**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Sneha Raikar  [sneharikar650@gmail.com](mailto:sneharikar650@gmail.com)  Individual Project |
| **Please paste the GitHub Repo link.** |
| Github Link:- <https://github.com/sneraikar/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 Statement**  The Corona Virus disease that currently affects the world, cause serious problems to the population at all levels: economic, emotional, status, planning, politics, etc., in addition to the complexity of traditions, ethics, individual psychology and social behavior of people. Therefore, it is required and necessary a people's attitudes analysis when adverse situations arise Identifying people's reaction to this threat can provide important information on how society behaves and reacts to unwanted and unexpected situations, which can be positive or negative, currently the Internet and social networks have become powerful tools to access people’s opinions and comments on various topics  The main objective is to make a predictive model, which could help in predicting the Sentiment of a tweets.  **Approaches involved**   1. Importing Analytical necessary library classes for future analysis. 2. Reading the csv data file from Google drive. 3. Setting figure size for future visualization. 4. Removing future warnings in seaborne plots. 5. Visualizing all the columns of the respective Data frame. 6. Viewing all data information 7. Checking the Unique values in the column (if any) 8. Converting the data types to similar objects as the Analysis Demands. 9. Formatting the “size” column into a single column in the dataset. 10. Eradicating special characters from the dataset columns.   **ALGORITHMS** **Naive Bayes****Stochastic Gradient Descent**  1. **Random Forest** 2. **Support Vector Machine** 3. **Logistic Regression**   **HYPER PARAMETER TUNING**   * Grid Search CV * Randomized Search CV * Bayesian Optimization   **CONCLUSION**   * Considering that the COVID-19 disease is global health problem and has affected most countries and their economies, this model focuses on analysing people’s reaction to the pandemic. * The main goal of the model is to deduce whether the sentiment of the public opinion is positive or negative by applying machine learning algorithms and NLP techniques. * Even though the analysis found variation of opinions, it seems that people mostly remain positive about the pandemic * January is the only month in which negative thoughts predominated, March is the month when the COVID-19disease was declared as a pandemic and many countries started to apply care measures and safety protocols, which coincides with the rise of positive thoughts. * To summarize, 62% of the users showed positive feelings and 38% of the users showed negative feelings. |