ABSTRACT

Title: Develop an Interpreter to Predict Language Using Machine Learning

Problem Statement:

Identifying the language of a given text is a complex task, especially with multiple languages and dialects. Current systems struggle with accuracy, especially with short text inputs or similar languages. This project aims to create a system that can reliably predict the language of any text.

Aim:

To develop an application that can predict the language of a given text using machine learning, improving multilingual support and communication tools.

Data Set: (Language Identification Dataset)

COCO (Common Objects in Context) dataset, specifically tailored for text detection and recognition. It contains over 63,000 images with annotations for more than 173,000 text instances, providing a rich source of real-world text in various contexts.

Algorithm Used:

The system will use machine learning models like **Naive Bayes**, **SVM**, and **Neural Networks** to classify languages based on linguistic features such as word frequency and character patterns. **Deep Learning** models, like **LSTM** or **Transformers**, will be explored for better accuracy.

Integration with AI:

Data Sources: A dataset of labelled multilingual text will train the model.

AI Models: Machine learning algorithms will analyse features such as n-grams and character-level patterns for language prediction.

Real-Time Application: The model will be deployed in an application for instant language detection in text-based inputs.

Batch-16

R. Jayaram-2320030244

S. Shanmukha-2320030361

P. Abhiram-2320030294