GitHub Link:
https://github.com/dheerajreddykomandla/NLP_TextEmotionDetection

CSCE 5290
Natural Language Processing
Project Proposal

1. Project Title and Members:

- **Project Title: Text Emotion Detection**
- **Team Number:** Team 10

Team Members:

Dheeraj Reddy Komandla (11526265)

Prem sai Vuppula (11528103)

Rahul Reddy Geereddy (11511846)

Sravan Boinapalli (11553755)

2. Goals and Objectives:

Motivation:

In this project we as a team decided to work on Text classification problem. In which we have decided to work on Text Classification problem. In recent times we have seen sentiment analysis has been using in a wide spectrum of the applications.

We as a team thought of challenging ourselves instead of working on the binary classification problem. The main motivation behind choosing the Text Emotion Detection is that the class labels are related to the sentiment analysis problem and more like the intent classification problem which is used in chat bots.

As per the research we have done, and the above reasons inclined us to choose Text Emotion Detection problem.

Significance:

As discussed earlier, the main important significance of this model is that it has a wide spectrum of the real-world applications.

If you consider an example of chatbot interactions, emotions play an important role in the communication between people of different cultures or regions. For example, let's say we have classified the emotions into 4 types.

They are: "joy, anger, sadness, fear"

In the below example we can observe the difference between single words and multiple words statements.

Single words

➢ Great

➢ Nope

➢ Recommend.

Multiple words statements

➢ This is great!

➢ Nope, won't buy again!

➢ Definitely recommend this product!

You can observe in both sentences, the emotion that represents are same, but the number of words varies. So, it is important for a chatbot model to classify the accurate emotion based on which the further communication must continue.

Let's try to understand the significance of the model with one more example. If you consider the below statements.

➢ This is not a friendly product.

➢ I hate this product; not useful at all.

➢ I want to return this product, maybe order a replacement.

If you observe the above 3 statements, all are having the sentiment negative. But the meaning, emotion, and the customer requirement for the 3 statements are different.

● **<u>Objectives</u>:**

Our team has several objectives that are required in order to successfully build an emotion detection model. The following objectives illustrates the workflow of this project.

1. **<u>Data Selection:</u>** As we did some research, we have shortlisted some of the datasets available in the Kaggle. We would like to spend some more time to pick the right dataset with good number of emotion samples.

2. **<u>Text Cleaning:</u>** In any of the NLP tasks, it is important to clean the text dataset. Because the raw dataset may contain lot of missing values, duplicate values etc. We need to handle all the cases. If the text is not cleaned, then the model may give bad results, or it fails.

3. **<u>Text Preprocessing:</u>** It is also an important phase in all NLP related tasks. In this phase, we generally remove the white spaces, remove unnecessary special characters, stop words removal etc.

4. **<u>Text Vectorization:</u>** This phase is also known as text featurization. In this phase we generally convert the text into vectors. There are some popular Vectorization techniques available such as Bag of Words, Tf-IDF, Word2Vec etc. By using these techniques, we will be able to convert the text into n dimensional vector.

5. **<u>Modelling:</u>** Once we converted the text into vector, then we are good to go for modelling. Here we want to try this various machine learning algorithms. Of all the models that are trained, which ever gives the best accuracy, we will choose that model as best model

6. **Visualization & Evaluation of Results:** In this phase, we basically compare the results of each model and plot the results to understand the performance of the model. Since it is a classification model, we would like to compare the performance with accuracy. If the dataset is imbalanced, then we will also look into f1-score, AUC score.

- **Features:**

  As discussed earlier we would like to spend some more time on picking the dataset, but in most of the text classification datasets, it contains 2 features
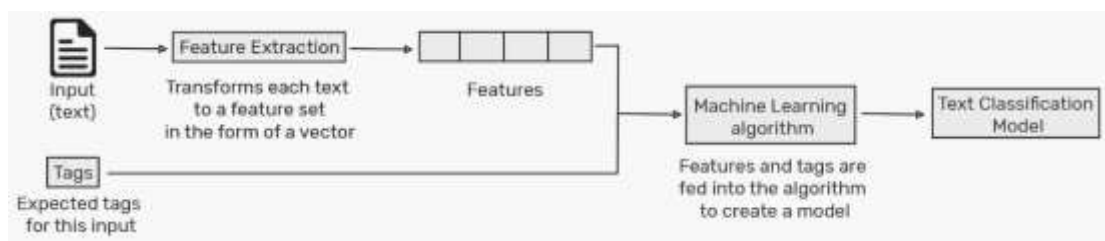
  1. **Text:** This feature mainly contains the raw text. This text represents the either review or statement for which we need to predict the emotion of the statement.

  2. **Label:** This feature represents the class label of the given sentence.In most of the datasets we basically have 4 classes. They are:

     I.   Fear

     II. Joy

     III. Anger

     IV. Sadness

  There are some import machine learning libraries are required to solve this problem. They are listed below:

     I.   Pandas

     II. SKlearn

     III. Matplotlib

     IV. XGboost

Workflow:

2. References :

https://www.kaggle.com/datasets/anjaneyatripathi/emotion-classification-nlp

https://www.kaggle.com/code/dorgavra/emotion-classification-nlp

https://www.degruyter.com/document/doi/10.1515/jisys-2022-0001/html?lang=en