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

BAIYES

Model	Training Loss	Validation Loss	Accuracy	F1 Score
No Changes	0.4197	0.4882	0.0714	0.0469
Removed words (164)	0.4106	0.4914	0.6815	0.485
Removed frequent tokens (23)	0.4896	0.4786	0.644	0.389
Removed Stopwordss & tokens	0.4489	0.4189	0.655	0.437
Removed Punctuations	0.627	0.6702	0.63	0.489

SVM

Model	Training Loss	Validation Loss	Accuracy	F1 Score
Removed frequent tokens (23)	0.567	0.4786	0.644	0.389
Removed Stopwordss & tokens	0.645	0.5189	0.455	0.08
Removed Punctuations	0.781	0.793	0.149	0.136

Sentimental Analysis

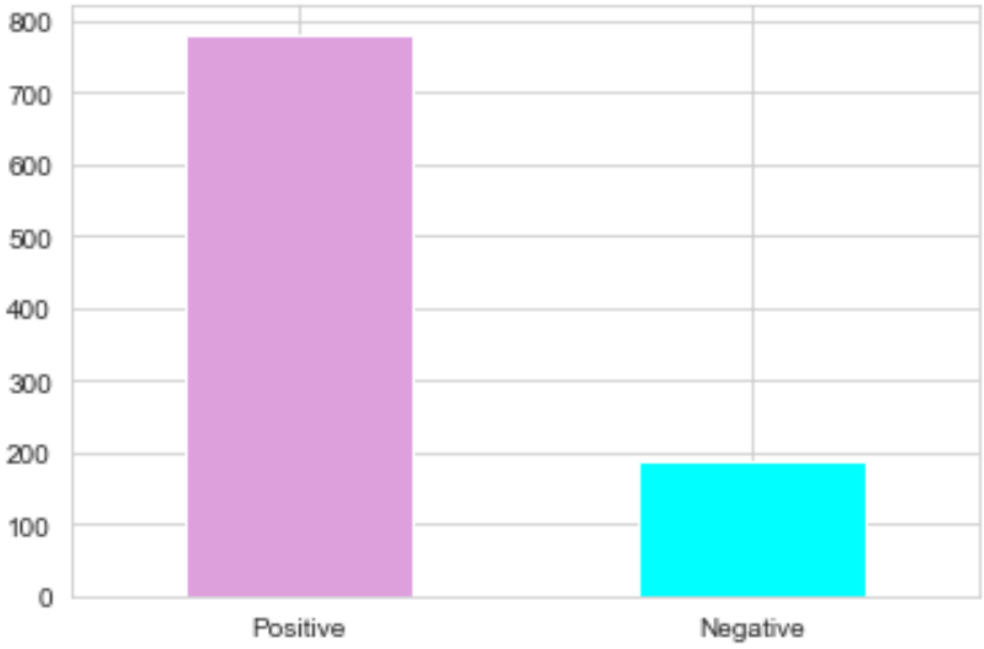
Out[41]:

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

Positive or Negative

Out[43]:

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



Positive / Negative quotes

Sentiment	Quote	Character	Anime	tokens	frequency	naive_bayes	polarity	subjectivity
Negative	188	188	188	188	188	188	188	188
Positive	782	782	782	782	782	782	782	782