kdizz / Group9Phase4

Code Issues Pull requests 1 Actions Projects Wiki Security Insights

Final Project for phase 4 moringa school

0 stars 2 forks 1 watching Branches Activity Tags

Public repository

main 4 Branches 0 Tags Go to file t Go to file Add file Code

kdizz more edits on README.md ac39c2e · 2 minutes ago

.ipynb_checkpoints	merged to main notebook before starting on Vec...	5 days ago
Team A	tested & added lemmatizzation and tokenization	5 days ago
Team B	changed the dataset to the correct final	last week
Team C	added feature importance graphs	2 days ago
TeamC	compiled all group members models	3 days ago
rando	deleted 2018 5-star reviews to eliminate skewen...	last week
.DS_Store	added feature importance graphs	2 days ago
GROUP 9 PHASE 4 PROJECT.pptx	deleted some slides	50 minutes ago
Main project.ipynb	reviewing against project checklist	38 minutes ago
Presentation.pdf	Powerpoint Presentation pdf	27 minutes ago
README.md	more edits on README.md	2 minutes ago
git repo code.txt	first commit	last week
reviews3.csv	changed the dataset to the correct final	last week

README

Group9Phase4 Project

Sentiment Analysis of Amazon Fine Food Reviews

Navigating the Repository

There are a total of 4 branches (main, team, teamb, teamc), and the final complete project is in the main branch. The team(a,b,c) branches were created for task allocation for pair-wise collaborations for the group members; however due to constraints, we eventually abandoned the branches and worked together on the main branch.

****

Project Introduction

The Project we are working on is based on a dataset from Amazon's Fine Food Reviews, The goal is to build a model(s) that can accurately determine sentiment from a review in text form and classify it as either positive, negative, or neutral.

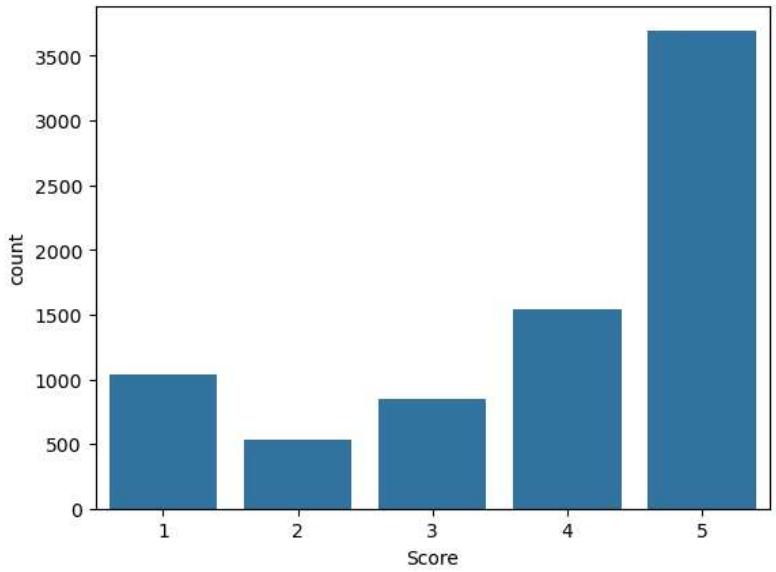
With the model(s) chosen, one can use them to rate reviews quickly and improve them for much better performance or tune them to determine the sentiment on a comment from other sources.

Analysis was conducted on;

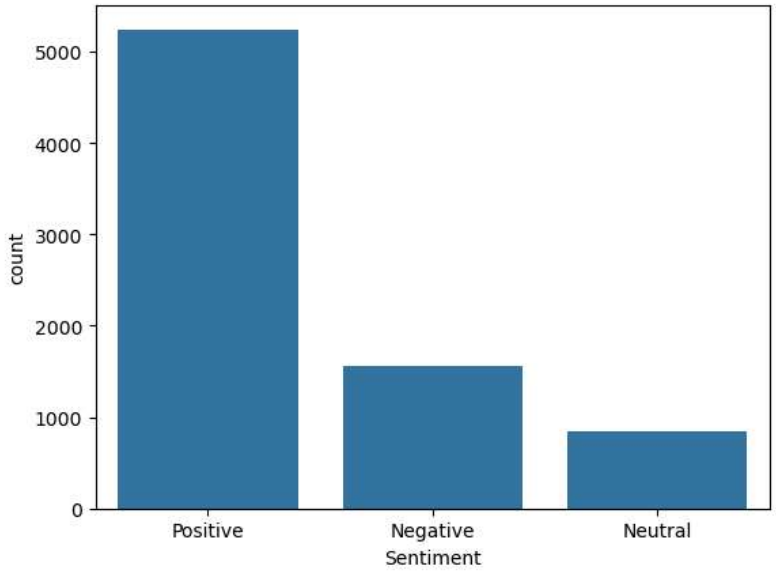
1. Reviews(Text)
2. Ratings(Scores)
3. Sentiments(Positive, Negative, Neutral)

Packages Used pandas, seaborn, matplotlib.pyplot, numpy, warnings, sklearn.linear_model(LogisticRegression), sklearn.metrics(accuracy_score, recall_score, precision_score, f1_score, classification_report, confusion_matrix), sklearn.model_selection(train_test_split), sklearn.model_selection(GridSearchCV), sklearn.ensemble(RandomForestClassifier), sklearn.preprocessing(LabelEncoder), nltk(word_tokenize, stopwords, WordNetLemmatizer, TreebankWordTokenizer), sklearn.feature_extraction.text(TfidfVectorizer,CountVectorizer), sklearn.naive_bayes(MultinomialNB), xgboost(XGBClassifier), tensorflow.keras(models,layers), transformers(DistilBertTokenizer, DistilBertForSequenceClassification, Trainer, TrainingArguments).

Insights

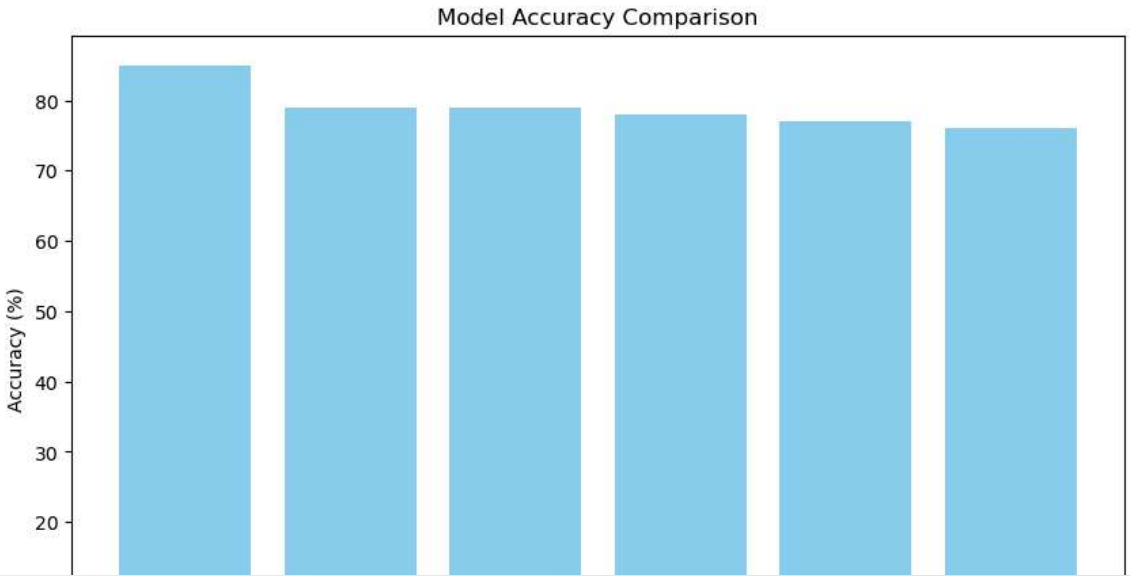


We can see the distribution of Ratings(Scores)



We can see the distribution of sentiments.

Model Performance by overall accuracy.



Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 5

Languages

- Jupyter Notebook 100.0%