Group 2 – Project 4 Proposal

Topic: Analyzing COVID-19 Variants and Their Impact on Outbreaks

Team Members:

- Amit Kumar
- Muneet Pruthi
- Rushad Confectioner

Objective:

Our goal is to develop a chatbot that analyzes the impact of different COVID-19 variants on populations. By studying how SARS-CoV-2 variants and mutations influence the duration and spread of local outbreaks, we aim to provide valuable insights for researchers and data scientists. This project will combine variant and epidemiological data to identify key factors shaping the course of the pandemic.

Dataset:

Kaggle COVID-19 Variants Survival Data

This dataset offers a resource for analyzing the global dynamics of COVID-19. It focuses on how different variants and mutations affect the length of local outbreaks. By integrating variant data with epidemiological trends, we can uncover significant patterns influencing the trajectory of the pandemic.

Research Questions & Assignments:

Amit Kumar

- Can we develop a machine-learning model to predict the duration of an epidemic?
- What features have the best predictive power?

Rushad Confectioner

- Can we identify clusters of variants or regions with similar outbreak patterns?
- Are there interactions between variables that explain complex epidemic trends?

Muneet Pruthi

- Is there a geographical pattern in outbreak duration?
- Does the number of variants in a country impact the length of an outbreak?

Methodology:

1. Data Collection

- Retrieve relevant datasets from Kaggle.

2. Data Cleaning

- Process and clean data using Pandas to ensure consistency and handle missing values.

3. Data Exploration

- Analyze key trends using Matplotlib to gain insights into relationships between variables.

4. Data Visualization

- Develop an interactive web-based visualization using Flask to present findings.

5. Machine Learning Model

- Utilize Scikit-learn (sklearn) to train a model predicting which COVID-19 variants had the most impact.

6. Model Evaluation

- Assess model accuracy and performance using standard evaluation metrics.

7. Conclusion

- Determine the scale of impact in terms of affected populations and the geographical spread of COVID-19 variants.

Tools & Technologies:

- **Data Processing & Analysis:** SQL, Pandas, GeoPandas
- **Visualization:** Matplotlib
- **Web Development:** HTML, CSS, Bootstrap, Flask
- **Machine Learning:** Scikit-learn (sklearn), ML models

GitHub Repository:

Group 2 Project 4 Repository

This project aims to provide a deeper understanding of COVID-19 variants, their geographical spread, and their influence on outbreak duration using data science and machine learning techniques.