**Predictive Modelling for COVID-19**

*Overview*

This repository houses the code and documentation for a machine learning project centered on predicting COVID-19 cases. The project delves into the influence of different symptoms, age factors, and known contacts on the likelihood of testing positive for COVID-19.

*Project Structure*

The project is structured into key sections:

1. Data Exploration and Preprocessing: Initial exploration of the dataset, handling missing values, and preprocessing steps.

2. Feature Selection: Utilizing SelectKBest and Recursive Feature Elimination (RFE) to select the most relevant features for predictive modelling.

3. Model Evaluation: Assessing the performance of various machine learning models, including Logistic Regression, Decision Tree, Random Forest, and K-Nearest Neighbors.

4. Model Selection and Optimization: Justifying the selection of the final model (Random Forest Classifier) based on metrics and ensemble learning benefits. Hyperparameter tuning using RandomizedCV is also discussed.

5. Trend Analysis Over Time: Incorporating a trend analysis over time, considering the monthly count of COVID-19 positive cases and their ranks.

6. Conclusion: Summarizing key findings, hypotheses, and the overall significance of the project. Highlighting insights beyond numerical metric