**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

| **Team Member’s Name, Email and Contribution:** |
| --- |
| **Suraj Kumar** ([surajkumar0892@gmail.com](mailto:surajkumar0892@gmail.com))   1. Data Understanding  * Data Wrangling * Data Analysis  1. Data Visualization: BarPlot, Line-Plot, Count-Plot, WordCloud 2. Feature Engineering: Data Preprocessing, TfidfVectorizer 3. Classification Algorithm Implementation:  * Logistic Regression * Random Forest Classifier * Naive Bayes Classifier * Support Vector Classifier (SVC)   **Shreya Ranjan** ([shreyasrivastav15@gmail.com](mailto:shreyasrivastav15@gmail.com))   1. Data Understanding  * Data Wrangling * Data Analysis  1. Data Visualization: BarPlot, Line-Plot, Count-Plot, WordCloud 2. Feature Engineering: Data Preprocessing, TfidfVectorizer 3. Classification Algorithm Implementation:  * Logistic Regression * Random Forest Classifier * Naive Bayes Classifier * Support Vector Classifier (SVC) |
| **GitHub Repo link:** |
| [**https://github.com/surajkumar089/Coronavirus-Tweet-Sentiment-Analysis**](https://github.com/surajkumar089/Coronavirus-Tweet-Sentiment-Analysis)  **https://github.com/Shreyaranjan16** |
| **Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)** |
| The virus (more precisely, a coronavirus) known as 2019 Novel Coronavirus (2019-nCoV) has been determined to be the origin of an outbreak of respiratory illnesses that was originally discovered in Wuhan, China. On March 11, 2020, the World Health Organization declared the virus to be pandemic. While some of us tried to resist this new strain, known as Covid-19, many of us did.  Sentiment analysis is the technique of computationally identifying and classifying opinions stated in a piece of text, particularly to ascertain if the person has a Positive, Negative, or Neutral attitude toward a given topic.  The sentiments of COVID-19 tweets will be predicted in this project. In order to implement this project, I'll use the data gathered from Tweeter & use the python environment to implement this project.  **Approach:**   1. Importing the data set was the first step in applying it to perform a descriptive analysis to learn more about the information in the data. 2. checked the provided data set for any missing values. 3. Performed EDA and Data Preprocessing. 4. Applied different ML Classification Algorithm:  * Logistic Regression * Random Forest Classifier * Naive Bayes Classifier * Support Vector Classifier (SVC)   Words like COVID19, grocery, supermarket, shop, price, etc. are frequently used in tweets, indicating that throughout the pandemic, individuals were mostly concerned about food supply and their costs.  Support vector classifier(SVC) has performed slightly better than the Logistic regression and got the highest test accuracy score around 60%.  Multinomial Naive Bayes performed the worst with a test accuracy score of just 0.35. |