Emotion Recognition by Textual Tweets Classification Using Voting Classifier LR-SGD

ABSTRACT

The proliferation of user-generated content on social media has made opinion mining an arduous job. As a microblogging platform, Twitter is being used to collect views about products, trends, and politics. Sentiment analysis is a technique used to analyze the attitude, emotions and opinions of different people towards anything, and it can be carried out on tweets to analyze public opinion on news, policies, social movements, and personalities. By employing Machine Learning models, opinion mining can be performed without reading tweets manually. Their results could assist governments and businesses in rolling out policies, products, and events. Seven Machine Learning models are implemented for emotion recognition by classifying tweets as happy or unhappy. With an in-depth comparative performance analysis,

it was observed that proposed voting classifier(LR-SGD) with TF-IDF produces the most optimal result with 79% accuracy and 81% F1 score. To further validate stability of the proposed approach on two more datasets, one binary and other multi-class dataset and achieved robust results.

**EXISTING SYSTEM**

* Sarlan *et al.* [2] established a sentiment analysis through extracting number of tweets with the help of prototyping and the results organized customers' views via tweets into positive and negative. Their research divided into two phrases. The first part is based on literature study which involves the Sentiment analysis techniques and methods that nowadays are used. In the second part, the application necessities and operations are described preceding to its development.
* In another research Alsaeedi and Zubair Khan [3] analyzed various kinds of sentiment analysis that is applied on to Twitter dataset and its conclusions. The distinct approaches and conclusions of algorithm performance were compared. Methods were used which were supervised ML based,, lexicon-based, ensemble methods. Authors used four methods that were Twitter sentiment Analysis using Supervised ML Approaches; Twitter sentiment Analysis using Ensemble Approaches. Twitter sentiment Analysis is using lexicon based Approaches.
* Lexicon based approaches have been explored by many researchers for emotion classification. Bandhakavi *et al.* [4] performed emotion-based feature extraction using domain specific lexicon generation.

Disadvantages

* The existing model which is ensemble of LR and SGD is not applied on both dataset and the results.
* Voting Classifier(VC) is not a cooperative learning which engages multiple individual classifiers.

**PROPOSED SYSTEM**

* In the proposed system, different techniques have been used for methodology in ML for its objectives. Versatile experiments were examined using different methods and techniques. Multiple classifiers applied on the dataset, but the Voting classifier is an ensemble of Logistic Regression and Stochastic Gradient Descent outperforms than all other ML models in terms of accuracy, recall, precision and F1-score.
* Twitter dataset used in this experiment is scrapped from Kaggle repository. First the dataset is pre-processed by removing unwanted data. Then, the data was split into two sets: training set and testing set. The training set was given the percentage of 70% while the test set portion is 30%. After that feature engineering techniques are applied on the training set. Multiple machine learning classifiers are trained on the training set and tested using the test set. The evaluation parameters used in this experiment are: (a) Accuracy (b) Recall (c) Precision (d) F1-score.

**Advantages**

* The proposed system presents a voting classifier (LR-SGD) and aims to estimate the performance of famous ML classifiers on twitter datasets.
* Data Visualization helps to understand the hidden patterns lying inside the dataset. It helps to qualitatively get more details about the dataset by visualizing the characteristics of the attributes.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).