Viva Questions with Diagrams & Highlights

Q1. Why did you choose Random Forest instead of boosting algorithms like XGBoost or AdaBoost?

Random Forest is simpler, needs less hyperparameter tuning, and handles categorical + numerical data natively. Boosting is more accurate in some cases but harder to tune and more prone to overfitting.

(Diagram could not be displayed - image file missing)

Q2. Why are we not performing the correlation matrix?

Random Forest does not require independent or uncorrelated features since it selects feature subsets at each split. Correlation analysis is more important in linear models.

Q3. Is Eigenvalue and Eigenvector important in this algorithm?

No. They are used in PCA or dimensionality reduction. Random Forest is a non-linear, rule-based model that builds trees using thresholds, not matrix decomposition.

Q4. Why are we not performing any data cleansing method?

Random Forest handles missing values well and is robust to noise. The dataset was relatively clean, and minimal preprocessing (like encoding) was sufficient.

Q5. What is the bias-variance trade-off in Random Forest?

Random Forest lowers variance by averaging over many uncorrelated trees. Individual trees might overfit, but the ensemble prediction is more stable and generalizable.

(Diagram could not be displayed - image file missing)