**Hate Speech Detection**

**Overview**

**This project focuses on building a machine learning model to detect hate speech in textual data. It preprocesses data, visualizes class distributions, and trains and evaluates models to classify text into categories like "Hate Speech," "Offensive Language," and "Normal Speech."**

**Features**

* **Text Preprocessing: Includes removing mentions, URLs, punctuation, and extra spaces while normalizing text.**
* **Data Visualization: Insights into the class distribution of the dataset.**
* **Model Training: Implementation of machine learning models using:**
  + **Support Vector Machine (SVM)**
  + **Linear SVC with hyperparameter tuning via Grid Search CV.**
* **Evaluation: Comprehensive evaluation with accuracy scores and classification reports.**

**Technologies Used**

* **Python**
* **Pandas**
* **Scikit-learn**
* **Matplotlib**
* **TF-IDF Vectorization**

**Project Workflow**

1. **Data Loading: Input CSV files are dynamically read into the project.**
2. **Preprocessing: Text is cleaned and normalized for model compatibility.**
3. **Exploratory Data Analysis: Visualize the dataset's class distribution.**
4. **Feature Extraction: Textual data is converted into numerical representations using TF-IDF.**
5. **Model Building: Train and tune SVM and Linear SVC models.**
6. **Evaluation: Models are evaluated based on accuracy and other metrics.**

**Key Results**

* **Achieved a classification accuracy of 90.6% using the SVM model.**
* **Optimized Linear SVC with hyperparameter tuning to improve performance further.**