CHARLES W. DAVIDSON
COLLEGE OF ENGINEERING

SENTIMENT MINING OF AMAZON PRODUCT REVIEWS

CMPE-256 Term project

Under the Guidance of – Prof M.Eirinaki

Team Senti-Map - GitHub

Mrinmayi Gavali(012431588)

FNU Aprajita(012489074)

Priyanka Kumar(013749866)

DATASET

8.9 M Records

Size: 10 GB

22 M Records

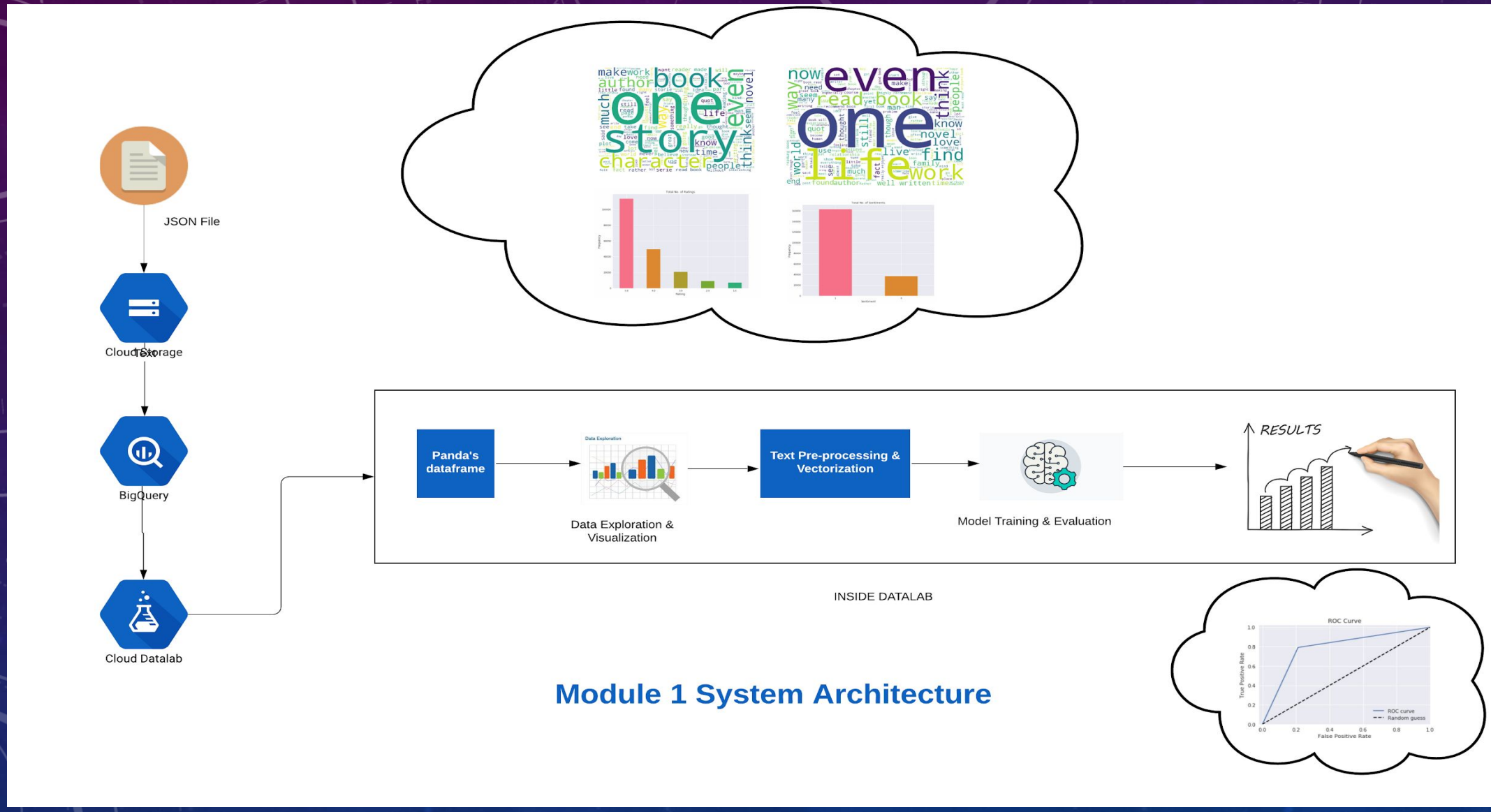
Size: 874 MB

PROBLEM SOLVED

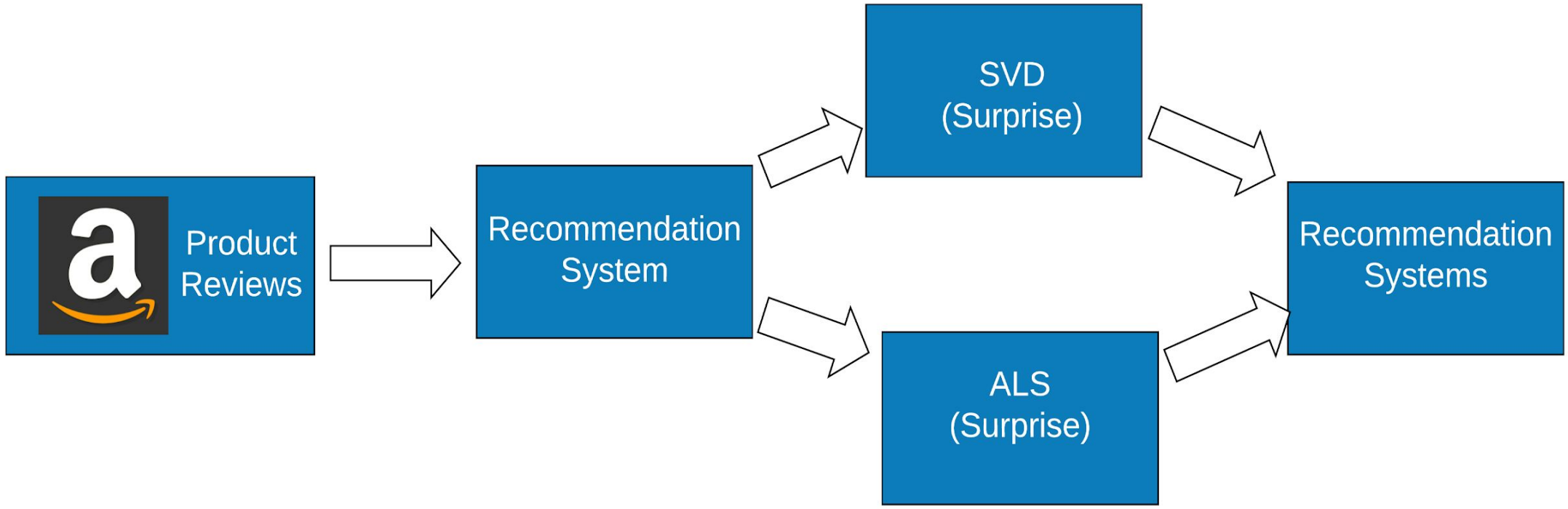
Predict User Sentiment Using Reviews

Recommend items to users

ARCHITECTURE - 1



ARCHITECTURE - 2



Module 2- System Architecture

SOLUTIONS

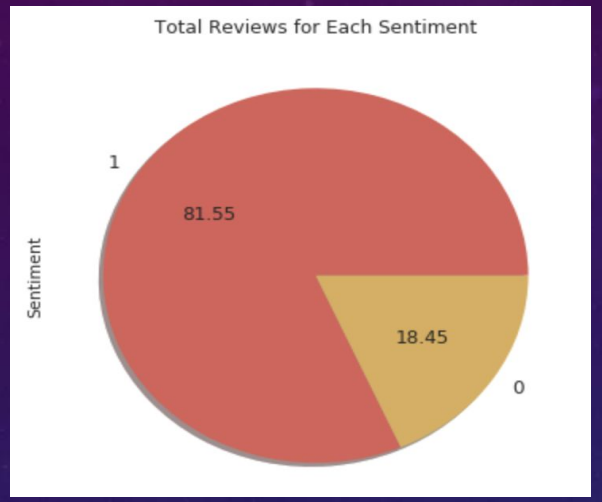
SENTIMENT ANALYSIS

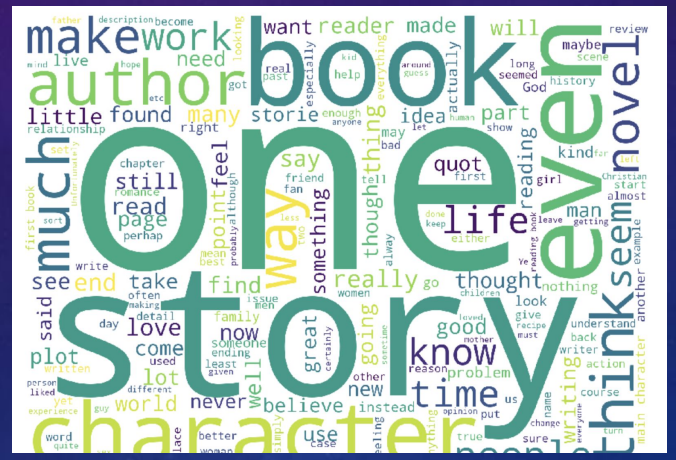
- Text- Processing
- TF-IDF
- Classification

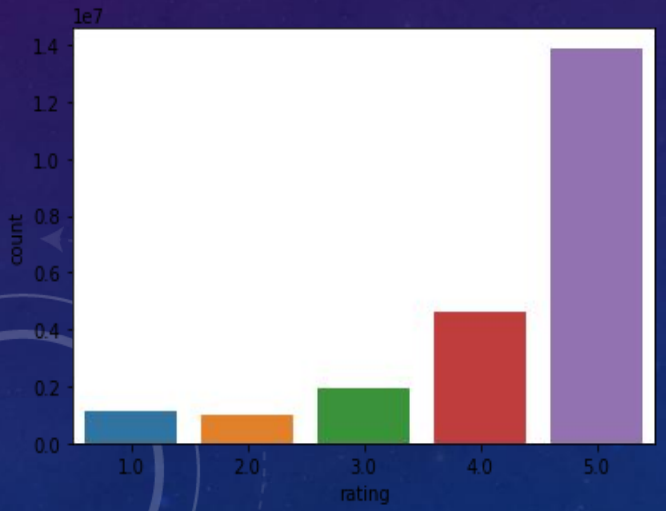
RECOMMENDER SYSTEM

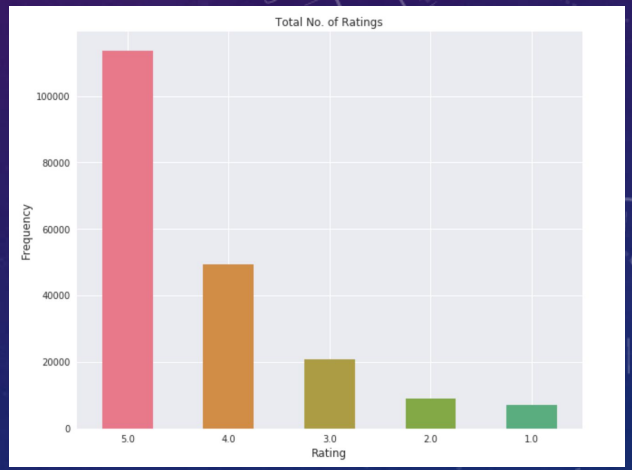
- Singular Value Decomposition
- Alternating Least Squares

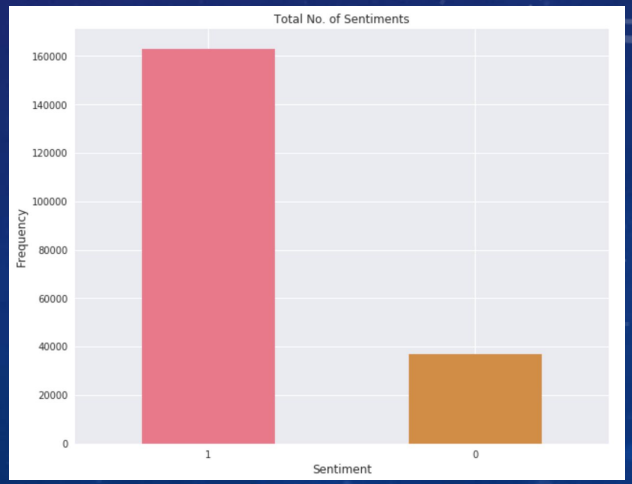
VISUALIZATIONS



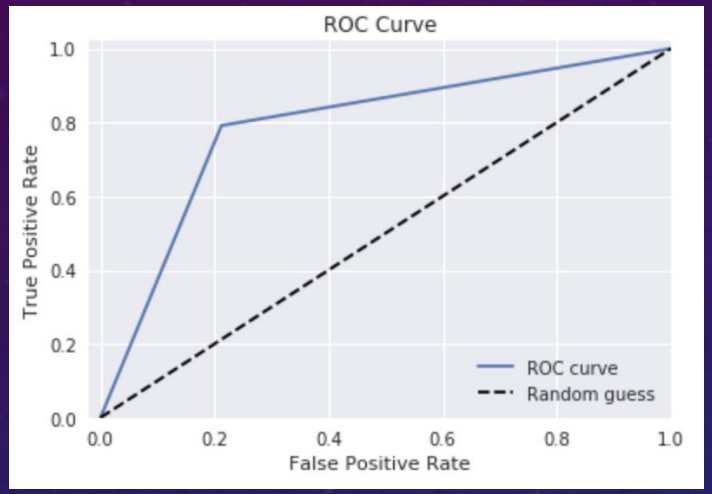




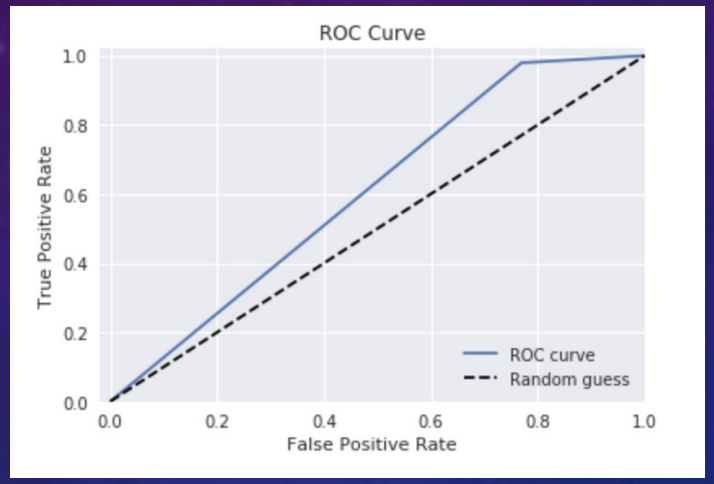




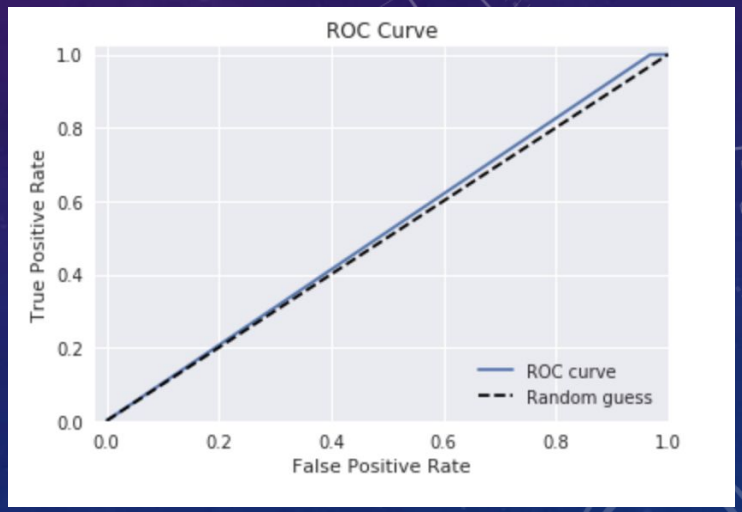
EVALUATIONS & RESULTS



ROC Curve for Logistic Regression



ROC Curve for Random Forests



ROC Curve for Multinomial Naive Bayes

CONCLUSIONS

Module 1

- Very large dataset
- Imbalanced Dataset
- Inconsistent Google Datalab session
- Data Visualization, Word Cloud, Text-Clustering, Model Training & Evaluations

Module 2

- Google Colab crashed for dataset > 1M
- Faced issues while implementing Apache Spark
- ALS & SVD both performed well for sample size of 1M
- ALS was faster on dataset size of 10M

Thank you for listening!

QUESTIONS?
ALWAYS WELCOME!