Machine Learning Models for Categorization of Tweet for suicide prevention

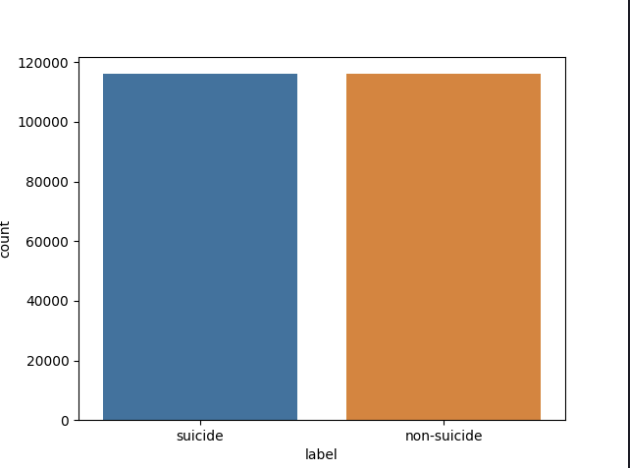
# Background/Reasoning

With the recent increase in mass shootings, and depressional problems, this Dataset seemed relevant. Often people who contemplate suicide or plan to do some type of malicious intent often use social media as a platform to express themselves. Many tech companies had contemplated the use of Machine Learning to predict depression; however, this in turn has brough controversy, as the 2015 San Bernadino shooting in which Apple™ decided not permit the US Government to access personal data in suspects iPhones. How well does Machine learning predict the possibility of extreme action, noted by people’s Tweets?

This dataset was procured from Kaggle, and uploaded by Nikhileswar Komati, and maybe localized at <https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch>.

# Data

The data consisted of 232074 totals tweets labeled as either ‘Suicide” or “Non-suicide”. The data had three attributes, none of which were null. Utilizing Seaborn Count Plot, it can be seen that the dataset is fairly distributed if not equal classes of Suicide labeled Tweets in comparison to Non- Suicide Tweets. This is classic classification of text into suicide and non-suicide labeling.



# Models

Three models were utilized to detect a binary response if the Tweets that are categorized as Suicide Watch. The models include: Logistic Regression, Naives\_Bayes, and a Neural Network Classifier all of which are OF the SKLearn API.

Their Data are as follows:

**Naives-Bay Report:**

precision recall f1-score support

0.0 0.97 0.81 0.88 23355

1.0 0.83 0.97 0.90 23060

accuracy 0.89 46415

macro avg 0.90 0.89 0.89 46415

weighted avg 0.90 0.89 0.89 46415

**Logistic Regression Report:**

precision recall f1-score support

0.0 0.93 0.94 0.94 23355

1.0 0.94 0.93 0.93 23060

accuracy 0.94 46415

macro avg 0.94 0.94 0.94 46415

weighted avg 0.94 0.94 0.94 46415

**Neural Network Classifier:**

Precision recall f1-score support

0.0 0.93 0.94 0.94 23355

1.0 0.94 0.93 0.93 23060

accuracy 0.94 46415

macro avg 0.94 0.94 0.94 46415

weighted avg 0.94 0.94 0.94 46415

It can be seen that both the Neural Network Classifier and Logistic Regression models had similar scores across all metrics. They had the highest Precision (93/ 94%) and Recall (94/93%) indicating that the model was able to predict the two classes distinctly. However, I suspect that the model is overfitting. The Navies Bayes score is not too far behind the two higher metric scoring models, the accuracy was 89%, although recall was poor for the suicide (81%) than non-suicide tweets (97%)

As an additional step I checked the length of the tweets to be compared to being suicide and not suicide. Although in general the length of the tweet was lengthier for non -suicidal Tweets, the scatter plot shows that these long tweets are outlier and there seems to be an even distribution of the length of tweet to the chances of being either suicidal or non-suicidal tweets.

