

Sentiment Analysis API

Preprocessing and cleaning

- 1) Lowered case all the text
- 2) Removed all hashtags, stop words, extra space, digits
- 3) Removed punctuations and lemmatizing text
- 4) Changed positive sentiment 1 and negative sentiment to 0.

Model tried:

1. Logistic Regression
2. RandomForestClassifier
3. BiLSTM (bidirectional LSTM)

Logistic Regression

testing accuracy: 0.91

Performed hyper tuning of parameters using GridSearchCV. Tried parameters like 'penalty', learning rate, and 'solver' etc.

RandomForestClassifier

testing accuracy: 0.90

Performed hyper tuning of parameters using GridSearchCV. Tried parameters like 'bootstrap', 'max_depth', 'max_features', 'min_samples_leaf', 'min_samples_split', 'n_estimators' etc.

BiLSTM (bidirectional LSTM)

- training accuracy: 0.98
- validation accuracy: 0.91
- testing accuracy: 0.89

Performed hyper tuning of parameters. Tried parameters like different neural levels, hidden unit levels, learning rate, activation function, optimizers, dropout, regularizers, batch normalization, etc.

Deployed this model on Heroku.

All the tries tested models and methods are committed to file **Sentiment-Analysis.ipynb** in git repository with all metric and chart.