# **CSCE 5290 - Natural Language Processing**

**Project Title**

**Twitter Sentiment and Emotion Classification System**

**Team Members**

| Ajay Jatprol |
| --- |
| Mohammed Mohammed Sabith Abul |
| Rahman Mehmood |
| Deepak Adimoolam |

**Introduction** :

Social media sites like Twitter, which provide real-time insights into public mood and emotional responses have become essential in influencing public discourse in the digital age. A challenge as well as an opportunity arises from the daily generation of millions of tweets: the extraction of relevant insights from this massive sea of textual data. Utilizing Natural Language Processing (NLP) techniques, this project, "**Twitter Sentiment and Emotion Classification System using NLP Techniques**," seeks to assess and categorize the feelings and emotions conveyed in tweets.

The study of natural language processing, which lies at the interface of linguistics and computer science, gives us the means to analyze and comprehend human language. This project aims to create a strong classification system that can identify the underlying sentiment positive, negative, or neutral by utilizing sophisticated NLP approaches.

**1. Motivation**

The increase of social media has resulted in increased Volumes, Variety of data with Twitter being major platforms for people discussion. Examining sentiments and feelings in tweets can come up with valuable understanding in people's beliefs, social trending topics, and user actions. However, directly classifying emotions and sentiments from small, casual texts like tweets is challenging due to components such as dialect, acronym, and context. This project proposal focuses on these components by creating and training a robust classification model for Twitter sentiment and emotion analysis. The understanding from this idea can be used in areas such as mental health and marketing a product and public relations.

**2. Significance**

This project is important as it provides access to the increasing field of natural language processing (NLP) by taking care of real-world data from social media platforms. Interpretation of people's sentiment is critical for firms, business, governments, and Observers to make educated selections. The possible influences include increasing users' assistance through analyzing sentiments of the users, forecasting shopping trends, and observing people's health through analyzing the emotions. By gaining the project's idea, we can provide a more refined understanding of social media discussion, which is crucial for multiple sectors depending on public opinion and frame of mind data.

**3. Objectives**

The leading objectives of this project are:

· To create a text classification model the ability to detect sentiment (positive, negative, neutral) and emotions (happiness,depressed, sadness, anger, etc.) in tweets posted by people in the twitter platform.

· To maximize the model's act using machine learning models or deep learning techniques, guarantying it precisely handles the refinement of Twitter language.

· To preliminary processing the dataset productively, addressing problems like noise dataset, abbreviations, and hashtags, using natural language processing techniques like lemmatizations, stop word removal etc.

· To determine the model's accuracy using precise metrics and guaranteed to generalize well to unseen data.

**4. Features**

· **Technical Features**: The structure will use natural language processing techniques, including stop-word removal, tokenization and sentiment lexicons,stemming combined with machine learning models such as Logistic Regression or deep learning techniques like LSTM (Long Short-Term Memory),Recurrent neural network etc.

· **Product**: A practical sentiment and emotion classification model, project documentation, and a final report specifying the process, visualization of the dataset,results, and future improvements.

· **Milestones**:

o Data Gathering.

o Data Analysis.

o Data processing and visualization.

o Multiple Model development and training.

o Performance evaluation and optimization.

o Comparing the models.

o Final report and presentation

· **Unique Elements**: This project’s distinctiveness lies in a blend of both sentiment and emotion classification on Twitter data, providing a more complete analysis of social media content.

**5. Dataset**

· **Dataset 1**: *Twitter Emotion Classification Dataset* - This dataset holds labeled tweets that act for various emotional states like joy, sadness, anger, and fear.

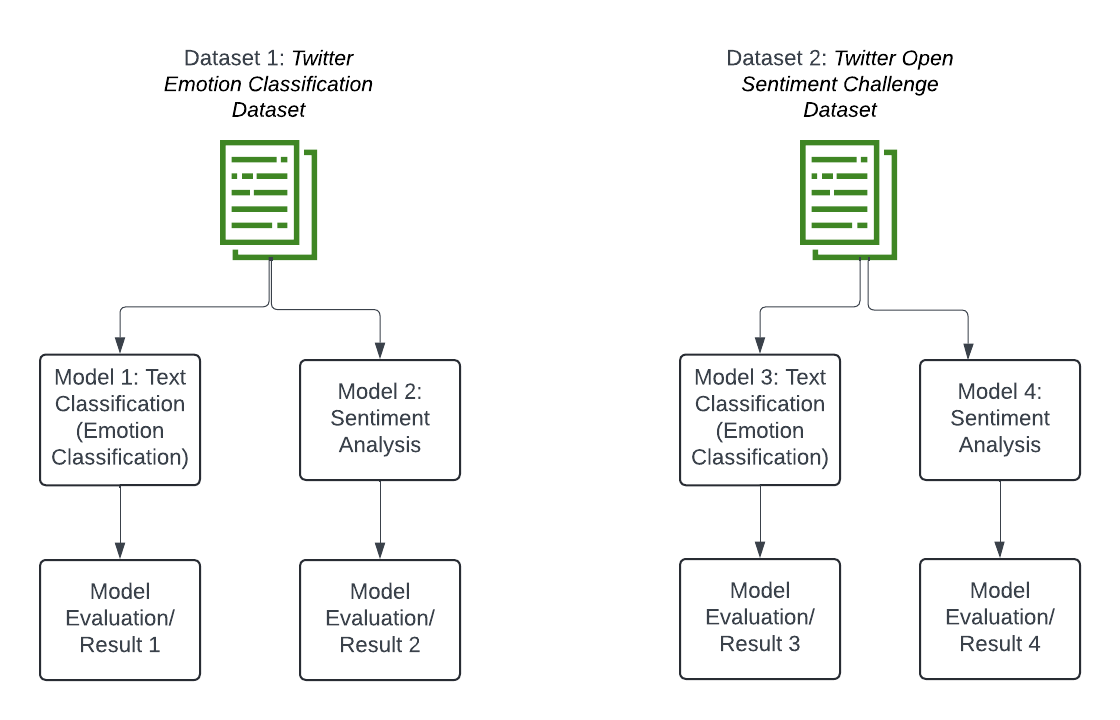
· **Dataset 2**: *Twitter Open Sentiment Challenge Dataset* - This dataset contains labeled tweets for sentiment classification (positive, negative, and neutral).

o **Size**: Both datasets have over 50,000 entries.

o **Sources**: openly accessible Twitter data for sentiment and emotion analysis challenges.

o **Preprocessing**: Data will be cleaned by pull out stop words and user mentions,hashtags. Tokenization, stemming, and lemmatization techniques will be applied to make the data suitable for the model.

**6.Visualization**



**Workflow Diagram**

Data gathering → Preprocessing → Feature Extraction → Model Training → Sentiment and Emotion Classification → Model Evaluation → Visualization → Feedback

**Tables**: To present classification accuracy, precision, recall, and F1-scores.