Week 8

Dataset Analysis

Team Members:

Member no.	Name	Email	Country	college	Specialization
1	Zyad Hussein	Zyadrashad262@yahoo.com	Egyptian	University of York	NLP

Problem Description:

According to a recent survey conducted by Zachary Laub, from Council of foreign relations, in 2019, online hate speech has been associated with a rise in violence against minority groups worldwide, leading to incidents such as mass shootings, lynchings, and acts of ethnic cleansing. Hate speech could appear in many forms such as a word, a sentence or a paragraph. The identification and detection of hate speech should be implemented using Deep Learning techniques as such implementation could contribute to the elimination of hate speech across social media platforms to achieve a safe space platform where consumers can interact freely, and consumers of all age groups can join the platforms. Hate speech can target consumers with different ethnicities, different backgrounds, different sex, different nationalities, different race, different colour or different ancestry which is likely to leave adverse long-term effects on the targeted group. Hence, increase of hate and crimes in communities. Detection of hate speech using deep learning techniques will yield great results as it is trained on more diverse dataset that includes most types of hate speech in most forms. alongside this, the utilization of Deep learning techniques will outperform any other techniques due to its high accuracy results and its architecture.

Data Understanding:

To implement the proposed technique, it was needed to acquire a dataset to train the chosen network to achieve hate speech prediction. The best dataset to train the network for such a task is a dataset that includes variety of hate speech in many forms such as numbers, letters, letters& numbers, special characters, words, sentences, or paragraphs. Furthermore, the analysis of the problem yielded that diversity in hate speech is required to achieve the ability of identification of all types of different hate speech. A dataset that includes twitter tweets, Facebook posts, Instagram captions or any other platform source that is accessed would wide is a perfect dataset as it ensures diversity and variety in data. Fortunately, the dataset provided consists of twitter tweets that has a sheer number of tweets unrelated to each other and from different unknown users. Ethical procedures were considered when acquiring the dataset as the usernames were all removed, and any information related to any user were all removed before publishing the dataset. Dataset was acquired from Kaggle.com.

Data Type:

The data provided consists of twitter tweets and labels of 1 or 0 which is interpreted as positive or negative. Hence, the dataset is considered as categorical dataset as tweets would fall under positive category or negative category. The dataset can also be considered descriptive as it describes forms of hate speech or normal speech.

Issues in Dataset:

Issues have also appeared when investigating the dataset such as unidentified characters as shown in figure 1 and figure 2. Alongside this, some tweets in the dataset had missing letters which could result in failing to interpret the words' meaning such as in figure 3. Moreover, some of the data acquired in the dataset are misplaced and misinterpreted as positive speech when it should be considered negative speech.

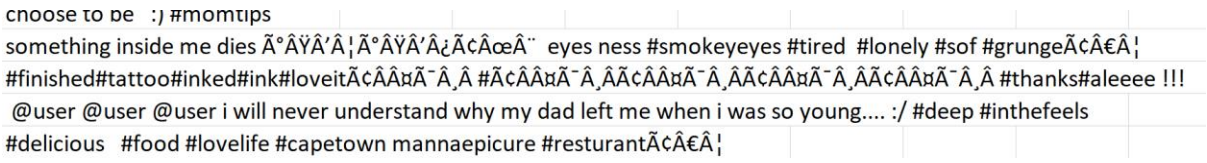

Fig. 1

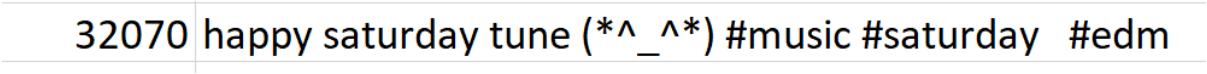

Fig. 2

Fig. 3

Approaches to overcome issues:

To overcome such issues, several approaches have been considered such as scanning the dataset to identify tweets with anomalies, noisy tweets or unidentified characters using replacing method such as in figure 4.

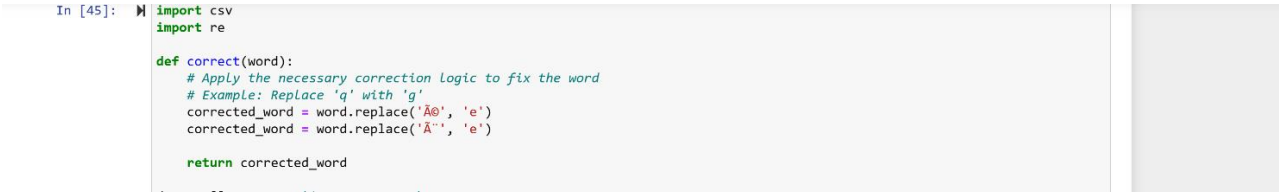

Fig. 4

Investigation is still being carried out to identify the possible solutions to words with missing letters. Some outliers have been detected as well and it is believed that these outliers can be removed from the dataset or a confidence score that can apply as a threshold to identify whether a sentence can be considered one of both categories and considered as an outlier. Another method that can be used to identify outlier is sentiment analysis of each tweet to ensure the adhering to their assigned category. There are more methods to achieve perfect preprocessing of dataset to meet training requirements and such methods are being investigated to identify the ideal one for the provided dataset.

References:

Laub, Z., 2019. Hate speech on social media: Global comparisons. *Council on foreign relations*, *7*.