

Capstone Project Submission

Team Member's Name, Email, and Contribution:

1 Sandesh Gaikwad
Email:- sandeshg3.14159@gmail.com
Contribution: - Everything(individual Project)

Please paste the GitHub Repo link.

Github
Link:-<https://github.com/GaikwadSandesh/Coronavirus-Tweet-Sentiment-Analysis>

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)

Problem: Use the tweet data and build a machine learning model to classify the tweets based on the sentiments.

Approach: First Handel all null values. Remove irrelevant columns. Remove user tags, and links, Remove stop words and punctuations. Then use stems to reduce vector columns. use vectorizer to create vectorized data. Build machine learning models.

Conclusions / Recommendations:

- Naive Bayes was used as the baseline model, Catboost gives the best performance among all models with 59% accuracy
- Catboost also gives the smallest RMSE and 76% accuracy if we consider only 3 labels
- The models from scikit learn do not support ordinal errors. The current classification weights every error in the same way but some sentiments are closer to other that needs to be considered. The use of classification models with ordinal errors will improve performance.