

Project Report

Team Member's Details

Group Name: Scalable Minds

Members:

Name: Igwebuike Eze and potential other participants
Email: samson12193@gmail.com
Country: UK
College/Company: Queen Mary
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 tweet 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.

Technology: PyTorch GPU, Azure

Introduction:

In this project, we aim to tackle the problem of hate speech detection using various machine learning approaches. Specifically, we will implement and compare two models:

1. NanoGPT
2. xLSTM (Extended Long Short-Term Memory)

We will train these models and compare their performance against state-of-the-art pre-trained models such as GPT, BERT, and ILAMA.

Objective:

The primary objective is to develop lightweight models that require significantly less memory and computational power, making them suitable for deployment on devices with limited resources like smartphones and tablets.

Business Understanding

Background:

Hate speech on social media platforms and other online forums is a growing concern. Detecting and mitigating such harmful content is crucial to maintaining a safe and inclusive digital environment. Traditional large language models (LLMs) like GPT, BERT, and ILAMA are highly effective but require significant computational resources, making them impractical for deployment on resource-constrained devices. Traditional large language models (LLMs) like GPT, BERT, and ILAMA are highly effective but

require significant computational resources, making them impractical for deployment on resource-constrained devices.

Goal:

Our goal is to create efficient and lightweight models for hate speech detection that can be deployed on low-resource devices. This will enable real-time detection and filtering of harmful content without the need for powerful computational infrastructure. In addition to efficiency, another aim is to build models with bespoke architectures that are relatively simple. These models should be easy to adjust to different platforms and scalable through parallelization. This flexibility will ensure that the models can be adapted for various deployment environments, enhancing their utility and longevity. For future development beyond this project, the model should also be designed to accommodate not only language data but also visual data enriched by physical reality. This extension will involve integrating neural operators to handle continuous data and complex dependencies across different data modalities. By doing so, we aim to create a comprehensive solution capable of processing and understanding multi-modal data inputs, paving the way for more sophisticated applications.

Benefits:

- **Accessibility:** Enable hate speech detection on a wider range of devices, including smart phones and tablets.
- **Cost-Efficiency:** Reduce the need for expensive computational resources, making the technology more accessible to smaller organizations and developers.
- **Scalability:** Facilitate the deployment of models in resource-constrained environments, allowing for wider adoption and impact.
- **Real-Time Processing:** Allow for the real-time detection and mitigation of hate speech, enhancing the user experience and safety on digital platforms.
- **Flexibility:** The bespoke architecture ensures the model can be easily adjusted to different platforms, improving adaptability and deployment efficiency.
- **Future-Proofing:** The capability to integrate neural operators for handling both language and visual data ensures the model remains relevant and expandable for future applications involving multi-modal data inputs.

Project Life-cycle Along with Deadline

Phases and Timeline

Problem Definition and Business Understanding (May 16 - May 20)

- Define the problem.
- Understand the business context and objectives.
- Generate a data intake report.

Data Collection and Preparation (May 21 - May 31)

- Collect and preprocess hate speech data.
- EDA and data interpretation

Detection of problems in the data (number of NA values, outliers , skewed etc) and how to overcome them.

Model Development (June 1 - June 15)

- Implement NanoGPT and xLSTM models.
- Train each model using the prepared dataset.

Evaluation and Comparison (June 16 - June 25)

- Evaluate the models using key performance metrics.
- Compare the results with state-of-the-art pre-trained models (GPT, BERT, ILAMA).

Documentation and Final Submission (June 26 - June 30)

- Compile the final report.
- Submit the project by July 1.

Deadline

The final project submission deadline is July 1st. And separate reports and deliverables will be provided according to the specified schedule of LISUM32.