## **Title:**

**Analyzing sentiments with machine learning by decoding tweets**

## **Team Members:**

1. **Venkatsai Maddala –** [**venkatsaimaddala@my.unt.edu**](mailto:venkatsaimaddala@my.unt.edu)
2. **Ranjitha Bejjarapu -** [**ranjithabejjarapu@my.unt.edu**](mailto:ranjithabejjarapu@my.unt.edu)
3. **Keerthi Annem -** [**keerthiannem@my.unt.edu**](mailto:keerthiannem@my.unt.edu)

## **GitHub link:** <https://github.com/vnkym96/NLP>

## **Motivation:**

In the current ‘digital’ society, tweets play a major role in identifying the personality and their influence over the topic. Tweets sometimes provoke a person or community and sometimes share useful information. Understanding a tweet is crucial in interpreting the context. This can be further developed to understand the motive of behind the tweet. Hence, sentiment analysis is performed to describe the emotion of the tweet. By this emotion analysis, decision-making is made easier. As it is a tedious process to analyze large volume of data, machine learning algorithms can be embedded to compile the data and output the result as desired. The main objective is to create the most accurate algorithm that can analyze the voluminous data and its usage in decision-making process. It may be in the marketing, business strategies, ad-campaigns, social trends, public opinion, accurate emotion description is necessary to make informed decisions in a time-to-time manner. The algorithm can be a base to understand the patterns of emotions in the tweets. This project can be further deployed in texts, facial recognition, etc.

## **Significance:**

This project works on the Natural Language Processing and sentiment analysis and provide great importance. constantly analyzing the tweets and identifying the sentiment in them. This process of analyzing the tweets is included with real-time monitoring of the public opinion, Allows the sponsors to manage the decisions and responses run by data in a quick way. The goal of the project is to enhance brand management, customer satisfaction, campaign effectiveness, research insights, and policymaking strategies. In Addition to it, the project's results provide enhancements and upgrades in machine learning techniques and applications, which can be adopted in innovation in sentiment analysis methods.

## **Objectives:**

The primary objective of this project is to upgrade machine learning model that is efficient to determine the sentiments of tweets and obtaining sentimental phrases. The purpose includes attaining higher accuracy in sentiment categorizing, stability to manage wide range of tweet content and flexibility to process huge number of tweets in instance. The output can be displayed by calculating performance metrics like accuracy, precision, F1- score and recall for the chosen dataset.

## **Features:**

This project involves the integration of Deep Convolutional Neural Networks (DCNN) framework of sentimental analysis and making use of word embeddings and usage of convolutional layers and retrieval text from tweets. The results consist of machine learning models that are trained and handles the categorized tweet sentiment and determining sentimental phrases. The milestones for this project are executing model training, analysis, fine-tuning stages that leads to the implementation of an adaptable sentiment analysis solution.

## **Dataset:**

The dataset consists of labeled tweets from Figures eight data containing tweets with existing sentimental labels like positive, negative, neutral. We are going do text cleaning by removing noise data and special tags. The cleaned text is tokenized by breaking it down into individual tokens. Finally those tokens are encoded into numerical representation for training the model.

## **Visualization:**

A diagram of a diagram

Description automatically generated

## Workflow explanation:

|  |  |
| --- | --- |
| **Step** | **Description** |
| Data Upload | Loading tweet dataset from train.csv and test.csv |
| Data cleaning | Removing unnecessary columns & handling missing values |
| Text Tokenization | Clean Text data, tokenize using spicy |
| Model Building | Designing DCNN architecture using Tensor flow |
| Training | Train the model on preprocessed data |
| Model Evaluation | Evaluating model performance on text data |
| Sentimental Forecast | Make predictions |