Name: Tanvin Nova

ID: 172-115-026

Batch: 44 (D)

Department of Computer Science and Engineering

Metroplitan University, Sylhet.

Name: Tanjima Anjum Prianka

ID: 181-115-129

Batch: 44 (D)

Department of Computer Science and Engineering

Metroplitan University, Sylhet.

Twitter Sentiment Analysis

Abstract: Sentiment analysis on twitter data was examined in this project. The purpose of this project was to (1) learn about natural language processing (2) working experience on Python natural language processing libraries.

Introduction: Sentiment analysis involves natural language processing because it deals with human-written text. The capacity to classify conclusions communicated within the content of tweets—and particularly to decide whether the writer's demeanor is positive, negative, or neutral —is highly valuable. In this direct, we'll utilize the method known as assumption examination to classify the conclusions of individuals on Twitter towards a speculative point called #hashtag. There are diverse ordinal scales utilized to classify tweets. A five-point ordinal scale incorporates five categories: Exceedingly Negative, Marginally Negative, Impartial, Slightly Positive, and Profoundly Positive. A three-point ordinal scale includes Negative, Impartial, and Positive; and a two-point ordinal scale incorporates Negative and Positive. In this direct, we are going utilize a three-point ordinal scale to classify tweets with #hashtag.

Background study:

Dataset: Sentiment analysis involves natural language processing because it deals with human-written text. For the purpose of training the model a twitter dataset containing 1.6 million tweets was used. This dataset encoded the target variable with a 3-point ordinal scale: 0 = negative, 2 = neutral, 4 = positive.

Here is a link of our dataset

https://drive.google.com/folderview?id=1R5tM6tDczrI7ELEN9s56sdAgckV-6sa4

Pre-processing: This is a combination of steps where data are filtered to make them ready to use.

- Letter casing: All the letters were converted to lower case.
- Tokenizing: Transforming the tweets into tokens. Tokens are words isolated by spaces in a sentence.
- Eliminating noise: Eliminating unwanted characters, such as HTML tags, punctuation marks, special characters, white spaces etc.
- Stopword removal: Some words do not contribute much to the machine learning model, so it's good to remove them. A list of stopwords can be defined by the nltk library, or it can be business-specific.
- Normalization: Normalization generally refers to a series of related tasks meant to put all text on the same level. Converting text to lower case, removing special characters, and removing stopwords will remove basic inconsistencies. Normalization improves text matching.

- Stemming: Eliminating affixes (circumfixes, suffixes, prefixes, infixes) from a word in order to obtain a word stem. Porter Stemmer is the most widely used technique because it is very fast. Generally, stemming chops off end of the word, and mostly it works fine i.e., working->work
- Lemmatization: The goal is same as with stemming, but stemming a word sometimes loses the actual meaning of the word. Lemmatization usually refers to doing things properly using vocabulary and morphological analysis of words. It returns the base or dictionary form of a word, also known as the *lemma* i.e., better->good.
- Vectorizing Data: Vectorizing is the process to convert tokens to numbers. It is an important step because the machine learning algorithm works with numbers and not text. Vectorization using tf-df was used in this project.

Prediction: Naïve bias and logistic regression algorithm was used for prediction.

Accuracy:

- Naïve bias: 76%
- Logistic regression 79%

Conclusion: This project was a basic implementation of natural language processing using Python NLTK library. There is more efficient technique available that may give higher accuracy.

References:

Sentiment Analysis Explained. (n.d.). Retrieved December 27, 2020, from https://www.lexalytics.com/technology/sentiment-analysis

How to Do Twitter Sentiment Analysis with Machine Learning. (2019, June 07). Retrieved from https://monkeylearn.com/blog/sentiment-analysis-of-twitter/

Wikiwand. 2020. Sentiment Analysis | Wikiwand. [online] Available at: <https://www.wikiwand.com/en/Sentiment_analysis> [Accessed 27 December 2020].