**Task 2: - Language Identification**

**Introduction**

This project aims to build a system that can detect the language of a given text and translate it into English, French, Hindi, or Spanish. The system uses text preprocessing, a machine learning model for language detection, and Google Translate for translation. A graphical user interface (GUI) is created using Tkinter for ease of use.

**Key Components**

Data Preprocessing: Removing punctuation and digits, and converting text to lowercase.

Feature Extraction: Using TF-IDF Vectorizer to convert text into numerical features.

Model Training: Training a Logistic Regression model using a pipeline.

Language Translation: Using Google Translate API for translation.

Graphical User Interface: Tkinter-based interface for input, detection, and translation.

**Libraries Used in the Language identification**

* numpy
* pandas
* matplotlib
* seaborn
* string
* re
* scikit-learn
* tkinter
* googletrans

**Data Preprocessing**

* Loading and Cleaning Data

The dataset is loaded from a CSV file and cleaned by removing punctuation and digits, and converting text to lowercase.

* Feature ExtractionTF-IDF Vectorizer is used to convert text data into numerical features.
* Model Training

A Logistic Regression model is trained using a pipeline.

* Language TranslationGoogle Translate API is used for translation. The detected language and the translated text are displayed in the GUI.

## Graphical User Interface

## The Tkinter library is used to create a GUI for language detection and translation.

### GUI Elements

* **Text Entry**: Input field for entering text to be detected and translated.
* **Language Detection Button**: Button to detect the language of the entered text.
* **Language Label**: Label to display the detected language.
* **Translate To Label**: Label and dropdown menu for selecting the target language.
* **Translate Button**: Button to translate the text.
* **Translated Text Entry**: Entry field to display the translated text.
* **Accuracy Label**: Label to display the model accuracy.
* **Confusion Matrix Button**: Button to display the confusion matrix.

**Results**

* Model Accuracy: The accuracy of the Logistic Regression model on the test set is displayed.
* Confusion Matrix: The confusion matrix can be displayed to visualize the performance of the model.
* Language Detection: The system can detect the language of the entered text.
* Translation: The system can translate the detected text into the selected target language.

**Conclusion**This system effectively detects the language of a given text and translates it into English, French, Hindi, or Spanish. The combination of TF-IDF feature extraction, Logistic Regression model, and Google Translate API provides a robust solution for language detection and translation tasks. The Tkinter GUI makes the system user-friendly and easy to use