

Twitter Sentiment Analysis

[DAB303 Final Project]

SMIT HARESHKUMAR RANA [792056]

SECTION 002

Background of Project

Sentiment analysis involves identifying the feelings, judgments, opinions, and attitudes and taking them into account as aspects of how people think. Industries are becoming interested in using textual data for semantic analysis to uncover consumer perceptions of their goods and services. Knowing the level of customer satisfaction is crucial so they can adjust their services as necessary. They attempt to extract the data from social media platforms to work on the text data. Many social media platforms, including Google Plus, Facebook, and Twitter, allow users to express their thoughts, feelings, and emotions regarding issues and events.

Website for microblogging with approximately 200 million users, Twitter is rapidly expanding among all other online social media networking sites. Twitter was founded in 2006 and is now the most popular microblogging platform. In one hour in 2017, 2 million users shared 8.3 million tweets.

Objective Proposal

Twitter users post their thoughts, emotions, and messages on their profiles in the form of tweets. A single tweet has a word limit of 140 characters. For tweet text, we use NLP techniques such as tokenization and removing stop words such as I, me, my, our, your, is, was, and so on. Natural language processing also contributes to data preprocessing by cleaning the text and removing special characters and punctuation marks.

Phase of Work

Python is being used in the implementations along with DeepNotes to facilitate tensorflow model development, data cleaning, preprocessing, transformation, and EDAs. Using the Twitter API, 1.6 million tweets were extracted to create Sentiment140 dataset (<http://help.sentiment140.com/for-students>). I pruned to 100000 records to improve model construction and sentiment analysis. I will utilise that dataset and divide it into 30% for testing and 70% for training. Following model training, we will assess the model to gauge the effectiveness of the trained model.

Expected Result

In order to overcome the difficulties in identifying the text sentiments (positive, negative) in tweets from, a sentimental analyzer employing a neural network and tensorflow must be created. We use the assessment metrics after model training to see how well the model predicts the upcoming. To assess the effectiveness of the models, we'll utilise the evaluation metrics listed below:

- Accuracy
- Confusion matrix with plot
- ROC Curve

Duration: 4 Weeks