

Machine Learning – Assignment 2

Bank Marketing Classification – End-to-End ML Deployment

Name: Ravindra Kasukurthi
BitsId: 2025AA05599
Program: M.Tech (AIML/DSE)
Course: Machine Learning

GitHub Repository Link:

<https://github.com/2025aa05599-WILP-Bits-Pilani/MachineLearning-Assignment-2.git>

Live Streamlit Application Link:

<https://2025aa05599machinelearning-assignment-2-zemb2csvshwxvnbkvk8appj.streamlit.app/>

Problem Statement

The objective of this assignment is to implement and compare multiple classification models on a real-world dataset. The models are evaluated using Accuracy, AUC, Precision, Recall, F1 Score, and Matthews Correlation Coefficient (MCC). The project is deployed as an interactive Streamlit web application.

Dataset Description

Dataset: Bank Marketing Dataset (UCI Repository)
Total Instances: 41,188
Total Features: 20 input features + 1 target variable
Target Variable: y (yes/no)

Models Implemented

1. Logistic Regression
2. Decision Tree Classifier
3. K-Nearest Neighbors
4. Naive Bayes (Gaussian)
5. Random Forest (Ensemble Model)
6. XGBoost (Ensemble Model)

Model Comparison Table

Model	Accuracy	AUC	Precision	Recall	F1	MCC
Logistic Regression	0.9139	0.9370	0.7001	0.4127	0.5193	0.4956
Decision Tree	0.9180	0.9334	0.6794	0.5161	0.5866	0.5484
KNN	0.9053	0.8616	0.6267	0.3943	0.4841	0.4491
Naive Bayes	0.8536	0.8606	0.4023	0.6174	0.4872	0.4189
Random Forest	0.9196	0.9521	0.7180	0.4719	0.5695	0.5414
XGBoost	0.9207	0.9543	0.6811	0.5571	0.6129	0.5728

Observations

XGBoost achieved the best overall performance with highest AUC, F1 Score, and MCC. Random Forest also performed strongly. Naive Bayes showed higher recall but lower precision. Logistic Regression and Decision Tree provided balanced baseline performance. The dataset imbalance makes MCC and Recall important metrics for evaluation.