Week 8 Deliverables

Orkun

26 June 2024

Team Member's Details

- **Group Name:** Orkun
- **Name:** Orkun Kınay
- **Email:** orkunkinay@sabanciuniv.edu
- **Country:** Turkey
- **College/Company:** Sabancı University
- **Specialization:** NLP

Problem Description

The goal of this project is to develop a hate speech detection model using Twitter data. The dataset consists of tweets labeled as hate speech or non-hate speech, which will be used to train and evaluate machine learning models. The primary objective is to accurately classify tweets and mitigate the spread of hate speech on social media platforms.

Data Understanding

The dataset used for this project contains tweets collected from Twitter, labeled as either hate speech or non-hate speech. The primary aim is to understand the distribution, quality, and characteristics of the data to identify potential issues and plan for data preprocessing steps.

Data Details

- **Total Number of Observations:** 31962
- **Total Number of Features:** 3
- **Features:**
  - **tweet:** The text of the tweet.
  - **label:** The label indicating whether the tweet is hate speech (1) or non-hate speech (0).
  - **id:** A unique identifier for each tweet.
- **Base Format of the File:** .csv
- **Size of the Data:** 4.7 MB

Data Problems

- **Number of NA Values:**
  - There are no missing values in the dataset.
- **Outliers:**
  - The dataset primarily consists of textual data, so outliers in the traditional numerical sense are not applicable. However, there are 9 tweets with excessively long text that could be considered outliers.
- **Skewed Data:**
  - The dataset is imbalanced, with a significantly higher number of non-hate speech tweets compared to hate speech tweets. Specifically, there are 29720 non-hate speech tweets and 2242 hate speech tweets.

Approaches to Handle Data Problems

- **Handling Imbalanced Data:**
  - Use techniques such as oversampling the minority class (hate speech) or undersampling the majority class (non-hate speech).
  - Apply advanced techniques like Synthetic Minority Over-sampling Technique (SMOTE) to create synthetic examples of the minority class.
- **Handling Text Outliers:**
  - Filter out tweets that are excessively long as they might be noise in the dataset.

GitHub Repo Link

https://github.com/orkunkinay/Hate-Speech-Detection/tree/main/data_preprocessing