Group Name: 8bit

Group Members:
Name: Prajesh Tejani
Email ID - prajeshtejani107@gmail.com / parthupatel467@gmail.com
Country - India
College - Technishe Universität Chemnitz
Specialization - NLP

Name - Jigar Borad
Email Id - jigarborad608@gmail.com
Country - India
College - Technishe Universität Chemnitz
Specialization - NLP

Problem Description:

The term hate speech is understood as any type of verbal, written or behavioural communication that attacks or uses derogatory or discriminatory language against a person or group based on what they are, in other words, based on their religion, ethnicity, nationality, race, colour, ancestry, sex or another identity factor. In this problem, We will take you through a hate speech detection model with Machine Learning and Python.

Hate Speech Detection is generally a task of sentiment classification. So for training, a model that can classify hate speech from a certain piece of text can be achieved by training it on data that is generally used to classify sentiments. So for the task of hate speech detection model, We will use the Twitter tweets to identify tweets containing Hate speech.

Business Understanding:

Understanding Problem —--> Building Machine Learning Model —-> Delivering Solution

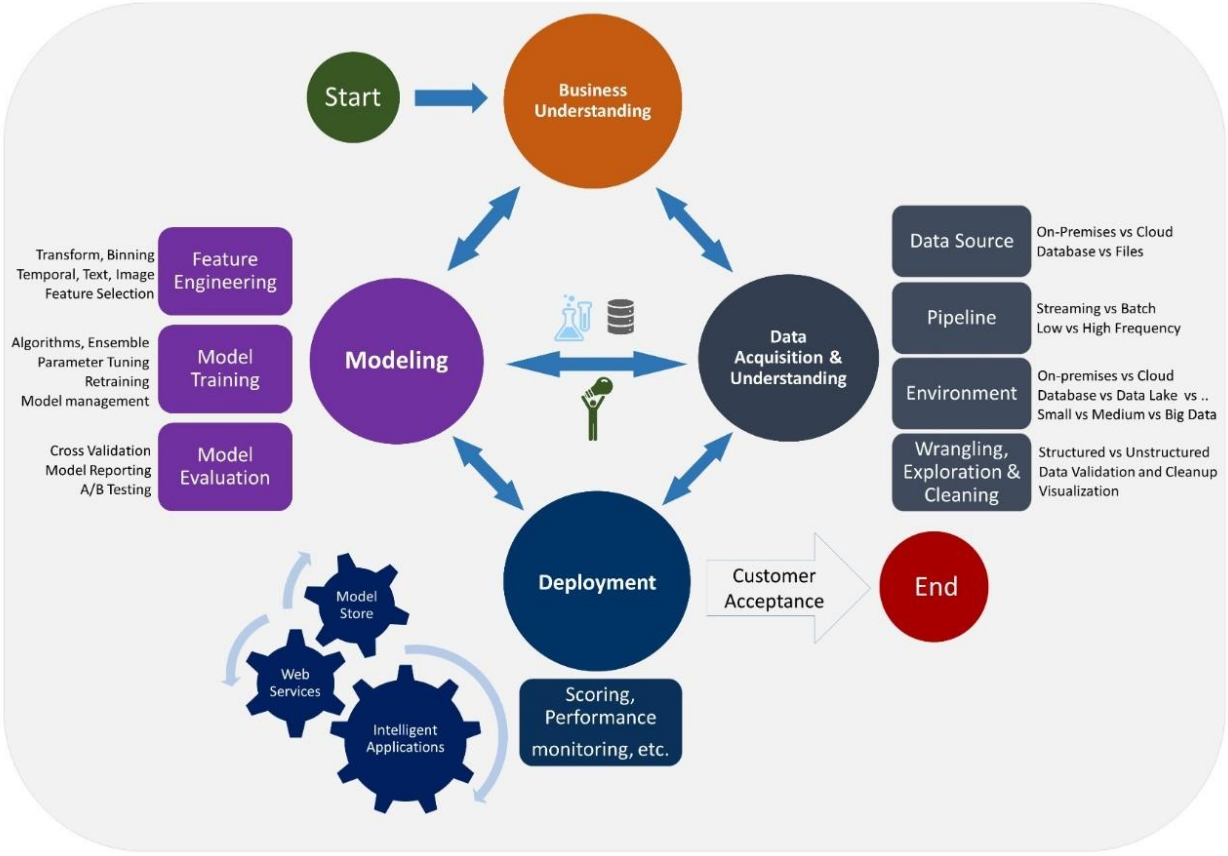
Aim: Our goal is to identify relevant data sources that this problem may have. Our aim is to create a model to deal with the detection of hate speech. Which we use in turn to find a solution. The model should be precise enough to make the solution meaningful, otherwise, we will make too many assumptions and approximations which will make the solution far from real and meaningless.

How To do this?

1. **Define objectives**
   A central objective of this step is to identify the hate speech keywords that the analysis needs to predict. We refer to these keywords as the model target and we determine that with the use of Natural Language Processing(Deep Learning) then we use the accuracy matrix associated with them to determine success of the project.
2. **Identify Data Sources**
   We will use the data of twitter comments to identify hate speech keywords. We will take a dataset 'Twitter Hate Speech' from kaggel.

Artifacts:

1. **Charter Document**



2. **Data Sources**
   Data Intek Report contains the data sources. This section specifies the original and destination location for the raw data.

Project LifeCycle:

1. **Initiating Phase** (1 weeks) :
    - Problem Understanding
    - Business Understanding
    - Collecting Data
    - Planning Further Timeline
2. **Planning/Data Cleansing Phase** (2 Week)
    - Data Understanding
    - Data Cleaning
    - Data Visualization
    - Planning work breakdown structure
    - Project Scope
3. **Executing Phase** (2 weeks)
    - Determining Model
    - Model Building
    - Model training
4. **Testing/Closing Phase** (1 Week)
    - Testing Data on selected model
    - Determining model accuracy
    - Delivering project

Expected Deadline : We will start the project from 19th July 2022 and expect to complete it in 6 weeks. So, the estimated deadline is 30th August 2022.