

USING MACHINE LEARNING TO TACKLE NEGATIVE SOCIAL MEDIA POSTS.

Lateefa Tihamiyu
DECEMBER 2019

ABSTRACT

There is so much to love about social media, but we can agree social media bullying is on the rise.

The purpose of this project is to tackle this problem by identifying what combination of words make up negative comments on social media platforms and creating a model that can predict if a user's post is negative before it is posted online.

INTRODUCTION

Over the last decade, social media have been a part of our daily lives. Many of us have profiles on these platforms where we connect with friends, family and share pictures of our adventures. But platforms created to connect with one another are slowly turning into pits of hate.

With the use of Machine Learning algorithms, we can use patterns of negative comments used by bullies to train models that can predict if a post is negative before it is posted online.

The following section will elaborate on the approaches that will be taken in order to achieve the best possible results.

METHODOLOGY

This section will elaborate on the dataset as well as the technical skills that will be applied in this project.

Data source

The dataset that will be used for this project was collected from Kaggle. It consists of 31 962 tweets from several Twitter users. Each tweet in the dataset has been previously labelled as being a positive (0) or negative (1) statement.

Pre-processing Techniques

To transform the text into useful data features, the Natural Language Processing library in python (NLTK) will be used to pre-process the dataset. Below is a flow chart of how the pre-processing steps will be applied.

Data pre-processing steps using the python NLTK library

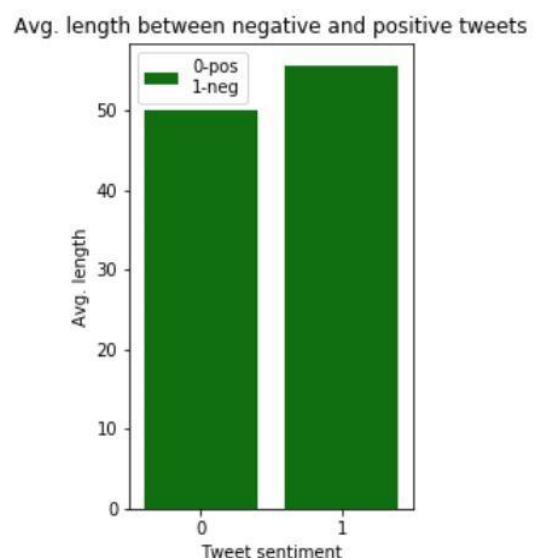
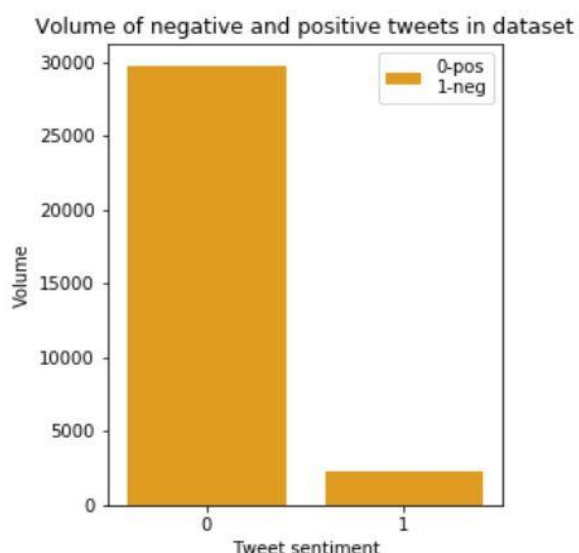


Feature Engineering

By creating a corpus, the tweets are blown out into individual tokens. The occurrence of each token in the corpus can then be vectorised, leading to the creature of new features. Some other features that can be engineered into the data include; the length of the original tweet or the average length of words in a tweet.

Exploratory Analysis

A brief exploration of the dataset shows and imbalance between the volume of negative and positive tweets in the data. This will have to be put into consideration when fitting the model. Also there isn't much difference in avg. length between positive and negative tweets.



At the end of this project, it should be possible to identify words commonly used in negative social media posts, and predict if a social media post is negative or not.