After conducting three experiments using Logistic Regression, Decision Tree, and Random Forest on the Pima Indians Diabetes dataset, we observed the following:

- Random Forest performed best overall with the highest accuracy and balanced F1 scores

- Experiment 1 (Feature Scaling) had little to no effect on Logistic Regression and Random Forest, but slightly improved Decision Tree performance

- Experiment 2 (Train-Test Split Ratio) helped Decision Tree accuracy and F1 score, but decreased performance in Random Forest

- Experiment 3 (Regularization) showed negligible changes in Logistic Regression performance where L1 and L2 regularization did not significantly improve results

- Decision Trees and Random Forests cannot use L1/L2 regularization as they don't rely on coefficients

- Challenges included :

    understanding which hyperparameters act as regularization in tree-based models

- Choosing the right model and tuning strategy based on data characteristics is more impactful than applying standard preprocessing steps blindly