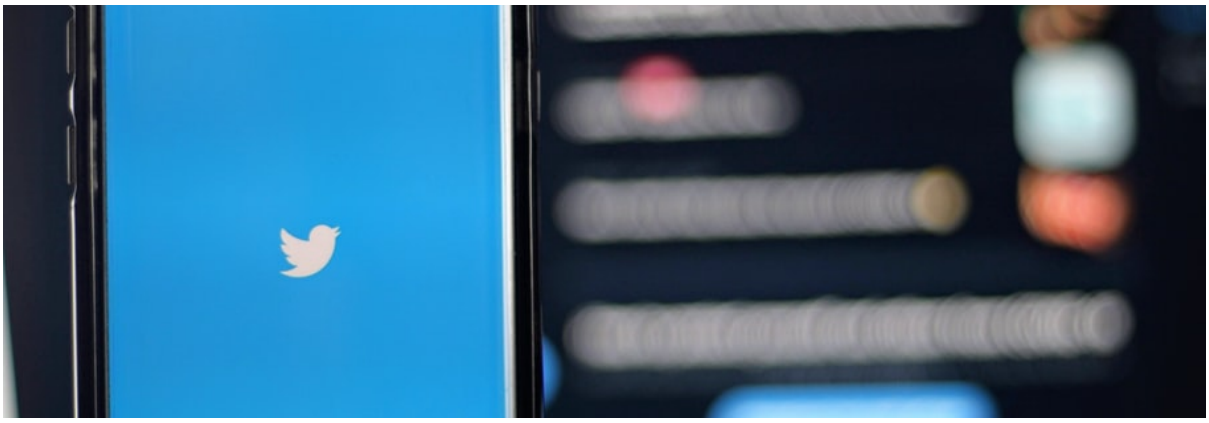
T5 Bootcamp Project Proposal
Submitted by: Raghad Abdullah Altowairgi

Detecting Hate-Speech and Cyberbullying



Cyberbullying is considered one of today's social plagues associated with the cyberspace. The invisible nature of people using social media and any micro-blogging platforms such as Twitter, makes it even easier for bullies to harass their victims. Due to that, it is required to obtain a solution to handle the abusive behavior on the cyberspace. Detecting such behavior in the cyberspace requires highly intelligent systems to process and understand the language used. However, detecting this kind of sentences is considered a challenging task due to the various morphology and dialects of one language, it gets more challenging when applied on micro-blogging data. This project aims to build a machine learning model that has the ability to classify a given text into whether it's offensive or not.

Dataset description

The first step in developing any supervised machine learning algorithm is data collection. This twitter corpus is available online for research community. The text is classified as: hate-speech, offensive language, and neither. The dataset also contains

text that can be considered racist, sexist, hurtful, or generally offensive. It should be mentioned that the dataset has more than 24K tweets and 7 columns. It can be found on Kaggle please see link attached:
https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset

Languages and Tools

The whole project will be implemented using Python 3.8.8 along with Jupyter Notebook, and existing libraires will be used such as numpy, pandas, matplotlib, seaborn, sklearn and more when needed.

Next Step

1. Perform Exploratory Data Analysis on the corpus to identify interesting insights from this it.
2. Merge related datasets, for better understanding.
3. Create a classification model that uses text data features and predict which tweet is offensive or not.
4. Evaluate the performance of the model using the standard metrics.