# detect hate speech in tweets

# **Project Motivation**

One of the major breakthroughs in internet is social media .it can allow to the people to share their view, idea on various topic. This project aims to automate content moderation to identify hate speech using machine learning binary classification algorithms.

**Question/need:**

* Are the algorithms that we will use distinguish heat of speech and non-heat of speech tweet and when detect them classified it correctly?

**Data Description:**

* **Dataset source:**

https://www.kaggle.com/mrmorj/hate-speech-and-offensive-language-dataset

* **Description of the Dataset:**
* count = number of CrowdFlower users who coded each tweet (min is 3, sometimes more users coded a tweet when judgments were determined to be unreliable by CF).
* hate\_speech = number of CF users who judged the tweet to be hate speech.
* offensive\_language = number of CF users who judged the tweet to be offensive.
* neither = number of CF users who judged the tweet to be neither offensive nor non-offensive.
* class = class label for majority of CF users where " heat of speech’’ labelled as 1
* The dataset contains tweets categorized as " heat of speech’’ labelled as 0 and ‘’ offensive\_language " labelled as 1 , and “ neither: labeled as 2 .

the labels on this dataset were voted on by crowdsource and determined by majority-rules. The “class” column labels each tweet as 0 for hate speech, 1 for offensive language or 2 for neither. I will be treating the data as a binary classification problem.

Therefore, the final model will be predicting whether a tweet is hate speech or not. To prepare the data for this, I will be manually replacing existing 1 and 2 values as 0, and replacing 0 as 1 to indicate hate speech.

* **Dataset info**
  + The dataset consists of 24783 rows and 7 columns. I am expecting on working on tweet features

Text

Description automatically generated

### **Algorithms**

* Using SVM , Decision tree and neural network to classify data
* Confusion matrix will be used to evaluated these algorithms

**Tools:**

* Panda ,Numby Sklearn, seaborn, NLP library such as nltk

**MVP Goal:**

### **Basic Data Exploration such as:**

* The shape of the dataset
* Head of the dataset
* Info of the dataset
* Summary of the dataset

## **Statistical Insight**

### **Cleaning data from URL and space**

* Using machine learning algorithm or deep learning algorithm to classified data
* Seaborn for visualizing data

Visualizations and answers after building the model to:

* Are the algorithm that we will use distinguish offensive and heat of speech tweet and classified it correctly?