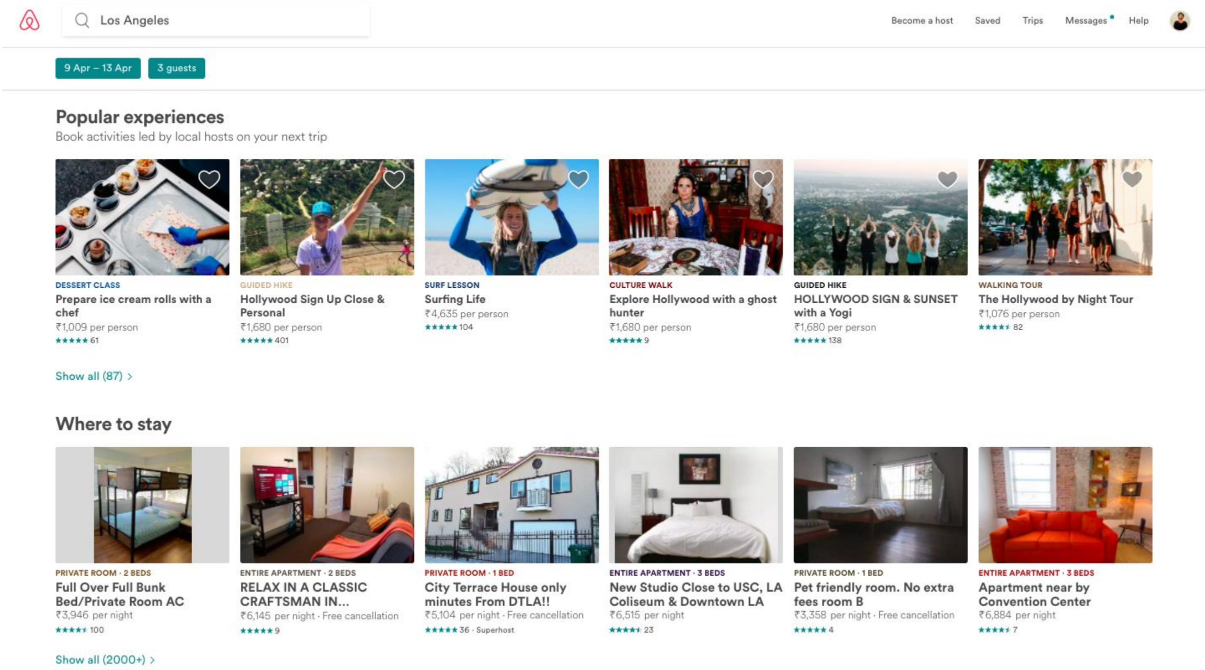
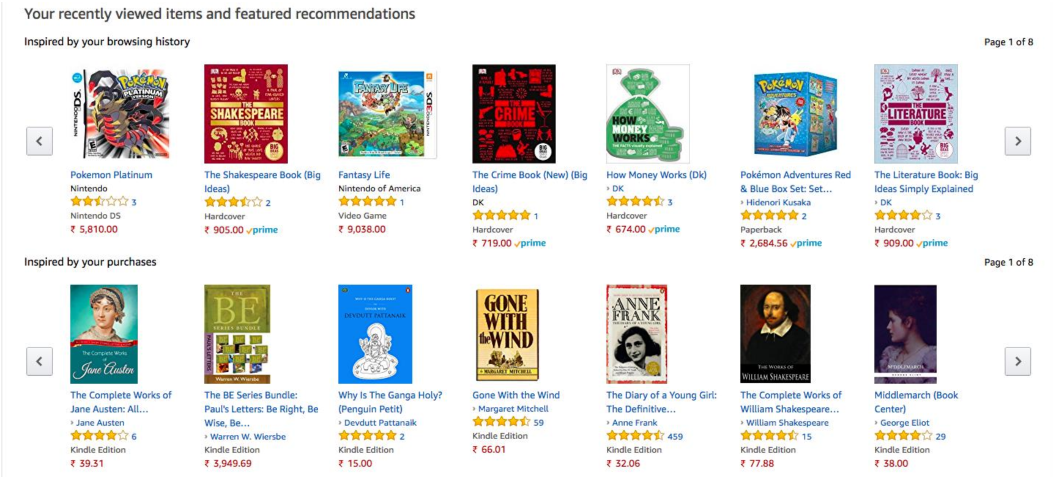
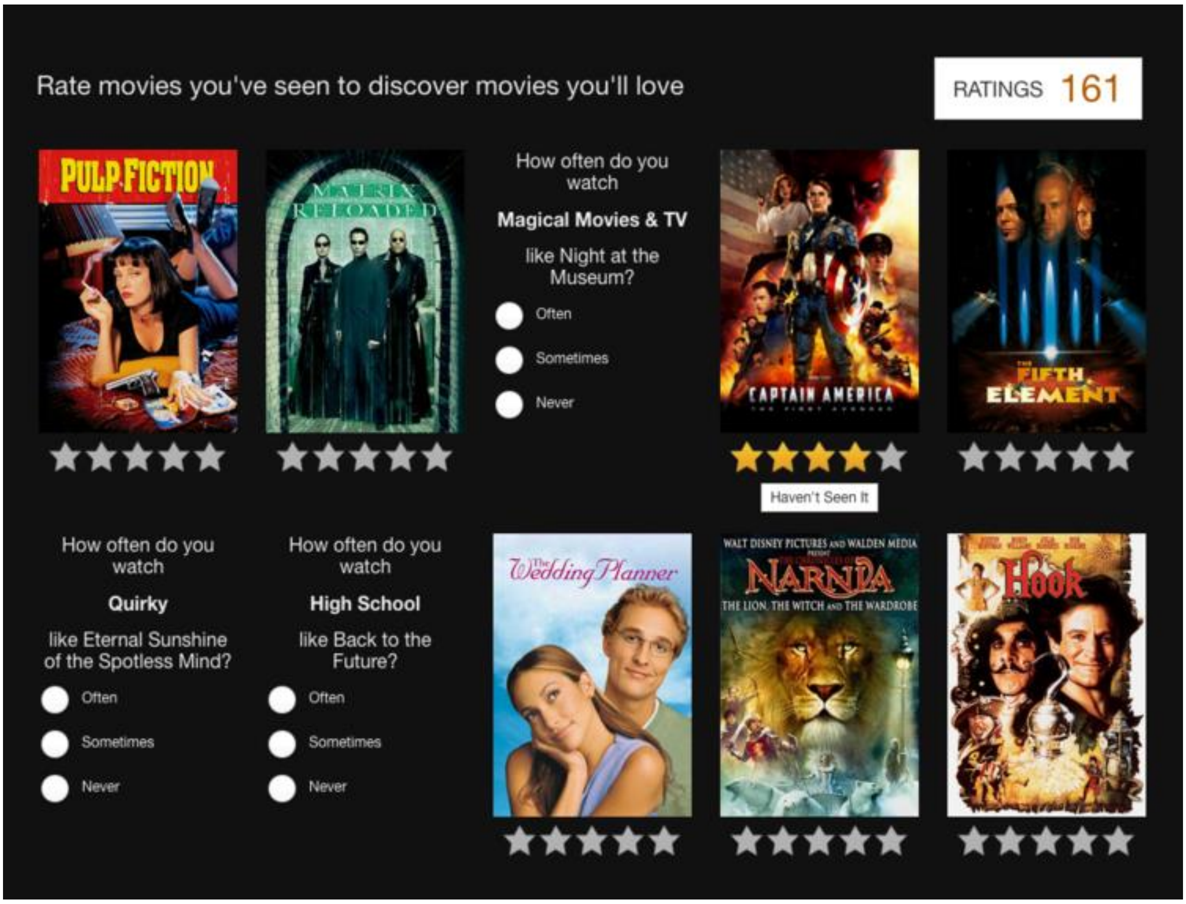
Chapter 1: Getting Started with Recommender Systems

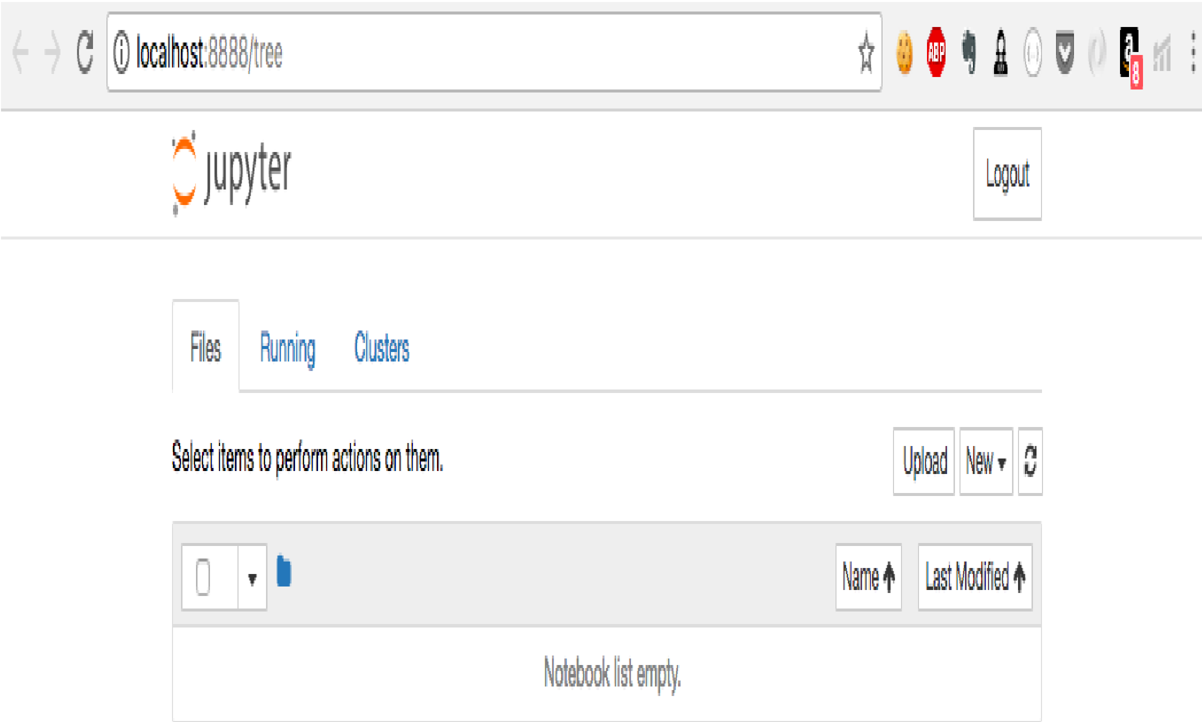
	i1	i2	i3	i4	i5	i6
U1	4	?	3	?	5	?
U2	?	2	?	?	4	1
U3	?	?	1	?	2	5
U4	?	?	3	?	?	1
U5	1	4	?	?	2	5
U6	5	?	2	1	?	4
U7	?	2	3	?	4	5

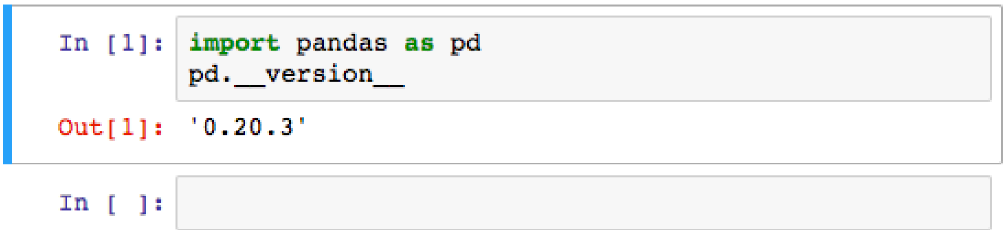



Your recently viewed items and featured recommendations
Inspired by your browsing history
Page 1 of 8
Pokemon Platinum
Nintendo
Nintendo DS
₹ 5,810.00
The Shakespeare Book (Big Ideas)
Hardcover
₹ 905.00 prime
Fantasy Life
Nintendo of America
Video Game
₹ 9,038.00
The Crime Book (New) (Big Ideas)
DK
Hardcover
₹ 719.00 prime
How Money Works (Dk)
DK
Hardcover
₹ 674.00 prime
Pokémon Adventures Red & Blue Box Set: Set...
Hidenori Kusaka
Paperback
₹ 2,684.56 prime
The Literature Book: Big Ideas Simply Explained
DK
Hardcover
₹ 909.00 prime
Inspired by your purchases
Page 1 of 8
The Complete Works of Jane Austen: All...
Jane Austen
Kindle Edition
₹ 39.31
The BE Series Bundle: Paul's Letters: Be Right, Be Wise, Be...
Warren W. Wiersbe
Kindle Edition
₹ 3,949.69
Why Is The Ganga Holy? (Penguin Petit)
Devdutt Pattanaik
Kindle Edition
₹ 15.00
Gone With the Wind
Margaret Mitchell
Kindle Edition
₹ 66.01
The Diary of a Young Girl: The Definitive...
Anne Frank
Kindle Edition
₹ 32.06
The Complete Works of William Shakespeare...
William Shakespeare
Kindle Edition
₹ 77.88
Middlemarch (Book Center)
George Eliot
Kindle Edition
₹ 38.00

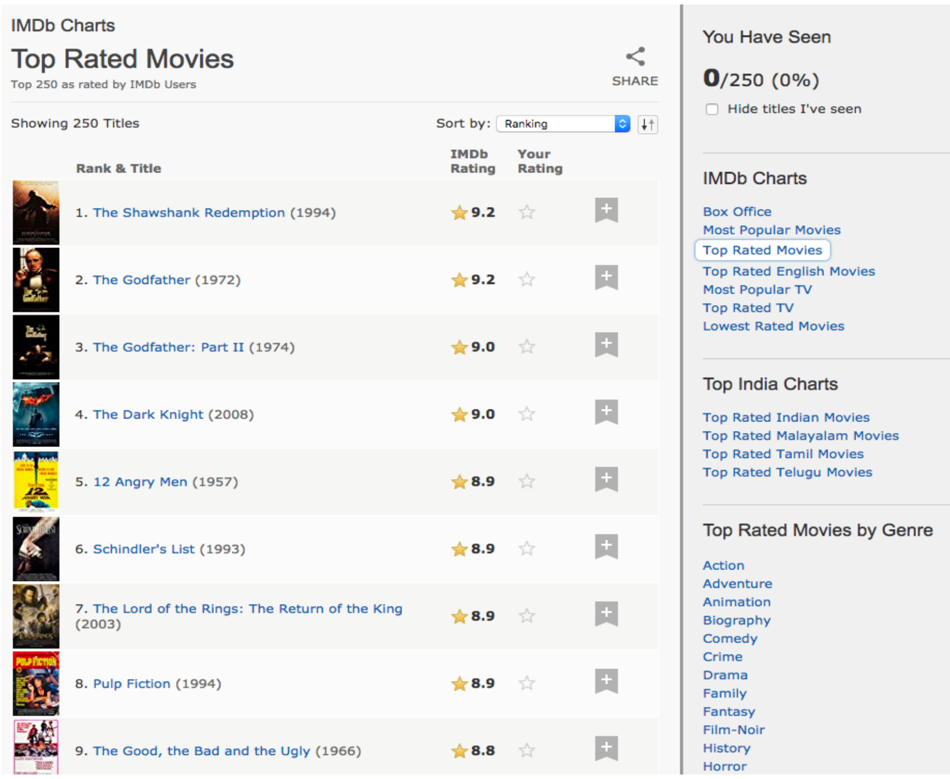

Rate movies you've seen to discover movies you'll love
RATINGS 161
PULP FICTION
MATRIX RELOADED
How often do you watch
Magical Movies & TV
like Night at the Museum?
Often
Sometimes
Never
CAPTAIN AMERICA
FIFTH ELEMENT
Haven't Seen It
How often do you watch
Quirky
like Eternal Sunshine of the Spotless Mind?
Often
Sometimes
Never
How often do you watch
High School
like Back to the Future?
Often
Sometimes
Never
The Wedding Planner
NARNIA
THE LION, THE WITCH AND THE WARDROBE
Hook

Chapter 2: Manipulating Data with the Pandas Library





Chapter 3: Building an IMDB Top 250 Clone with Pandas



	title	vote_count	vote_average	score	runtime
10309	Dilwale Dulhania Le Jayenge	661.0	9.1	8.855148	190.0
314	The Shawshank Redemption	8358.0	8.5	8.482863	142.0
834	The Godfather	6024.0	8.5	8.476278	175.0
40251	Your Name.	1030.0	8.5	8.366584	106.0
12481	The Dark Knight	12269.0	8.3	8.289115	152.0
2843	Fight Club	9678.0	8.3	8.286216	139.0
292	Pulp Fiction	8670.0	8.3	8.284623	154.0
522	Schindler's List	4436.0	8.3	8.270109	195.0
23673	Whiplash	4376.0	8.3	8.269704	105.0
5481	Spirited Away	3968.0	8.3	8.266628	125.0
2211	Life Is Beautiful	3643.0	8.3	8.263691	116.0
1178	The Godfather: Part II	3418.0	8.3	8.261335	200.0
1152	One Flew Over the Cuckoo's Nest	3001.0	8.3	8.256051	133.0
1176	Psycho	2405.0	8.3	8.245381	109.0
351	Forrest Gump	8147.0	8.2	8.184252	142.0
1184	Once Upon a Time in America	1104.0	8.3	8.183804	229.0
1154	The Empire Strikes Back	5998.0	8.2	8.178656	124.0
18465	The Intouchables	5410.0	8.2	8.176357	112.0
289	Leon: The Professional	4293.0	8.2	8.170276	110.0
3030	The Green Mile	4166.0	8.2	8.169381	189.0
1170	GoodFellas	3211.0	8.2	8.160414	145.0
2216	American History X	3120.0	8.2	8.159278	119.0
1161	12 Angry Men	2130.0	8.2	8.140785	96.0
9698	Howl's Moving Castle	2049.0	8.2	8.138499	119.0
2884	Princess Mononoke	2041.0	8.2	8.138264	134.0

	title	runtime	vote_average	vote_count	year	genre
0	Toy Story	81.0	7.7	5415.0	1995	animation
0	Toy Story	81.0	7.7	5415.0	1995	comedy
0	Toy Story	81.0	7.7	5415.0	1995	family
1	Jumanji	104.0	6.9	2413.0	1995	adventure
1	Jumanji	104.0	6.9	2413.0	1995	fantasy

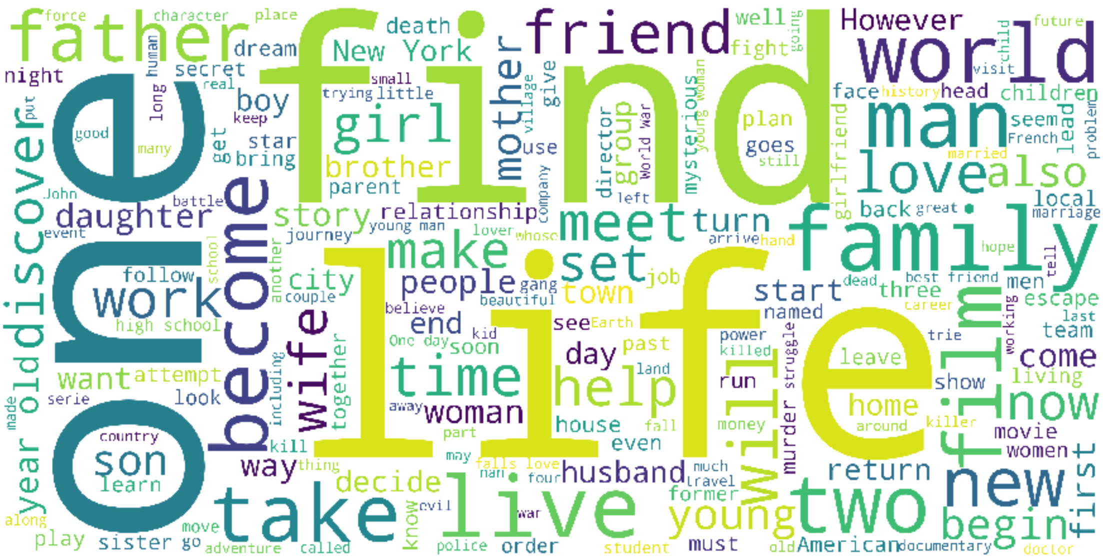
In [114]:
```
#Generate the chart for top animation movies and display top 5.
build_chart(gen_df).head()
```

```
Input preferred genre
animation
Input shortest duration
30
Input longest duration
120
Input earliest year
1990
Input latest year
2005
```

Out[114]:

	title	runtime	vote_average	vote_count	year	genre	score
9698	Howl's Moving Castle	119.0	8.2	2049.0	2004	animation	7.994823
359	The Lion King	89.0	8.0	5520.0	1994	animation	7.926672
0	Toy Story	81.0	7.7	5415.0	1995	animation	7.637500
6232	Finding Nemo	100.0	7.6	6292.0	2003	animation	7.549423
546	The Nightmare Before Christmas	76.0	7.6	2135.0	1993	animation	7.460500

Chapter 4: Building Content-Based Recommenders




```
34682    How the Lion Cub and the Turtle Sang a Song
9353                                  The Lion King 1½
9115                    The Lion King 2: Simba's Pride
42829                                             Prey
25654                                   Fearless Fagan
17041                                     African Cats
27933                 Massaï, les guerriers de la pluie
6094                                         Born Free
37409                                       Sour Grape
3203                                  The Waiting Game
Name: title, dtype: object
```

	cast	crew	id
0	[{'cast_id': 14, 'character': 'Woody (voice)',...	[{'credit_id': '52fe4284c3a36847f8024f49', 'de...	862
1	[{'cast_id': 1, 'character': 'Alan Parrish', '...	[{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...	8844
2	[{'cast_id': 2, 'character': 'Max Goldman', 'c...	[{'credit_id': '52fe466a9251416c75077a89', 'de...	15602
3	[{'cast_id': 1, 'character': "Savannah 'Vannah...	[{'credit_id': '52fe44779251416c91011acb', 'de...	31357
4	[{'cast_id': 1, 'character': 'George Banks', '...	[{'credit_id': '52fe44959251416c75039ed7', 'de...	11862

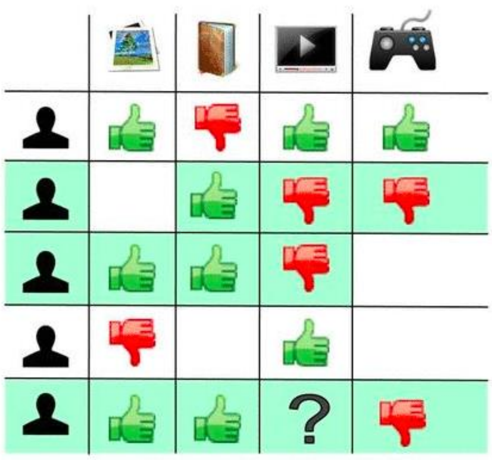
	id	keywords
0	862	[{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1	8844	[{'id': 10090, 'name': 'board game'}, {'id': 1...
2	15602	[{'id': 1495, 'name': 'fishing'}, {'id': 12392...
3	31357	[{'id': 818, 'name': 'based on novel'}, {'id':...
4	11862	[{'id': 1009, 'name': 'baby'}, {'id': 1599, 'n...

	title	genres	runtime	vote_average	vote_count	year	overview	id	cast	crew	keywords
0	Toy Story	['animation', 'comedy', 'family']	81.0	7.7	5415.0	1995	Led by Woody, Andy's toys live happily in his ...	862	[{'cast_id': 14, 'character': 'Woody (voice)',...	[{'credit_id': '52fe4284c3a36847f8024f49', 'de...	[{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1	Jumanji	['adventure', 'fantasy', 'family']	104.0	6.9	2413.0	1995	When siblings Judy and Peter discover an encha...	8844	[{'cast_id': 1, 'character': 'Alan Parrish', '...	[{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...	[{'id': 10090, 'name': 'board game'}, {'id': 1...
2	Grumpier Old Men	['romance', 'comedy']	101.0	6.5	92.0	1995	A family wedding reignites the ancient feud be...	15602	[{'cast_id': 2, 'character': 'Max Goldman', 'c...	[{'credit_id': '52fe466a9251416c75077a89', 'de...	[{'id': 1495, 'name': 'fishing'}, {'id': 12392...
3	Waiting to Exhale	['comedy', 'drama', 'romance']	127.0	6.1	34.0	1995	Cheated on, mistreated and stepped on, the wom...	31357	[{'cast_id': 1, 'character': "Savannah 'Vannah...	[{'credit_id': '52fe44779251416c91011acb', 'de...	[{'id': 818, 'name': 'based on novel'}, {'id':...
4	Father of the Bride Part II	['comedy']	106.0	5.7	173.0	1995	Just when George Banks has recovered from his ...	11862	[{'cast_id': 1, 'character': 'George Banks', '...	[{'credit_id': '52fe44959251416c75039ed7', 'de...	[{'id': 1009, 'name': 'baby'}, {'id': 1599, 'n...

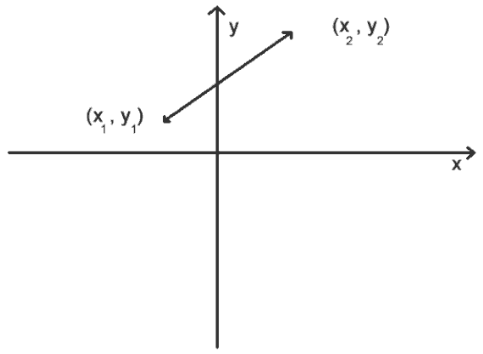
	title	cast	director	keywords	genres
0	Toy Story	[Tom Hanks, Tim Allen, Don Rickles]	John Lasseter	[jealousy, toy, boy]	[animation, comedy, family]
1	Jumanji	[Robin Williams, Jonathan Hyde, Kirsten Dunst]	Joe Johnston	[board game, disappearance, based on children'...	[adventure, fantasy, family]
2	Grumpier Old Men	[Walter Matthau, Jack Lemmon, Ann-Margret]	Howard Deutch	[fishing, best friend, duringcreditsstinger]	[romance, comedy]
3	Waiting to Exhale	[Whitney Houston, Angela Bassett, Loretta Devine]	Forest Whitaker	[based on novel, interracial relationship, sin...	[comedy, drama, romance]
4	Father of the Bride Part II	[Steve Martin, Diane Keaton, Martin Short]	Charles Shyer	[baby, midlife crisis, confidence]	[comedy]

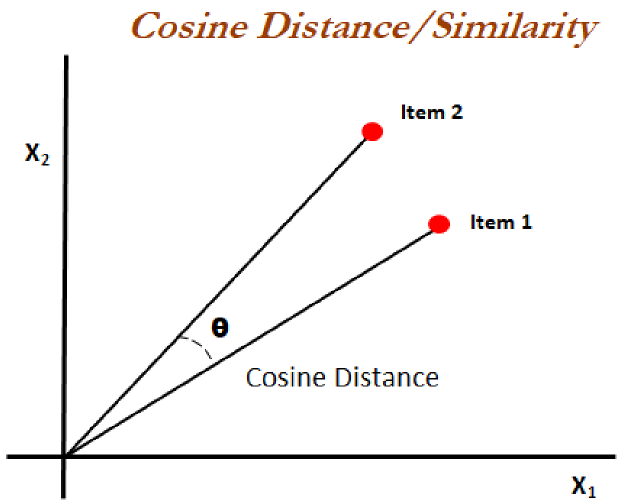
```
29607                                         Cheburashka
40904                     VeggieTales: Josh and the Big Wall
40913    VeggieTales: Minnesota Cuke and the Search for...
27768                                  The Little Matchgirl
15209                  Spiderman: The Ultimate Villain Showdown
16613                             Cirque du Soleil: Varekai
24654                                    The Seventh Brother
29198                                        Superstar Goofy
30244                                                My Love
31179                  Pokémon: Arceus and the Jewel of Life
Name: title, dtype: object
```

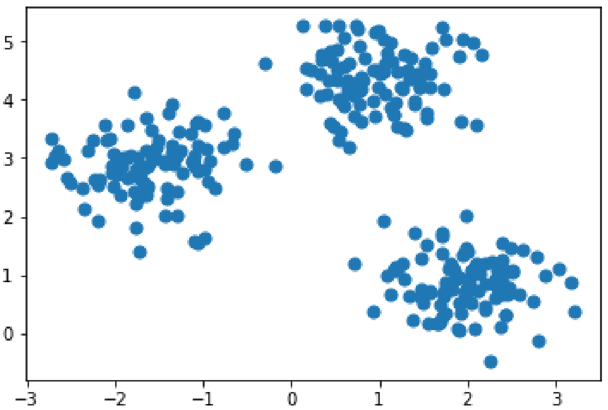
Chapter 5: Getting Started with Data Mining Techniques

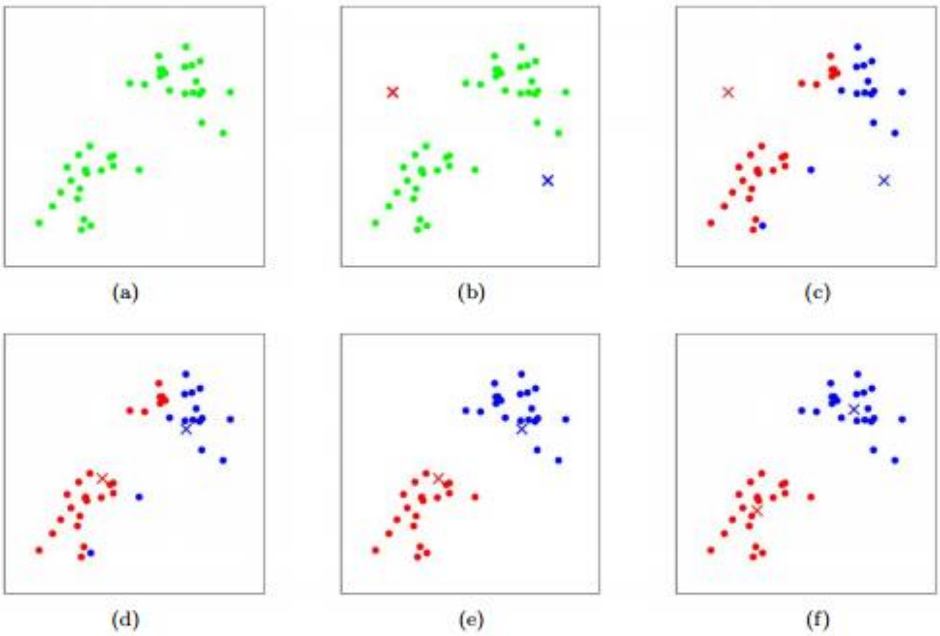


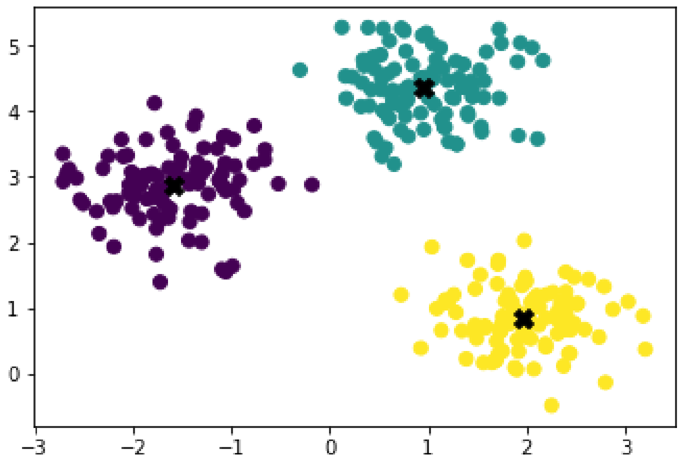
$$d=\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

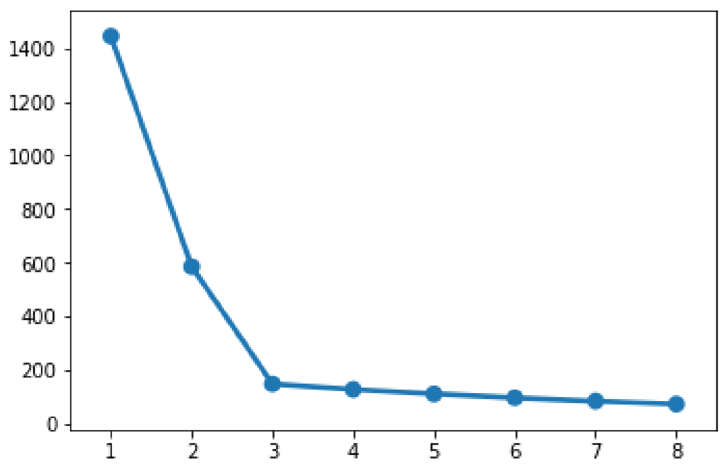


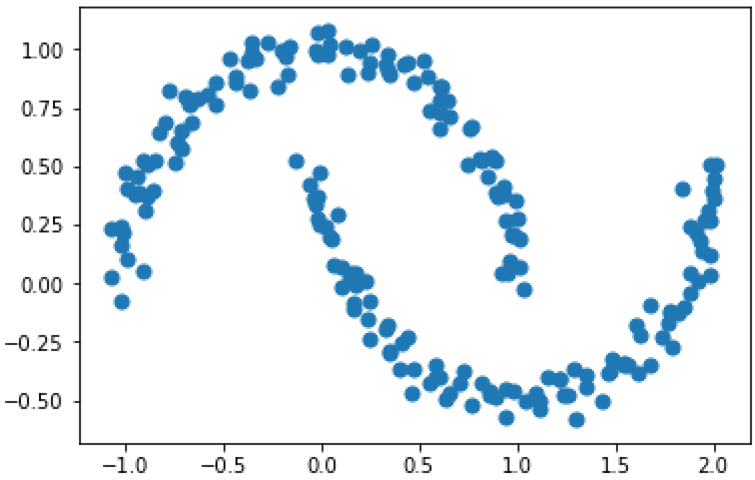




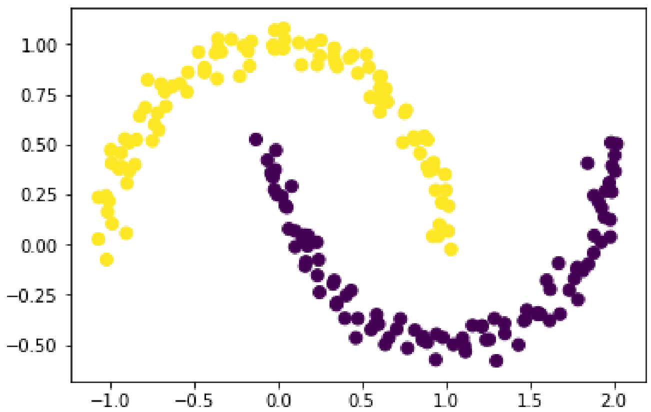

(a)
(b)
(c)
(d)
(e)
(f)


5
4
3
2
1
0
−3
−2
−1
0
1
2
3

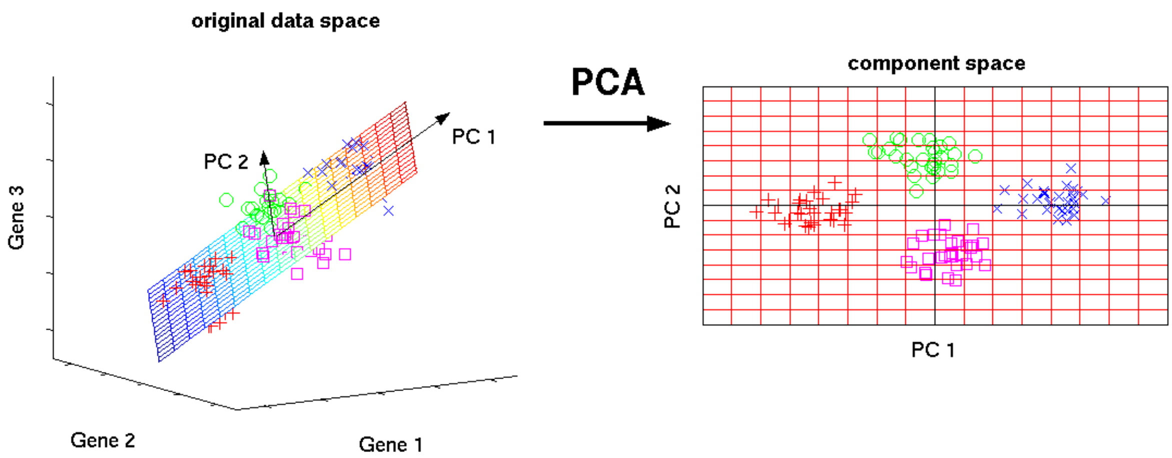

1400
1200
1000
800
600
400
200
0
1
2
3
4
5
6
7
8



1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0



1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

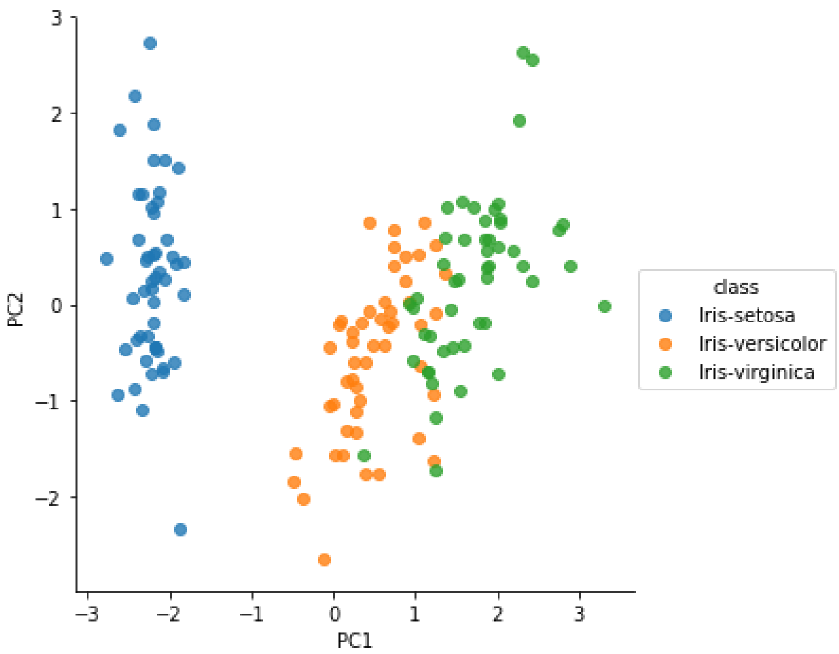


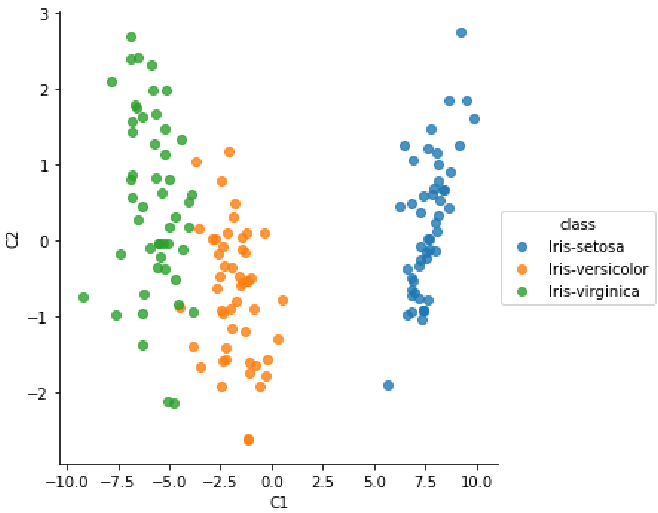
original data space
PCA
component space
PC 2
PC 1
Gene 3
Gene 2
Gene 1
PC 2
PC 1

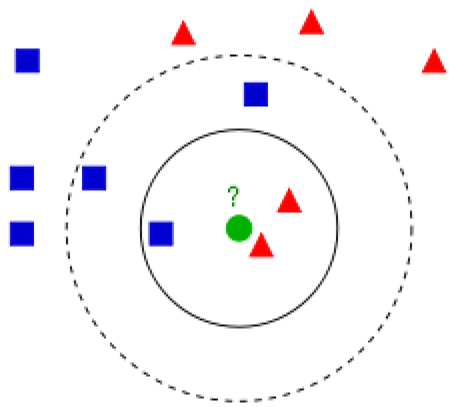
	sepal_length	sepal_width	petal_length	petal_width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

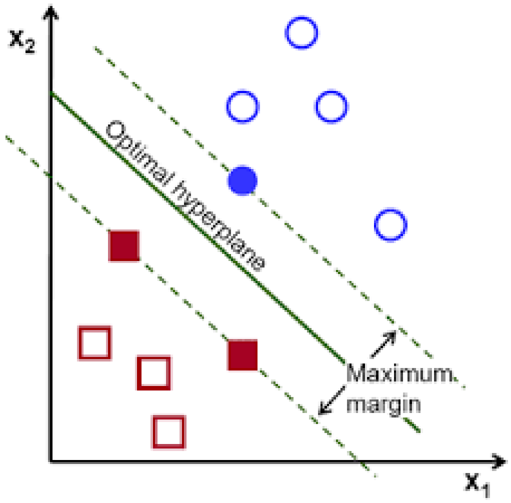
	sepal_length	sepal_width	petal_length	petal_width
0	-0.900681	1.032057	-1.341272	-1.312977
1	-1.143017	-0.124958	-1.341272	-1.312977
2	-1.385353	0.337848	-1.398138	-1.312977
3	-1.506521	0.106445	-1.284407	-1.312977
4	-1.021849	1.263460	-1.341272	-1.312977

	PC1	PC2
0	-2.264542	0.505704
1	-2.086426	-0.655405
2	-2.367950	-0.318477
3	-2.304197	-0.575368
4	-2.388777	0.674767

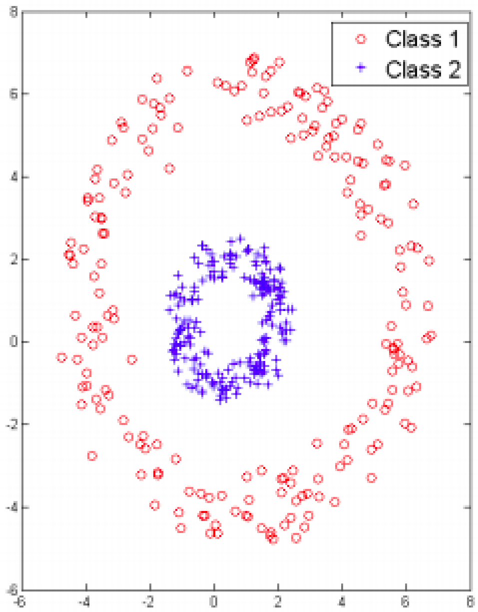

3
2
1
0
−1
−2
PC2
−3
−2
−1
0
1
2
3
PC1
class
Iris-setosa
Iris-versicolor
Iris-virginica


3
2
1
0
−1
−2
C2
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
C1
class
Iris-setosa
Iris-versicolor
Iris-virginica

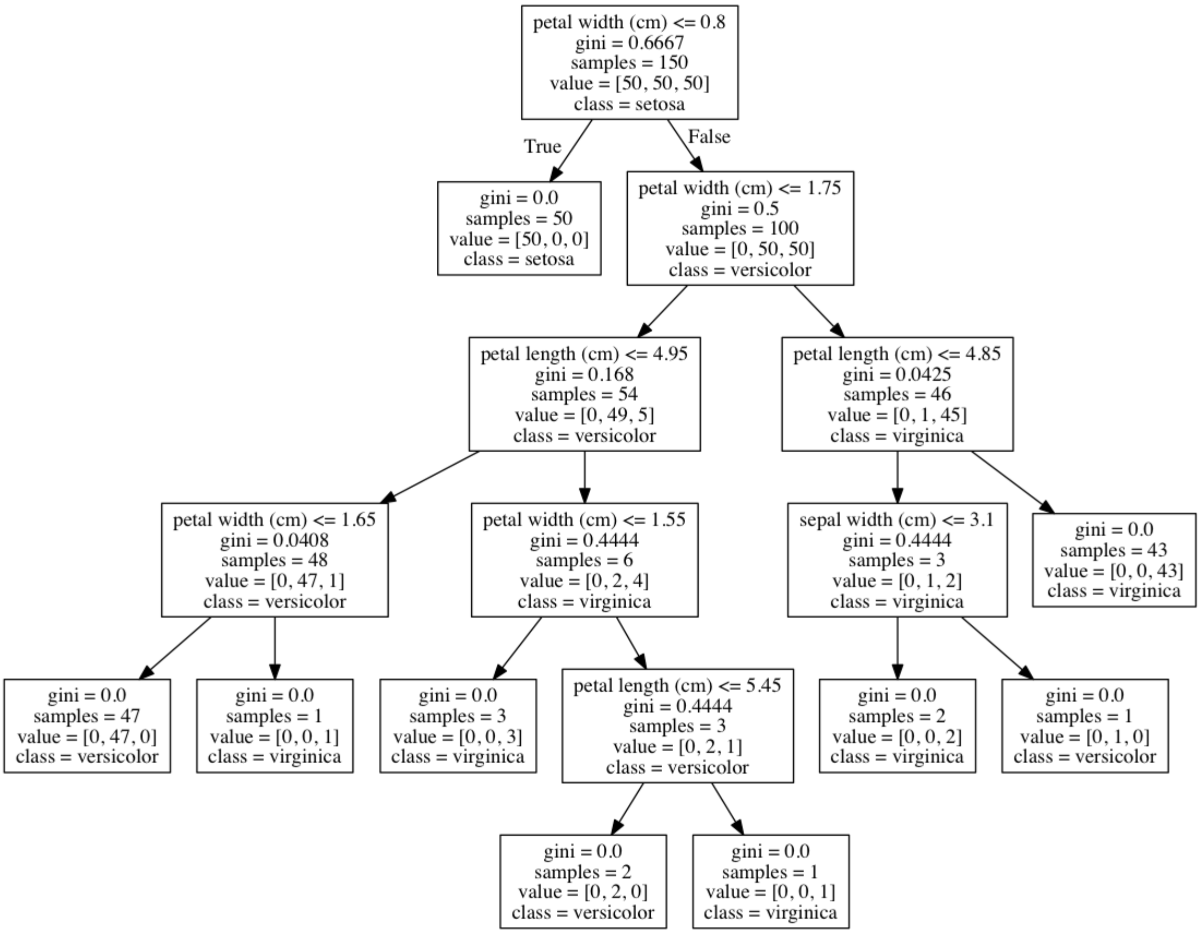

?

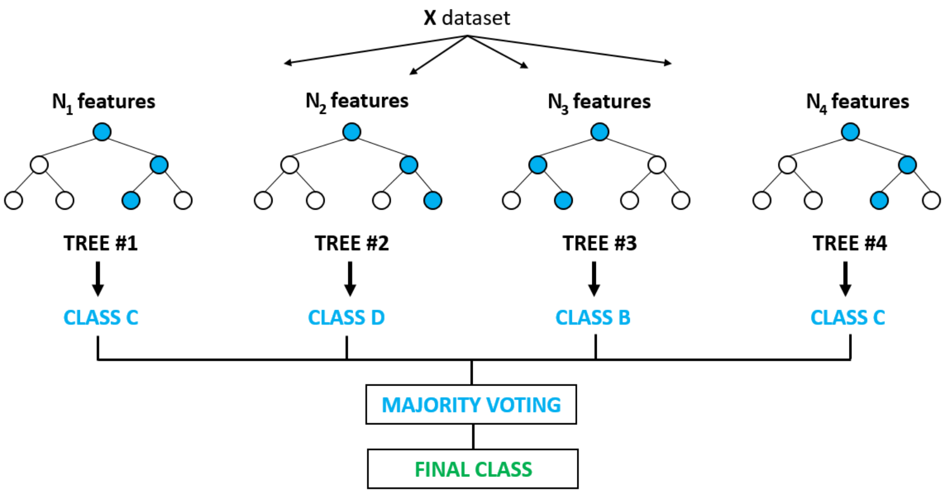


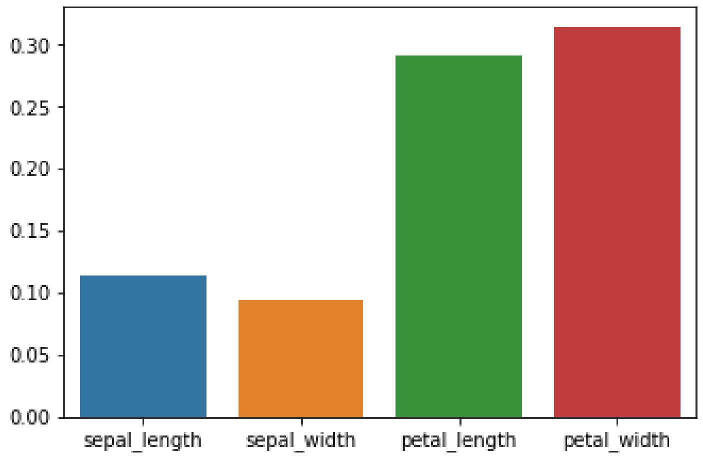
x2
Optimal hyperplane
Maximum margin
x1



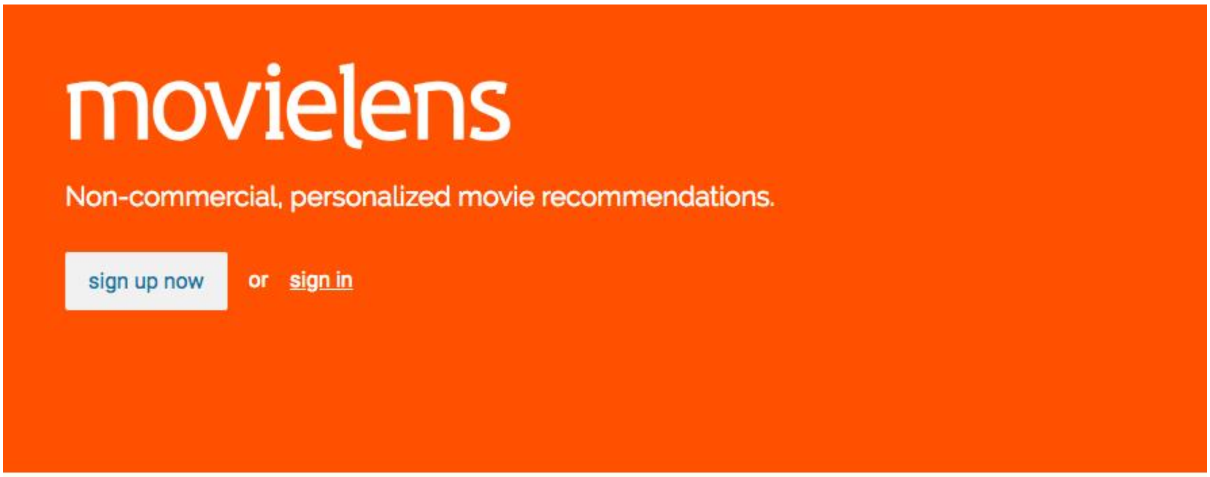
Class 1
Class 2
8
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8

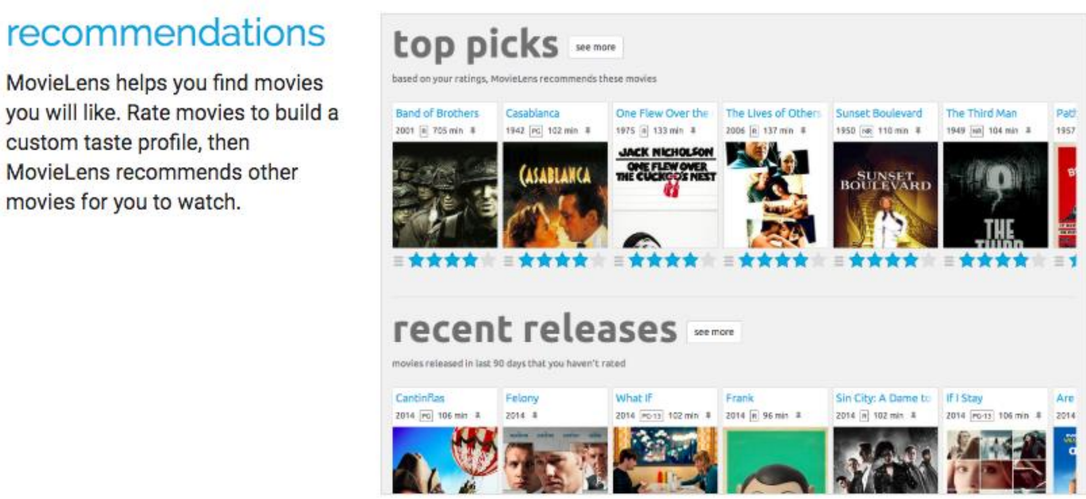

petal width (cm) <= 0.8
gini = 0.6667
samples = 150
value = [50, 50, 50]
class = setosa
True
False
gini = 0.0
samples = 50
value = [50, 0, 0]
class = setosa
petal width (cm) <= 1.75
gini = 0.5
samples = 100
value = [0, 50, 50]
class = versicolor
petal length (cm) <= 4.95
gini = 0.168
samples = 54
value = [0, 49, 5]
class = versicolor
petal length (cm) <= 4.85
gini = 0.0425
samples = 46
value = [0, 1, 45]
class = virginica
petal width (cm) <= 1.65
gini = 0.0408
samples = 48
value = [0, 47, 1]
class = versicolor
petal width (cm) <= 1.55
gini = 0.4444
samples = 6
value = [0, 2, 4]
class = virginica
sepal width (cm) <= 3.1
gini = 0.4444
samples = 3
value = [0, 1, 2]
class = virginica
gini = 0.0
samples = 43
value = [0, 0, 43]
class = virginica
gini = 0.0
samples = 47
value = [0, 47, 0]
class = versicolor
gini = 0.0
samples = 1
value = [0, 0, 1]
class = virginica
gini = 0.0
samples = 3
value = [0, 0, 3]
class = virginica
petal length (cm) <= 5.45
gini = 0.4444
samples = 3
value = [0, 2, 1]
class = versicolor
gini = 0.0
samples = 2
value = [0, 0, 2]
class = virginica
gini = 0.0
samples = 1
value = [0, 1, 0]
class = versicolor
gini = 0.0
samples = 2
value = [0, 2, 0]
class = versicolor
gini = 0.0
samples = 1
value = [0, 0, 1]
class = virginica


X dataset
N1 features
N2 features
N3 features
N4 features
TREE #1
TREE #2
TREE #3
TREE #4
CLASS C
CLASS D
CLASS B
CLASS C
MAJORITY VOTING
FINAL CLASS


0.30
0.25
0.20
0.15
0.10
0.05
0.00
sepal_length
sepal_width
petal_length
petal_width

Chapter 6: Building Collaborative Filters





	user_id	age	sex	occupation	zip_code
0	1	24	M	technician	85711
1	2	53	F	other	94043
2	3	23	M	writer	32067
3	4	24	M	technician	43537
4	5	33	F	other	15213

	movie id	movie title	release date	video release date	IMDb URL	unknown	Action	Adventure	Animation	Children's	...	Fantasy	Film-Noir	Horror	Musical	Myst
0	1	Toy Story (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Toy%20Story%2...	0	0	0	1	1	...	0	0	0	0	
1	2	GoldenEye (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?GoldenEye%20(...	0	1	1	0	0	...	0	0	0	0	
2	3	Four Rooms (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Four%20Rooms%...	0	0	0	0	0	...	0	0	0	0	
3	4	Get Shorty (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Get%20Shorty%...	0	1	0	0	0	...	0	0	0	0	
4	5	Copycat (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Copycat%20(1995)	0	0	0	0	0	...	0	0	0	0	

5 rows × 24 columns

	user_id	movie_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

movie_id	1	2	3	4	5	6	7	8	9	10	...	1669	1670	1671	1673	1674	1675	1676	1679	1681	1682
user_id																					
1	5.0	3.0	4.0	3.0	3.0	5.0	4.0	1.0	5.0	3.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

5 rows × 1647 columns

user_id	1	2	3	4	5	6	7	8	9	10	...	934	935	936	937
user_id															
1	1.000000	0.099097	0.107680	0.034279	0.152789	0.086705	0.078864	0.068940	0.092399	0.098726	...	0.259636	0.289092	0.318824	0.149105
2	0.099097	1.000000	0.252131	0.026893	0.062539	0.039767	0.089474	0.078162	0.037670	0.031866	...	0.019031	0.065417	0.055373	0.086503
3	0.107680	0.252131	1.000000	0.000000	0.045543	0.078812	0.095354	0.059498	0.053879	0.074209	...	0.050703	0.056561	0.107294	0.098892
4	0.034279	0.026893	0.000000	1.000000	0.202843	0.299619	0.163724	0.038474	0.153021	0.290192	...	0.048524	0.048312	0.022202	0.091910
5	0.152789	0.062539	0.045543	0.202843	1.000000	0.375963	0.131795	0.110944	0.400758	0.181573	...	0.080312	0.162988	0.182856	0.114262
6	0.086705	0.039767	0.078812	0.299619	0.375963	1.000000	0.211282	0.107795	0.328923	0.253871	...	0.074170	0.094619	0.084235	0.115620
7	0.078864	0.089474	0.095354	0.163724	0.131795	0.211282	1.000000	0.037040	0.183375	0.126203	...	0.066843	0.058766	0.068759	0.087159
8	0.068940	0.078162	0.059498	0.038474	0.110944	0.107795	0.037040	1.000000	0.155435	0.032419	...	0.000000	0.101710	0.034568	0.045002
9	0.092399	0.037670	0.053879	0.153021	0.400758	0.328923	0.183375	0.155435	1.000000	0.164532	...	0.049310	0.153506	0.065471	0.060088
10	0.098726	0.031866	0.074209	0.290192	0.181573	0.253871	0.126203	0.032419	0.164532	1.000000	...	0.074822	0.092575	0.098653	0.136230

	user_id	movie_id	rating	age	sex	occupation	zip_code
0	889	684	2	24	M	technician	78704
1	889	279	2	24	M	technician	78704
2	889	29	3	24	M	technician	78704
3	889	190	3	24	M	technician	78704
4	889	232	3	24	M	technician	78704

Surprise

A Python scikit for recommender systems.

Home
Documentation
GitHub page
Star 1,879 Fork 373
Follow

Maintained by Nicolas Hug
Page built with Jekyll and Hyde

Overview

Surprise is a Python scikit building and analyzing recommender systems.

Surprise **was designed with the following purposes in mind**:

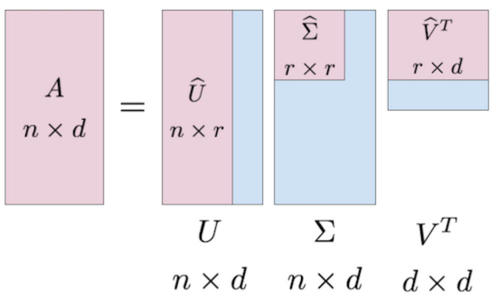
- Give users perfect control over their experiments. To this end, a strong emphasis is laid on documentation, which we have tried to make as clear and precise as possible by pointing out every detail of the algorithms.
- Alleviate the pain of Dataset handling. Users can use both *built-in* datasets (Movielens, Jester), and their own *custom* datasets.
- Provide various ready-to-use prediction algorithms such as baseline algorithms, neighborhood methods, matrix factorization-based (SVD, PMF, SVD++, NMF), and many others. Also, various similarity measures (cosine, MSD, pearson...) are built-in.
- Make it easy to implement new algorithm ideas.
- Provide tools to evaluate, analyse and compare the algorithms performance. Cross-validation procedures can be run very easily using powerful CV iterators (inspired by scikit-learn excellent tools), as well as exhaustive search over a set of parameters.

The name *SurPRISE* (roughly :)) stands for Simple Python RecommendatIon System Engine.

Getting started, example

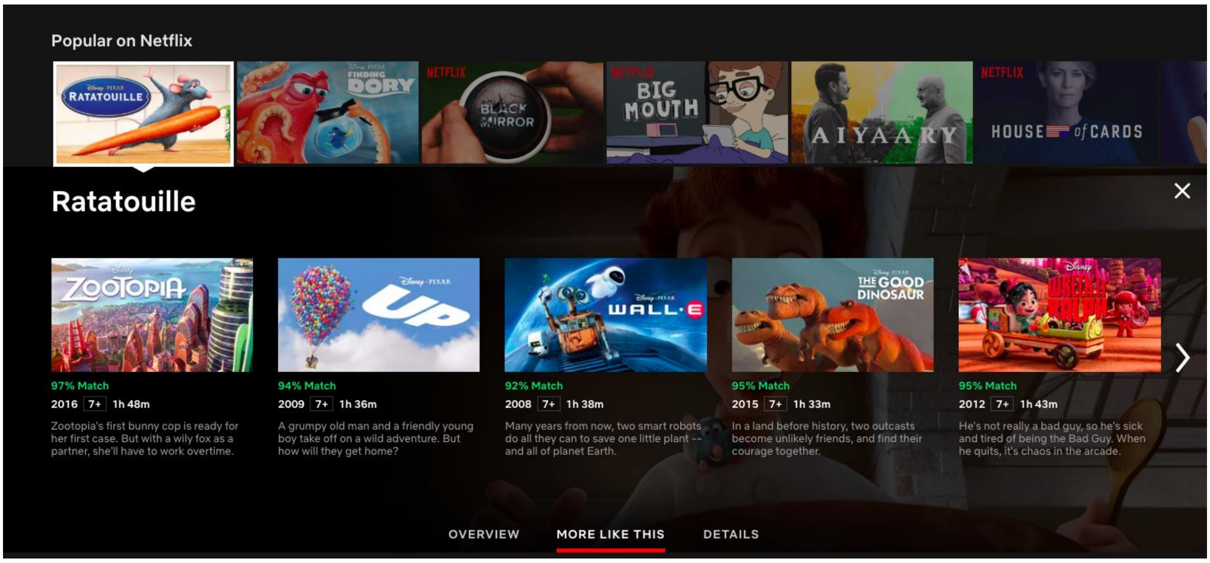
Here is a simple example showing how you can (down)load a dataset, split it for 5-fold cross-validation, and compute the MAE and RMSE of the SVD algorithm.

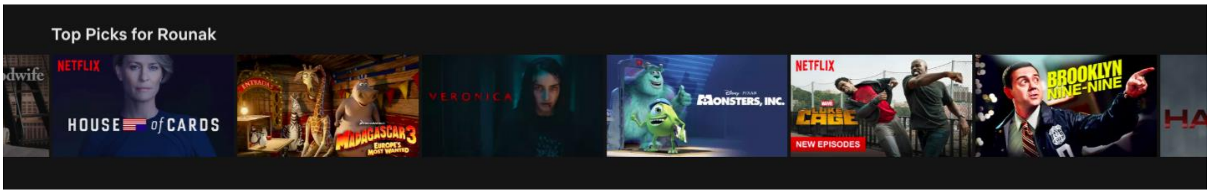
```
Evaluating RMSE of algorithm KNNBasic.

------------
Fold 1
Computing the msd similarity matrix...
Done computing similarity matrix.
RMSE: 0.9776
------------
Fold 2
Computing the msd similarity matrix...
Done computing similarity matrix.
RMSE: 0.9789
------------
Fold 3
Computing the msd similarity matrix...
Done computing similarity matrix.
RMSE: 0.9695
------------
Fold 4
Computing the msd similarity matrix...
Done computing similarity matrix.
RMSE: 0.9810
------------
Fold 5
Computing the msd similarity matrix...
Done computing similarity matrix.
RMSE: 0.9849
------------
------------
Mean RMSE: 0.9784
------------
------------
```



```
Evaluating RMSE of algorithm SVD.

------------
Fold 1
RMSE: 0.9371
------------
Fold 2
RMSE: 0.9417
------------
Fold 3
RMSE: 0.9289
------------
Fold 4
RMSE: 0.9379
------------
Fold 5
RMSE: 0.9379
------------
------------
Mean RMSE: 0.9367
------------
------------
```

Chapter7: Hybrid Recommenders





	title	vote_count	vote_average	year	id	est
1011	The Terminator	4208.0	7.4	1984	218	3.140748
974	Aliens	3282.0	7.7	1986	679	3.126947
8401	Star Trek Into Darkness	4479.0	7.4	2013	54138	3.079551
7705	Alice in Wonderland	8.0	5.4	1933	25694	3.054995
3060	Sinbad and the Eye of the Tiger	39.0	6.3	1977	11940	3.028386
8658	X-Men: Days of Future Past	6155.0	7.5	2014	127585	2.997411
2014	Fantastic Planet	140.0	7.6	1973	16306	2.957614
522	Terminator 2: Judgment Day	4274.0	7.7	1991	280	2.914548
1621	Darby O'Gill and the Little People	35.0	6.7	1959	18887	2.844940
1668	Return from Witch Mountain	38.0	5.6	1978	14822	2.804012

	title	vote_count	vote_average	year	id	est
522	Terminator 2: Judgment Day	4274.0	7.7	1991	280	3.943639
2834	Predator	2129.0	7.3	1987	106	3.866272
8401	Star Trek Into Darkness	4479.0	7.4	2013	54138	3.858491
1011	The Terminator	4208.0	7.4	1984	218	3.856029
7705	Alice in Wonderland	8.0	5.4	1933	25694	3.701565
922	The Abyss	822.0	7.1	1989	2756	3.676465
974	Aliens	3282.0	7.7	1986	679	3.672303
1621	Darby O'Gill and the Little People	35.0	6.7	1959	18887	3.628234
1668	Return from Witch Mountain	38.0	5.6	1978	14822	3.614118
2014	Fantastic Planet	140.0	7.6	1973	16306	3.602051