NLP & RECOMMENDATION SYSTEM

“FOR AMAZON FINE FOOD REVIEWS”

BY: Abrar & Amal

Table of Contents

Introduction

- We've search in Natural Language Processing techniques to manipulate reviews of fine foods from amazon. Then reviews will be prepared (cleaning and preprocessing) to feed the unsupervised model .In addition ,our aims to build recommendation system
- The Reviews include product and user information, ratings, and a plain text review.

Basic information about the download dataset

The data source used from Kaggle:
https://www.kaggle.com/snap/amazon-fine-food-reviews

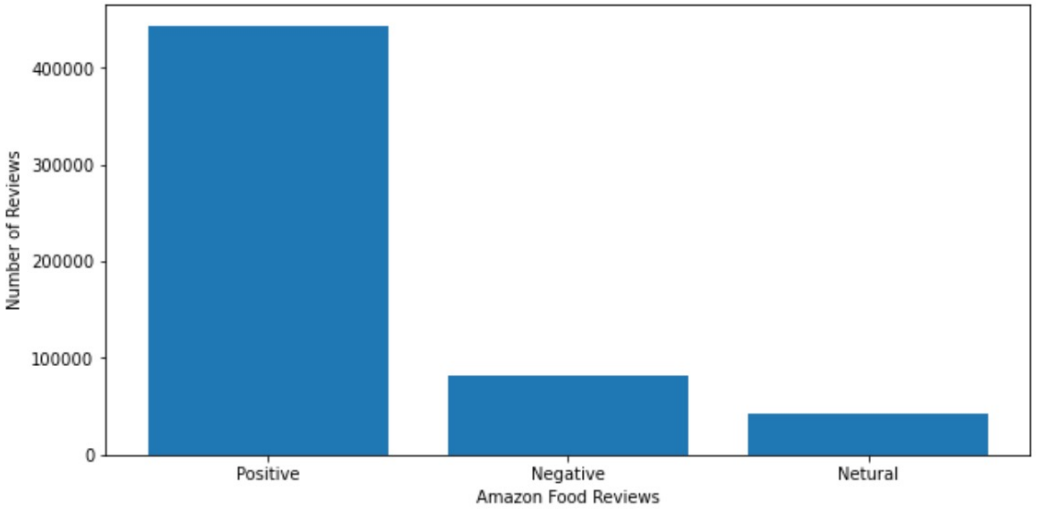
- Number of reviews: 568454
- Number of user: 256059
- Number of products: 74258
- Timespan: Oct 1999- Oct 2012
- Number of Attributes/Columns in data: 10

TOOLS

- Pandas library will be used to create data frames.
- NLTK toolkit to perform common NLP tasks.
- Sklearn library will be to implement the classification and clustering models.
- Matplotlib and Seaborn to visualize and discuss the results of the analys
- Spacy

EDA

- Remove Unneeded it columns and remove null values



Class Imbalance..

Feature Engineering

- We have engineered the features by binning technique (added column) which classifies a review to be positive if and only if the corresponding Score for the given review is 4 or 5.
- We classified a review to be negative if and only if the corresponding Score for the given review is 1 or 2.

word cloud of the most positive frequent words

Pre-processing

Data Cleaning

Remove symbols, hashtags, mentions

Spelling Correction

Correcting the misspelled words in tweet

Lemmatization

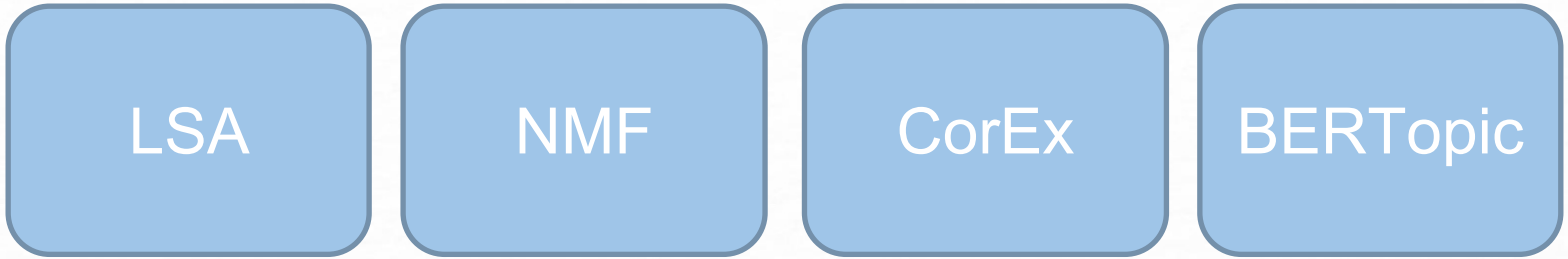
Lemmitize the review words

Vectorization

Count Vectorizer, TF-IDF Vectorizer

Topic Modelling

Trying different topic Model with different number of topic

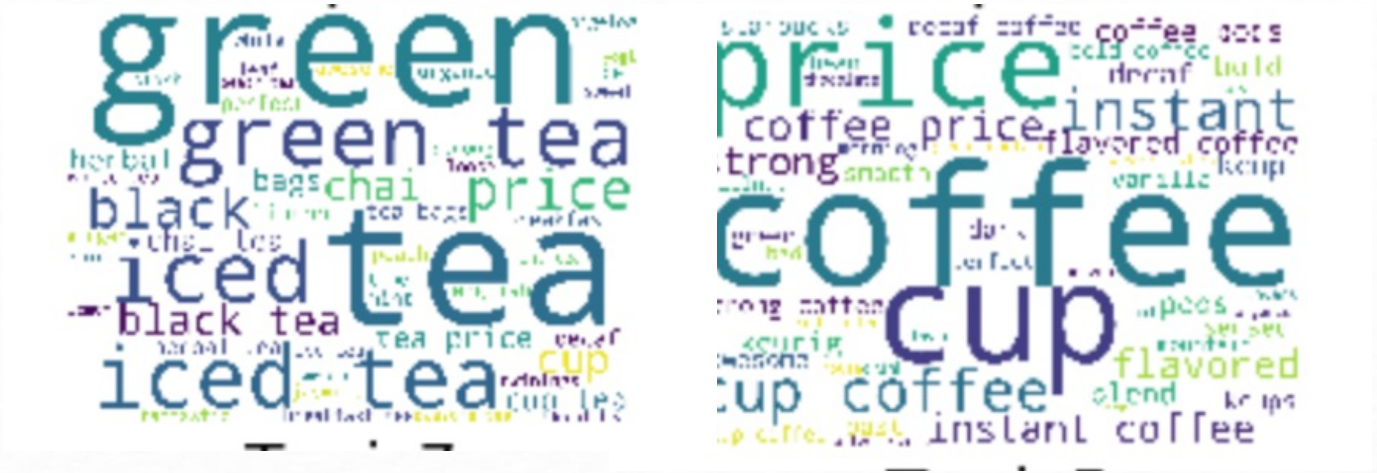


The best model is BERTopic

K-mean Clustering

Clustering is a type of unsupervised learning method of machine learning. it doesn't contain labelled output variable.

Topic Modelling



Topic:-

Tea

Coffee

Topic Modelling

Topic:- Healthy Snack Doog Foods

Recommendation System

- Sort the products on recommendation score
- Generate a recommendation rank based upon score
  #Get the top 5 recommendations, including : Get and sort the user's ratings
- sorted_user_ratings
- sorted_user_predictions
- Use popularity based recommender model to make predictions
- Add user_id column for which the recommendations are being generated
  Enter 'userID' and 'num_recommendations' for the user
  E.g :

- **The final result is :**
  Enter 'userID' and 'num_recommendations' for the user,
  userID = 121

CONCLUSION

We conclude,we achieve the satisfied result with K-means cluster, where the text is meaningful and in the nearest future our next step is applying what we have learned in nlp techniques in supervised learningn model

THANKS