CSYE 7200: Big-Data Systems Engineering Using Scala SEC 01

Final Project Proposal

"Mobile Phone Recommender System"

Team 2

Suoge Zhu

Ke Wang

Silu Zhao

Use Cases

Case 1 (Brand navigation)

System default front page: Display popular brands based on average rating(show information including average price, average rating, review counts and **sentiment score** of review text)

User input: Customer brand preference

Output:

1. Top ten products under input brand
2. Important words extracted from review texts
3. **Rater bias** between review sentiment score and review rating (>, =, <) for input brand

Rater Bias: To check whether there is a tendency for customers to give a higher rating than the normalized review text sentiment score, or vice versa.

Use Cases

Case 2 (Price navigation)

System default front page: Display average ratings for each price range (show Information including average price, review count, and average **sentiment score** of review text)

User input: Customer price range preference

Output:

1. Top ten products under input brand
2. Important words extracted from review texts
3. **Rater bias** between review sentiment score and review rating (>, =, <) for input brand

Methodology

1. Read data from .csv file, clean and filter those attributes which are not required.

2. Use Spark to perform statistical analysis on the data, e.g. data summarization, data aggregation.

3. Use Spark Stanford NLP Library to calculate the sentiment score of each review text.

4. Use Apache Zeppelin for the visualization to display our statistical and sentiment analyses results.

Data Set

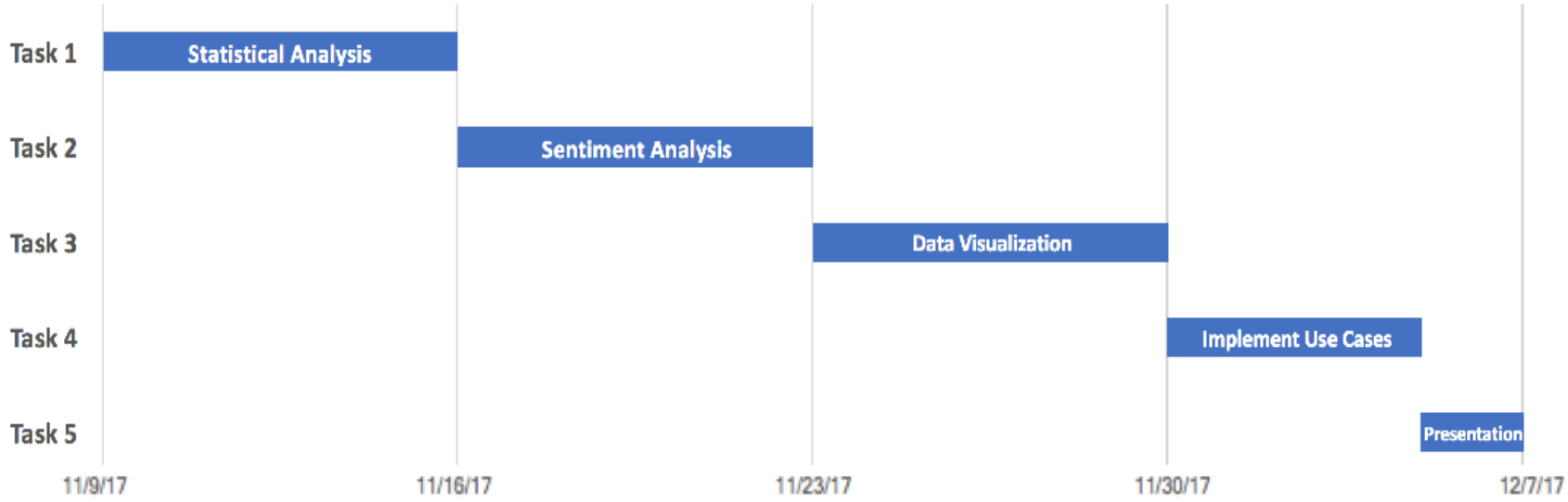
Amazon Reviews: Unlocked Mobile Phones (34.78 MB .csv file)

Description : More than 400,000 reviews from Amazon's unlocked mobile phone category
(Data was acquired in December, 2016)

Given below are the fields:

a) Product Title
b) Brand
c) Price
d) Rating
e) Review text
f) Number of people who found the review helpful

Milestones


Task 1
Statistical Analysis
Task 2
Sentiment Analysis
Task 3
Data Visualization
Task 4
Implement Use Cases
Task 5
Presentation
11/9/17
11/16/17
11/23/17
11/30/17
12/7/17

What Will We Do Using Scala & Repository

- Clean data
- Statistical analysis, e.g. compute ratings, list top 10 products
- Sentiment analysis combining with some other methodologies
- Repository: https://github.com/nolanzsg/Scala_Final_Project
  a) Source code
  b) ReadMe which will be updated frequently
  c) Presentation slides

Acceptance criteria

- Verify recommender system based on brand

  - Realize reviews numerical

  - Compare results to smartphone sales by brand in 2016

  - Similarity should be 75%

- Verify recommender system based on price

  - Compare results to smartphone sales by price range in 2016

  - Similarity should be 75%

Goals of the project

- Process Amazon Mobile Phone Reviews dataset by performing both statistical and sentiment analyses to extract useful information.
- Compare rating and review text sentiment score to analyze Rater Bias.
- Develop a recommendation system with the analysis results above.