

Sharadchandra Pawar College of Engineering and Technology , Someshwarnagar

Name : Vaibhav Arvind Jamdade

Roll no : CO327

Mini Project Name

Classify tweets into positive and negative tweets Introduction

- ✚ Sentiment analysis or opinion mining refers to identifying as well as classifying the sentiments that are expressed in the text source.
- ✚ Tweets are often useful in generating a vast amount of sentiment data upon analysis.
- ✚ These data are useful in understanding the opinion of people on social media for a variety of topics.
- ✚ Twitter sentiment analysis analyzes the sentiment or emotion of tweets.
- ✚ It uses natural language processing and machine learning algorithms to classify tweets automatically as positive, negative, or neutral based on their content.

- ✚ It can be done for individual tweets or a larger dataset related to a particular topic or event.
- ✚ In this project, we try to implement an NLP Twitter sentiment analysis model that helps to overcome the challenges of sentiment classification of tweets.
- ✚ We will be classifying the tweets into positive or negative sentiments.
- ✚ The necessary details regarding the dataset involving the Twitter sentiment analysis project are:

The dataset provided is the **Sentiment140 Dataset** which consists of **1,600,000 tweets** that have been extracted using the Twitter API.

Project Pipeline

- ✚ Import Necessary Dependencies
- ✚ Read and Load the Dataset
- ✚ Exploratory Data Analysis
- ✚ Data Visualization of Target Variables
- ✚ Data Preprocessing
- ✚ Splitting our data into Train and Test sets.
- ✚ Transforming Dataset using TF-IDF Vectorizer
- ✚ Function for Model Evaluation
- ✚ Model Building
- ✚ Model Evaluation

Algorithms used :

Machine Learning algorithms like

- ✚ Naive Bayes
- ✚ Logistic Regression
- ✚ SVM

Deep learning algorithms like

- ✚ RNN can be used to create Twitter Sentiment Analysis

Advantages :

Sentimental Analysis models are used in various industries for different purposes. Some examples are:

- ✚ Using these models, we can get people's opinions on social media platforms or social networking sites regarding specific topics.
- ✚ Companies use these models to know the success or failure of their product by analyzing the sentiment of the product reviews and feedback from the people.

- ✚ Health industries use these models for the text analysis of patients' feedback and improve their services based on that.
- ✚ We can also find new marketing trends and customer preferences using these models.

Disadvantages :

- ✚ Lexicon-based sentiment analysis methods usually do not identify sarcasm, negation, grammar mistakes, misspellings, or irony. Thus, it may not be suitable for analyzing data gathered from social media platforms.
- ✚ As the whole classification is based on tags and rules, companies should have sufficient data to create a reliable dictionary.
- ✚ They are very strict and domain-dependent in that a word is labeled as the same no matter the context. For instance, the term “amazing” can be either positive or negative, depending on the context.
- ✚ They are prone to human bias. For instance, if the people preparing the dictionary don't have sufficient domain knowledge, the method won't yield accurate results.

‡ As the labeling is handled manually, data preparation can be time-consuming.

Conclusion :

We hope through this article, you got a basic of how Sentimental Analysis is used to understand public emotions behind people's tweets. As you've read in this article, Twitter Sentimental Analysis helps us preprocess the data (tweets) using different methods and feed it into ML models to give the best accuracy.