

Investigating the Effectiveness of Multilingual NLP Models for Sentiment Analysis

Authors : Othmane Touri, Sanaa El Filali, El Habib Benlahmar

Abstract : Natural Language Processing (NLP) has gained significant attention lately. It has proved its ability to analyze and extract insights from unstructured text data in various languages. It is found that one of the most popular NLP applications is sentiment analysis which aims to identify the sentiment expressed in a piece of text, such as positive, negative, or neutral, in multiple languages. While there are several multilingual NLP models available for sentiment analysis, there is a need to investigate their effectiveness in different contexts and applications. In this study, we aim to investigate the effectiveness of different multilingual NLP models for sentiment analysis on a dataset of online product reviews in multiple languages. The performance of several NLP models, including Google Cloud Natural Language API, Microsoft Azure Cognitive Services, Amazon Comprehend, Stanford CoreNLP, spaCy, and Hugging Face Transformers are being compared. The models based on several metrics, including accuracy, precision, recall, and F1 score, are being evaluated and compared to their performance across different categories of product reviews. In order to run the study, preprocessing of the dataset has been performed by cleaning and tokenizing the text data in multiple languages. Then training and testing each model has been applied using a cross-validation approach where randomly dividing the dataset into training and testing sets and repeating the process multiple times has been used. A grid search approach to optimize the hyperparameters of each model and select the best-performing model for each category of product reviews and language has been applied. The findings of this study provide insights into the effectiveness of different multilingual NLP models for Multilingual Sentiment Analysis and their suitability for different languages and applications. The strengths and limitations of each model were identified, and recommendations for selecting the most performant model based on the specific requirements of a project were provided. This study contributes to the advancement of research methods in multilingual NLP and provides a practical guide for researchers and practitioners in the field.

Keywords : NLP, multilingual, sentiment analysis, texts

Conference Title : ICAIDM 2023 : International Conference on Artificial Intelligence and Data Mining

Conference Location : London, United Kingdom

Conference Dates : May 15-16, 2023