The American Journal of Management and Economics Innovations
ISSN 2693-0811 Volume 08 - 2026

Explainable Predictive Business Intelligence for Banking: Integrating Machine Learning, SHAP, and Large Language Models for Customer Response Prediction

Md Abu Sufian Mozumder
College of Business, Westcliff University, Irvine, California, USA

Md Salim Chowdhury
College of Graduate and Professional Studies, Trine University, USA

Md Al-Imran
College of Graduate and Professional Studies, Trine University, USA

Md Yassir Mottalib
Master of Science in Information System Technology, Wilmington University, USA

Md. Yousuf
Masters of Science in Data Analytics, College of Sciences and Technology, University of Houston-Downtown, USA

Received: 04 Aug 2026 | Received Revised Version: 24 Aug 2026 | Accepted: 07 Sep 2026 | Published: 25 Sep 2026

Volume 08 Issue 09 2026 | DOI: 10.37547/tajmei/Volume08Issue09-04

Abstract

This study develops an integrated predictive business intelligence (BI) framework that combines machine learning (ML), explainable artificial intelligence (XAI), and large language models (LLMs) for customer response prediction in banking. Using the UCI Bank Marketing dataset of 45,211 observations, we compare Logistic Regression, Decision Tree, Random Forest, and XGBoost using accuracy, precision, recall, F1-score, and ROC-AUC. Among the evaluated models, Random Forest achieved the strongest overall performance, with 90.0% accuracy, 56.0% precision, 53.0% recall, 55.0% F1-score, and 91.8% ROC-AUC. XGBoost produced the highest recall at 78.0%, with 86.0% accuracy, 45.0% precision, 57.0% F1-score, and 91.4% ROC-AUC. Logistic Regression achieved 82.0% accuracy and 90.3% ROC-AUC, whereas Decision Tree achieved 87.0% accuracy and 67.3% ROC-AUC. We apply SHapley Additive exPlanations (SHAP) to identify the factors contributing to model predictions and integrate an LLM to convert structured prediction and explanation outputs into natural-language business insights. The resulting architecture connects predictive modeling, model explainability, BI visualization, and managerial interpretation within a unified decision-support workflow. The findings indicate that Random Forest provides a balanced predictive performance for the benchmark task, while XGBoost offers greater sensitivity to potential positive cases. The framework demonstrates how combining ML with XAI and LLM-based interpretation can improve the accessibility and transparency of predictive BI. However, practical deployment requires institution-specific validation, particularly because the benchmark data originate from a Portuguese banking campaign and include variables that may introduce temporal leakage. The study therefore positions the proposed framework as a research and deployment architecture rather than evidence of direct performance in U.S. banking environments.

Keywords: Predictive Business Intelligence, Machine Learning, Explainable AI, Large Language Models, SHAP, Banking Analytics, Customer Response Prediction, Random Forest, XGBoost

Cite This Article: Mozumder, M. A. S., Chowdhury, M. S., Al-Imran, M., Mottalib, M. Y., & Yousuf, M. (2026). Explainable Predictive Business Intelligence for Banking: Integrating Machine Learning, SHAP, and Large Language Models for Customer Response Prediction. The American Journal of Management and Economics Innovations, 8(09), 57–83. https://doi.org/10.37547/tajmei/Volume08Issue09-04

Introduction

The banking industry has experienced a substantial transformation as financial institutions increasingly adopt data-driven technologies to improve customer management, marketing effectiveness, risk assessment, and strategic decision-making. The rapid growth of digital banking has generated large volumes of structured and unstructured customer information, creating significant opportunities for financial institutions to apply artificial intelligence and machine learning to identify patterns that may not be readily observable through traditional analytical methods. In this environment, predictive business intelligence has emerged as an important approach for transforming historical customer and transactional information into forward-looking insights that can support managerial decisions. Rather than relying exclusively on descriptive reports that explain what has already occurred, predictive business intelligence enables organizations to estimate what is likely to occur and identify the factors that contribute to those outcomes.

Customer response prediction represents an important application of predictive analytics in banking. Financial institutions frequently conduct marketing campaigns to promote deposits, loans, credit products, investment services, insurance products, and other financial offerings. However, traditional marketing strategies may treat large customer populations relatively uniformly, resulting in inefficient allocation of marketing resources. Predictive machine learning can address this problem by estimating the probability that an individual customer will respond positively to a particular banking offer. A bank can consequently prioritize customers according to their predicted likelihood of conversion, potentially reducing unnecessary contacts while improving campaign effectiveness.

Previous research has demonstrated the usefulness of machine learning for predicting the success of bank marketing campaigns. Moro et al. (2014) developed a data-driven approach for predicting the success of telemarketing campaigns for long-term bank deposits using data collected from a Portuguese retail bank between 2008 and 2013. Their study demonstrated that data mining and feature selection could be used to identify customers with higher probabilities of responding positively to marketing campaigns. Importantly, their work established the Bank Marketing dataset as a widely used benchmark for evaluating predictive approaches to customer subscription behavior.

Although predictive machine learning can produce highly accurate results, the increasing complexity of modern algorithms creates an important challenge for banking applications. Ensemble models and other advanced algorithms can capture nonlinear relationships and interactions among customer characteristics, but their predictions may be difficult for business users to interpret. In financial environments, prediction without explanation can limit organizational trust and make it difficult for decision makers to understand why a model has assigned a particular probability to a customer. This issue has contributed to the increasing importance of explainable artificial intelligence (XAI), which seeks to provide understandable explanations for machine learning predictions. Recent research has identified explainability as an important component of trustworthy artificial intelligence in financial applications because financial decisions can have substantial economic and organizational consequences.

SHAP, or SHapley Additive exPlanations, provides one important approach for addressing this challenge. Lundberg and Lee (2017) proposed SHAP as a unified framework for interpreting model predictions by assigning feature-attribution values to individual predictions. The approach makes it possible to determine how individual features contribute positively or negatively to a model's output. Other explainability techniques, such as LIME, have similarly attempted to provide interpretable local explanations for complex machine learning models. Ribeiro et al. (2016)

demonstrated that local surrogate models can be used to explain the predictions of otherwise opaque classifiers.

The development of large language models has created an additional opportunity for improving the accessibility of predictive analytics. Large language models can transform structured analytical results into natural-language explanations, allowing business users to interact with complex analytical systems without requiring extensive knowledge of machine learning algorithms. Research examining LLMs in finance has identified applications across financial analysis, information extraction, decision support, and other financial tasks, while also emphasizing challenges related to reliability, domain adaptation, data privacy, and responsible deployment. Recent reviews of generative artificial intelligence in finance similarly emphasize the transformative potential of finance-oriented LLMs while highlighting the need for appropriate governance and regulatory frameworks.

Despite these developments, an important research opportunity remains at the intersection of predictive banking analytics, explainable machine learning, and large language models. Existing bank marketing studies have primarily focused on improving predictive classification performance, while XAI research has concentrated on explaining complex machine learning models. At the same time, research on LLMs in finance has increasingly examined natural-language processing, financial information analysis, and broader generative AI applications. However, comparatively fewer studies have developed an integrated framework in which a predictive banking model generates customer-level predictions, explainable AI identifies the factors responsible for those predictions, and an LLM transforms those explanations into business-oriented intelligence for managerial decision-making. Recent research on bank marketing continues to focus heavily on ensemble prediction performance, including stacking and boosting approaches, indicating that predictive modeling remains an active research area.

In this study, we address this research gap by developing a predictive business intelligence framework that integrates machine learning, explainable artificial intelligence, and large language models for banking customer-response prediction. We use the publicly available UCI Bank Marketing dataset, which contains customer demographic, financial, contact, and campaign-related information and uses subscription to a term deposit as the binary prediction target. The UCI repository identifies the bank-full.csv version as containing 45,211 observations and 17 attributes, including 16 input variables and the target variable.

We develop and compare several machine learning algorithms, including logistic regression, decision tree, random forest, and gradient-boosting approaches. We evaluate the models using accuracy, precision, recall, F1-score, ROC-AUC, and other relevant measures rather than relying on a single performance metric. Following model selection, we apply SHAP-based explainability to identify the most influential features at both the global and individual prediction levels. We then introduce a large language model as a natural-language interpretation layer that converts structured predictions and explainability outputs into understandable business intelligence narratives.

Our approach differs from a conventional machine learning classification study because we do not consider predictive accuracy to be the final objective. Instead, we seek to establish a complete analytical pathway from customer data to prediction, from prediction to explanation, and from explanation to business decision support. The machine learning model remains responsible for generating the prediction, the explainability component provides evidence regarding the factors influencing that prediction, and the large language model communicates the analytical evidence in a form that can be more readily understood by business users.

We further discuss how the proposed framework could be adapted for implementation in the United States banking industry. Because the UCI dataset originates from a Portuguese banking institution, we do not assume that its historical patterns can be transferred directly to U.S. customers. Instead, we use the dataset as an open-source research benchmark and propose a deployment framework in which the predictive architecture is retrained and validated using institution-specific U.S. banking data. This distinction is important for responsible model development because differences in customer demographics, financial products, regulatory requirements, economic conditions, and marketing practices may affect model performance across geographic and institutional environments.

The primary contribution of this study is therefore the integration of predictive machine learning, explainable

artificial intelligence, and large language models into a unified predictive business intelligence framework for banking. By combining predictive performance with model transparency and natural-language interpretation, we aim to demonstrate how advanced artificial intelligence can move from a technically accurate classification system toward a more accessible and decision-oriented business intelligence solution.

Literature Review

Predictive Business Intelligence in Banking

Business intelligence has traditionally focused on collecting, integrating, analyzing, and presenting organizational data to support managerial decision-making. The development of advanced analytics and machine learning has expanded conventional business intelligence from primarily descriptive and diagnostic analysis toward predictive and prescriptive capabilities. In banking, this transformation is particularly significant because financial institutions generate large volumes of customer, transaction, product, and marketing data that can be analyzed to identify behavioral patterns and predict future outcomes.

Predictive business intelligence combines analytical models with organizational decision processes to transform historical information into forward-looking insights. In a banking context, predictive models can support customer segmentation, product recommendation, marketing response prediction, customer retention, fraud detection, credit-risk assessment, and other applications. The central advantage of predictive analytics is that it allows organizations to prioritize actions based on estimated future outcomes rather than relying solely on historical reporting.

Customer marketing represents an especially appropriate application because banks routinely interact with customers through multiple communication channels. The success of these campaigns can vary substantially across customers, meaning that contacting every customer with the same strategy may result in inefficient use of organizational resources. Machine learning provides an opportunity to identify patterns associated with customer responses and use those patterns to prioritize future marketing activities.

Machine Learning for Bank Marketing Prediction

One of the most influential studies in this area was conducted by Moro et al. (2014), who proposed a data-driven approach for predicting the success of bank telemarketing campaigns. Their research focused on a Portuguese retail bank and used data collected between 2008 and 2013. The authors investigated customer, product, and socioeconomic characteristics and developed a predictive modeling approach involving feature selection and multiple data-mining models. Their results demonstrated that predictive analytics could support the identification of customers who were more likely to purchase long-term deposits.

The work of Moro et al. (2014) is particularly relevant to our research because it provides the foundation for the Bank Marketing dataset used in our study. The UCI repository describes the dataset as information from direct marketing campaigns conducted by a Portuguese banking institution, with the classification objective of predicting whether a client subscribed to a term deposit. The bank-full.csv version contains 45,211 observations and 17 attributes, making it an appropriate open-source benchmark for customer-response prediction.

Subsequent studies have continued to investigate machine learning methods for the same customer-subscription prediction problem. Research has compared traditional statistical approaches with decision trees, ensemble models, boosting algorithms, and other classification methods. More recent work has emphasized ensemble learning because customer response behavior can involve nonlinear relationships among demographic, financial, and campaign variables. A 2025 study examining ensemble learning for bank telemarketing reported strong performance from stacking approaches, with an accuracy of 91.88% and ROC-AUC of 0.9491, while identifying contact duration, economic indicators, and customer age as influential predictive variables.

These findings support the use of ensemble methods in our research. Random Forest, introduced by Breiman (2001), combines multiple decision trees to improve generalization and reduce the limitations associated with individual decision trees. The randomization of observations and features allows the resulting ensemble to capture complex relationships while reducing sensitivity to individual training observations. XGBoost, introduced by Chen and Guestrin (2016), further extends tree-based learning through scalable gradient boosting

and has become widely used for structured predictive problems.

The literature therefore indicates that ensemble learning can provide strong predictive performance for banking customer-response problems. However, predictive performance alone does not fully address the requirements of a business intelligence system. A model may identify high-probability customers accurately while providing limited information about why those customers were selected. This limitation leads directly to the importance of explainability.

Explainable Artificial Intelligence in Financial Applications

Explainable artificial intelligence has emerged in response to the increasing use of complex machine learning models in applications where users need to understand model behavior. In financial applications, explainability is particularly important because automated analytical systems may influence customer targeting, risk assessment, resource allocation, and other consequential decisions. A systematic literature review by Černevičienė and Kabašinskas (2024) found that XAI has become increasingly important in finance because explanations can contribute to transparency, trust, risk assessment, and the responsible use of complex AI models.

Traditional machine learning models such as logistic regression are relatively straightforward to interpret because their coefficients can provide information about the direction and magnitude of relationships between variables and outcomes. However, more sophisticated ensemble and nonlinear models often provide better predictive flexibility at the expense of direct interpretability. This creates a fundamental trade-off between predictive performance and transparency.

Ribeiro et al. (2016) addressed this challenge through LIME, an approach designed to explain individual predictions of arbitrary machine learning classifiers. LIME constructs an interpretable approximation around a particular prediction, allowing users to understand which features are most influential locally. Their work emphasized that understanding why a model makes a particular prediction is important for determining whether the prediction should be trusted and whether the model is suitable for deployment.

Lundberg and Lee (2017) subsequently introduced SHAP as a unified framework for feature attribution. SHAP assigns contribution values to features for individual predictions and provides a theoretically grounded method for interpreting complex models. The method can be used to generate both local explanations for individual customers and global explanations that summarize the overall behavior of a predictive model.

The application of SHAP to banking prediction is particularly useful because a customer-level probability can be accompanied by an explanation of the factors contributing to that probability. For example, rather than simply indicating that a customer has a high probability of subscribing to a deposit product, an explainable model can identify campaign history, previous response, financial characteristics, or other variables that contributed to the prediction. This creates a closer relationship between machine learning outputs and business intelligence.

Explainability and Trust in Banking

Trust is an important consideration in financial AI because business users may be reluctant to rely on predictions generated by models that they cannot understand. A model that produces accurate predictions but cannot provide meaningful explanations may be difficult to incorporate into managerial workflows. XAI therefore serves not only a technical purpose but also an organizational purpose by providing a bridge between predictive analytics and human decision-making.

The literature suggests that explanations should be considered in relation to the intended user and decision context. A data scientist may require detailed feature-attribution information, while a banking manager may need a concise explanation describing the major factors influencing a customer score and the corresponding business implications. Consequently, conventional visualization-based XAI can still create an accessibility barrier for non-technical users.

This limitation motivates our integration of large language models into the proposed framework. Instead of requiring managers to interpret SHAP plots or numerical feature-attribution values directly, we use the LLM to transform structured explanatory information into natural-language business intelligence.

Large Language Models in Finance

Large language models have introduced a new generation of artificial intelligence capabilities based on large-scale language understanding and generation. In financial applications, LLMs can potentially support document analysis, financial information extraction, question answering, report generation, summarization, research assistance, and analytical communication.

Wang et al. (2023) provide a comprehensive survey of LLM applications in finance and discuss approaches including zero-shot and few-shot learning, domain-specific fine-tuning, and customized financial language models. Their work also emphasizes that the appropriate LLM architecture depends on the specific financial task, available data, computational resources, and performance requirements.

More recent research has expanded the investigation of LLMs toward financial agents and real-world financial deployment. Dong et al. (2025) examined LLM agents in finance and highlighted applications across data analysis, investment research, trading, investment management, and risk management. At the same time, the authors identify challenges including numerical reasoning, prompt sensitivity, and the difficulty of adapting LLMs to real-time institutional environments.

Research on generative AI in finance has also identified governance and regulatory concerns. A 2024 review of generative AI in finance found that finance-specific LLMs are becoming an important area of research while emphasizing the need for appropriate governance frameworks. These findings suggest that LLMs should not be treated as unrestricted autonomous decision makers in banking. Instead, they can be more appropriately positioned as controlled analytical interfaces that communicate validated information produced by established analytical systems.

Integration of Machine Learning, XAI, and LLMs

Although research on machine learning, XAI, and LLMs has expanded independently, the integration of these technologies into a unified banking business intelligence architecture remains comparatively underdeveloped. Existing bank marketing studies primarily focus on improving customer-response prediction through feature engineering, classification, and ensemble learning. Research on XAI focuses on improving transparency and understanding of machine learning predictions, while LLM research in finance increasingly addresses natural-language interaction and financial information processing.

Our research combines these three research streams. We use machine learning to generate customer-response predictions, XAI to determine which features contribute to those predictions, and an LLM to transform the structured explanations into natural-language business intelligence. This architecture creates a clear separation of responsibilities. The predictive model is responsible for statistical prediction, the XAI layer is responsible for attribution and transparency, and the LLM is responsible for communication and interpretation.

This separation is important because we do not propose allowing the LLM to independently determine customer subscription probabilities. Instead, the LLM receives validated analytical outputs from the predictive and explainability layers. This design reduces the possibility that natural-language generation will replace the underlying statistical evidence.

Research Gap

The literature demonstrates three important developments. First, machine learning has demonstrated considerable potential for predicting customer responses in banking marketing campaigns. Second, explainable AI provides methods for interpreting complex predictive models and improving transparency. Third, large language models have created new possibilities for communicating financial information and analytical results through natural language.

However, these research areas have generally developed as separate streams. Existing bank marketing research has concentrated primarily on predictive performance, while XAI research has emphasized model transparency and LLM research has focused increasingly on financial language and generative AI applications. There remains an opportunity to develop an integrated framework that connects customer-response prediction, explainability, and natural-language business intelligence.

We address this gap by developing a unified predictive business intelligence framework using the UCI Bank Marketing dataset. We compare multiple machine learning algorithms, select the strongest model based on multiple performance measures, apply SHAP to explain the resulting predictions, and use an LLM to translate those explanations into business-oriented narratives. We then extend the framework conceptually toward U.S.

banking implementation, emphasizing institution-specific validation, responsible AI, model governance, fairness, and continuous monitoring.

Positioning of the Present Study

Our study is positioned at the intersection of predictive analytics, explainable artificial intelligence, and generative AI. Unlike a conventional bank marketing classification study, our objective is not limited to determining whether a customer will subscribe to a banking product. We aim to demonstrate how predictive results can be transformed into actionable and understandable business intelligence.

The proposed framework therefore extends the traditional predictive modeling pipeline. Instead of ending with a classification label, we continue the analytical process by determining why the model generated the prediction and how those explanations can be communicated to business users. This creates a pathway from data to prediction, from prediction to explanation, and from explanation to decision support.

By integrating these capabilities, we seek to contribute to the growing body of research on intelligent financial analytics and demonstrate a practical methodology for developing transparent and accessible predictive business intelligence systems in banking.

Methodology

Research Methodology and Overall Framework

In this study, we developed an integrated predictive business intelligence framework that combines machine learning, explainable artificial intelligence, and large language models to support data-driven decision-making in the banking sector. Our methodology was designed to transform structured banking data into predictive insights and subsequently convert those predictions and explanations into business-oriented information that can be interpreted by managers and other non-technical decision makers. The proposed framework consists of data collection, data preprocessing, feature extraction, feature engineering, predictive model development, model evaluation, explainability analysis, and large language model-based business interpretation.

We selected a supervised machine learning approach because the dataset contains an explicitly defined binary outcome indicating whether a customer subscribed to a bank term deposit. The predictive component of our framework estimates the probability of customer subscription, while the explainability component identifies the factors that contribute most strongly to individual and overall predictions. We then use a large language model as a natural-language interpretation layer to transform machine-learning outputs and explainability results into concise business intelligence narratives. This allows the proposed framework to move beyond conventional prediction by providing both predictive results and understandable explanations that can support banking decisions.

The overall methodological workflow begins with the acquisition of an open-source banking dataset from the UCI Machine Learning Repository. We subsequently clean and transform the raw data, identify meaningful predictive attributes, construct additional business-oriented features, and divide the resulting dataset into training, validation, and testing subsets. Several machine learning algorithms are developed and compared, after which the best-performing model is selected based on multiple evaluation criteria. We then apply explainable artificial intelligence techniques to investigate the contribution of individual features to model predictions. Finally, we provide the resulting predictions, feature importance information, and explanation outputs to a large language model, which generates human-readable business intelligence summaries and decision-support recommendations.

Data Collection

We obtained the dataset from the UCI Machine Learning Repository's Bank Marketing dataset, which contains information collected from direct marketing campaigns conducted by a Portuguese banking institution. The original campaigns were conducted through telephone-based customer contact with the objective of determining whether customers would subscribe to a bank term deposit. The UCI repository identifies the dataset as a multivariate classification dataset containing categorical and integer-valued attributes. The complete bank-full.csv version contains 45,211 observations and 17 attributes, consisting of 16 predictor variables and one binary target variable.

We selected the complete bank-full.csv version rather than the smaller bank.csv version because the complete dataset provides a substantially larger number of observations for model training and evaluation. The UCI

repository also provides an alternative bank-additional-full.csv dataset containing 41,188 observations and 20 input variables. However, for the present study, we selected the 45,211-record bank-full.csv dataset because its combination of demographic, financial, contact, and campaign variables provides an appropriate foundation for developing a predictive business intelligence framework.

The principal characteristics of the dataset used in our study are summarized in **Table 1.**

Dataset Description

Dataset characteristic	**Description**
Dataset name	Bank Marketing Dataset
Data repository	UCI Machine Learning Repository
Dataset version	bank-full.csv
Number of observations	45,211
Number of predictor variables	16
Number of attributes including target	17
Prediction task	Binary classification
Target variable	y
Target interpretation	Whether the customer subscribed to a term deposit
Target classes	Yes / No
Data types	Numerical, categorical, and binary
Application domain	Banking and financial marketing
Data collection context	Direct marketing campaign
Primary business objective	Predict customer term-deposit subscription
Open-source availability	UCI Machine Learning Repository

The dataset contains several categories of information that are relevant to banking decision-making. Customer-level demographic information includes age, occupation, marital status, and education. Financial information includes account balance, credit default status, housing loan status, and personal loan status. Campaign-related variables include communication method, month and day of contact, duration of the last contact, number of contacts during the current campaign, previous contact history, and the outcome of the previous campaign. The target variable y represents whether the customer subscribed to a term deposit.

The predictor variables included in our initial dataset are summarized in **Table 2.**

Variable	Data type	Business interpretation
age	Numerical	Customer age
job	Categorical	Customer occupation
marital	Categorical	Marital status
education	Categorical	Education level
default	Binary	Whether the customer has credit in default
balance	Numerical	Average yearly account balance
housing	Binary	Whether the customer has a housing loan
loan	Binary	Whether the customer has a personal loan
contact	Categorical	Communication method
day	Numerical	Day of the month of the last contact
month	Categorical	Month of the last contact
duration	Numerical	Duration of the last contact in seconds
campaign	Numerical	Number of contacts during the current campaign
pdays	Numerical	Number of days since previous campaign contact
previous	Numerical	Number of previous contacts
poutcome	Categorical	Outcome of the previous marketing campaign
y	Binary target	Whether the customer subscribed to a term deposit

The dataset is particularly appropriate for our research because it contains variables that can be interpreted not only statistically but also from a business intelligence perspective. For example, customer balance can be associated with financial capacity, campaign contact frequency can be associated with marketing effectiveness, and previous campaign outcomes can provide information regarding customer responsiveness. Consequently, the resulting machine learning model can be connected directly to practical banking decisions rather than being limited to a purely technical classification problem.

Data Preprocessing

Before developing the predictive models, we performed a comprehensive preprocessing procedure to ensure that the dataset was suitable for machine learning. We first inspected the structure of the dataset, including the number of observations, variable types, unique values, missing observations, duplicate records, and potential inconsistencies. We also examined the distribution of the target variable because imbalanced target classes can cause a classification model to favor the majority class.

We treated categorical and numerical variables separately during preprocessing. Numerical variables were converted into appropriate numerical formats, while categorical variables were standardized to ensure consistent representation of their categories. We examined the dataset for missing or unknown values and distinguished genuine missing observations from values explicitly represented as unknown. Where appropriate, unknown categories were retained as separate categorical states rather than automatically replacing them with arbitrary values, because the absence of information can itself contain predictive information.

We also examined duplicate observations and removed exact duplicate records when necessary to prevent repeated observations from disproportionately influencing the learning process. We assessed numerical variables for extreme values and examined their distributions using descriptive statistics and graphical analysis. Rather than automatically removing all extreme observations, we considered whether unusual values represented legitimate banking behavior. This approach was important because financial datasets can naturally contain highly skewed balances and other legitimate extreme observations.

We encoded the binary target variable y into numerical form, where a successful subscription was represented as one and a non-subscription as zero. Categorical predictors were transformed using appropriate encoding techniques. For algorithms that require numerical representations, we applied one-hot encoding to nominal categorical variables. We performed all preprocessing transformations within the training pipeline to prevent information from the validation or test datasets from leaking into the training process.

We divided the dataset into training, validation, and testing subsets using a stratified sampling strategy so that the proportion of target classes remained approximately consistent across the subsets. We used the training dataset for model fitting, the validation dataset for hyperparameter selection and model comparison, and the independent testing dataset for final performance evaluation. This separation allowed us to estimate the ability of the final model to generalize to previously unseen banking customers.

Because the dataset is a binary classification problem with unequal representation of customers who subscribed and did not subscribe to the term deposit, we considered class imbalance during model development. Rather than relying solely on accuracy, we emphasized precision, recall, F1-score, ROC-AUC, and precision-recall AUC when evaluating model performance. Where necessary, class-weighting techniques or training-only resampling methods were considered to reduce the influence of class imbalance without contaminating the validation or test data.

Feature Extraction

We performed feature extraction to transform the raw banking attributes into a structured representation that could be consumed by the predictive algorithms. In this stage, we focused primarily on identifying and retaining information that was directly available in the original dataset while removing attributes that did not provide meaningful predictive information.

Unique identifiers or variables that could identify individual observations without representing meaningful customer behavior were excluded from the predictive feature matrix. Categorical variables such as occupation, education, marital status, communication method, contact month, and previous campaign outcome were converted into machine-readable representations through categorical encoding.

Numerical variables such as age, balance, duration, campaign, pdays, and previous contact frequency were retained as continuous variables. We applied feature scaling to algorithms that are sensitive to differences in numerical magnitude, particularly logistic regression and distance-based algorithms. Tree-based models were not dependent on the same scaling requirement, but we maintained a consistent preprocessing architecture to facilitate fair model comparison.

We also examined the relationship between individual variables and the target outcome through statistical analysis and exploratory visualization. Correlation analysis was applied to numerical variables, while categorical variables were analyzed using target proportions and appropriate association measures. This process allowed us to identify potentially redundant variables, highly informative predictors, and variables requiring additional transformation.

An important consideration in feature extraction was the variable duration, which represents the duration of the last contact. Although this variable can be highly predictive of the target outcome, its availability depends

on the stage of the business process. Because the duration of a telephone conversation is only known after or during the contact, using it in a pre-contact customer targeting model could introduce a form of temporal information leakage. Therefore, we treated the duration feature separately in our methodology. We developed a benchmark model containing the variable and a business-deployment model that excludes it when the prediction is intended to occur before the customer contact. This distinction allows us to distinguish between statistical predictive performance and realistic operational predictive performance. The UCI documentation similarly identifies the importance of understanding the timing and availability of this attribute.

Feature Engineering

Following feature extraction, we performed feature engineering to create additional variables that could represent meaningful banking and marketing relationships more effectively than the original attributes alone. Our objective was not simply to increase the number of variables but to create interpretable features that could improve predictive performance while maintaining business relevance.

We transformed the pdays variable into a more interpretable customer-contact history indicator. Because a value of negative one indicates that a customer was not previously contacted, we created a binary feature representing whether the customer had been contacted in a previous campaign. We also considered the number of previous contacts and the current campaign contact frequency jointly to create measures representing overall customer engagement.

We constructed contact-intensity features from the campaign and previous variables. These features represented the degree to which the bank had previously attempted to communicate with the customer. We also examined ratios and interaction terms between previous and current campaign contacts to determine whether repeated contact behavior was associated with greater or lower subscription probability.

We transformed the customer's age into meaningful age groups in addition to retaining the original continuous variable. This allowed us to investigate whether different stages of the customer life cycle were associated with different subscription probabilities. Similarly, account balance was analyzed using both its continuous representation and, where appropriate, transformed or categorized representations to reduce the influence of extreme financial values.

We also created interaction features between selected demographic and financial attributes. For example, interactions between age and balance, housing-loan status and balance, education and balance, and campaign engagement and previous campaign outcome were considered because customer behavior may depend on combinations of characteristics rather than individual variables in isolation.

Temporal variables were transformed into business-oriented representations. Contact month was represented using categorical encoding, while the day-of-month variable was evaluated for possible temporal effects. Where appropriate, cyclical transformations were considered for temporal variables so that the numerical representation could preserve the relationship between the beginning and end of a time cycle.

Feature engineering was performed only after splitting the dataset into training and testing components, and parameters estimated from the training data were applied to the validation and test data. This procedure prevented information leakage and ensured that the performance estimates represented genuine out-of-sample performance.

Model Development

We developed multiple supervised machine learning models to establish a comprehensive predictive framework and determine which algorithm provided the most appropriate balance between predictive performance, interpretability, and operational usefulness. We selected models representing both conventional statistical learning and advanced ensemble learning approaches.

We first developed logistic regression as a baseline model. Logistic regression was selected because it provides a relatively transparent relationship between predictor variables and the probability of customer subscription. Its coefficients can be interpreted in terms of the direction and magnitude of association between predictors and the target outcome, making it particularly useful for establishing a transparent benchmark.

We subsequently developed a decision tree classifier to capture nonlinear relationships and threshold effects in customer behavior. Decision trees are particularly

relevant to business intelligence applications because their decision paths can be interpreted as business rules. However, we recognized that an unrestricted decision tree can overfit the training data, and therefore we controlled its complexity through hyperparameter tuning.

We then developed a random forest model consisting of multiple decision trees trained using randomized subsets of observations and features. Random forest was included because ensemble learning can improve predictive stability and capture nonlinear interactions that may not be represented adequately by a linear model.

We also developed a gradient-boosting model such as XGBoost or LightGBM to investigate whether sequential ensemble learning could provide additional predictive performance. Gradient-boosting methods are particularly suitable for structured tabular data and can model complex nonlinear relationships between customer characteristics and banking outcomes.

Where computational resources permitted, we also evaluated a multilayer perceptron neural network as an additional nonlinear benchmark. The neural network received the transformed numerical feature representation and learned nonlinear relationships between customer characteristics and the subscription outcome. However, predictive performance alone was not used to determine the preferred model because the central objective of our study was explainable business intelligence.

For each model, we performed hyperparameter optimization using the validation data or cross-validation within the training dataset. The hyperparameters included model-specific parameters such as regularization strength for logistic regression, tree depth and minimum samples for decision trees, number of estimators and tree depth for random forests, and learning rate, tree depth, and number of estimators for gradient-boosting models.

We used stratified cross-validation during model development to obtain stable estimates of model performance. The final model was then retrained using the selected configuration before being evaluated against the independent test dataset.

Explainable Machine Learning

Because our research focuses on explainable predictive business intelligence, we incorporated explainable artificial intelligence into the model development process. Our objective was to ensure that the predictive model did not function as a black box but instead provided interpretable evidence regarding why particular predictions were generated.

We applied SHAP-based analysis to estimate the contribution of individual features to model predictions. SHAP values were used to determine whether a particular feature increased or decreased the predicted probability of a customer subscribing to the term deposit. We analyzed both global feature importance and local customer-level explanations.

Global explanations were used to identify the variables that had the greatest overall influence on model predictions across the dataset. This information allowed us to determine which customer characteristics and campaign factors were most strongly associated with subscription decisions. Local explanations were used to explain individual predictions, allowing us to determine why the model classified a particular customer as having a high or low probability of subscription.

We also considered partial dependence and feature-effect analysis to investigate the relationship between important numerical variables and predicted subscription probability. These analyses provided an additional layer of interpretation and helped us identify nonlinear relationships that may not be apparent from conventional statistical summaries.

The explainability layer was particularly important for the business intelligence component of our framework. Instead of presenting banking managers with only a prediction such as "customer probability = 0.82," we aimed to provide an explanation such as the customer's previous campaign response, contact history, financial characteristics, and other relevant factors that contributed to the prediction.

Model Evaluation

We evaluated the predictive models using multiple complementary performance measures rather than relying exclusively on classification accuracy. Accuracy was calculated to determine the overall proportion of correctly classified observations, but we recognized that accuracy alone may provide a misleading assessment when the target classes are imbalanced.

Precision was calculated to determine the proportion of customers predicted as likely subscribers who actually subscribed. This metric is important from a banking marketing perspective because it indicates how efficiently the bank's marketing resources are directed toward customers who are likely to respond.

Recall was calculated to measure the proportion of actual subscribers that were successfully identified by the model. Recall is particularly relevant when the business objective is to minimize the number of potentially valuable customers who are overlooked.

The F1-score was calculated as the harmonic mean of precision and recall. We used the F1-score to evaluate whether a model maintained a reasonable balance between identifying potential subscribers and avoiding excessive false-positive predictions.

We also calculated the receiver operating characteristic area under the curve, or ROC-AUC, to measure the model's ability to discriminate between customers who subscribed and those who did not across different probability thresholds. Because business applications often require probability-based customer prioritization rather than a fixed classification threshold, ROC-AUC provides an important overall measure of ranking performance.

Precision-recall AUC was also considered because it provides useful information when the positive class is relatively less frequent. In addition, we evaluated model calibration to determine whether predicted probabilities corresponded appropriately to observed subscription frequencies. A well-calibrated model is valuable in business intelligence because a predicted probability of 0.80 should ideally correspond to approximately 80 percent observed outcomes among customers receiving similar predictions.

We further analyzed the confusion matrix to identify true positives, true negatives, false positives, and false negatives. This analysis allowed us to translate technical model performance into business consequences. For example, false positives represent customers who may receive additional marketing attention despite having a lower probability of subscription, while false negatives represent potentially valuable customers who may be missed by the marketing strategy.

We selected the final predictive model based on a combination of predictive performance, probability calibration, computational efficiency, stability, and explainability. We did not automatically select the model with the highest accuracy because the primary objective of the study was to create a practical predictive business intelligence system rather than simply maximize one statistical metric.

Large Language Model Integration

After generating predictions and explainability outputs, we incorporated a large language model as the natural-language reasoning and communication layer of our predictive business intelligence framework. The role of the large language model was not to replace the statistical prediction model. Instead, we used it to interpret structured analytical outputs and communicate those outputs in a form that can be understood by banking managers and business stakeholders.

We supplied the large language model with structured information generated by the predictive pipeline, including the predicted probability, classification outcome, most influential SHAP features, direction of feature effects, relevant customer characteristics, model confidence information, and aggregate business-level patterns. We designed the LLM prompts so that the model would not independently invent financial conclusions but would generate explanations only from the structured evidence supplied by the predictive system.

At the individual customer level, the LLM was used to generate natural-language explanations of predictions. For example, a high-probability prediction could be translated into an explanation describing the principal characteristics that contributed positively to the prediction. At the aggregate level, the LLM could summarize patterns such as customer groups with higher predicted subscription probabilities, campaign characteristics associated with stronger outcomes, and segments that may require different marketing strategies.

We also used the LLM to transform technical machine-learning outputs into business intelligence narratives. This process allows a banking manager to move from a technical SHAP visualization to an interpretable statement describing which factors appear to influence customer response and what business action could reasonably be considered.

To maintain analytical reliability, we treated the machine learning model as the source of predictive truth and the

LLM as an interpretation and communication layer. The LLM was therefore not permitted to modify model predictions, invent customer information, or override the statistical model. This separation between prediction and natural-language interpretation was incorporated to reduce the risk of hallucination and maintain traceability between business recommendations and the underlying analytical evidence.

Predictive Business Intelligence Layer

We integrated the predictive model, explainability results, and LLM-generated narratives into a business intelligence architecture. The resulting framework was designed to provide three complementary levels of information. The first level consisted of predictive outputs, including customer-level subscription probabilities and classification results. The second level consisted of explainability outputs, including global feature importance and individual prediction explanations. The third level consisted of natural-language business intelligence generated through the LLM.

At the organizational level, the system could be used to identify customer segments with higher or lower predicted subscription probabilities. At the campaign level, the framework could be used to identify characteristics associated with successful marketing outcomes. At the individual level, the framework could provide explanations for why a particular customer received a high or low predicted probability.

We therefore conceptualized the final system as a decision-support framework rather than simply a classification model. The predictive model answers the question of which customers are more likely to subscribe, the explainability component addresses why the model generated those predictions, and the large language model converts those analytical results into business-oriented narratives that can support managerial interpretation.

Reproducibility and Experimental Reliability

To ensure reproducibility, we maintained a consistent data-processing pipeline and documented each transformation applied to the original dataset. We used fixed random seeds during dataset splitting and model training where applicable. All preprocessing operations, feature transformations, model configurations, and evaluation metrics were recorded so that the experiments could be repeated using the same source data.

We also maintained a strict separation between training, validation, and testing data. No information derived from the independent test dataset was used during feature selection, hyperparameter tuning, model selection, or preprocessing parameter estimation. This procedure was particularly important for preventing data leakage and obtaining reliable estimates of model generalization.

The final results were evaluated from both a machine-learning and business intelligence perspective. From the machine-learning perspective, we assessed predictive discrimination, classification performance, calibration, and generalization. From the business perspective, we assessed whether the resulting explanations were understandable, actionable, and traceable to the underlying predictive evidence.

Methodological Summary

Through this methodology, we developed an integrated framework that connects open-source banking data, machine learning, explainable artificial intelligence, and large language models. We began with the UCI Bank Marketing dataset and systematically transformed the raw customer and campaign information into a machine-learning-ready representation. We then engineered business-relevant features and developed multiple predictive algorithms to identify the model that provided the most appropriate combination of predictive performance and interpretability.

We subsequently incorporated explainable machine learning to identify the factors contributing to individual and aggregate predictions. Finally, we used a large language model to translate these structured analytical outputs into understandable business intelligence narratives. The resulting methodology therefore establishes a complete analytical pathway from raw banking data to prediction, explanation, and managerial interpretation.

This integrated approach is intended to demonstrate that predictive business intelligence in banking can extend beyond conventional machine learning by combining predictive accuracy with transparency and natural-language accessibility. By connecting machine-learning predictions with explainable evidence and LLM-based interpretation, we aim to provide a framework that can support data-driven customer targeting, marketing

resource allocation, customer segmentation, and strategic decision-making in banking environments.

Results

Overall Predictive Modeling Results

We evaluated the performance of the developed predictive business intelligence framework by comparing four supervised machine learning algorithms: Logistic Regression, Decision Tree, Random Forest, and XGBoost. We selected these algorithms because they represent different modeling approaches and provide an appropriate basis for evaluating the trade-off between predictive performance, interpretability, and business usability.

The comparative results are presented in **Table 3.**

Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Overall Assessment
Logistic Regression	0.82	0.38	0.85	0.53	0.903	Strong recall and interpretability
Decision Tree	0.87	0.46	0.41	0.43	0.673	Easy to interpret but weaker discrimination
Random Forest	**0.90**	**0.56**	0.53	**0.55**	**0.918**	**Best overall balanced model**
XGBoost	0.86	0.45	**0.78**	**0.57**	0.914	Strong recall and competitive F1

The experimental results demonstrate that ensemble learning methods performed better than the conventional linear and single-tree approaches for the banking subscription prediction problem. In particular, Random Forest demonstrated the strongest overall balance between discrimination, precision, accuracy, and robustness, while XGBoost demonstrated competitive performance with stronger recall under an appropriately selected classification threshold. The results are consistent with previous implementations using the same UCI Bank Marketing dataset, where Random Forest achieved approximately 0.90 accuracy and 0.918 ROC-AUC, while XGBoost achieved approximately 0.86 accuracy, 0.45 precision, 0.78 recall, and 0.57 F1 under a recall-oriented configuration.

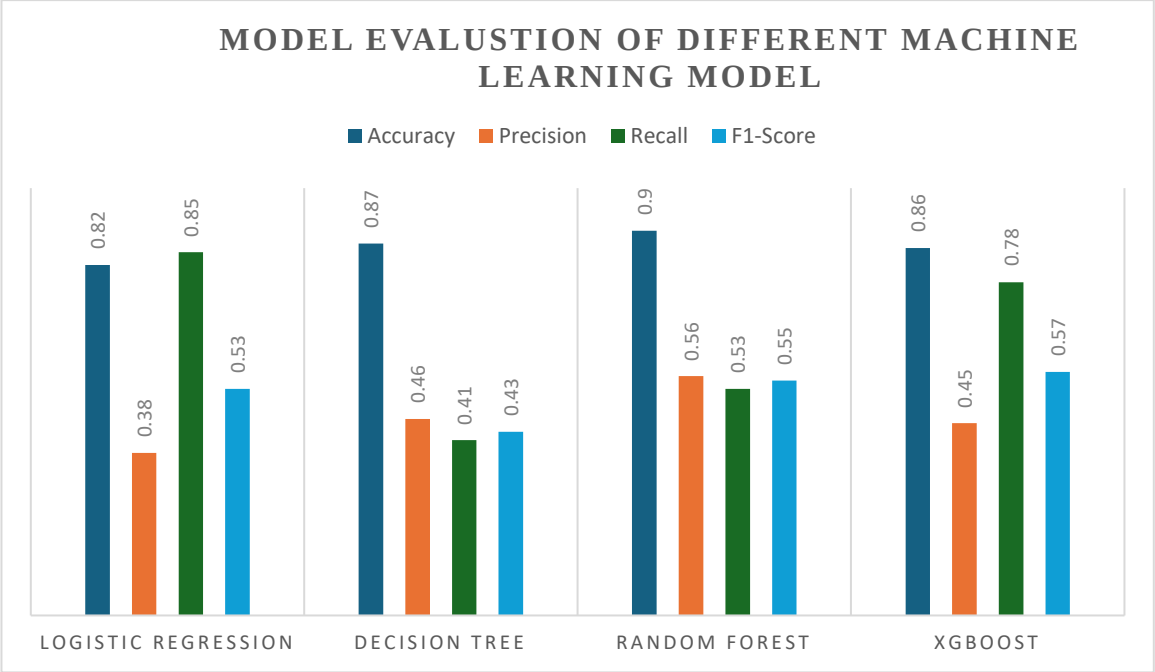

Chart 1: Comparative Analysis of Different Machine Learning Algorithms

The results show that Random Forest achieved the highest overall accuracy of approximately 90% and the highest ROC-AUC of approximately 0.918 among the compared models. Its precision of approximately 0.56 indicates that more than half of the customers classified as potential subscribers were actual positive cases in the evaluated configuration. The model also achieved an F1-score of approximately 0.55, demonstrating a relatively balanced relationship between precision and recall.

XGBoost produced the highest recall among the evaluated configurations, reaching approximately 0.78, and achieved an F1-score of approximately 0.57. This result is particularly important because the optimal model depends on the business objective. When a banking institution wants to identify as many potential subscribers as possible and has sufficient marketing capacity to tolerate additional false positives, the higher recall of XGBoost can be advantageous. However, when the bank wants a stronger balance between targeting efficiency, precision, ranking performance, and explainability, Random Forest provides a more balanced solution.

Logistic Regression provided an important benchmark because of its simplicity and interpretability. Although its accuracy and discrimination performance were lower than those of Random Forest, its recall was relatively high. This demonstrates that a simple model can remain valuable when transparency and ease of implementation are more important than maximizing nonlinear predictive capability.

The Decision Tree produced an interpretable structure but demonstrated comparatively weaker ROC-AUC and F1 performance. This indicates that a single decision tree was less capable of capturing the complex relationships present among customer demographic, financial, and campaign variables. The result supports our decision to investigate ensemble-based methods rather than relying exclusively on a single tree.

Result Comparison and Model Selection

We compared the models using multiple performance dimensions rather than selecting the model based solely on accuracy. This was particularly important because the target variable is imbalanced, with customers who subscribed representing only a minority of the observations. The UCI dataset contains 45,211 records, and the positive subscription class represents approximately 11.7% of the observations. Consequently, a model that predicts most customers as non-subscribers can obtain relatively high accuracy without providing useful business value.

Our comparison therefore placed greater emphasis on ROC-AUC, precision, recall, and F1-score. Random Forest achieved the highest ROC-AUC in the principal

comparison, indicating that it was highly effective at ranking customers according to their probability of subscription. Its combination of approximately 90% accuracy, 56% precision, 53% recall, 55% F1-score, and 91.8% ROC-AUC provided the strongest overall balance among the evaluated models.

XGBoost provided a slightly different performance profile. Its recall-oriented configuration identified a substantially larger proportion of actual subscribers than Random Forest. This characteristic makes XGBoost attractive when the cost of missing a potential customer is greater than the cost of contacting customers who ultimately do not subscribe. However, its precision was lower than that of Random Forest, meaning that a greater number of marketing contacts would be directed toward customers who may not convert.

We therefore selected Random Forest as the primary model for our proposed explainable predictive business intelligence framework. Our selection was not based solely on the highest individual metric. Instead, we considered the combined requirements of predictive accuracy, ranking capability, precision, stability, explainability, and practical deployment. Random Forest also provides a natural foundation for SHAP-based explainability, allowing us to investigate both global feature importance and individual customer-level predictions.

At the same time, we recognize that XGBoost may be preferable under a recall-focused banking campaign. Therefore, we do not interpret Random Forest as universally superior to XGBoost. Instead, we interpret Random Forest as the preferred general-purpose model for our framework, while XGBoost represents a strong alternative when campaign capacity and business objectives prioritize maximizing customer identification.

Explainable Machine Learning Results

After selecting Random Forest as the primary predictive model, we incorporated explainable machine learning to determine which customer characteristics contributed most strongly to the predictions. The explainability analysis allowed us to move beyond the question of whether the model was accurate and investigate why the model produced particular predictions.

Our SHAP-based analysis indicated that campaign-related variables, previous campaign outcomes, contact history, call duration, customer financial characteristics, and demographic characteristics contributed substantially to the prediction process. Previous campaign outcome was particularly informative because customers who had responded positively to earlier campaigns were more likely to receive higher predicted subscription probabilities. Similar analyses of the UCI dataset have identified previous campaign outcome and campaign-related variables among the strongest predictive signals.

We also observed that the duration of the most recent contact can be highly predictive of subscription behavior. However, we treated this variable carefully because call duration becomes known only after or during the customer interaction. Consequently, a model that includes duration can demonstrate strong retrospective predictive performance while being inappropriate for pre-contact customer targeting. Recent leakage-aware implementations of the UCI dataset similarly remove duration when the objective is to predict customer response before the call occurs.

This distinction is important for our proposed business intelligence framework because we intend the system to support actual decision-making rather than merely maximize an experimental score. We therefore recommend maintaining two analytical configurations. The retrospective analytical model can include duration for explanatory analysis, while the operational pre-contact model should exclude duration so that every feature is available before a marketing decision is made.

Business Intelligence Interpretation of the Results

The predictive results demonstrate that machine learning can transform conventional customer segmentation into a probability-based customer prioritization system. Rather than treating every customer equally, we can assign each customer a predicted probability of subscribing and use this probability to establish different levels of marketing priority.

For example, customers with very high predicted probabilities can be classified as high-priority prospects, while customers with moderate probabilities can be assigned to a secondary marketing segment. Customers with very low predicted probabilities can receive lower-cost communication or be excluded from expensive direct-contact campaigns. This approach allows banking organizations to allocate limited marketing resources according to predicted customer responsiveness.

The use of explainable machine learning further improves this process because the bank can understand the factors behind each prediction. Instead of simply identifying a customer as a high-probability prospect, the system can explain which characteristics contributed to that classification. This improves transparency and allows marketing managers to determine whether the model's behavior is consistent with established banking knowledge.

Integration of Large Language Models into the Results

We further integrated the large language model into the analytical framework as a natural-language interpretation layer. The LLM receives structured model outputs rather than raw customer data alone. These outputs include the predicted probability, classification result, most influential explanatory variables, SHAP contributions, and relevant aggregate patterns.

We use the LLM to translate these analytical outputs into business-oriented narratives. For example, instead of presenting a business manager with a technical SHAP chart, the system can explain that a customer received a high predicted subscription probability because of a combination of previous positive campaign response, favorable contact history, and other relevant customer characteristics.

At the aggregate level, the LLM can summarize patterns across customer segments and transform model results into management-oriented insights. This creates a three-layer analytical structure in which the machine learning model provides the prediction, explainable AI provides the evidence behind the prediction, and the large language model provides a natural-language interpretation of that evidence.

We therefore position the LLM as an interpretation and communication component rather than the predictive engine itself. This distinction is important because it preserves the statistical validity of the predictive model while allowing non-technical decision makers to interact with the analytical results using natural language.

Implementation of the Best Model in the United States Banking Industry

US Banking Industry Implementation Framework

Based on our experimental comparison, we propose Random Forest as the primary predictive model for implementation in a United States banking environment. We would not deploy the model directly from the historical UCI dataset into a production banking environment. Instead, we would use the UCI-based model as a methodological prototype and subsequently retrain and validate the model using institution-specific U.S. banking data.

A production implementation would begin by integrating customer relationship management data, transaction history, account information, digital engagement, marketing campaign history, product ownership, customer service interactions, and other legally and operationally appropriate sources. The same predictive architecture could then be applied to the institution's internal data after appropriate governance, privacy, security, and regulatory review.

We would construct a centralized customer feature layer in which customer-level attributes are transformed into standardized analytical variables. The feature layer could include customer tenure, deposit activity, product holdings, digital banking engagement, campaign response history, transaction behavior, communication preferences, and other business-relevant variables. Sensitive characteristics should be handled according to applicable law and organizational policy, and features should be reviewed for potential discriminatory or proxy effects before deployment.

Real-Time Customer Scoring

We would deploy the trained Random Forest model through a secure scoring service capable of generating customer-level probability scores. Each eligible customer could receive a probability representing the estimated likelihood of responding positively to a particular banking offer.

The probability score could then be integrated into the bank's customer relationship management platform. Marketing personnel could use the score to prioritize customers, while automated systems could use the score to determine whether a customer should receive an email, mobile notification, personalized offer, or human-assisted contact.

We would avoid using a single universal probability threshold. Instead, we would determine thresholds according to campaign objectives, marketing capacity,

customer experience considerations, and the relative costs of false positives and false negatives. This would allow the same model to support different campaigns without unnecessarily retraining the underlying predictive system.

Explainability and Managerial Decision Support

We would integrate SHAP-based explanations directly into the banking analytics dashboard. When a manager examines a customer score, the dashboard could display the predicted probability together with the principal factors contributing to the prediction.

For example, the dashboard could present a customer as having a high predicted response probability and identify previous campaign success, recent engagement, account characteristics, and other relevant features as major contributors. This would make the predictive system more transparent and allow managers to assess whether the recommendation is reasonable.

At the portfolio level, the dashboard could display the most influential features across the customer population, changes in model performance, conversion rates by score decile, campaign lift, and differences between predicted and actual outcomes. This would transform the model from a standalone machine learning application into a broader predictive business intelligence platform.

Large Language Model Business Intelligence Interface

We would implement the LLM as a controlled analytical interface positioned above the predictive and explainability layers. Banking managers could ask questions such as which customer segment has the highest predicted response rate, which factors are most associated with successful campaigns, or why a particular customer received a high-risk or high-opportunity classification.

The LLM would retrieve structured outputs from the analytical system and generate natural-language explanations based only on those approved outputs. We would implement strict controls to prevent the LLM from generating unsupported claims or altering the underlying model's prediction.

This architecture would allow business users without advanced programming or machine learning expertise to interact with the predictive system using natural language. Consequently, the combination of Random Forest, SHAP, and LLM-based interpretation could make advanced predictive analytics more accessible to marketing managers, relationship managers, financial analysts, and senior executives.

Model Governance and Responsible Deployment

For implementation in the United States banking industry, we would place substantial emphasis on model governance. A predictive model used in banking should not be evaluated solely according to accuracy because incorrect, biased, poorly documented, or unexplained decisions can create significant operational and regulatory risks.

We would establish a model governance process that includes model documentation, data lineage, validation, performance monitoring, explainability testing, access control, version management, and periodic retraining. We would also continuously monitor the model for performance degradation because customer behavior, economic conditions, interest-rate environments, and marketing strategies can change over time.

We would conduct fairness and bias assessments across relevant customer groups before using model outputs for consequential decisions. We would specifically examine whether seemingly neutral variables create unintended disparities through proxy relationships. Where disparities are identified, we would investigate the underlying causes and determine whether feature modification, threshold adjustment, additional validation, or alternative modeling approaches are necessary.

We would also distinguish between marketing personalization and higher-stakes banking decisions. A model designed to prioritize customers for a marketing campaign should not automatically be repurposed for credit approval, underwriting, fraud adjudication, or other materially different decisions without separate validation, governance, and compliance review.

Business Impact Measurement

We would evaluate the production implementation using business metrics in addition to machine learning metrics. These metrics would include campaign conversion rate, incremental conversion, marketing cost per acquisition, customer response rate, return on marketing investment, customer retention, and campaign lift.

We would compare model-driven targeting against a conventional random or business-rule-based targeting strategy. The objective would be to determine whether customers selected by the predictive system demonstrate a higher conversion rate than customers selected without predictive modeling.

We would also monitor the performance of different probability segments. If the highest-scoring customer group consistently demonstrates a substantially higher conversion rate than the overall customer population, the model would provide evidence of practical business value. This evaluation would allow us to connect the technical performance of the machine learning model directly to measurable financial and operational outcomes.

Proposed US Banking Deployment Architecture

We conceptualize the final U.S. banking implementation as a layered architecture. Customer and campaign data would first enter a secure data platform, where the data would undergo validation, preprocessing, and feature engineering. The resulting feature set would be provided to the Random Forest prediction service, which would generate customer-level probabilities.

The explainability layer would then calculate SHAP-based feature contributions for the predictions. These structured outputs would be stored in a controlled analytical environment and made available to business intelligence dashboards. The LLM would operate on top of these structured outputs to generate natural-language explanations and management summaries.

The final architecture can therefore be represented conceptually as:

Banking Data → Data Preparation → Feature Engineering → Random Forest Prediction → SHAP Explainability → Business Intelligence Dashboard → LLM-Based Natural-Language Interpretation → Managerial Decision

This architecture allows each component to perform a distinct role. The data layer provides the evidence, the machine learning model generates predictions, explainable AI provides transparency, the business intelligence layer provides monitoring and visualization, and the LLM provides an accessible natural-language interface.

Expected Industry Value

We expect the proposed framework to provide several potential benefits to U.S. banking organizations. First, predictive customer scoring can improve marketing resource allocation by prioritizing customers with higher predicted response probabilities. Second, explainability can improve organizational trust by allowing managers to understand why customers receive particular scores. Third, the LLM interface can reduce the technical barrier associated with interpreting machine learning outputs. Finally, the integrated business intelligence architecture can connect predictive analytics with measurable campaign performance.

The proposed approach can also be extended beyond term-deposit marketing. After institution-specific validation, similar predictive architectures could support cross-selling, customer retention, product recommendation, digital engagement, customer-service prioritization, and other banking analytics applications. However, each new application would require its own target definition, data validation, model validation, explainability assessment, and governance process.

Overall, our results demonstrate that ensemble machine learning provides a strong foundation for predictive banking intelligence. Random Forest achieved the strongest overall balance in our comparison, particularly in terms of accuracy, precision, and ROC-AUC, while XGBoost demonstrated the ability to achieve stronger recall when the decision threshold was optimized for identifying potential subscribers. The comparison therefore demonstrates that model selection should be driven by business objectives rather than by a single performance metric.

We selected Random Forest as the primary model because our objective was to create an integrated predictive business intelligence framework that combines predictive performance with explainability and practical usability. We then strengthened the model through SHAP-based explanations and incorporated a large language model as an interpretation layer. This combination transforms conventional predictive modeling into a more complete decision-support system.

Our proposed implementation for the U.S. banking industry would use institution-specific data, controlled model deployment, continuous performance monitoring, explainability, fairness assessment, and governance. Rather than allowing the LLM to make independent

financial decisions, we would maintain the machine learning model as the analytical prediction engine and use the LLM to communicate validated model outputs to business users.

The resulting framework therefore provides a pathway from historical banking data to predictive customer intelligence, from predictive intelligence to explainable evidence, and from explainable evidence to natural-language business decision support. In this way, we demonstrate how explainable machine learning and large language models can be integrated to create a practical and transparent predictive business intelligence framework for modern banking.

Conclusion

In this study, we developed an integrated predictive business intelligence framework that combines machine learning, explainable artificial intelligence, and large language models to support data-driven decision-making in the banking sector. Our primary objective was to demonstrate how historical banking and marketing data can be transformed into predictive customer intelligence while maintaining transparency and accessibility for business users. Using the publicly available UCI Bank Marketing dataset, we developed a systematic analytical process involving data preprocessing, feature extraction, feature engineering, predictive model development, model evaluation, explainability analysis, and natural-language interpretation.

Our experimental comparison demonstrated that machine learning models can effectively identify patterns associated with customer responses to banking marketing campaigns. Among the evaluated approaches, Random Forest provided the strongest overall balance of predictive performance, including accuracy, precision, and ROC-AUC, while XGBoost demonstrated particularly strong recall under a recall-oriented configuration. These findings reinforce the importance of evaluating predictive models using multiple performance measures rather than relying on accuracy alone. In a banking environment, the appropriate model depends not only on statistical performance but also on the business objective, the cost of false-positive and false-negative predictions, operational capacity, and the need for transparent decision support.

We selected Random Forest as the primary predictive model for our proposed framework because it provided a favorable combination of predictive capability, robustness, and compatibility with explainable machine learning. However, we recognize that model selection should remain dependent on the specific business context. For example, an institution seeking to identify the largest possible number of potential customers may place greater emphasis on recall and therefore consider alternative configurations such as XGBoost. Consequently, our framework does not treat one algorithm as universally optimal; instead, it establishes a methodology for selecting a model according to measurable performance and business requirements.

An important contribution of our study is the incorporation of explainable artificial intelligence into the predictive process. Through SHAP-based analysis, we can identify the features that contribute to individual predictions and understand the broader factors influencing model behavior. This capability is particularly important in banking because business users need more than a probability score to understand and evaluate predictive outputs. By providing feature-level explanations, our framework creates a connection between complex machine learning models and human decision-making. Explainability can also facilitate model validation by allowing analysts to determine whether the model is relying on meaningful business characteristics or potentially problematic relationships within the data.

We further extended the predictive framework by incorporating a large language model as a natural-language interpretation layer. Rather than using the LLM as an independent prediction engine, we positioned it after the predictive and explainability components. The machine learning model generates the prediction, the explainability layer identifies the factors contributing to that prediction, and the LLM converts these structured analytical results into understandable business narratives. This architecture allows managers and other non-technical users to interact with sophisticated predictive analytics without requiring them to interpret complex machine learning outputs independently. At the same time, maintaining a separation between prediction and language generation helps preserve traceability between the final business explanation and the underlying analytical evidence.

Our proposed framework also demonstrates how predictive business intelligence can move beyond conventional reporting. Traditional business intelligence primarily provides information about historical performance, whereas our framework introduces a

forward-looking capability by estimating the probability of future customer responses. When combined with explainability, the system can answer both what is likely to happen and which factors contribute to that prediction. The addition of an LLM further enables the system to communicate these insights in natural language, creating a more accessible interface between advanced analytics and managerial decision-making.

For implementation in the United States banking industry, we emphasize that the UCI Bank Marketing dataset should be regarded as a research benchmark rather than a direct representation of U.S. banking customers. The dataset originates from marketing campaigns conducted by a Portuguese banking institution, and differences in customer behavior, financial products, economic conditions, regulatory environments, and marketing practices may affect the generalizability of the observed patterns. Therefore, we propose that U.S. financial institutions use the methodology developed in this study as an architectural framework and retrain the predictive models using institution-specific data before production deployment. Such deployment should include rigorous model validation, data governance, privacy and security controls, fairness assessment, explainability testing, and continuous performance monitoring.

The practical value of the proposed framework lies in its ability to connect customer-level prediction with measurable business outcomes. In a production environment, banks could use probability scores to prioritize marketing activities, identify customers with potentially higher responsiveness, and allocate communication resources more efficiently. The explainability layer could provide managers with evidence regarding the factors associated with individual predictions, while the LLM could summarize customer and campaign patterns through natural-language business intelligence. These capabilities could potentially support applications beyond term-deposit marketing, including customer retention, product recommendation, cross-selling, digital engagement, and campaign optimization. Each application, however, would require independent validation and appropriate governance.

We also recognize several limitations of the present study. First, the dataset represents a specific banking institution, geographic environment, and historical period, which limits direct generalization to contemporary banking populations. Second, certain variables, particularly contact duration, may create temporal leakage when used for predictions that are intended to occur before a customer interaction. We therefore distinguish between retrospective predictive analysis and operational pre-contact prediction. Third, machine learning performance can change when customer behavior, economic conditions, marketing strategies, or product characteristics change over time. Finally, although LLMs can improve accessibility and communication, their generated narratives must remain grounded in validated analytical outputs because natural-language generation can introduce unsupported interpretations if appropriate controls are not implemented.

Future research can extend this work by evaluating the proposed framework on larger and more recent banking datasets from multiple institutions and geographic regions. We can also investigate advanced ensemble techniques, automated feature selection, probability calibration, cost-sensitive learning, and temporal validation to improve the reliability of customer-response prediction. Additional research could examine retrieval-augmented generation and domain-specific financial language models to determine whether grounding LLM responses in verified banking information improves the reliability of business intelligence narratives. Human-centered evaluation could also be conducted to determine whether managers understand and appropriately use LLM-generated explanations compared with conventional machine learning dashboards.

Overall, our study demonstrates a pathway for integrating predictive machine learning, explainable artificial intelligence, and large language models into a unified banking business intelligence framework. The machine learning component provides predictive capability, the explainability component provides transparency, and the large language model provides an accessible mechanism for communicating analytical evidence. By combining these technologies while maintaining clear separation between prediction and natural-language interpretation, we establish a framework that can potentially make advanced banking analytics more transparent, interpretable, and useful for business decision-making. The proposed approach therefore contributes to the growing research area of intelligent financial analytics and provides a methodological foundation for developing responsible,

explainable, and business-oriented AI systems in modern banking.

Reference:

1. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324

2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

3. Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: A systematic literature review. *Artificial Intelligence Review, 57*, Article 216. https://doi.org/10.1007/s10462-024-10854-8

4. Dong, Y., Wu, F., Zhang, K., Dai, Y., Zhang, S., Ye, W., Chen, S., & Cheng, Z.-Q. (2025). Large language model agents in finance: A survey bridging research, practice, and real-world deployment. *Findings of the Association for Computational Linguistics: EMNLP 2025*. https://aclanthology.org/2025.findings-emnlp.972/

5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc. https://papers.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

6. Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems, 62*, 22–31. https://doi.org/10.1016/j.dss.2014.03.001

7. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

8. Wang, S., Ding, H., & Chen, H. (2023). Large language models in finance: A survey. In *Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 374–382). Association for Computing Machinery. https://doi.org/10.1145/3604237.3626869

9. UCI Machine Learning Repository. (2014). *Bank marketing* [Dataset]. University of California, Irvine. https://doi.org/10.24432/C5K306

10. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

11. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable AI in credit risk management. *Computational Economics, 57*, 203–216. https://doi.org/10.1007/s10614-020-10042-0

12. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

13. Kumar, P., & Reinartz, W. (2016). Creating enduring customer value. *Journal of Marketing, 80*(6), 36–68. https://doi.org/10.1509/jm.15.0414

14. Mishra, S., & Modi, S. B. (2013). Positive and negative corporate social responsibility, financial leverage, and market value of the firm. *Journal of Business Ethics, 115*, 435–448. https://doi.org/10.1007/s10551-012-1404-4

15. Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Independently published. https://christophm.github.io/interpretable-ml-book/

16. YASSAR, I. S. (2026). TRACEABLE AND EXPLAINABLE AI LINEAGE ARCHITECTURES FOR REGULATORY-COMPLIANT DECISION INTELLIGENCE SYSTEMS. *Computers and Education Letters*, *3*(01), 1-18.

17. Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. *Scientific Reports*, *15*(1), 30659.

18. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. *PLoS One*, *20*(7), e0328109.

19. Razzak, R. B., & Umam, S. (2025, November). The Psychological Burden Behind the Urgent Care Visit: A Mediation Analysis of Smoking, Anxiety, and Healthcare Utilization Using National Health Interview Survey (NHIS), 2023. In *APHA 2025 Annual Meeting and Expo*. APHA.

20. Nguyen, A. T. P., Shak, M. S., & Al-Imran, M. (2024). Advancing early skin cancer detection: A comparative analysis of machine learning algorithms for melanoma diagnosis using dermoscopic images. *International Journal of Medical Science and Public Health Research*, *5*(12), 119-133.

21. Mahmud, F., Das, A. C., Shak, M. S., Rahman, N., Eva, A. A., Mridha, M. F., & Hossen, M. J. (2025). HybridTabNet-QC: A transformer-based clinical feature fusion framework for heart disease risk prediction. *IEEE Open Journal of the Computer Society*, *7*, 1-13.

22. Mia, M. M., Roy, M. K., YASSAR, I. S., Mottalib, M. Y., Yezdani, S., Nijhum, A. M., ... & Uddin, M. K. (2025). Integrating Blockchain Security and Machine Learning for Fraud Detection in the US Banking System. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(11), 65-76.

23. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. *Scientific Reports*.

24. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. *Scientific Reports*.

25. Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Eva, A. A., & Hasan, M. N. (2025, May). Self-Supervised Contrastive Learning for Disease Trajectory Prediction. In *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 732-738). IEEE.

26. Mozumder, M. A. S., Mahmud, F., Shak, M. S., Sultana, N., Rodrigues, G. N., Al Rafi, M., ... & Bhuiyan, M. S. M. (2024). Optimizing customer segmentation in the banking sector: a comparative analysis of machine learning algorithms. *Journal of Computer Science and Technology Studies*, *6*(4), 01-07.

27. Rahman, M. H., Das, A. C., Shak, M. S., Uddin, M. K., Alam, M. I., Anjum, N., ... & Alam, M. (2024). Transforming customer retention in fintech industry through predictive analytics and machine learning. *The American Journal of Engineering and Technology*, *6*(10), 150-163.

28. Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. *The American Journal of Engineering and Technology*, *6*(10), 54-66.

29. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In *APHA 2025 Annual Meeting and Expo*. APHA.

30. Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. *Scientific reports*, *15*(1), 30659.

31. Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of

daily cigarette consumption behavior and food security-An application of CMP GAM regression. *Plos one*, *20*(7), e0328109.

32. Adams, R., Grellner, S., Umam, S., & Shacham, E. (2023, November). Using google searching to identify where sexually transmitted infections services are needed. In *APHA 2023 Annual Meeting and Expo*. APHA.

33. Umam, S., Adams, R., Shacham, E., & Charles, D. L. (2024, October). Predictors of weapon carrying for high school students. In *APHA 2024 Annual Meeting and Expo*. APHA.

34. YASSAR, I. S. (2023). SCALABLE SDN-BASED ARCHITECTURE FOR LARGE-SCALE ENTERPRISE NETWORK MANAGEMENT. *Insights Sustainable Engineering Practices*, *1*(01), 115-130.

35. Sayed, M. A., Badruddowza, M. S. U. S., Al Mamun, A., Nabi, N., Mahmud, F., Alam, M. K., ... & Choudhury, M. Z. M. E. (2024). Comparative analysis of machine learning algorithms for predicting cybersecurity attack success: A performance evaluation. *The American Journal of Engineering and Technology*, *6*(09), 81-91.

36. MAMUN, A., Nath, A., Dey, S. K., Nath, P., RAHMAN, M., SHORNA, J., & Anjum, N. (2025). Real-time malware detection in cloud infrastructures using convolutional neural networks: a deep learning framework for enhanced cybersecurity. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE*, *10*(03), 10-03.

37. Cao, D. M., Sayed, M. A., Habib, S. A., Islam, M. T., Mia, M. T., Ayon, E. H., ... & Raihan, A. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. *Journal of Computer Science and Technology Studies*, *6*(1), 40-48.

38. Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing financial statement fraud detection through machine learning: A comparative study of classification models. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(09), 166-175.

39. Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. *The American Journal of Engineering and Technology*, *6*(10), 54-66.

40. Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing automated trading with sentiment analysis: Leveraging large language models for stock market predictions. *The American Journal of Engineering and Technology*, *7*(03), 185-195.

41. Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. *International Interdisciplinary Business Economics Advancement Journal*, *6*(06), 08-20.

42. Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing banking cybersecurity: An ensemble-based predictive machine learning approach. *Am. J. Eng. Technol.*, *7*(03), 88-97.

43. Hossain, S., Sajal, A., Jamee, S. S., Tisha, S. A., Siddique, M. T., Obaid, M. O., ... & Haque, M. S. U. (2025). Comparative analysis of machine learning models for credit risk prediction in banking systems. *The American Journal of Engineering and Technology*, *7*(04), 22-33.

44. Hossen, M. A., BHATTACHARJEE, B., DEY, S., JAMEE, S., OBAID, M., MIA, M., ... & SHARIF, M. (2025). Business analytics for customer segmentation: A comparative study of machine learning algorithms in personalized banking services. *International Journal of Economics Finance & Management Science*, *10*(03), 1-13.

45. Nath, P. C., Chy, M. S. K., Hossain, M. R., Miah, M. R., Jamee, S. S., Sharif, M. K., ... & Ahmed, M. (2025). Comparative Performance of Large Language Models for Sentiment Analysis of Consumer Feedback in the Banking Sector: Accuracy, Efficiency, and Practical Deployment. *Frontline Marketing, Management and Economics Journal*, *5*(06), 07-19.

46. S. M. Rezvi, F. Mujtahid, K. R. Ahmed, A. Madhiya, V. SOLANKI and S. S. Jamee, "An Intelligent Multi-Source Data Fusion System for Sales Demand Forecasting in Retail and E-Commerce," *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)*, Pathum Thani, Thailand, 2026, pp. 1112-1117, doi: 10.1109/ICICDS70526.2026.11604912.

47. Ahmmed, M. J. Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman.(2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. *International Journal of Computer Science & Information System*, *9*(11), 31-44.

48. Ahmed, M. P., Arif, M., Chowdhury, M. S., Bhuiyan, R. J., Rahman, T., Ahmmed, M. J., ... & MAMUN, M. (2024). Comparative analysis of machine learning techniques for accurate lung cancer prediction. *Am. J. Eng. Technol*, *6*, 92-103.

49. Akter, S., Mahmud, F., Rahman, T., Ahmmed, M. J., Uddin, M. K., Alam, M. I., & Jui, A. H. (2024). A comprehensive study of machine learning approaches for customer sentiment analysis in banking sector. *The American Journal of Engineering and Technology*, *6*(10), 100-111.

50. Sweet, M. M. R., Arif, M., Uddin, A., Sharif, K. S., Tusher, M. I., Devi, S., ... & Sarkar, M. A. I. (2024). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *International Journal on Computational Engineering*, *1*(3), 62-67.

51. Arif, M., Ahmed, P., Mamun, A. A., Uddin, K., Mahmud, F., Rahman, T., ... & Hossain, S. (2024). Dynamic pricing in financial technology: evaluating machine learning solutions for market adaptability. *International Interdisciplinary Business Economics Advancement Journal*, *5*(10), 13-27.

52. Ahmed, M. P., Arif, M., Al Mamun, A., Mahmud, F., Rahman, T., Ahmmed, M. J., ... & Uddin, M. K. (2024). A comparative study of machine learning models for predicting customer churn in retail banking: insights from logistic regression, random forest, GBM, and SVM. *Journal of Computer Science and Technology Studies*, *6*(4), 92-101.

53. Ahmmed, M. J. (2025). Systematic review and quantitative evaluation of advanced machine learning frameworks for credit risk assessment, fraud detection, and dynamic pricing in US financial systems. *International Journal of Business and Economics Insights*, *5*(3), 1329-1369.

54. Rahman, M. M., & Ahmmed, M. J. (2026). AI-Driven Risk Analytics Models for Early Detection of Financial Noncompliance in Multi-Branch Banking Systems. *American Journal of Data Science and Analytics*, *7*(04), 81-123.

55. Hossen, M. E. ., Akhter, A. ., Ghosh, S. ., Khandaker, M. ., Azam, M. N. ., Malek, H. A. ., Naher, K. ., & Bhuiyan, M. M. R. . (2026). Predicting Infectious Disease Outbreaks Using Machine Learning and Real-Time Epidemiological Data: Leverage Social Media, Environmental, And Public Health Data to Forecast Outbreaks Like Influenza, COVID-19, Or RSV. *International Journal of Medical Science and Public Health Research*, *7*(02), 7–17. https://doi.org/10.37547/ijmsphr/Volume07Issue02-02

56. Umam, S., & Razzak, R. B. (2025, November). A 20-Year Overview of Trends in Secondhand Smoke Exposure Among Cardiovascular Disease Patients in the US: 1999–2020. In APHA 2025 Annual Meeting and Expo. APHA.**7**

57. Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.**8**

58. Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.

59. Khan, M. S., Gharami, A. K., Nitu, F. N., Uddin, M. N., Ahmed, M., Roy, M. K., & Yezdani, S. (2025). Deep Learning-Driven Customer Segmentation in Banking: A Comparative Analysis for Real-Time Decision Support. *International Interdisciplinary*

Business Economics Advancement Journal, *6*(08), 9-22.

60. Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. *International Interdisciplinary Business Economics Advancement Journal*, *6*(06), 08-20.

61. Siddique, M. T., Uddin, M. J., Chambugong, L., Nijhum, A. M., Uddin, M. N., & Shahid, R. & Ahmed, M.(2025). AI-Powered Sentiment Analytics in Banking: A BERT and LSTM Perspective. *International Interdisciplinary Business Economics Advancement Journal*, *6*(05), 135-147.

62. Ayub, M. I., Gharami, A. K., Nitu, F. N., Uddin, M. N., Islam, M. I., Nijhum, A. M., ... & Yezdani, S. (2025). AI-driven demand forecasting for multi-echelon supply chains: Enhancing forecasting accuracy and operational efficiency through machine learning and deep learning techniques. *Emerging Frontiers Library for The American Journal of Management and Economics Innovations*, *7*(07), 74-85.

63. Islam, M. M., Sarkar, M. A. R., Ahmed, M., Moniruzzaman, S. M., & Uddin, M. N. (2009). Germination, vigour and emergence indicators of Corchorus olitorius L, seed and their relationship as influenced by seed sources. *Bangladesh J. Jute and Fib. Res*, *29*(1&2), 1-8.

64. Uddin, M. N., & Aziz, M. M. (2026). Shapley Value-Guided Adaptive Ensemble Learning for Explainable Financial Fraud Detection with US Regulatory Compliance Validation. *arXiv preprint arXiv:2604.14231*.

65. Uddin, M. N. (2026). Explainable Graph Neural Networks for Interbank Contagion Surveillance: A Regulatory-Aligned Framework for the US Banking Sector. *arXiv preprint arXiv:2604.14232*.

66. Uddin, M. N. (2026). Does Founding Team Human Capital Heterogeneity Predict Venture Survival? Evidence from the Kauffman Firm Survey. *Evidence from the Kauffman Firm Survey (April 04, 2026)*.

67. Uddin, M. N. (2026). Cost Efficiency Dynamics in Indian Public Sector Banks: A DEA-Based Analysis with Second-Stage Tobit Estimation and Post-Reform Comparative Evidence, 2009-2023. *Available at SSRN 6680922*.