

Machine Learning – Assignment 2

Bank Marketing Classification – End-to-End ML Deployment

Name: Ravindra Kasukurthi

BitsId: 2025AA05599

[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)

Table 1: Model Comparison

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

Table 2: Model Observations

Model	Observation
Logistic Regression	High AUC but lower recall for minority class.
Decision Tree	Improved recall compared to Logistic Regression.
KNN	Moderate performance with lower AUC.
Naive Bayes	High recall but low precision (more false positives).
Random Forest	Balanced precision and recall.
XGBoost	Best overall performance with highest AUC, F1 and MCC.

BITS Virtual Lab Screenshots

This screenshot shows a browser window with multiple tabs. The main content area displays a Streamlit application titled "Bank Marketing Classification App". On the left, there's a sidebar with options like "Select Classification Model" (set to "Logistic Regression") and "Upload Test Dataset (CSV)". Below that, a file named "bank-additional-full.csv" is listed with a size of 5.8MB. A green banner at the bottom indicates "Model in Use: Logistic Regression". The right side of the screen shows a code editor with an "app.py" file containing Python code for a Streamlit app. The code imports Streamlit, pandas, joblib, and various scikit-learn models. It sets up a Streamlit interface with a title, markdown, and info sections. It also defines a function to load models from files. The Streamlit configuration section includes settings for page title, cache resources, and gathering usage statistics. A note at the bottom says "You can now view your Streamlit app in your browser".

```
import streamlit as st
import pandas as pd
import joblib
from sklearn.metrics import (
    accuracy_score,
    roc_auc_score,
    precision_score,
    recall_score,
    f1_score,
    matthews_corrcoef,
    confusion_matrix,
    classification_report
)
from sklearn.preprocessing import LabelEncoder
```

```
# python
# motions, and the occasional swag, please enter your email address below. Otherwise leave this field blank.
Email: 2025aa05599@ilp.bits-pilani.ac.in
You can find our privacy policy at https://streamlit.io/privacy-policy
Summary:
- This open source library collects usage statistics.
- We cannot see and do not store information contained inside Streamlit apps, such as text, charts, images, etc.
- Telemetry data is stored in servers in the United States.
- If you'd like to opt out, add the following to ./streamlit/config.toml, creating that file if necessary:
[browser]
gatherUsageStats = false

You can now view your Streamlit app in your browser.
Local URL: http://localhost:8501
Network URL: http://172.31.96.6:8501
/home/cloud/anaconda3/lib/python3.12/site-packages/sklearn/base.py:463: InconsistentVersionWarning: Trying to unpickle estimator LogisticRegression from version 1.6.1 when using version 1.8.0. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability
warnings.warn(
/home/cloud/anaconda3/lib/python3.12/site-packages/sklearn/base.py:463: InconsistentVersionWarning: Trying to unpickle estimator LogisticRegression from version 1.6.1 when using version 1.8.0. This might lead to breaking code or invalid results. Use at your own risk. For more info please refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability
Ln1, Col1 Spaces:4 UTF-8 LF (Python 3.9.19)
```

This screenshot shows the deployed Streamlit application at localhost:8501. The title "Bank Marketing Classification App" is visible at the top. The "Compare Multiple Machine Learning Models" section contains instructions to "Upload the dataset once and switch between models to compare performance". Below this, there are two sections: "Select Classification Model" (set to "Logistic Regression") and "Upload Test Dataset (CSV)". A file named "bank-additional-full.csv" is listed with a size of 5.8MB. A green banner at the bottom indicates "Model in Use: Logistic Regression". The "Unloaded Data Preview" section is shown below the upload area.

The screenshot shows a web browser window with multiple tabs open. The active tab is titled "2025AA05599 ML Assign..." and displays a Streamlit application. The application has a header with the title "Compare Multiple Machine Learning Models". Below the header, there is a blue button with the text "Upload the dataset once and switch between models to compare performance.". A dropdown menu labeled "Select Classification Model" is open, showing "Logistic Regression" as the selected option. Below this, there is a section for "Upload Test Dataset (CSV)" with a file input field containing "bank-additional-full.csv" (5.8MB). At the bottom of the application, a green bar indicates "Model in Use: Logistic Regression".

Compare Multiple Machine Learning Models

Upload the dataset once and switch between models to compare performance.

Select Classification Model

Logistic Regression

Upload Test Dataset (CSV)

Drag and drop file here
Limit 200MB per file • CSV

bank-additional-full.csv 5.8MB

Model in Use: Logistic Regression