Reuben Martor

DSC 680 – Applied data Science

Milestone 2/Writeup

Ten (10) Potential questions audience may ask

1. What is the primary objective of this project?

2. What kind of data did you use for this model, and where was it sourced from?

3. How did you preprocess the data before using it for model training?

4. Which machine learning models did you evaluate, and why did you choose RandomForestClassifier?

5. How does the Random Forest Classifier help avoid overfitting?

6. What performance metrics were used to evaluate the model?

7. Did you encounter any challenges during the model development phase?

8. How did you handle symptoms that were not present in every disease case in the dataset?

9. What role does Principal Component Analysis (PCA) play in this project?

10. How do you ensure the system’s global applicability, especially in different geographical regions?