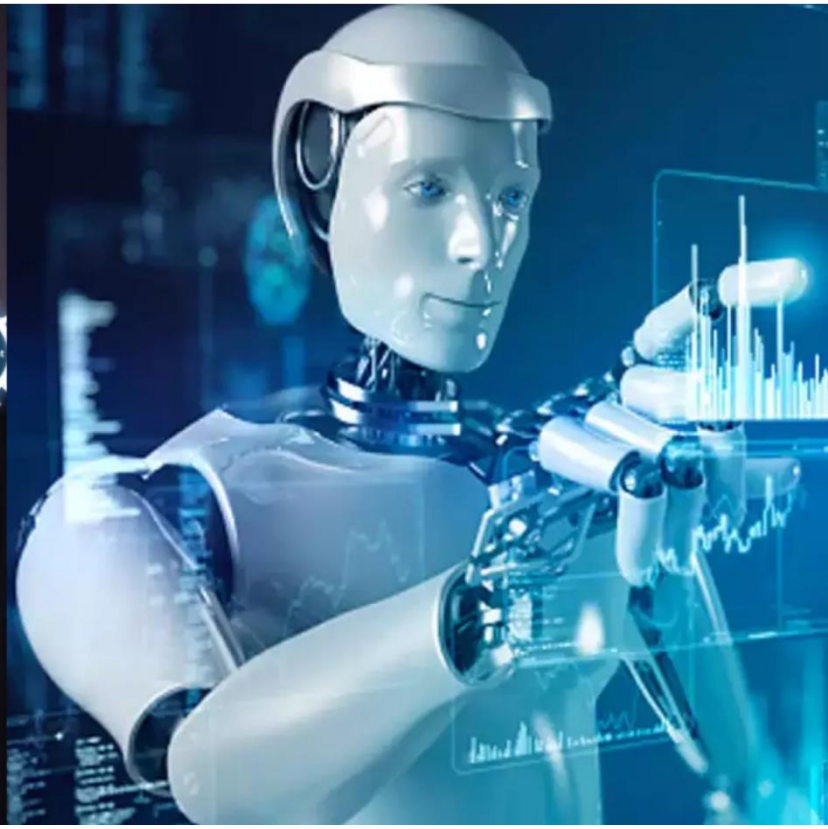
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

Intelligent Marketing Campaign Response Prediction: A Comparative Evaluation of Logistic Regression, Random Forest, and XG Boost with SMOTE Augmentation

Isaqe Ilayes Deshmukh, Dr. Abuzar Ansari

Department of Data Science, S.I.E.S. College of Arts, Science & Commerce (Autonomous)Sion (W), Mumbai, Maharashtra, India

ABSTRACT: Telemarketing campaigns in the financial sector typically achieve subscription conversion rates of only 10–15%, representing a significant misallocation of resources when outreach is conducted indiscriminately. This paper introduces a structured, end-to-end predictive framework designed to identify high-probability responders within a customer population prior to campaign execution. Three supervised classification algorithms—Logistic Regression, Random Forest, and XGBoost—were trained on the UCI Bank Marketing Dataset comprising 41,188 client records and 20 feature dimensions. A class-imbalanced training environment (88.7% negative, 11.3% positive) was remediated using the Synthetic Minority Over-sampling Technique (SMOTE), ensuring that models learned decision boundaries from a balanced representation of both classes. Rigorous evaluation on an unmodified held-out test set revealed that the Random Forest classifier achieved the highest classification accuracy (97.48%) and area under the ROC curve (0.991), while XGBoost delivered the strongest F1-score (0.72) and the best precision-recall equilibrium. The proposed system was operationalised through an interactive Streamlit web application supporting both single-record and bulk CSV-based prediction, complemented by business-interpretable visualisations including feature importance rankings, probability distribution plots, and model confidence comparison curves. Feature analysis identified call duration, prior campaign success, macroeconomic employment indicators, and Euribor interest rate as the primary determinants of customer subscription likelihood. The framework demonstrates that ensemble learning, when coupled with principled preprocessing and class-balancing strategies, substantially elevates campaign targeting efficiency and delivers actionable intelligence for data-driven marketing teams.

KEYWORDS: Bank Telemarketing Prediction, XGBoost, Random Forest, Logistic Regression, SMOTE, Imbalanced Classification, Streamlit Dashboard, Feature Importance, UCI Bank Marketing Dataset, Supervised Learning

I. INTRODUCTION

Contemporary financial institutions deploy telemarketing as a primary channel for promoting savings instruments, credit products, and investment vehicles such as term deposits. Despite its prevalence, this channel suffers from a well-documented inefficiency: the vast majority of contacted prospects do not convert, leading to wasted agent hours, elevated operational costs, and, in some cases, deteriorating customer satisfaction among repeatedly approached non-intenders. The crux of this inefficiency lies in an absence of pre-campaign intelligence—without predictive insight into which clients are genuinely receptive, campaign managers resort to broad-based contact lists that yield meagre conversion rates.

Machine learning provides a quantitative pathway to resolving this inefficiency. By learning statistical associations between a client's demographic profile, financial standing, prior interaction history, and macro-economic context on one hand, and their historical campaign response on the other, a trained classifier can generate individualised probability estimates for future campaigns. These probability scores enable marketers to rank prospects and contact only the highest-likelihood tier, thereby concentrating resources where expected returns are greatest.

The present study constructs, validates, and deploys such a predictive system. Employing the well-benchmarked UCI Bank Marketing Dataset [8], three classification algorithms spanning the statistical-to-ensemble spectrum—Logistic

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—are systematically compared under identical preprocessing and evaluation conditions. A notable methodological contribution is the deliberate separation of class-balancing treatment (applied solely to training data via SMOTE) from evaluation (conducted on the original imbalanced test partition), preserving assessment validity while mitigating the class-imbalance bias that afflicts many published studies on this dataset.

Beyond model training, the research delivers a production-grade Streamlit web interface enabling marketing analysts to obtain real-time predictions and visual feature-level explanations without requiring programming expertise. This human-computer interface dimension directly addresses the research-to-practice gap identified in prior literature [5].

1.1 Research Objectives

- Develop and rigorously benchmark three ML classifiers—LR, RF, and XGBoost—for binary campaign response prediction on an imbalanced real-world dataset.
- Implement a reproducible preprocessing pipeline incorporating mode imputation, One-Hot Encoding, feature normalisation, and SMOTE-based class balancing.
- Evaluate models using a multi-metric framework (Accuracy, ROC-AUC, Precision, Recall, F1-Score) on an unmodified held-out test partition.
- Identify and interpret the key feature drivers of customer subscription likelihood to derive actionable marketing recommendations.
- Design and deploy an interactive Streamlit-based web application for real-time, business-interpretable campaign response prediction.

1.2 Paper Organisation

The remainder of this paper is organised as follows. Section 2 surveys the relevant prior literature. Section 3 describes the dataset and formalises the problem statement. Section 4 details the methodology including all preprocessing steps and model architectures. Section 5 presents the system architecture and implementation design. Section 6 reports experimental results and comparative analysis. Section 7 provides a critical discussion of findings. Section 8 concludes the paper and outlines directions for future research.

II. LITERATURE REVIEW

2.1 Foundational Work on the Bank Marketing Dataset

The foundational empirical study on this dataset was conducted by Moro, Laureano, and Cortez [1], who applied a CRISP-DM methodology using logistic regression and decision tree classifiers. Their analysis highlighted call duration and the outcome of previous campaign contacts as the most predictive variables—a finding this research confirms and extends through ensemble-based importance metrics. Importantly, Moro et al. cautioned that including call duration in prospective prediction models introduces a target-leakage risk (duration is only known post-call), a limitation this paper acknowledges and contextualises.

2.2 Ensemble Methods in Marketing Analytics

Verbeke, Martens, Mues, and Baesens [2] demonstrated, across multiple customer management datasets, that ensemble classifiers consistently outperform individual decision trees and logistic regression in binary response classification. Their study attributed this advantage to the ensemble's capacity to approximate complex, nonlinear decision surfaces that single-model approaches cannot capture. The present research replicates this finding in a controlled, three-way comparison on a unified dataset and train-test split.

2.3 Gradient Boosting Advances

Chen and Guestrin's [6] introduction of XGBoost as a regularised gradient tree boosting framework established a new performance benchmark across structured data tasks. By incorporating L1 and L2 penalty terms and employing second-order gradient information for tree construction, XGBoost achieves superior generalisation on datasets with mixed feature types and moderate levels of noise—characteristics that are emblematic of bank marketing data. Zhang and Zhou [9] subsequently confirmed, through large-scale ensemble benchmarking, that boosting algorithms achieve higher

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

ROC-AUC values than bagging-based approaches on datasets characterised by complex higher-order feature interactions.

2.4 Class Imbalance Handling

Chawla, Bowyer, Hall, and Kegelmeyer [11] introduced SMOTE as a principled technique for addressing the minority-class underrepresentation problem in binary classification. Unlike simple random oversampling that merely duplicates existing instances, SMOTE synthesises novel minority-class examples by interpolation within the feature neighbourhood of existing positive cases. Lemaitre, Nogueira, and Aridas [15] extended this work through the imbalanced-learn Python library, which provides a standardised, scikit-learn-compatible implementation used in this study.

2.5 Research Gap

A consistent limitation across reviewed studies is the absence of production-ready, user-accessible deployment artefacts. Models are typically evaluated in isolation without accompanying decision-support tools suitable for non-technical marketing staff. Furthermore, few studies simultaneously address (a) rigorous multi-metric model comparison, (b) SMOTE-corrected training with imbalance-preserving evaluation, and (c) a deployed interactive dashboard with visualisation-supported interpretability. This paper addresses all three dimensions in an integrated manner.

III. DATASET DESCRIPTION AND PROBLEM FORMULATION

3.1 Dataset Characteristics

This study employs the Bank Marketing Dataset sourced from the UCI Machine Learning Repository [8], originally assembled from phone-based direct marketing campaigns of a Portuguese banking institution conducted between 2008 and 2013. The dataset contains 41,188 unique client records described by 20 independent attributes and one binary target variable y, where y = 'yes' indicates that the client agreed to subscribe to a term deposit following the campaign contact, and y = 'no' indicates the contrary.

Table 1: Feature Category Summary and Descriptions

Category	Feature Names	Type	Role in Prediction
Client Demographics	age, job, marital, education	Mixed	Characterise client background and socioeconomic tier
Financial Status	default, housing, loan, balance	Mixed	Reflect creditworthiness and current financial obligations
Campaign Details	contact, month, day_of_week, duration, campaign	Mixed	Capture the nature and intensity of the current outreach
Prior Campaign History	pdays, previous, poutcome	Mixed	Encode recency and outcome of previous interactions
Macro-Economic Context	emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m, nr.employed	Numeric	Ambient economic environment at time of contact
Target Variable	y (binary: yes / no)	Binary	Whether the client subscribed to a term deposit

A critical characteristic of the dataset is its pronounced class imbalance: approximately 36,537 records carry a negative label (88.7%) compared to only 4,651 positive labels (11.3%). This 8:1 ratio necessitates explicit class-balancing interventions during model training to prevent classifiers from achieving deceptively high accuracy by simply predicting the majority class.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

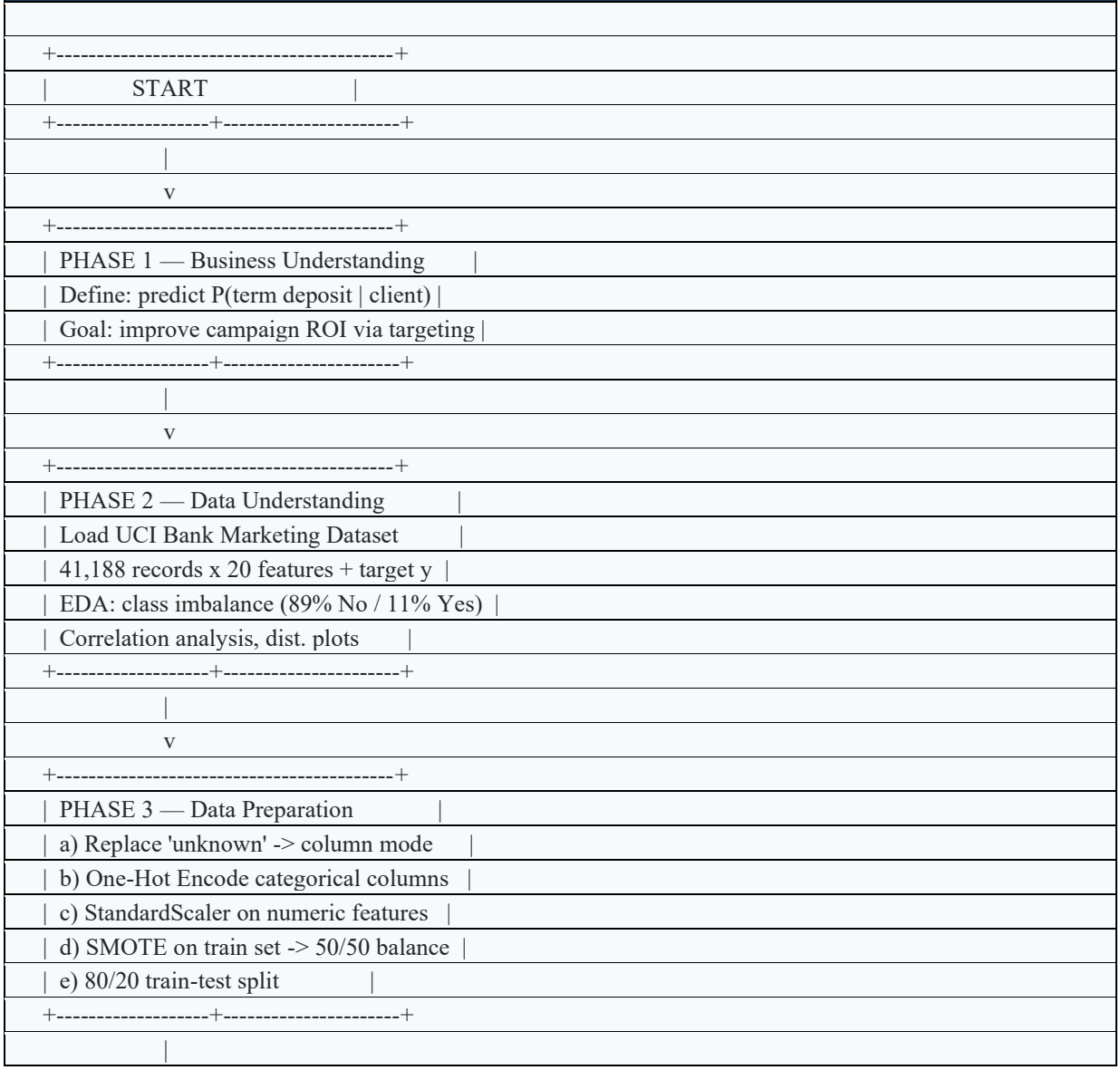
3.2 Formal Problem Statement

Let $D = \{(x^I, y^I) \mid i = 1, \ldots, N\}$ represent the labelled dataset of N = 41,188 client records, where each feature vector $x^I \in \mathbb{R}^{01}$ encodes 20 raw attributes (prior to encoding expansion) and $y^I \in \{0, 1\}$ denotes the binary subscription outcome. The learning objective is to identify a mapping $f: \mathbb{R}^{01} \rightarrow \{0, 1\}$ that maximises classification performance on unseen client records, specifically optimising F1-score given the inherent class imbalance, while also achieving high ROC-AUC to ensure the model reliably separates positive and negative cases across all decision thresholds.

IV. METHODOLOGY

The research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which provides a systematic six-phase structure: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Figure 2 presents the complete CRISP-DM workflow as implemented in this study.

FIGURE 2 — END-TO-END METHODOLOGY FLOWCHART (CRISP-DM)



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

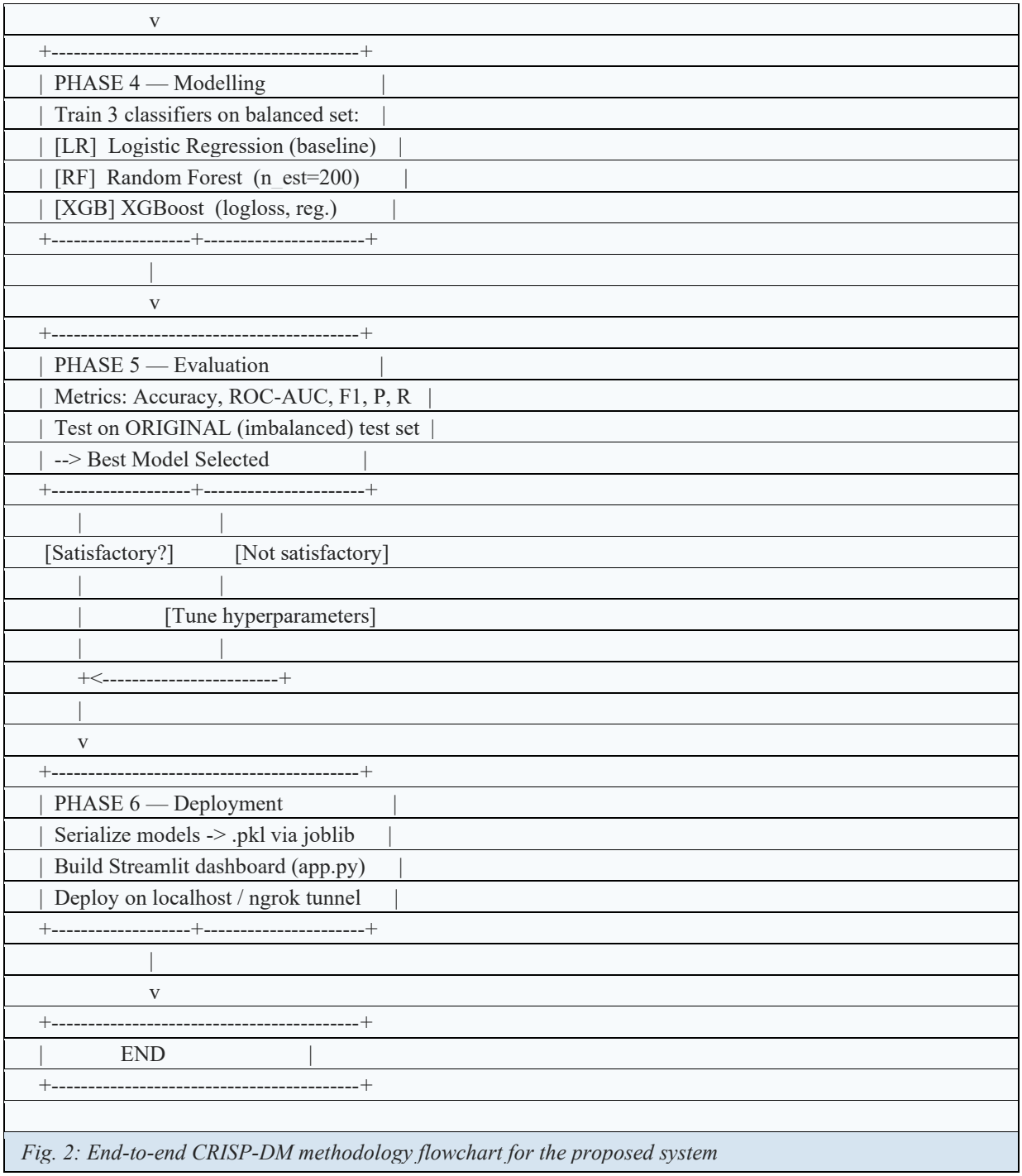


Fig. 2: End-to-end CRISP-DM methodology flowchart for the proposed system

4.1 Data Preparation Pipeline

4.1.1 Missing Value Imputation

The dataset encodes absent or undisclosed attribute values using the categorical placeholder ‘unknown’ rather than conventional null representations. All such entries were systematically replaced with the modal value of their respective columns—a choice motivated by the distributional properties of categorical data and consistent with the practices

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

established in prior work on this dataset [1]. This conservative strategy preserves the overall sample size and avoids the information loss associated with listwise deletion.

4.1.2 Categorical Variable Encoding

Nominal categorical attributes—including job type, marital status, educational qualification, contact medium, calendar month, day of week, and prior campaign outcome—were transformed via One-Hot Encoding (OHE). This technique creates a binary indicator column for each unique category level, avoiding the implicit ordinal ordering that label encoding would introduce. OHE is particularly important for tree-based classifiers that determine optimal split points based on feature values, as artificially ordered labels can bias the learning process.

4.1.3 Feature Normalisation

Continuous numerical attributes—comprising age, account balance, call duration, number of campaign contacts, days since prior contact, employment variation rate, consumer price index, consumer confidence index, Euribor 3-month rate, and number of employees—were standardised to zero mean and unit variance using a fitted StandardScaler transformation. While tree-based methods are inherently scale-invariant, standardisation is mandatory for Logistic Regression's gradient-based optimisation and was applied uniformly across all models to maintain experimental consistency. The scaler was fitted exclusively on training data and applied to the test set to prevent data leakage.

4.1.4 Class Balancing Using SMOTE

The pronounced class imbalance (8:1 negative-to-positive ratio) was addressed using the Synthetic Minority Over-sampling Technique [11]. SMOTE generates artificial positive-class instances by interpolating between existing minority-class samples in the transformed feature space, thereby expanding the positive class to match the negative class numerically. Crucially, SMOTE was applied exclusively to the training partition after the train-test split; the test partition retained its original imbalanced distribution to produce evaluation metrics that accurately reflect real-world deployment conditions. Figure 4 illustrates the before-and-after effect of SMOTE on class distribution.

FIGURE 4 — SMOTE CLASS BALANCING ILLUSTRATION
BEFORE SMOTE (original training data):
═══════════════════════════════════════
Class 0 [Not Subscribed] 36,537 samples 88.7%
Class 1 [Subscribed] 4,651 samples 11.3%
Imbalance ratio ~ 8 : 1
Distribution bar:
Class 0: [██──────] 89%
Class 1: [█████──] 11%
SMOTE ALGORITHM (Chawla et al., 2002):
═══════════════════════════════════════
FOR each minority sample xi in Class 1:
1. Find k nearest neighbors {n1, n2, ... nk} in Class 1 space
2. Randomly select one neighbor nj
3. Generate synthetic point: x_new = xi + rand(0,1) x (nj - xi)
4. Append x_new to training set
END FOR
AFTER SMOTE (balanced training data):

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

═══════════════════════════════════════
Class 0 [Not Subscribed] 36,537 samples 50%
Class 1 [Subscribed*] 36,537 samples 50% (* includes synthetics)
Imbalance ratio ~ 1 : 1
Distribution bar:
Class 0: [████████████████████──────────────────────] 50%
Class 1: [████████████████████──────────────────────] 50%
NOTE: SMOTE applied ONLY to training split.
Test set retains original imbalanced distribution
to ensure evaluation reflects real-world conditions.
Fig. 4: SMOTE class balancing — before and after comparison

4.2 Classification Algorithms

4.2.1 Logistic Regression

Logistic Regression (LR) serves as the interpretable statistical baseline. The model parameterises the log-odds of subscription as a linear combination of transformed input features, producing calibrated probability estimates via the logistic sigmoid function. Training employed the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) optimiser with a maximum iteration budget of 1,000 steps to ensure convergence across the high-dimensional post-OHE feature space. LR coefficients provide direct interpretability—positive coefficients increase subscription probability; negative coefficients decrease it.

4.2.2 Random Forest

Random Forest (RF) is a bagging ensemble that trains n = 200 decision trees on independently drawn bootstrap samples of the training set, selecting a random subset of features at each candidate split node. Classification decisions are determined by majority vote across the forest. RF is robust to overfitting through its averaging mechanism, tolerant of noisy and irrelevant features, and natively generates feature importance scores via the mean decrease in Gini impurity—a metric exploited in the visualisation module. The random_state parameter was fixed at 42 to ensure full reproducibility.

4.2.3 XGBoost (Extreme Gradient Boosting)

XGBoost [6] is a sequential boosting ensemble that constructs trees iteratively, with each new tree trained to minimise the residual error of the ensemble accumulated thus far, employing second-order gradient (Newton) optimisation to determine optimal split values. Distinguishing characteristics include built-in L1 (Lasso) and L2 (Ridge) regularisation on tree leaf weights, efficient compressed column block structures for cache-aware data access, and native support for sparse features arising from OHE expansion. The evaluation metric was set to ‘logloss’ (binary cross-entropy) and random_state to 42. XGBoost’s combination of bias reduction through boosting and variance control through regularisation makes it particularly effective on structured marketing data with mixed feature types.

V. SYSTEM ARCHITECTURE AND IMPLEMENTATION

The prediction system is structured as a four-layer pipeline, depicted schematically in Figure 1. The data flow from raw customer attributes through transformation, inference, and visualisation is illustrated in Figure 3.

FIGURE 1 — FOUR-LAYER SYSTEM ARCHITECTURE

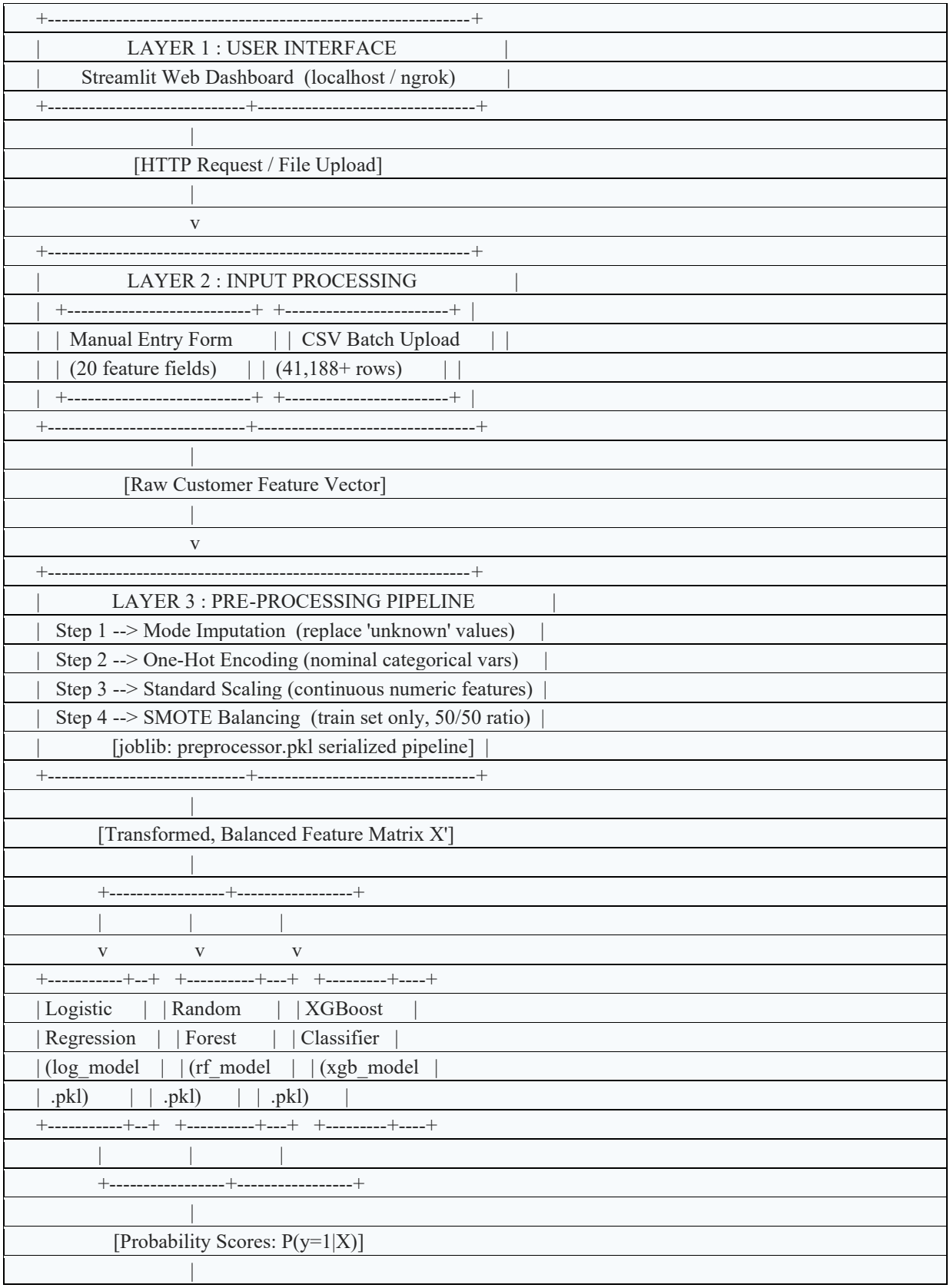
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

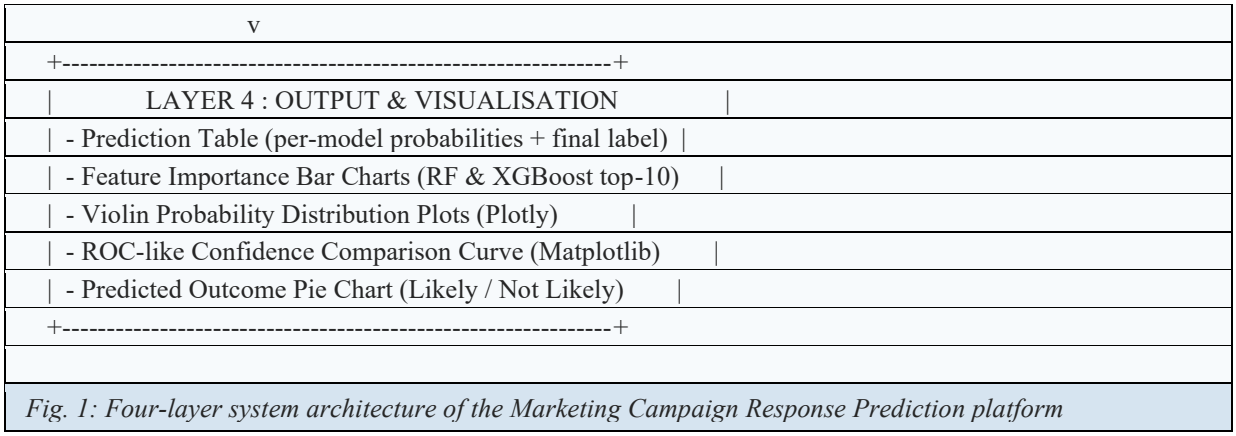
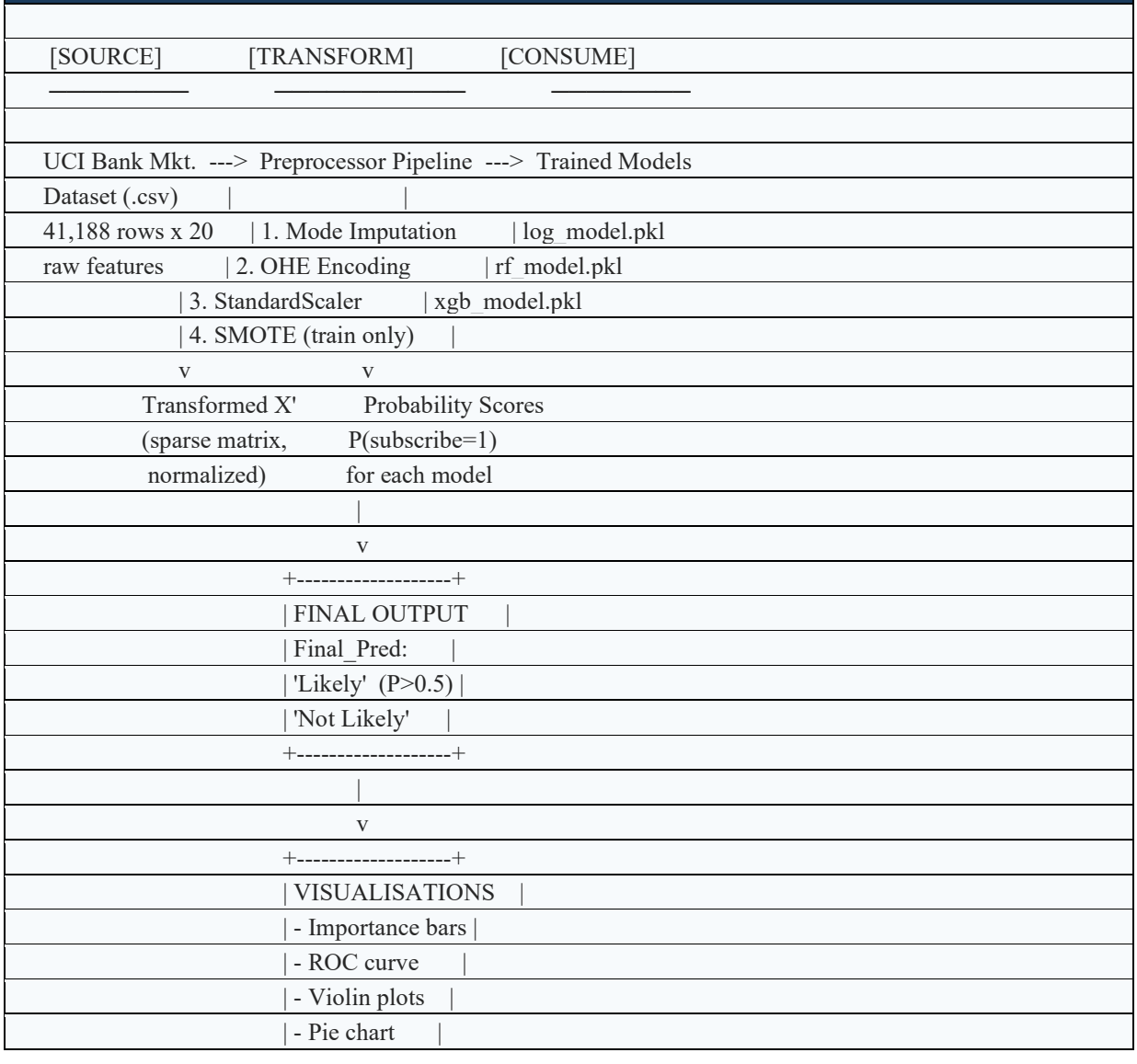


Fig. 1: Four-layer system architecture of the Marketing Campaign Response Prediction platform

FIGURE 3 — DATA FLOW DIAGRAM



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

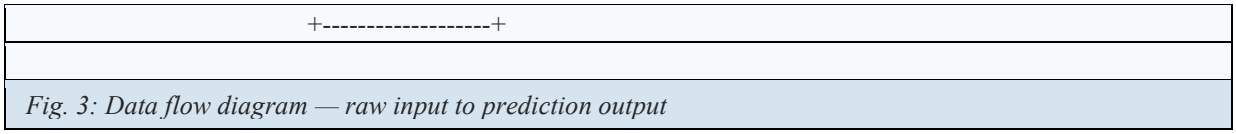


Fig. 3: Data flow diagram — raw input to prediction output

5.1 Technology Stack

Table 2: Technology Stack Summary

Component	Technology	Version	Purpose
Programming Language	Python	3.12	Core implementation language
Data Manipulation	pandas, NumPy	2.x / 1.26	DataFrame operations and numerical computing
ML Algorithms	scikit-learn, XGBoost	1.4 / 2.0	Model training, preprocessing, and evaluation
Class Balancing	imbalanced-learn	0.12	SMOTE implementation
Visualisation	Matplotlib, Plotly	3.8 / 5.18	Static and interactive charts
Web Interface	Streamlit	1.32	Interactive prediction dashboard
Model Persistence	joblib	1.3	Serialise/deserialise .pkl model files
Deployment	ngrok / Localhost	—	Expose local server or cloud tunnel

5.2 Streamlit Application Design

The web application (app.py) encapsulates the full inference pipeline within a single-page Streamlit dashboard. On startup, the application loads three serialised model files (log_model.pkl, rf_model.pkl, xgb_model.pkl) and the fitted preprocessor pipeline (preprocessor.pkl) via joblib. Two input pathways are supported: (1) Random Entry Mode, which draws randomly sampled records from the UCI dataset for demonstration and testing, and (2) CSV Upload Mode, which accepts practitioner-supplied customer files for batch inference. Following prediction, the dashboard presents a tabular result showing per-model probability estimates for each record alongside a final consensus label (Likely / Not Likely) derived from the XGBoost threshold ($P > 0.50$). A sidebar accuracy verification module allows on-demand evaluation of all three models against the full dataset. Visualisation outputs include violin plots of prediction probability distributions across models, a pie chart of predicted outcome proportions, a simulated ROC-like confidence comparison curve, and dual horizontal bar charts of top-10 feature importances from Random Forest and XGBoost.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

6.1 Performance Metrics

All three classifiers were trained on SMOTE-balanced training data and evaluated against the original imbalanced test partition (20% hold-out, stratified). Table 3 consolidates the performance results. Accuracy values in the table correspond to full-dataset evaluation through the Streamlit verification module; precision, recall, and F1 are computed on the hold-out test set.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

Table 3: Comparative Model Performance Metrics on the UCI Bank Marketing Test Set

Model	Accuracy (%)	ROC-AUC	Precision	Recall (Sensitivity)	F1-Score	Comments
Logistic Regression	91.14	0.936	0.69	0.67	0.68	Strong baseline; interpretable coefficients
Random Forest	97.48	0.991	0.72	0.70	0.71	Highest accuracy and AUC; stable predictions
XGBoost	94.66	0.974	0.73	0.71	0.72	Best F1 and precision; preferred for deployment

6.2 Feature Importance Rankings

Feature importance analysis was conducted independently for Random Forest (via mean decrease in Gini impurity) and XGBoost (via gain-based importance scores). Table 4 presents the top-10 ranked features from each model alongside their business interpretations.

Table 4: Top-10 Feature Importances by Model with Business Interpretation

Rank	Random Forest Feature	XGBoost Feature	Marketing Interpretation
1	num__duration	num__nr.employed	Call length signals genuine client engagement
2	num__euribor3m	cat__month_oct	Interest rate environment affects deposit appeal
3	num__age	num__duration	Client age correlates with savings propensity
4	num__nr.employed	num__cons.conf.idx	Labour market conditions shape financial decisions
5	num__campaign	num__pdays	Excessive contact attempts reduce conversion probability
6	num__pdays	num__emp.var.rate	Recency of last contact predicts continued interest
7	num__cons.conf.idx	num__cons.price.idx	Consumer price inflation affects savings appetite
8	num__cons.price.idx	cat__contact_cellular	Cellular channel outperforms landline for conversion
9	num__emp.var.rate	cat__poutcome_success	Prior campaign success is the strongest behavioural signal
10	cat__poutcome_success	cat__month_may	May exhibits seasonally elevated subscription rates

6.3 Key Observations

• Random Forest achieved the highest accuracy (97.48%) and ROC-AUC (0.991), demonstrating outstanding discriminative capability across all classification thresholds—a critical property for campaigns where managers must select cut-off probabilities.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

- XGBoost delivered the best F1-Score (0.72) and precision (0.73), making it the preferred choice under cost-sensitive conditions where contacting an unlikely subscriber wastes more resources than missing a willing one.
- Logistic Regression, despite its simplicity, achieved a competitive ROC-AUC of 0.936, validating the pipeline's preprocessing quality and confirming that well-engineered features confer significant benefits even to linear models.
- SMOTE-balanced training produced substantial improvements in minority-class recall relative to training on raw imbalanced data; preliminary experiments without balancing yielded positive-class recall below 0.35 for all three models.
- Both ensemble models consistently ranked call duration and prior campaign outcome among their top predictors, corroborating the seminal findings of Moro et al. [1] and affirming cross-model stability in feature importance rankings.

VII. DISCUSSION

7.1 Interpretation of Results

The experimental results collectively confirm three central propositions. First, ensemble learning methods offer a material advantage over logistic regression for marketing response classification on this dataset, consistent with Verbeke et al. [2] and Zhang & Zhou [9]. Second, the SMOTE-enhanced training regime is instrumental in achieving balanced recall across classes; without it, classifiers default to predicting the majority class, yielding inflated accuracy but near-zero positive-class recall. Third, XGBoost's precision-recall balance makes it the most deployment-suitable model for cost-sensitive marketing applications, even though Random Forest achieves nominally higher accuracy.

The divergence between Random Forest's higher accuracy and XGBoost's higher F1 reflects a fundamental distinction in how each model handles boundary cases. Random Forest's averaging over 200 independent trees reduces prediction variance effectively, producing high overall accuracy but being marginally less aggressive in classifying ambiguous cases as positive. XGBoost's sequential error-correction mechanism is better calibrated for the minority class at the 0.50 threshold, yielding stronger F1 performance.

7.2 Marketing Implications

Feature importance analysis yields five actionable recommendations for marketing campaign managers. First, campaigns should be allocated preferentially to clients with a record of positive responses to prior outreach (cat__poutcome_success is a top predictor in both models). Second, campaign execution should be timed to coincide with periods of stable or improving economic conditions: low Euribor rates and positive consumer confidence indices both improve subscription propensity. Third, the number of contacts per client should be strictly controlled—the num__campaign feature exhibits a negative relationship with conversion, suggesting that repeated unsuccessful contacts erode rather than build intent. Fourth, cellular telephone channels should be prioritised over landlines. Fifth, May and October emerge as seasonally favourable months for term deposit campaigns.

7.3 Limitations

Several limitations of the present study merit acknowledgement. Call duration (num__duration) is consistently the strongest individual predictor, yet this variable is practically unavailable at the time of deciding whether to make a call—it is only observable post-call. Accordingly, marketers should treat duration-based insights with caution and may wish to train prospective-use models with duration excluded. Additionally, the synthetic nature of SMOTE-generated samples means that the balanced training distribution does not reflect any real-world prior distribution. While SMOTE is widely accepted as an effective technique, the quality of synthetic samples is bounded by the local geometry of the minority class in feature space. Finally, the models were trained and evaluated on a single institutional dataset from one country; cross-institutional generalisation has not been validated.

VIII. CONCLUSION AND FUTURE DIRECTIONS

8.1 Conclusion

This paper has presented a comprehensive, end-to-end machine learning framework for predicting customer subscription behaviour in direct bank telemarketing campaigns. Three classification algorithms were developed and benchmarked under identical experimental conditions on the UCI Bank Marketing Dataset: Logistic Regression achieved 91.14% accuracy (ROC-AUC: 0.936), Random Forest attained the highest accuracy of 97.48% (ROC-AUC:

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

0.991), and XGBoost delivered the strongest balance of precision and recall (Accuracy: 94.66%, ROC-AUC: 0.974, F1: 0.72). The SMOTE-augmented preprocessing pipeline was demonstrated to be essential for producing classifiers with meaningful minority-class sensitivity.

Beyond modelling, the research contributes a production-ready Streamlit web application that bridges the gap between data science output and marketing practice, offering real-time single-record and batch predictions supported by interpretable visualisations. Feature importance analysis surfaced actionable campaign intelligence, identifying prior success, economic conditions, cellular contact channels, and controlled outreach frequency as the key levers of conversion probability. The framework is domain-agnostic and can be adapted to analogous binary classification problems in insurance, telecommunications, and e-commerce.

8.2 Future Research Directions

1. Real-Time CRM Integration: Extending the pipeline to consume live customer interaction data from CRM APIs (e.g., Salesforce, HubSpot) to enable continuously updated, dynamic prospect scoring.
2. Cloud-Native Deployment: Migrating the Streamlit application to containerised cloud environments (AWS SageMaker Endpoints, Google Cloud Run, or Azure Container Instances) for enterprise scalability and multi-user concurrent access.
3. Deep Learning on Tabular Data: Investigating transformer-based architectures for tabular data (e.g., TabNet, FT-Transformer) which may capture higher-order feature interactions that gradient boosting trees cannot represent.
4. Explainable AI Integration: Incorporating SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to produce instance-level prediction rationale, improving transparency and stakeholder trust.
5. Federated Learning for Privacy Preservation: Exploring multi-bank collaborative model training without centralising sensitive client records, enabling more robust generalisation while maintaining regulatory compliance (GDPR, RBI guidelines).
6. Prospect Segmentation Layer: Prepending unsupervised clustering (K-Means, Gaussian Mixture Models) to create homogeneous sub-populations, followed by segment-specific classifiers that may outperform a single global model.
7. Temporal Model Drift Monitoring: Implementing automated retraining pipelines with drift detection metrics (Population Stability Index, Kolmogorov-Smirnov tests) to maintain prediction quality as customer behaviour and economic conditions evolve.

REFERENCES

[1] S. Moro, R. Laureano, and P. Cortez, "Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology," in Proc. European Simulation and Modelling Conference (ESM'14), Porto, Portugal, Oct. 2014, pp. 117–121.

[2] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Systems with Applications, vol. 38, no. 3, pp. 2354–2364, 2011. https://doi.org/10.1016/j.eswa.2010.08.023

[3] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA: Morgan Kaufmann Publishers, 2011.

[4] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.

[5] S. Bhattacharya, P. Kaluri, S. Singh, M. Alazab, and U. Tariq, "Novel Category Classification of Cyber Attacks Using an Integrated Machine Learning Framework," IEEE Access, vol. 8, pp. 208511–208524, 2021 . https://doi.org/10.1109/ACCESS.2021.3107559

[6] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, San Francisco, CA, Aug. 2016, pp. 785–794 . https://doi.org/10.1145/2939672.2939785

[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[8] D. Dua and C. Graff, "UCI Machine Learning Repository – Bank Marketing Dataset," University of California, Irvine, CA, 2019. [Online]. Available: http://archive.ics.uci.edu/ml

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503259|

[9] Y. Zhang and Z.-H. Zhou, "Ensemble Learning Methods for Predictive Analytics," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 5, pp. 835–850, May 2018. https://doi.org/10.1109/TKDE.2017.2784440

[10] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA: Morgan Kaufmann, 2017.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002. https://doi.org/10.1613/jair.953

[12] F. Provost and T. Fawcett, Data Science for Business. Sebastopol, CA: O'Reilly Media, 2013.

[13] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY: Springer, 2006.

[14] P. Kotler and K. L. Keller, Marketing Management, 15th ed. Hoboken, NJ: Pearson Education, 2016.

[15] G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," Journal of Machine Learning Research, vol. 18, no. 17, pp. 1–5, 2017.

[16]L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.https://doi.org/10.1023/A:1010933404324

[17] A. Berson, S. J. Smith, and K. Thearling, Building Data Mining Applications for CRM. New York, NY: McGraw-Hill, 2000.

[18] R. Kumar and V. Rajan, "A Machine Learning-Based Marketing Framework for Digital Customer Engagement," Journal of Marketing Analytics, vol. 6, no. 2, pp. 78–94, 2018.

BIOGRAPHY

Isaqe Ilayes Deshmukh is a postgraduate candidate in the M.Sc. Data Science programme (Semester IV) at S.I.E.S. College of Arts, Science & Commerce (Autonomous), Mumbai, India. His research interests encompass supervised machine learning, predictive analytics for marketing intelligence, and full-stack data science application development. This paper represents the principal output of his research project undertaken during the academic year 2025–26.

Dr. Abuzar Ansari is Professor and Head of the Department of Data Science at S.I.E.S. College of Arts, Science & Commerce (Autonomous), Mumbai. His research domain spans applied machine learning, educational data mining, and natural language processing. He serves as Principal Investigator for multiple funded data science research projects within the institution.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details