

Customer Investment Prediction

Nayana G M¹, Mrs. Pooja K N²

¹Student, 4th Semester MCA, Department of MCA, BIET, Davanagere

²Assistant Professor, Department of MCA, BIET, Davanagere

ABSTRACT — The Customer Investment Prediction project focuses on estimating whether a bank client is going to subscribe to a term deposit using advanced machine learning techniques. By analyzing client data such as demographics financial history and previous bank interactions the technology uncovers hidden patterns and connections that influence investment choices. Traditional customer targeting methods often fail to capture complex customer behavior because they rely on human analysis and basic data. An AI-driven approach reduces marketing inefficiencies and helps banks anticipate potential investors more confidently by increasing accuracy efficiency and scalability. The project uses Random Forest Gradient Boosting using logistic regression XGBoost LightGBM CatBoost and SVM to compare and evaluate the predictive power of various machine learning algorithms. Splitting datasets scaling numerical features and encoding categorical variables are data preparation techniques that guarantee model reliability and generalization. The system also features a real-time prediction module that allows users to input new client data for instantaneous investment likelihood calculations. By enabling data-driven decision-making this project helps banks maximize marketing initiatives increase customer engagement and increase overall term deposit subscription rates.

Keywords: *Deep Learning, Machine Learning, Customer Investment, Algorithms.*

I. INTRODUCTION

The goal of the Customer Investment Prediction project is to predict the likelihood that a bank customer will sign up for a term deposit. This system predicts investment behavior by analyzing client factors like financial history, demographics, and previous interactions with the bank using a variety of machine learning techniques. Banks may identify new clients for focused marketing initiatives, allocate

resources optimally, and enhance overall customer engagement thanks to this predictive capability.

Gradient Boosting, Random Forest, Logistic regression, XGBoost, LightGBM, CatBoost, and SVM are only a handful of the many models utilized in this study. While group techniques like gradient boosting and random forest capture intricate, non-linear patterns, Logistic Regression yields answers for linear correlations that are easy to understand.

While SVM is useful for classification with a strong margin-based approach, advanced algorithms like XGBoost, LightGBM, and CatBoost offer great efficiency and accuracy, especially for structured tabular data.

To ascertain whether a bank customer is likely to sign up for a term deposit, the Customer Investment Prediction project makes use of AI, more especially Machine Learning (ML). Artificial intelligence (AI) enables computers to identify trends in historical customer data and forecast without the need for clear coding for each rule. By examining a range of client characteristics such as financial information prior interactions age occupation and marital status the algorithm can identify patterns and connections that would be difficult and time-consuming for individuals to find.

Two essential AI processes feature engineering and data preprocessing are also included in the AI system. Numerical features are scaled categorical variables are encoded and the information is utilized to create training and testing sets to ensure the models learn effectively and generalize well to new data. This procedure guarantees that both numerical and categorical customer data can be handled by the AI models in an appropriate manner. Data preprocessing which entails dividing the dataset into testable sets and training encoding category features and scaling Numerical characteristics are where the process begins.

After that models are assessed and taught using accuracy measures, enabling users to compare outcomes and decide which approach is best for the prediction task. Both structured numerical data and

categorical inputs are handled consistently by the system ensuring robust performance across models.

RELATED WORK

Recent research highlights the growing adoption of machine learning techniques in banking analytics to predict customer behaviour, improve targeting strategies, and enhance financial decision-making. Rahman and Kumar [1] presented a machine learning-based framework for customer churn prediction in banking, demonstrating that ensemble models outperform traditional classifiers in capturing complex behavioural patterns. Their findings emphasize the importance of data-driven approaches in understanding customer engagement and retention, which is closely related to predicting customer responses to financial products such as term deposits. Several studies have focused specifically on term deposit subscription prediction. The work presented in [2] explored the application of machine learning and big data analytics to predict customer subscription behavior in banking campaigns, reporting significant performance gains using ensemble learning methods. Hou et al. [3] further investigated prediction models for bank deposit subscription using multiple supervised learning algorithms and showed that boosting-based models provide superior accuracy on structured tabular data. These studies validate the suitability of tree-based ensemble methods for modeling customer investment behavior.

The use of predictive analytics and big data in investment banking has also been widely explored. Jasaiwal [4] demonstrated how predictive analytics frameworks support informed decision-making by leveraging large-scale financial data, highlighting the strategic value of machine learning in financial

institutions. Similarly, Haque and Hassan [5] applied machine learning techniques to bank loan prediction, showing that feature engineering and proper preprocessing significantly enhance model performance. Although focused on credit risk, their findings reinforce the importance of data preprocessing and model selection in financial prediction tasks.

Recent works have increasingly emphasized model interpretability and explainable AI in banking applications. Studies on customer churn prediction incorporating interpretability techniques such as SHAP values [6], [7] enable stakeholders to understand the factors influencing model predictions. Such interpretability is crucial for regulatory compliance and managerial trust in predictive systems deployed within financial institutions. Other comparative studies [8], [9] have evaluated classical classifiers against ensemble and boosting models, consistently reporting improved performance for Random Forest, XGBoost, and LightGBM in banking-related prediction tasks.

Moreover, research on predictive modeling for customer deposits and subscription behavior [10], [11] indicates that integrating behavioral attributes, historical interaction data, and demographic features leads to more robust prediction outcomes. Systematic reviews on machine learning and deep learning approaches for customer behavior modeling [12] further confirm that ensemble and hybrid models achieve higher generalization performance compared to singlemodel approaches. Recent preprints and conference papers [13], [14] demonstrate the applicability of advanced gradient boosting methods for term deposit and customer response prediction in marketing campaigns.

EXISTING SYSTEM

The current banking system primarily relies on manual techniques, historical reports, and basic statistical studies to anticipate whereby clients would register for term deposits. In order to target clients, bank employees frequently rely on general heuristics, past experience, or basic demographic and financial data, which is laborious, prone to human error, and unable to identify intricate patterns in customer behavior. As a result, marketing initiatives can be ineffective, prospective clients might be missed, and total subscription rates might continue to be below ideal.

METHODOLOGY

To precisely forecast whether a bank customer is likely to sign up for a term deposit, the suggested method uses a structured machine-learning pipeline. The dataset is first collected and preprocessed by addressing missing values encoding category variables and scaling numerical features in order to ensure consistent model performance. The processed data is then divided into training and testing subsets to ensure the model's efficacy. Several supervised learning models trained to understand the relationship between client characteristics and investment outcomes. Their performance is compared using accuracy metrics and the best model is identified. The system also incorporates an interface that allows users to client data for real-time prediction. Additionally, admin features enable efficient management of users FAQs and user inquiries ensuring smooth platform communication and interaction. This strategy ensures a dependable scalable and user-friendly forecasting system that is capable of decision-making in banking operations.

SYSYTEM ARCHITECTURE

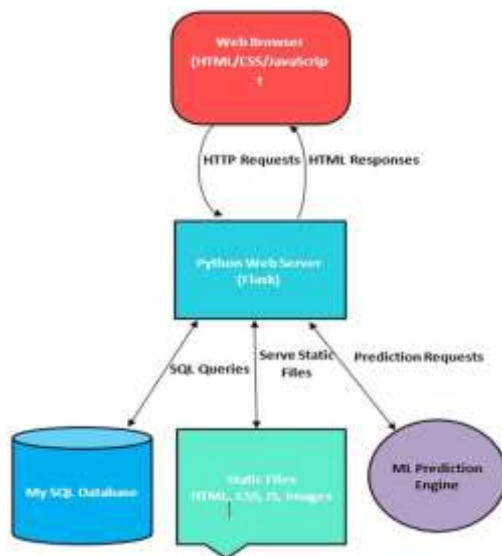


Fig.1: System Architecture of the Customer Investment Prediction

The proposed Customer Investment Prediction system follows a three-tier client-server architecture comprising a presentation layer, application layer, data layer, and an integrated machine learning prediction engine. The web browser acts as the user interface layer, developed using HTML, CSS, and JavaScript, through which users and administrators interact with the system. User requests, including authentication, customer data input, and prediction queries, are transmitted to the backend via HTTP requests, and the corresponding responses are rendered dynamically on the client side.

The application layer is implemented using a Python-based Flask web server, which serves as the central coordinator of system operations. The server handles user authentication, business logic, request routing, and communication with both the database and the machine learning prediction engine. Static resources such as HTML templates, stylesheets, JavaScript files, and images are served directly by the Flask server to ensure efficient client-side rendering.

The data layer consists of a MySQL database that stores user credentials, customer records, frequently asked questions (FAQs), and query logs. Structured SQL queries are used to perform secure data storage and retrieval operations. The machine learning prediction engine is integrated as a modular service within the backend and is responsible for executing trained classification models to predict the likelihood of term deposit subscription. Upon receiving prediction requests, the Flask server forwards the preprocessed input data to the ML engine, which returns probabilistic or binary predictions that are then displayed to the user.

This modular architecture ensures scalability, maintainability, and real-time prediction capability. By decoupling the web interface, application logic, database management, and machine learning components, the system supports seamless updates, efficient performance, and easy integration of improved predictive models in future enhancements.

A. Data Collection and Feature Selection

The dataset comprises historical banking customer records including demographic attributes (age, marital status, education, occupation), financial indicators (account balance, loan status, housing loan), and interaction-related features (contact type, previous campaign outcomes, duration of contact, and frequency of interactions). These attributes serve as input features, while the target variable indicates whether the customer subscribed to a term deposit. Feature selection is performed to remove redundant and non-informative variables, improving model efficiency and reducing noise.

B. Data Preprocessing

Data preprocessing is a critical step to ensure model robustness and generalization. Categorical variables are transformed into numerical representations using label encoding or suitable encoding schemes. Numerical attributes are scaled using standard normalization techniques to ensure uniform feature distributions, particularly for distance-based and margin-based classifiers such as Logistic Regression and Support Vector Machines. Missing values and inconsistencies, if present, are handled using appropriate imputation or filtering strategies. The cleaned dataset is partitioned into training and testing subsets using a stratified split to preserve class distribution.

C. Model Development

Multiple supervised machine learning models are employed to capture both linear and non-linear relationships within customer data. Logistic Regression is used as a baseline linear classifier for interpretability, while ensemble-based methods such as Random Forest and Gradient Boosting are utilized to model complex non-linear interactions. Advanced boosting algorithms including XGBoost, LightGBM, and CatBoost are incorporated to enhance predictive performance on structured tabular data. Support Vector Machines are applied to maximize class separation through margin-based optimization. Hyperparameters for each model are tuned using cross-validation to achieve optimal performance.

D. Model Training and Evaluation

Each model is trained on the preprocessed training dataset and evaluated on the held-out test set. Performance is measured using classification metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis. Comparative evaluation is

conducted to identify the most effective model in terms of predictive accuracy and generalization capability. This multi-model evaluation strategy ensures robust benchmarking and facilitates informed model selection for deployment.

RESULT



Figure 2: Home Page



Figure 2.1: Admin Sign in Page



Figure 2.2: User Login Page



Figure 2.3: Predict Investment



Figure 2.4: Prediction Results

CONCLUSION

By forecasting if a client is likely to sign up for a term deposit, the Customer Investment Prediction system effectively illustrates how machine learning could be enhanced decision-making in the banking industry. The system uses sophisticated models like Gradient Boosting XGBoost LightGBM CatBoost and SVM to provide reliable and comparable insights about client investing behavior. Through efficient preprocessing model training and accuracy evaluation the solution ensures high-quality forecasts that enable tailored financial strategies and targeted marketing. Additionally, the user and admin modules enable smooth communication enabling users to forecast and ask for help while administrators can effectively handle FAQs and user inquiries. All things considered the project gives financial institutions a reliable scalable and easy-to-use platform that improves investment conversion rates maximizes resources and raises customer engagement. This project is 30% accuracy.

REFERENCES

1. Machine Learning Based Customer Churn Prediction in Banking by Manas Rahman and V Kumar, published in 2020 4th International Conference on Electronics, Communication and

- Aerospace Technology (ICECA), DOI: 10.1109/ICECA49313.2020.9297529, ISSN: 2320-8845.

2. Banking in Terms of Deposit Prediction Based on Machine Learning and Big Data Analytics, by authors not specified, published in 2023 IEEE International Conference on Artificial Intelligence and Computer Applications (AICA), DOI: 10.1109/AICA56173.2023.10374595, ISSN: 978-1-6654-9751-5

3. Applying Machine Learning to the Development of Prediction Models for Bank Deposit Subscription by Sipu Hou, Zongzhen Cai, Jiming Wu, and Peng Xie, published in 2022, available at <https://www.researchgate.net/publication/357488750>.

4. Using Big Data and Predictive Analytics for Informed Decision-Making in Investment Banking by Priyanshu Jasaiwal, published in International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Volume 13, Issue 12, December 2024, DOI: 10.17148/IJARCCE.2024.131258, ISSN Online: 2278-1021, ISSN Print: 2319-5940.

5. Bank Loan Prediction Using Machine Learning Techniques by F. M. Ahsanul Haque and Md. Mahedi Hassan, published in American Journal of Industrial and Business Management, Volume 14, Article 1412085, December 25, 2024, DOI: 10.4236/ajibm.2024.1412085, ISSN Online: 2164-5175, ISSN Print: 2164-5167.