CSCE 5290 Section 002

**Group- 12**

**Text Classification using Spacy NLP**

**Team members:**

**Priyanka Kare**

**Sai Nikhil Kotte**

**Goals and Objectives:**

**1.Motivation**

Any data that is obtained from product reviews, movie reviews, news data or emails this data must be classified into which categories they fall. Suppose if we take product reviews data, we can classify the product reviews into positive reviews and negative reviews, similarly if we take news data, we can classify that into sports, business, education etc.., even mails can be classified as spam or normal. This type of classification is done to categories the data. This type of classification model must be built to make the classification easy.

**2.Significance**

The main significance of building the text classification model is to find and analyze how the data is populated and for reviews dataset we need to identify how many positive and negative feedback is given for the product. Based on the analysis we can find which products are sold well and stock them more for sales.

These reviews help us to analyze the problems in the data and based on the feedback the sales can be improved. The most efficient package which is used in NLP for classification is Spacy.

**3.Objectives**

The first objective is to categorize the text automatically into one or more different categories. There is a lot of research going now a days on text classification to improve the model and to give more accurate results. In this project we will take Amazon product reviews dataset and build the text classification model.

From the built model we will evaluate the performance of the model and analyze how accurately the model is used to classify the text. The final goal is to build the model with good performance and high accuracy.

**4.Features**

Extraction of features is considered as one of the important steps and is called as feature engineering.

There are many methods to extract the features such as count vectors as features, TF-IDF vectors as features which contains word level, N-Gram level, Character level, word embeddings as features, text/NLP based features and topic models as features. The above mentioned methods are used for feature extraction for any text classification model. Based on the chosen dataset.

Work flow

Graphical user interface, text, application, chat or text message

Description automatically generated

**Reference:**

[https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/#:~:text=The%20goal%20of%20text%20classification%20is%20to%20automatically,classification%20are%3A%20Understanding%20audience%20se](https://www.analyticsvidhya.com/blog/2018/04/a-comprehensive-guide-to-understand-and-implement-text-classification-in-python/#:~:text=The%20goal%20of%20text%20classification%20is%20to%20automatically,classification%20are%3A%20Understanding%20audience%20sentiment%20from%20social%20media%2C)

<https://www.machinelearningplus.com/nlp/custom-text-classification-spacy/>