SJSU SAN JOSÉ STATE UNIVERSITY

Project Title:

Diner Recommender

Team Name:

Data Buff's

Team members:
Anjali Bandaru - 013854126
Jaya Sindhu Manda - 013852332
Kavyashree Chandrashekar - 013856518

Project Description:

This project uses dataset provided by yelp to recommend restaurants to users. Recommender systems are the major tools which helps the customer to get the best out of the available choices and to generate revenue and make the product or organization popular for any firm owner. The diner recommender systems provide one such benefit, which gives a list of restaurants based on the ratings of the previous customers. Moreover, one single recommendation system for the whole country will not be efficient as the customers has very low tendency to travel long distances for food. So, we need to come up with the recommender system which is precise to a city or a relatively practical span of area so that customers will invest his time to travel that much. The customers opinion trends can be shared with the restaurant in order to get it better and also the restaurant positive looks will be published so that right customers get right thing.

Dataset:

Name : Yelp Data set
Link : https://www.yelp.com/dataset
Size : 8.65 GB (**>100Million**)

Approach:

1. **Data preprocessing** to process the dataset for removing outliers, noise and missing values.
2. **Dimensionality reduction** to obtain a smaller set of attributes that best describe the data.
3. **Similarity measures** to come up with recommendations.
4. **Evaluate** the accuracy.

Proposed Methodology/Techniques:

1. **Singular Vector Decomposition (SVD)** will be used for dimensionality reduction which helps in analysis and matrix computations.
2. **Cosine Similarity** is to calculate the similarity between restaurants to come up with similar restaurants.
3. **Stochastic Gradient Descent (SGD)** To learn the bias and to deal the text content in review for classification.
4. **Spark** is to compute huge datasets

References:

https://www.yelp.com/dataset
https://github.com/Yelp/dataset-examples
https://scikit-learn.org
https://spark.apache.org/documentation.html