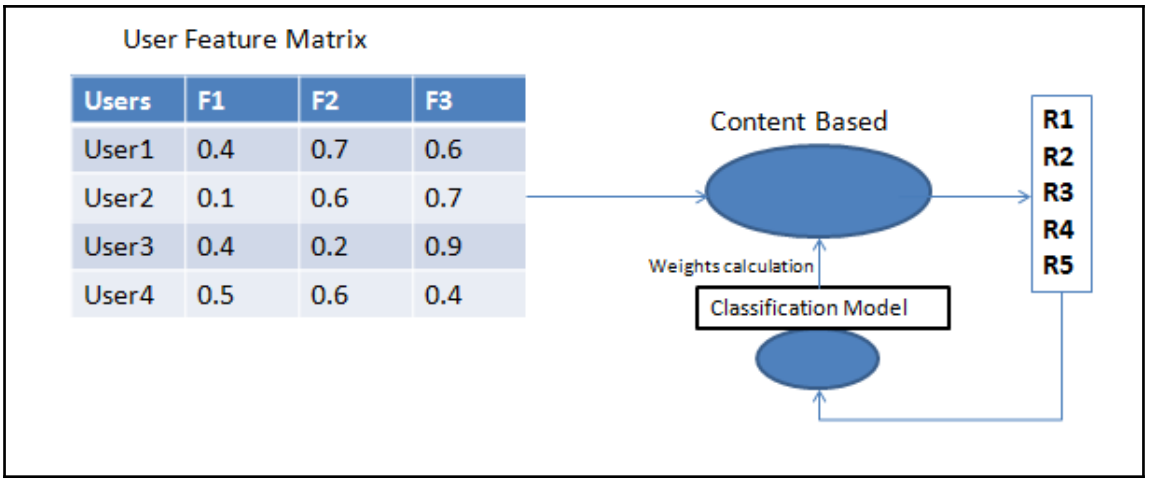

Context of recommendations are changing
different people at different places look for different things with different people and friends
Context for recommendations are changing
We will seeking different recommendations at different times and different groups
Moving from recommendations to Suggestions
Next best actions


Smart Manager
LG
FOOD MANAGER
Right Door
Add New Food
Cheese


https://news.google.co.in
Top Stories
Hillary Clinton
Mosul
Manchester United F.C.
Michel Aoun
FC Barcelona
Bhopal
Gareth Bale
Arsenal F.C.
Samsung Galaxy
Air pollution
Kochi, Kerala
India
World
Business
Technology
Entertainment
Sports
Science
More Business stories
Technology »
50 percent Samsung Galaxy Note 7 owners may switch to iPhone 7: IDC
Firstpost - 3 hours ago
Who will gain from Samsung's loss? This has been on everyone's mind ever since Samsung Note 7 fiasco hit headlines compelling the company to recall Note 7 completely (in a fireproof box) and offer refund to users.
Firstpost
Apple MacBook prices increased by Rs 10000
The Mobile Indian - 2 hours ago
Just last week Apple introduced the new MacBook Pro lineup with TouchBar and Touch ID functionality. While everyone had their eyes on the newly announced MacBook Pro, Apple silently increased the prices of MacBook lineup in India.
The Mobile ...
Reliance Jio launch doubles demand of 4G VoLTE devices
Firstpost - 3 hours ago
Demand for Voice over Long Term Evolution (VoLTE) enabled devices has doubled since the launch of Reliance Jio, according to a report in Economic Times.
Financial E...
LG may launch V20 at Rs 49990 in India by November-end
India Today - 57 minutes ago
Back in September, LG launched its V20 smartphone, the first phone to run Android Nougat at launch. The successor of LG V10, which never reached India, may soon launch in the country with a price tag of Rs 49,990.
India Today
More Technology stories


Book Recommendations based on previous interactions
MARY BALOGH
The ARRANGEMENT
GRACE
THE AUTUMN BRIDE
KRISTAN HIGGINS
the best man
BEC McMASTER
HEART OF IRON
SHERRY THOMAS
the Luckiest Lady in London


Recommendations
users
50%
50%


User Feature Matrix
Users F1 F2 F3
User1 0.4 0.7 0.6
User2 0.1 0.6 0.7
User3 0.4 0.2 0.9
User4 0.5 0.6 0.4
Content Based
R1
R2
R3
R4
R5
Weights calculation
Classification Model