Internship ML Tasks Report

NAME: Mohammad Mansoor Khan

(Data Science & Analytics)

DHC-1360

# Task 1: Exploring and Visualizing the Iris Dataset

This task involved exploring the classic Iris dataset. Visualizations such as scatter plots, histograms, and box plots were used to understand relationships between sepal and petal dimensions across different species. This analysis helps in understanding class separability and feature distribution.

# Task 2: Credit Risk Prediction

In this task, we worked on a credit risk prediction problem using a test dataset. The data underwent cleaning, feature engineering, and label encoding. A simulated binary target was created based on credit history. Logistic Regression and Decision Tree models were trained, and their performance was evaluated using accuracy, confusion matrix, and classification reports.

# Task 3: Customer Churn Prediction

Using the 'Churn\_Modelling.csv' dataset, this task predicted whether customers would leave a bank. After cleaning, encoding, and scaling, a Random Forest Classifier was trained. Key features affecting churn were visualized using feature importance and SHAP values. The model demonstrated reasonable predictive capability and helped identify high-risk customers.

# Task 4: Predicting Insurance Claim Amounts

This task aimed to predict medical insurance charges using the 'insurance.csv' dataset. The data was encoded and scaled. A Linear Regression model was trained, and performance was evaluated using MAE and RMSE. Residual plots and scatter plots of predicted vs actual charges helped in model diagnostics.

# Task 5: Personal Loan Acceptance Prediction

This task predicted whether a bank customer would accept a term deposit offer. After encoding and scaling the 'bank.csv' data, Logistic Regression and Decision Tree models were trained. Performance was assessed, and feature importance analysis was conducted. It was found that call duration, month of contact, and previous campaign success were strong predictors of acceptance.