Project Report: Predicting Client Subscription to Term Deposits

Objective

The goal of this project was to build a machine learning model capable of predicting whether a client would subscribe to a term deposit, based on a dataset from a direct marketing campaign of a banking institution. This solution aims to support marketing teams in making data-driven decisions and improving campaign targeting.

Data Overview

The dataset contained client information such as job type, marital status, education level, loan history, contact details, and previous campaign outcomes. Additional engineered features, such as age and balance groups, were introduced to enhance the model's predictive power.

Key preprocessing steps included:

- • Encoding categorical variables
- • Feature scaling
- • Feature engineering: grouping age and balance into meaningful bins
- • Balancing the dataset using SMOTE to address class imbalance

Methods and Models

Three classification algorithms were evaluated:

1. Logistic Regression – A baseline linear model to assess basic predictability
2. Random Forest Classifier – An ensemble tree-based method known for its robustness
3. XGBoost Classifier – A gradient boosting method optimized for performance

Two datasets were used to train and test the models. All models were trained on the same preprocessed data.

Evaluation Metrics Used:

- • Accuracy – Overall correctness
- • Precision – Correct positive predictions over all positive predictions
- • Recall – Correct positive predictions over all actual positives
- • F1 Score – Harmonic mean of precision and recall

Model Performance Summary

Model	Precision (0)	Recall (0)	F1-Score (0)	Precision (1)	Recall (1)	F1-Score (1)	Accuracy
Logistic Regression	0.94	0.85	0.90	0.35	0.60	0.44	0.82
Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00
XGBoost	0.96	0.94	0.95	0.62	0.72	0.67	0.92

Insights & Recommendations

• Random Forest achieved perfect accuracy but showed signs of overfitting, making it unreliable for generalization however, XGBoost demonstrated the best balance between sensitivity (recall) and specificity, making it a robust choice for deployment.
• Features like duration of the last call, previous campaign outcome, and balance group had a significant impact on predictions.
• The relatively lower recall suggests that while the model is good at precision (minimizing false positives), there is room to improve sensitivity (capturing all actual positives).

Deployment

The XGBoost model was deployed using Streamlit, creating an interactive web application that:
• Accepts user inputs (via dropdowns and sliders with descriptive labels)
• Returns real-time predictions
• Can support marketing officers in screening clients likely to subscribe

Conclusion

This project successfully built and deployed a robust predictive model for client subscription using real-world marketing data. With a 92% accuracy and balanced performance across key metrics, the model is well-suited for practical business use in customer segmentation and campaign strategy.