

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
HRITIK SHARMA Email- Hritik.2.sharma@gmail.com Contribution- <ol style="list-style-type: none">1. Data cleaning2. Data preprocessing3. Perform story generation from visualization4. Extracting features to perform analysis5. Perform counter vectorizer for multiclass classification6. Naive Bayes Classifier for multiclass Classification7. Stochastic Gradient Descent-SGD Classifier(MULTICLASS CLASSIFICATION)8. RANDOM FOREST CLASSIFIER (For Multiclass Classification)9. Extreme Gradient Boosting (For Multiclass Classification)10. Support vector machine(For Multiclass Classification)11. Logistic Regression(For Multiclass Classification)12. Presentation of the final report
Please paste the GitHub Repo link.
Github Link:- https://github.com/Link/to/Repo
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 Description

This challenge asks you to build a classification model to predict the sentiment of COVID-19 tweets. The tweets have been pulled from Twitter and manual tagging has been done then.

The names and usernames have been given codes to avoid any privacy concerns.

You are given the following information:

Approach To Analyze Various Sentiments

Before we proceed further, One should know what is mean by Sentiment Analysis. Sentiment Analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the people attitude towards a particular topic is Positive, Negative, or Neutral.

Following is the Standard Operating Procedure to tackle the Sentiment Analysis kind of project. We will be going through this procedure to predict what we supposed to predict!

1. **Exploratory Data Analysis.**
2. **Data Preprocessing.**
3. **Vectorization.**
4. **Classification Models.**
5. **Evaluation.**
6. **Conclusion.**

Conclusion

In this way, we can explore more from various textual data and tweets. Our models will try to predict the various sentiments correctly. I have used various models for training our dataset but some models show greater accuracy while some do not. **The best model for this dataset would be Stochastic Gradient Descent.**

