Multilingual Named Entity Recognition for Global Businesses

Abstract:
For global businesses, extracting valuable information from multilingual content is critical for decision-making and operations. This project focuses on **multilingual named entity recognition (NER)**, which involves identifying and classifying entities such as people, organizations, and locations from text in multiple languages.

Methodology:
The system will preprocess text in multiple languages using standard **NLP techniques**, including **tokenization**, **lemmatization**, and **part-of-speech tagging**. Multilingual **pre-trained transformer models** like **mBERT** or **XLM-R** will be employed for **entity recognition**. The system will be trained on multilingual NER datasets, including **conll-03**, to identify entities from a variety of languages. **Transfer learning** will be used to adapt models to specific business-related entity extraction tasks, such as extracting company names or product details from global data sources.

Outcome:
The expected outcome is a multilingual NER system that can extract relevant business-related entities from content in various languages. This system will enhance global business operations by enabling automatic processing of documents, contracts, emails, and other text data in multiple languages, improving efficiency in international markets and supporting better decision-making.