HomeWork 3:

Name: Malini Kottarappatt Bhaskaran

NET ID : Mxk152030

Q1:

Cluster results

```
Cluster0
 2506,Other Sister, The (1999),Comedy|Drama|Romance
 160,Congo (1995),Action|Adventure|Mystery|Sci-Fi
 1780,Ayn Rand: A Sense of Life (1997),Documentary
 1390,My Fellow Americans (1996),Comedy
 548,Terminal Velocity (1994),Action
Cluster1
 296,Pulp Fiction (1994),Crime|Drama
 50,Usual Suspects, The (1995),Crime|Thriller
 2858,American Beauty (1999),Comedy|Drama
 1617,L.A. Confidential (1997),Crime|Film-Noir|Mystery|Thriller
 527,Schindler's List (1993),Drama|War
Cluster2
 1193,One Flew Over the Cuckoo's Nest (1975),Drama
Cluster3
 344,Ace Ventura: Pet Detective (1994),Comedy 150,Apollo
 13 (1995),Drama
 368,Maverick (1994),Action|Comedy|Western
 1608,Air Force One (1997),Action|Thriller
 2054,Honey, I Shrunk the Kids(1989),Adventure|Children's|Comedy|Fantasy|Sci-
Fi
Cluster4
1127,Abyss, The (1989),Action|Adventure|Sci-Fi|Thriller
Cluster5
2455,Fly, The (1986),Horror|Sci-Fi Cluster6
1148,Wrong Trousers, The (1993),Animation|Comedy
1344,Cape Fear (1962),Film-Noir|Thriller
```

162,Crumb (1994),Documentary

52,Mighty Aphrodite (1995),Comedy

348,Bullets Over Broadway (1994),Comedy

Cluster7

2987,Who Framed Roger Rabbit? (1988),Adventure|Animation|Film-Noir
1197,Princess Bride, The (1987),Action|Adventure|Comedy|Romance

2997,Being John Malkovich (1999),Comedy Cluster8

1196,Star Wars: Episode V - The Empire Strikes
Back(1980),Action|Adventure|Drama|Sci-Fi|War

1198,Raiders of the Lost Ark (1981),Action|Adventure

1374,Star Trek: The Wrath of Khan (1982),Action|Adventure|Sci-Fi

1376,Star Trek IV: The Voyage Home (1986),Action|Adventure|Sci-Fi

541,Blade Runner (1982),Film-Noir|Sci-Fi

Cluster9

2804,Christmas Story, A (1983),Comedy|Drama

Q2:

Decision Tree: Accuracy varies based on the training data and testing data sample.

Reporting an average accuracy after running 5 times.

Accuracy: 97.8 + 91.2 + 97.6 + 92.2 + 95.04 = 94.768%

Naïve Bayes:

Accuracy: 89.7 + 82.35 +92.77+84.26 +83.13 = 86.452%

Q3:

MSE = 0.7964

References:

http://spark.apache.org/docs/latest/mllib-clustering.html#k-means

http://spark.apache.org/docs/latest/mllib-naive-bayes.html

http://spark.apache.org/docs/latest/mllib-decision-tree.html#classification

http://spark.apache.org/docs/latest/mllib-collaborative-filtering.html#examples

PorterStemmer library for the 4th one.