

Project Delivery 1

Group Name: The Survivor
Names: Abdelrahman Atef Abouzid
Email: Hemanysat@gmail.com
Country: Egypt
College: Ain Shams University
Specialization: NLP

Problem Statement

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 a 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

To develop a hate speech classifier that can accurately identify and flag hate speech on social media platforms, particularly Twitter. The goal is to mitigate the negative impact of hate speech on individuals and groups, and to create a safer and more inclusive online community. This problem is important for businesses and organizations that rely on social media for communication and marketing, as hate speech can damage their reputation and negatively impact their customer base. By developing a hate speech classifier, businesses can proactively address hate speech and take appropriate actions to ensure that their online community remains inclusive and respectful. Additionally, the hate speech classifier can be used by

social media platforms to automatically identify and remove hate speech content, which can improve user experience and mitigate the negative impact of hate speech on their platform.

Project lifecycle

Task	Deadline	Description
Problem Understanding	24 March	During this phase, the team will conduct a thorough analysis of the problem statement and determine the project's objectives and requirements.
Data Cleaning and Normalization	29 March	In this phase, the team will collect and preprocess the Twitter data to make it suitable for the machine learning model. This will involve tasks such as data cleaning, data normalization, and data transformation.
Representation Learning	5 April	In this phase, the team will use pre-trained transformer models such as BERT or RoBERTa to learn contextualized word embeddings from the preprocessed data. This phase is crucial for the success of the project, as the quality of the learned representations will significantly impact the model's performance.
Model Building & Training	12 April	During this phase, the team will build and train the hate speech classification model using the learned representations. This will involve selecting an appropriate machine learning algorithm, fine-tuning the pre-trained model, and optimizing the hyperparameters. The team will use a subset of the preprocessed data for training and validation

Performance Evaluation & Reporting	19 April	In this phase, the team will evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, and F1-score. The team will also report on the model's strengths and weaknesses and provide recommendations for future improvements.
Model Deployment	26 April	During this phase, the team will deploy the model on a cloud-based platform such as AWS or Google Cloud Platform. The team will also create an API for the model, allowing it to be integrated into various applications.
Model Inference	30 April	Finally, the team will perform a final test of the deployed model to ensure that it is functioning correctly and providing accurate results.

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 13/2/2023

Internship Batch: LISUM18

Version:<1.0>

Data intake by: Abdelrahman Atef

Data intake reviewer: na

Data storage location: [hemany1/Hate-speech-classification \(github.com\)](https://github.com/hemany1/Hate-speech-classification)

Tabular data details:

Total number of observations	49159
Total number of files	2
Total number of features	1
Base format of the file	csv
Size of the data	4.6 MB

