**Sunil Sah**

**Project 4 Loan Prediction Report**

The analysis aimed to predict loan approval status (Loan\_Status) using historical data by implementing Logistic Regression and Random Forest Classifier models. The dataset was explored, revealing missing values and a mix of numerical and categorical features, which were handled through imputation and encoding. Key features, such as Income\_to\_Loan\_Ratio and Family\_Size, were engineered to enhance model predictions.

Correlation analysis identified relationships between features, aiding in feature selection. After splitting the data into training and testing sets, Logistic Regression served as the baseline model, achieving moderate performance with an F1 score of {Logistic Regression F1 Score}. The Random Forest model outperformed Logistic Regression with higher accuracy, precision, recall, and an F1 score of {Random Forest F1 Score}, indicating better generalization and predictive power. Visualizations, including a confusion matrix heatmap, confirmed the model’s enhanced performance.

The results suggest that Random Forest is more suitable for this dataset due to its ability to capture complex patterns, making it a strong candidate for deployment in real-time loan prediction systems. Further improvements, such as hyperparameter tuning and advanced preprocessing, could further boost model accuracy.