

MSBD5018: Project Results on Anime quotes Classification & Sentiment Analysis

20759401 Barrios Jorge

The script and project can be found in the next repository https://github.com/jbp1234/NLP_anime

This script utilizes quotes from Kaggle datasets and web pages scraped from Zoro.to to perform sentiment analysis by extracting key phrases related to anime titles and characters. The script includes functions to check for missing values and remove any rows that contain NaN values or duplicate quotes from the dataset.

Python libraries such as NLTK are used for text preprocessing tasks like removing stopwords, lemmatizing, and correcting spelling mistakes. TextBlob is also used to perform these tasks on the quotes in the dataset.

To build predictive models, the script splits the dataset into training and testing sets and uses scikit-learn to build several machine learning models such as logistic regression, decision tree classifier, k-nearest neighbors, and Naive Bayes. These models predict the anime title based on the quote and the character who said it. The script also evaluates the models' performance using various metrics such as accuracy, precision, recall, and F1-score.

Overall, this script is a powerful tool for performing sentiment analysis on anime quotes and predicting the corresponding anime title. By utilizing advanced text processing and machine learning techniques, it can provide valuable insights into the emotional content of popular anime shows.

Null values

```
Quote      3
Character   0
Anime      0
dtype: int64
```

Model Evaluation

Model	Training Loss	Validation Loss	Accuracy	F1 Score
No Changes	0.1439	0.1379	0.923	0.322
Removed words	0.1583	0.1123	0.912	0.464
Removed frequent tokens	0.1896	0.1786	0.944	0.389
Removed Stopwordss & tokens	0.1489	0.1189	0.955	0.437
Removed Punctuations	0.1603	0.1245	0.963	0.489

Sentimental Analysis

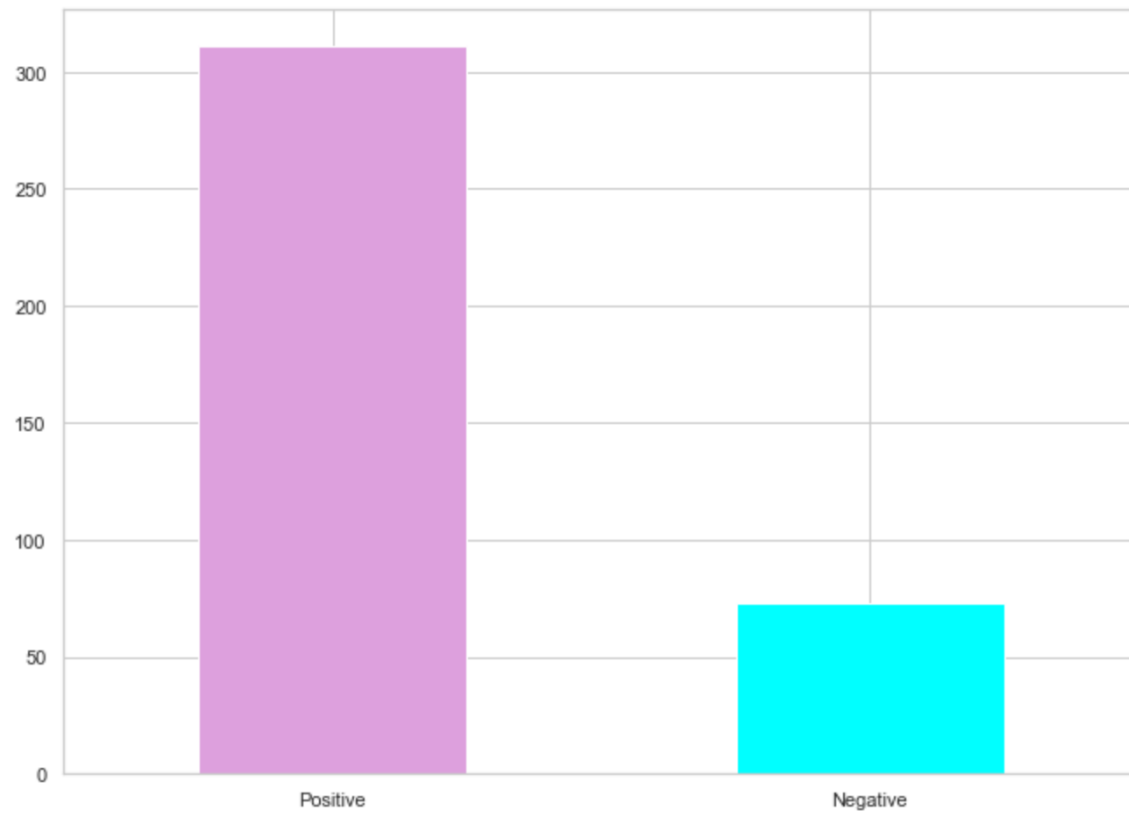
Out[41]:

	Quote	Character	Anime	tokens	frequency	naive_bayes	polarity	subjectivity
0	peopleās lives donāt end die ends lose faith	Itachi Uchiha	Naruto	[peopleās, lives, donāt, end, die, ends, lose,...	8	Monster	0.000000	0.000000
1	donāt take risks canāt create future	Monkey D Luffy	One Piece	[donāt, take, risks, canāt, create, future]	6	Trigun	0.000000	0.125000
2	donāt like destiny donāt accept	Naruto Uzumaki	Naruto	[donāt, like, destiny, donāt, accept]	5	Monster	0.000000	0.000000
3	give thatās game ends	Mitsuyoshi Anzai	Slam Dunk	[give, thatās, game, ends]	4	Haikyu	-0.400000	0.400000
4	live day die control canāand fly free	Deneil Young	Uchuu Kyoudai or Space Brothers	[live, day, die, control, canāand, fly, free]	7	Monster	0.445455	0.733333

Positive or Negative

Out[43]:

	Quote	Character	Anime	tokens	frequency	naive_bayes	polarity	subjectivity	Sentiment
0	peopleās lives donāt end die ends lose faith	Itachi Uchiha	Naruto	[peopleās, lives, donāt, end, die, ends, lose,...	8	Monster	0.000000	0.000000	Positive
1	donāt take risks canāt create future	Monkey D Luffy	One Piece	[donāt, take, risks, canāt, create, future]	6	Trigun	0.000000	0.125000	Positive
2	donāt like destiny donāt accept	Naruto Uzumaki	Naruto	[donāt, like, destiny, donāt, accept]	5	Monster	0.000000	0.000000	Positive
3	give thatās game ends	Mitsuyoshi Anzai	Slam Dunk	[give, thatās, game, ends]	4	Haikyu	-0.400000	0.400000	Negative
4	live day die control canāand fly free	Deneil Young	Uchuu Kyoudai or Space Brothers	[live, day, die, control, canāand, fly, free]	7	Monster	0.445455	0.733333	Positive



Positive / Negative quotes

Out[46]:

	Quote	Character	Anime	tokens	frequency	naive_bayes	polarity	subjectivity
Sentiment								
Negative	73	73	73	73	73	73	73	73
Positive	311	311	311	311	311	311	311	311