Classification for Polarity Score of Ukraine War's News

Group 1 / 108048110 郭玫均、108070023 李宜庭、108070015 張為淳

We decided only to use polarity as our target of classification since we have found out that our source data has similar subjectivity value, caused by the incorrect data source we use; we'll improve this part in the following process of our project.

For the reasons mentioned above, we also choose another online retail dataset to present a more complete and detailed results. Therefore, we'll submit another program file, which includes some extra visualizations and techniques.

In this report, we'll focus on the process and explanation of classification for our Ukraine War's news data.

The following is the processing steps:

1. Feature Engineering: Label the polarity score and
2. Spliting Data: Split data into train, test set, and fit the data into label encoder.
3. Label Encoder: Fit the training data into label encoder, that is, turn target variable (negative, positive) into 0,1.
4. TFIDF: Convert TFIDF vector, and compute TFIDF from training set.
5. Classification: Using decision tree algorithm as classifier. Fit training set on the classifier and predict labels on testing dataset.
6. Precision & Recall: Print out the accuracy and recall score of test data set for both negative and positive polarity.