

Iteration #3

Ishwarya Rajkumar - Data Preparation and Exploration (EDA)

1. **Task:** Finish cleaning and preprocessing the dataset, including handling missing values, encoding categorical variables, and feature scaling.
 - **Expected Output:** A clean dataset ready for model training.
2. **Task:** Conduct exploratory data analysis (EDA) to identify patterns, outliers, and correlations that may influence the model.
 - **Expected Output:** Visualizations and reports on data insights (including histograms, scatter plots, correlation heatmaps).

Afrah Fathima - Model Building and Evaluation

1. **Task:** Train and evaluate baseline models (Logistic Regression and Random Forest) on the preprocessed dataset, focusing on key evaluation metrics like accuracy, precision, recall, and ROC-AUC.
 - **Expected Output:** Trained models and evaluation results.
2. **Task:** Perform cross-validation to ensure the models are not overfitting and provide reliable predictions.
 - **Expected Output:** Cross-validation results (with metrics for each fold).

Lakshmi Mahadevan - Hyperparameter Tuning and Model Optimization

1. **Task:** Use GridSearchCV and RandomizedSearchCV to optimize hyperparameters of the Logistic Regression and Random Forest models.
 - **Expected Output:** Optimized models with the best-performing hyperparameters.
2. **Task:** Monitor for overfitting during model optimization and apply techniques like early stopping or regularization if necessary.
 - **Expected Output:** Optimized model ready for validation with reduced overfitting.

All Members - Integration and Final Testing

1. **Task:** Integrate all components (data preprocessing, model training, and evaluation) into a final pipeline that can be executed end-to-end.
 - **Expected Output:** A functioning, integrated pipeline ready for testing.
2. **Task:** Finalize project documentation and prepare for Iteration 3 submission.
 - **Expected Output:** Final report and code review complete.