Okan Çiftçi - Muhammed Yusuf Fırat

Big Data

12 January 2020

Okan Çiftçi - 151805041

Muhammed Yusuf Fırat - 161805056

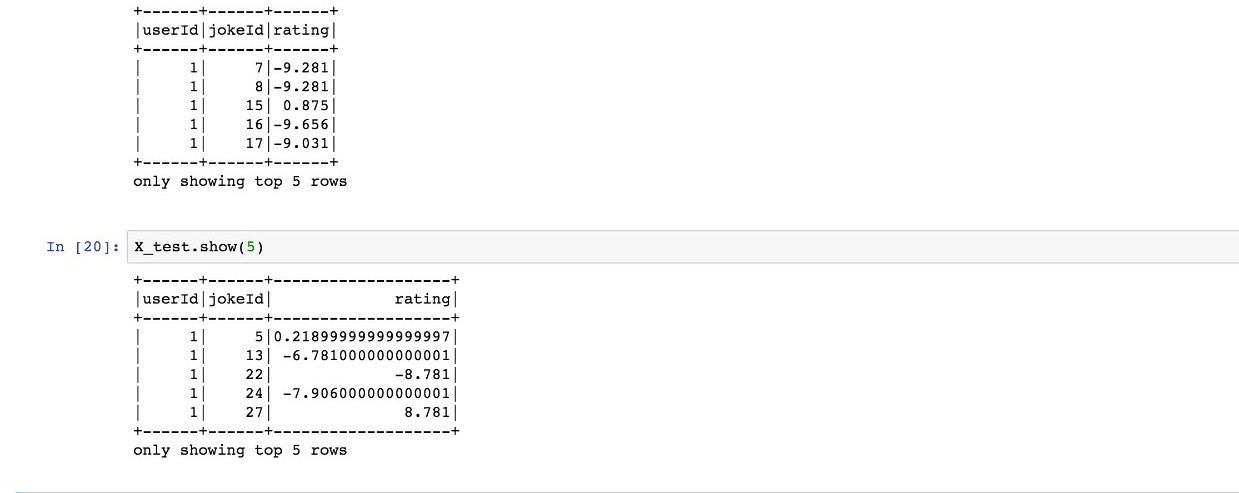
We find our dataset from berkeley.

* The dataset contains over 1.7 million continuous ratings (-10.00 to +10.00) of 150 jokes from 59,132 users.
* The dataset is collected between November 2006 - May 2009.
* The complete dataset has two CSV files:-
* jester\_ratings.csv: Each row is formatted as [User ID] [Item ID] [Rating]
* jester\_items.csv: Maps item ID's to jokes
* The ratings are real values ranging from -10.00 to +10.00.

Dataset source is <http://eigentaste.berkeley.edu/dataset/>

Rank : 10 - Iterations : 10 - Alpha : 0.01

MSE : 20.3934156829641



Rank : 50 - Iterations : 10 - Alpha : 0.01

MSE : 20.195725710425585

Rank : 200 - Iterations : 10 - Alpha(Lambda) : 0.01

MSE : 17.674535416592246

We can use last model according to MSE,our data looks like to complex according to mse we can use high rank to find patterns and we can also increase iterations number to reduce MSE value. According to higher rank give more less mse therefore we can increase rank value.