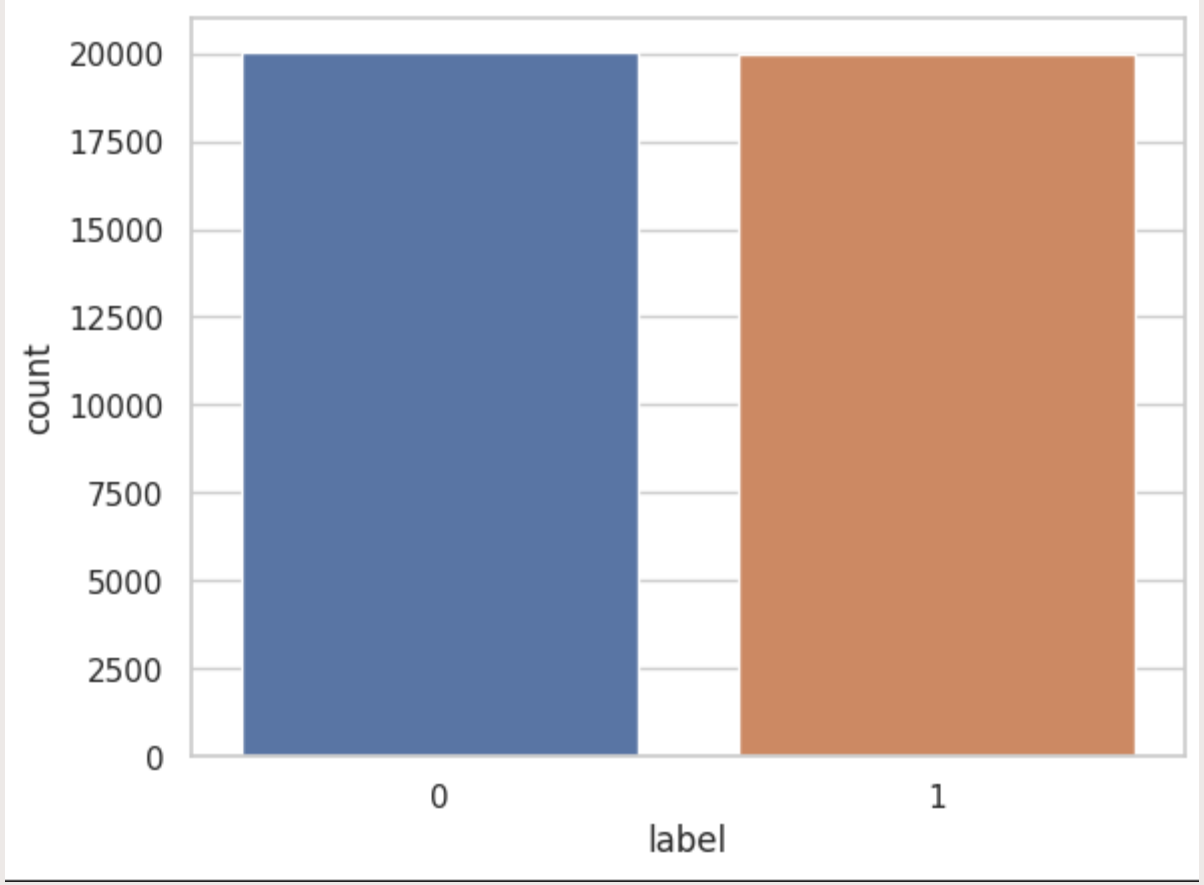
IMDB MOVIE RATINGS SENTIMENT ANALYSIS

OVERVIEW

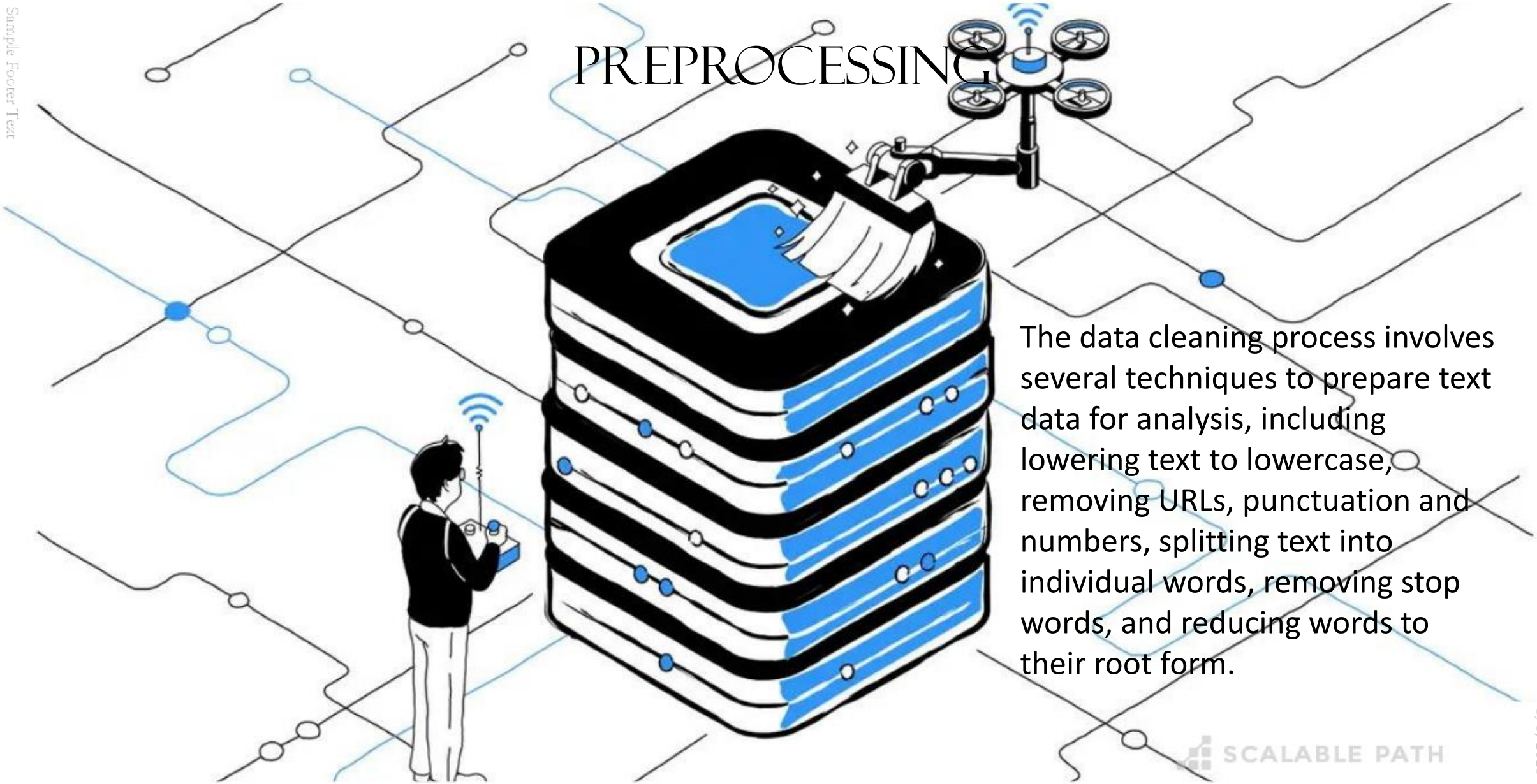
- Building a machine learning algorithm to predict whether the given review is positive or negative.
- My approach was to try three different model son the dataset and compare the results based on the metrics of F1 score, accuracy, etc.

DATA

- CSV file which included 50,000 data points.
- Features of the CSV file include:
  - *Text*
  - *Label*
- The datapoints were equally distributed and balanced.



PREPROCESSING



The data cleaning process involves several techniques to prepare text data for analysis, including lowering text to lowercase, removing URLs, punctuation and numbers, splitting text into individual words, removing stop words, and reducing words to their root form.

TRAINING

- The data points were split into 80-20 split for better training of the model and testing it upon the remaining data points.
- Three models were chosen to be worked upon for this sentiment analysis.
  - *Logistic Regression Classifier*
  - *RandomForestClassifier*
  - *XGBoost.*

PERFORMANCE

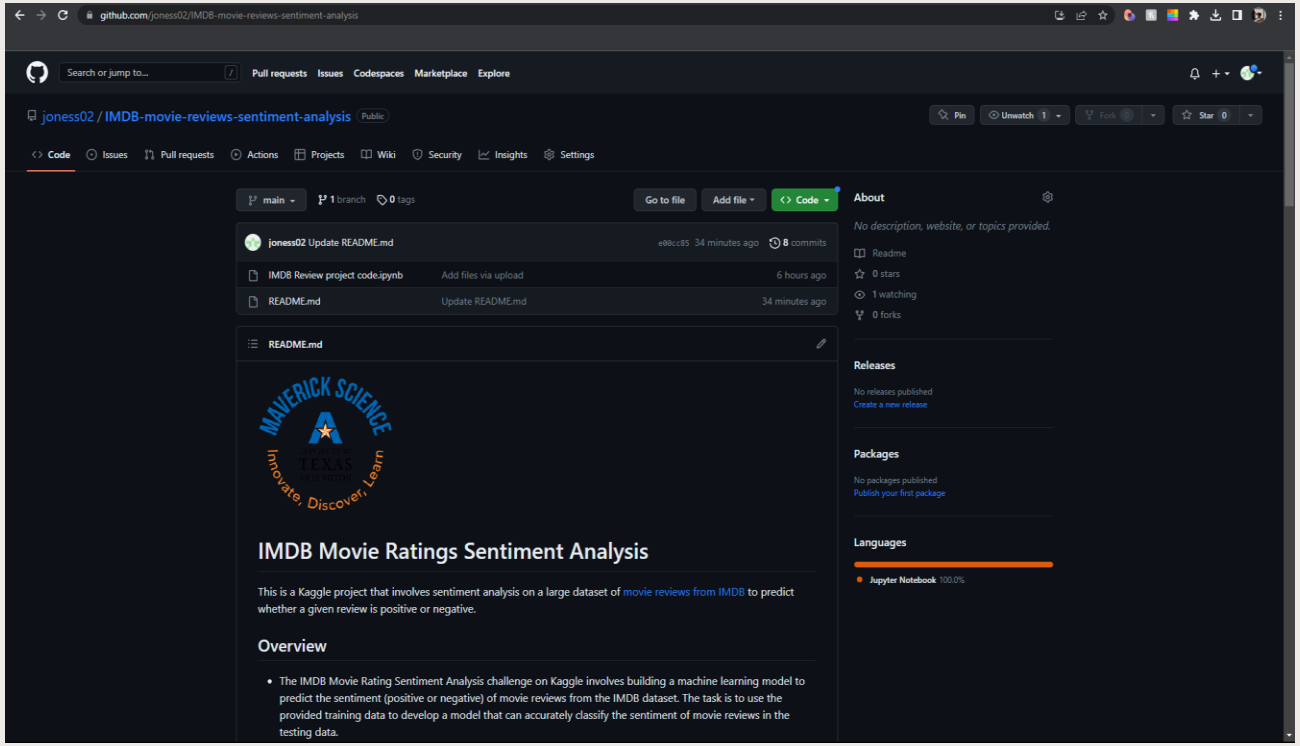
- Comparing the performance of each of the models used.
- Logistic Regression Classifier was the best out of all the three models used.
  - *With the highest accuracy, precision, recall and f1 score.*

	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.881183	0.871605	0.892767	0.882059
Random Forest	0.845437	0.847527	0.840668	0.844083
XGBoost	0.851982	0.840108	0.867729	0.853695

CONCLUSION

- As we can see the Logistic Regression Classifier was best in predicting the sentiments of the given dataset, it can be best used for the use of movie rating sentiment prediction.

REPOSITORY


github.com/joness02/IMDB-movie-reviews-sentiment-analysis
Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
joness02 / IMDB-movie-reviews-sentiment-analysis
Public
Pin
Unwatch 1
Star 0
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
main
1 branch
0 tags
Go to file
Add file
Code
joness02 Update README.md
8 commits
IMDB Review project code.ipynb
Add files via upload
6 hours ago
README.md
Update README.md
34 minutes ago
README.md
MAVERICK SCIENCE
TEXAS
Innovate, Discover, Learn
IMDB Movie Ratings Sentiment Analysis
This is a Kaggle project that involves sentiment analysis on a large dataset of movie reviews from IMDB to predict whether a given review is positive or negative.
Overview
The IMDB Movie Rating Sentiment Analysis challenge on Kaggle involves building a machine learning model to predict the sentiment (positive or negative) of movie reviews from the IMDB dataset. The task is to use the provided training data to develop a model that can accurately classify the sentiment of movie reviews in the testing data.
About
No description, website, or topics provided.
Readme
0 stars
1 watching
0 forks
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
Languages
Jupyter Notebook 100.0%