Team Project 13
Daniel Kim (01461497)
Haasitha Pidaparthi (012669254)
Simran Singh (010219339)

CMPE 256 Project Checkpoint

Project Meeting Dates:

03/21: Initial project meeting

All three of us met and went through the links to the dataset from the provided repositories. To decide on a project idea, each team member agreed to go through the links separately to find a good dataset.

03/25: Project Proposal Draft 1

Daniel found the amazon automotive products dataset and proposed a project idea for the amazon product recommender. Three of us met to go through the dataset Daniel found and wrote the project proposal based on Daniel's proposed project idea.

04/06: Project Proposal Draft 2

The team members went through the comments provided by the professor and brainstormed the attributes in the dataset to focus on. We discussed the algorithms to use that would best correspond with the data.

04/13: Data Preprocessing

Team members met to go through the initial JSON dataset and preprocess it into a format (CSV file) that each team member could use to devise the recommender system. The initial dataset was too large (~1.7m), so we dropped some unnecessary attributes, dropped some rows (ratings older than 2016 and not verified by amazon). We mitigated the row count to approximately .9m. Haasitha was the lead while Simran and Daniel assisted in the process.

04/17: Initial Data Modeling (rough draft)

In this project meeting, we preprocessed and manipulated the dataset for our project needs. We implemented a correlation map using the extracted dataset. We also dropped some null columns. We used the Logistic Regression model to get the baseline accuracy and decided to build on that. Simran was the lead while Haasitha and Daniel assisted in the Jupyter notebook creation.

04/22: Project Checkpoint / Distributing Algorithm Choice to work on

The team members collaborated on Project Checkpoint meeting dates. Additionally, each team member chose an algorithm to train the dataset on their own. We planned on comparing the test accuracy results in the next meeting.

Contributions

Daniel Kim – Helped look for the dataset on Amazon Product Recommender. He actively contributed on giving ideas during preprocessing steps.

Haasitha Pidaparthi – Worked on the jupyter notebook for the preprocessing step. To work better on the dataset, I extracted the data from the json file into a csv. I removed the columns unnecessary for the recommender system and also reduced the data size from 1.7 million to 900,000 by removing very old or unverified reviews.

Simran Singh – Worked on the logistic regression model using the preprocessed data. Actively participated in all the meetings.