Language Detection on NLP

INTRODUCTION AND RATIONALE

Language recognition is one of the most widely used tools on a global scale. Here I am with the most recent technological topic. I got this data set from Kaggle. Our dataset has 22000 rows and 2 columns, with words in 22 different languages in the text column. And first, I checked the null values in the data, and in our dataset, we have only two columns: text and language. We have text with words, and in "Language," it mentions the name of the language. For feature extraction, I used the countvectorizer, which helps convert a collection of text documents into a matrix of token counts.

MODEL

- The Multinomial Naive Bayes is one of the variants of Naive Bayes algorithm in machine learning.
- It is very useful to use on a dataset that is distributed multinomially. This algorithm is especially preferred in classification tasks based on natural language processing.
- In language detection, the feature would be a set of words observed in the text

CONCLUSION

The effectiveness of language detection on NLP depends on the accuracy of the algorithms and the data used. Here,i used a multinomial naive bayes model and got a 0.95 score, which is the best score for this problem. By using this technology, organizations can better communicate with customers and better understand their needs. In the end, language detection on NLP is an important tool for understanding the diverse nature of human language and its applications in natural language processing.

Karri Dharma Teja
dharmatej14june2001@gmail.com
Linkedin : karridharmateja
Github: dharmatejakarri