Emotion Detection From Tweets

Baseline Result

Kapuriya Janak (MT22032)
Kirtirajsinh Mahida (MT22104)
Mansi Patel (MT22109)
Tanay (2020140)
Vijay (2020582)
Abhisek (2020014)

➢ Updated Problem formulation

Social media platforms will feed similar exciting posts, such as Instagram suggest reels, YouTube recommends videos, and Twitter refers to all tweets that belong to a person's past tweets and interests. But sometimes recommendation system fails to recommend appropriate tweets. An AI system cannot understand the feeling behind a person's tweet as a human can. So, we aim to build an algorithm that can detect various emotions from tweets.

➢ Literature Review

The research paper uses Sentiment analysis on Twitter data. The study compares the different Machine Learning models like Naive Bayes, Random Forest, Decision Tree, and Voting Classifier. Voting Classifier is a cooperative learning model that combines the predictions of multiple classifiers for better performance. Here, the LR-SGD model is used for ensemble Logistic Regression and Stochastic Gradient Descent. The dataset used has every record labeled happy or unhappy, and data visualization is used to understand patterns in the dataset also; for feature extraction, TF, and TF-IDF techniques are used to convert textual features into vector form. After testing the above-mentioned algorithms using TF and TF-IDF features, it was found that the Voting Classifier(VC) as an Ensemble of LR and SGD gave the highest accuracy. From the comparison done, it was found that the highest accuracy was given by VC with 79%, followed by LR(Linear Regression) with 78%, the highest precision value of 79% was achieved by 79%, and VC achieved 78%. Vc

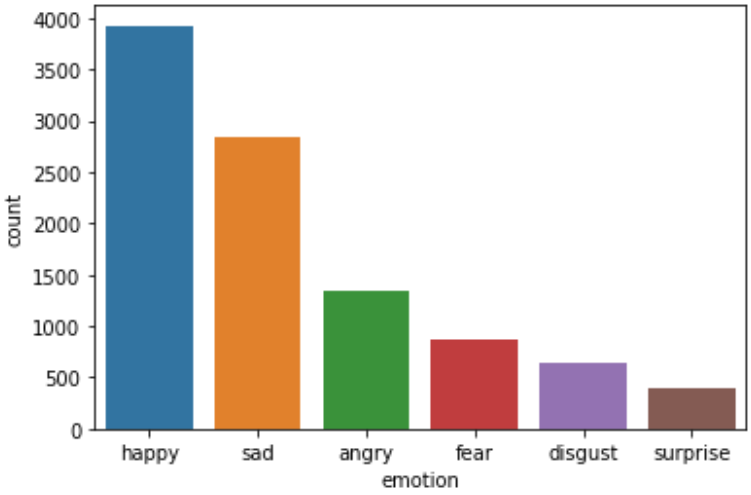
achieved the highest Recall and F1 scores with 84% and 81%, respectively, and LR showed 80% recall and 80% F1-Score. This study utilized combinations of ML models as VC, and it was clearly seen that VC(LR-SGD) outperformed the traditional ML-based models.

➢ Baseline Results

Data Preprocessing

The steps included are Lower casing the data, removing unwanted columns, removing all the URLs present in the tweet, removing punctuations, removing numbers and emoji's, removing stop words, and finally, applied lemmatization to the tweets.

Data Visualization



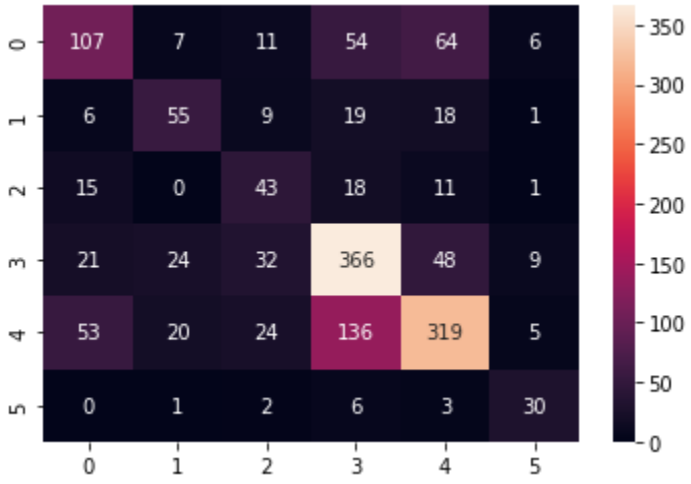
Methods and Models Result

We have used the Word2Vec method for converting tweet text into a vector of numbers. Word2vec uses context of neighboring words to make word embedding that will help model to learn semantic of word in the better way.

Naive Bayes

```
Accuracy: 0.5958549222797928
precision: 0.6095822417691428
recall: 0.5958549222797928
fscore: 0.5972450171800386
```

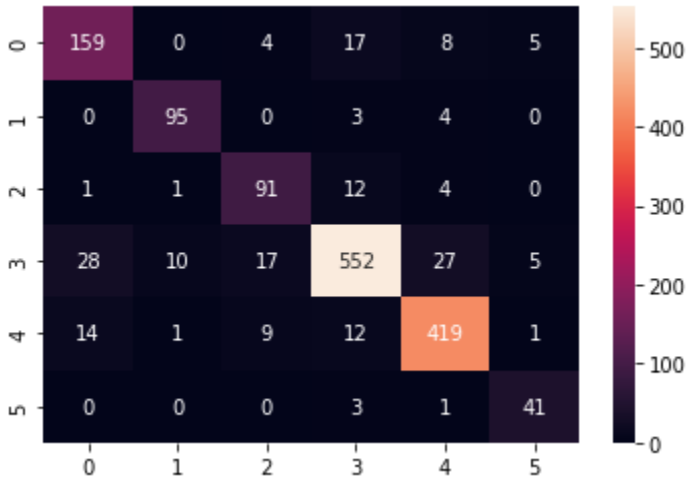
Confusion Matrix



SVM

```
Accuracy: 0.8788860103626943
precision: 0.8791097952411145
recall: 0.8788860103626943
fscore: 0.87821139603388
```

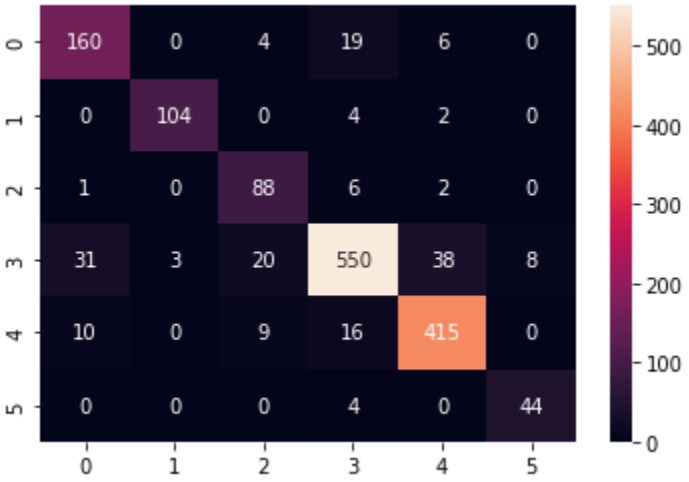
Confusion Matrix



Decision Tree

```
Accuracy: 0.8814766839378239
precision: 0.8830594771029311
recall: 0.8814766839378239
fscore: 0.8806872428799765
```

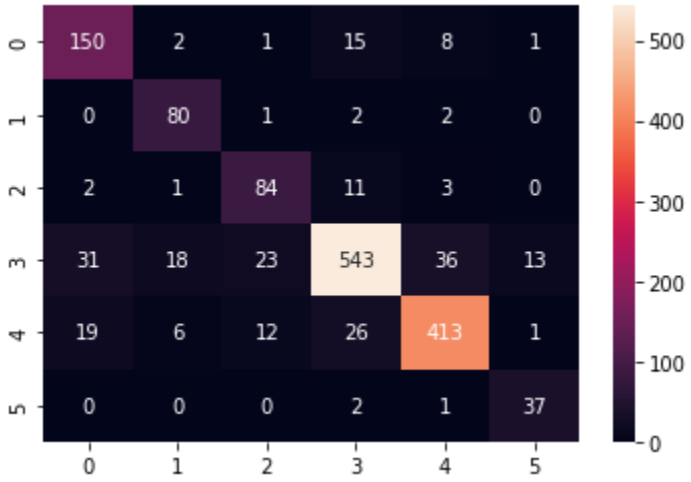
Confusion Matrix



Neural Network Model

```
Accuracy: 0.8465025906735751
precision: 0.8493194770880256
recall: 0.8465025906735751
fscore: 0.8447916763554727
```

Confusion Matrix



Out of all the models trained and tested, the Decision tree has the highest F1 score of 88%.

References

[1] A. Yousaf et al., "Emotion Recognition by Textual Tweets Classification Using Voting Classifier (LR-SGD)," in IEEE Access, vol. 9, pp. 6286-6295, 2021, doi: 10.1109/ACCESS.2020.3047831.