Hate Speech Detection

Presented by:
Nishat Salsabil Rainy -1812620042
Mohd. Istiaq Hossain Junaid-1821577642
Samreen sohail-1711648642

Introduction

With the advancement of Technology and Internet, people have got many platforms where they can freely share their thoughts and opinions. We know that everyone has the right to freedom of speech. However, this right is being misused to discriminate and attack others, physically or verbally.

We have thought about developing a system that could catch comments that encourage hate and conflict.

What is defined as hate speech?

The term hate speech is understood as any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.

Examples:

- “Queers are an abomination and need to be helped to go straight to hell!”
- “We have to kill all the Palestinians unless they are resigned to live here as slaves.”
- “Women shouldn’t talk sports on tv. They belong in the kitchen.”

Motivation

Over the years, there have been hundreds of incidents related to hate crimes that have taken place and have led to fights, riots, and multitudes of casualties. Hate speech is only powerful because of its ability to plant thoughts of discrimination without being noticed.

What if we could design an algorithm that could detect hate speech?

To create a hate-speech-detecting algorithm, we are going to use machine learning techniques.

Our Proposed System

- To create a hate-speech-detecting algorithm, we are going to use Python-based NLP machine learning techniques.
- Tf-Idf vectorization to extract keywords that convey importance within hate speech.
- Based on a machine learning technique called logistic regression, we'll train the models to classify hate speech.

What kind of ML system is it?

1.As it is trained with human supervision it is actually supervised learning.

2.In this the data we are feeding to the algorithm will automatically includes desired solutions,which called labels.

Data Collection:

It is important to collect dataset to test train and tuning the model of the project.

We will use some dataset which link in given below.

Kaggle:

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

2.https://www.kaggle.com/naurosromim/bengali-hate-speech-dataset/version/1

Methodology

The main goal of this project is to implement a framework to detect hate speech in the spoken content of videos as shown below in figure 1.

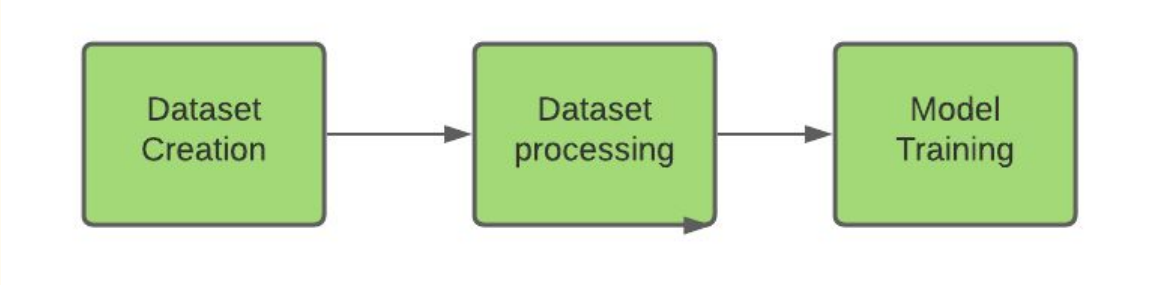


Fig1:Hate speech detection framework

The entire process can be divided into three main parts:

1. Build the video dataset.
2. Extract audio from the video dataset and convert into textual format.
3. Train machine learning models over the dataset and classify videos as normal or hateful.

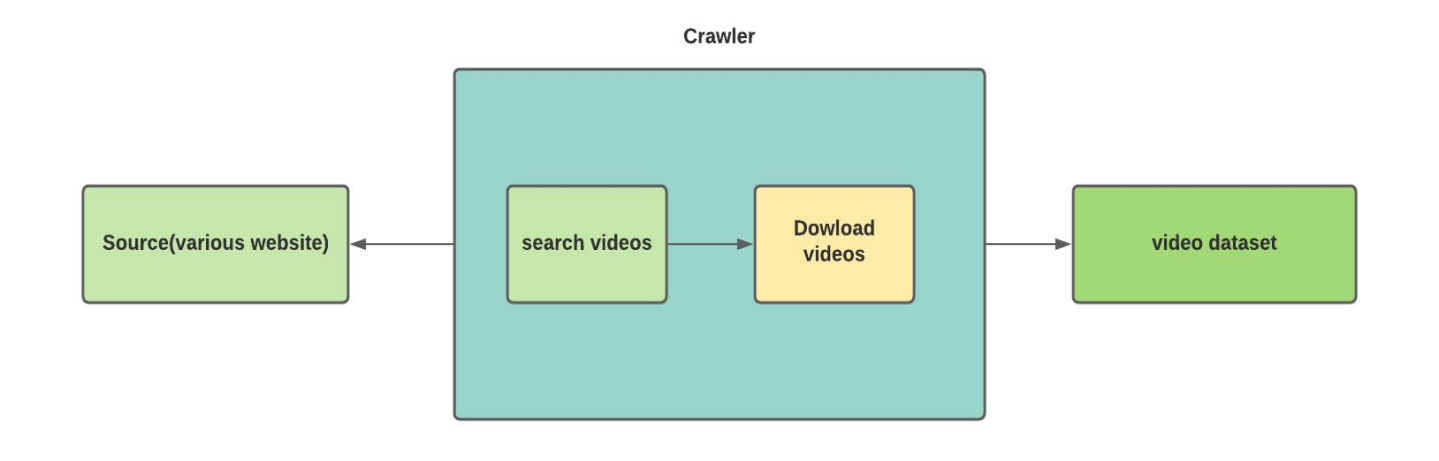


Fig 2:Dataset Creation Module

Dataset Processing

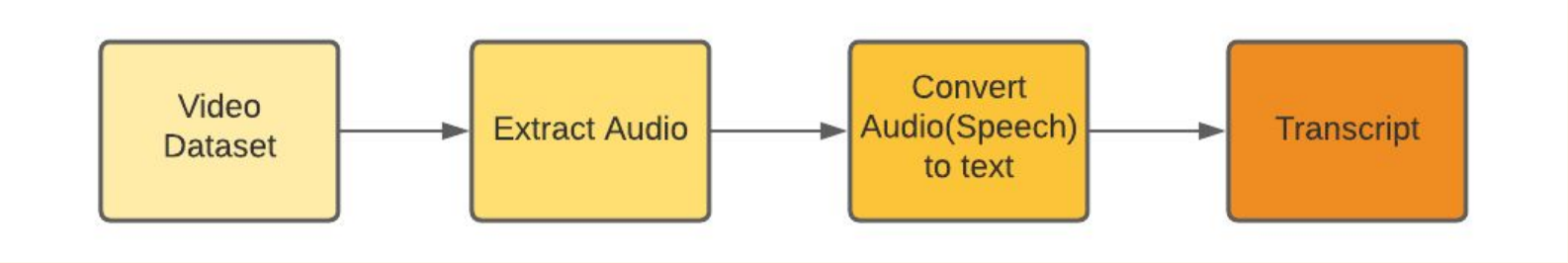
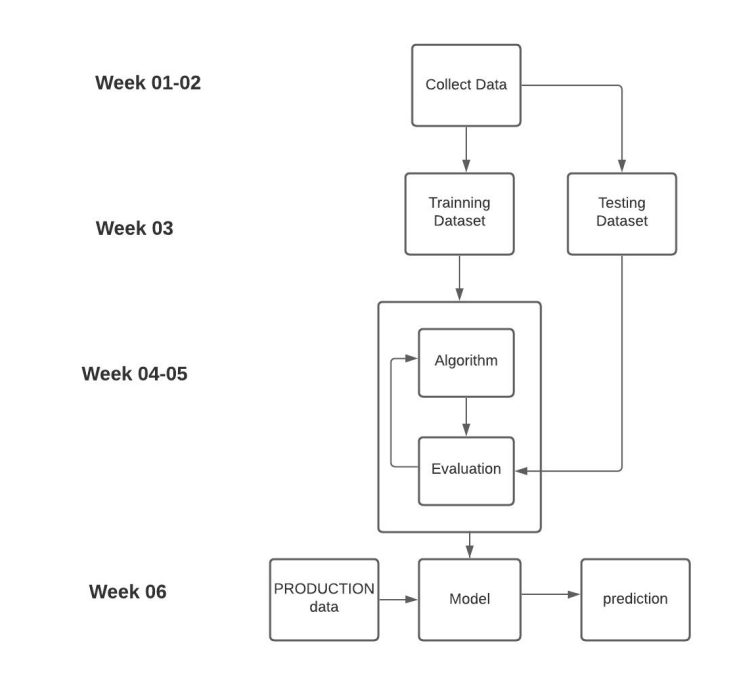


Fig 3:Dataset Processing Model

Workflow


Week 01-02
Collect Data
Trainning Dataset
Testing Dataset
Week 03
Week 04-05
Algorithm
Evaluation
Week 06
PRODUCTION data
Model
prediction

Gaan Chart

Task	Week1	Week2	Week3	Week4	Week5	Week6
1. Data collection						
2. Training Data						
3.Testing Data						
4. Algorithm						
5. Evaluation						
6. Model prediction						

Referencess

1.https://www.researchgate.net/publication/346638333_Hate_Speech_detection_in_the_Bengali_language_A_dataset_and_its_baseline_evaluation.
2.https://paperswithcode.com/paper/detecting-hate-speech-and-offensive-language

Conclusion

Hate speech on social media has been increasing significantly in recent years. As more people are turning towards video-sharing sites, people tend to post opinionated videos which might not always be peaceful. There is a need for optimal hate speech detection system. The existing hate speech detection methods focus on text data. Hence, there is a need to find an optimal approach to detect hate speech in videos. Our project deals with converting the video into text format before passing it as input to machine learning models. We will use an NLP technique to extract keywords that convey importance within hate speech. And Finally, we'll train the models to classify hate speech based on a machine learning technique.