

Bank Customer Churn Prediction Using Machine Learning and Interactive Streamlit Dashboard

Tosif Raza Mansoori

Data Analytics Intern Unified Mentor Pvt. Ltd

Abstract — Customer retention has become one of the most significant challenges faced by modern banking organizations. Due to increasing competition in the financial sector, customers can easily switch from one bank to another if they are dissatisfied with the services provided. Therefore, predicting customer churn has become an important business problem, as retaining existing customers is generally more cost-effective than acquiring new ones. This project presents a Machine Learning-based Bank Customer Churn Prediction Dashboard developed using Python and Streamlit. The objective of the project is to analyze customer information and accurately predict whether a customer is likely to discontinue banking services. Along with prediction, the dashboard provides interactive visualizations and business insights that assist organizations in making informed decisions. The project begins with data collection and preprocessing, where duplicate records and unnecessary attributes are removed. Categorical variables are converted into numerical values using Label Encoding, and numerical features are standardized using StandardScaler. The cleaned dataset is then used to train multiple Machine Learning classification models. Three Machine Learning algorithms were implemented and compared, namely Logistic Regression, Decision Tree Classifier, and Random Forest Classifier. Their performances were evaluated using Accuracy Score, Precision, Recall, F1-Score, and Confusion Matrix. Experimental results showed that the Random Forest classifier achieved the highest prediction accuracy of 86.25%, making it the final model selected for deployment. To improve usability, the trained model was integrated into an interactive Streamlit dashboard. Users can enter customer details and instantly receive churn predictions along with prediction confidence, customer risk level, and business recommendations. The dashboard also includes interactive data visualizations, customer analytics, dataset exploration, feature importance analysis, and model comparison charts. Overall, this project demonstrates how Machine Learning and Data Analytics can support banking organizations in reducing customer churn, improving customer retention strategies, and making data-driven business decisions.

Keywords— Machine Learning, Customer Churn Prediction, Data Analytics, Banking Analytics, Random Forest, Streamlit, Business Intelligence, Predictive Analytics, Python.

I. INTRODUCTION

The banking industry is one of the largest sectors that generates enormous amounts of customer data every day. Banks continuously collect information related to customer demographics, financial transactions, account activities, and banking services. This data contains valuable information that can help organizations understand customer behavior and improve business performance.

One of the major challenges faced by banks is customer churn, which refers to customers leaving the bank and discontinuing its services. Customer churn negatively affects organizational profitability because acquiring a new customer requires significantly higher investment compared to retaining an existing one. Consequently, predicting customer churn has become an essential application of Data Analytics and Machine Learning in the banking industry.

Machine Learning enables organizations to identify hidden patterns in historical customer data and predict whether a

customer is likely to leave the organization in the future. These predictive insights allow banks to take preventive actions such as offering personalized services, loyalty programs, promotional offers, and improved customer support before valuable customers decide to leave.

The primary objective of this project is to develop a Machine Learning-based Customer Churn Prediction Dashboard capable of accurately predicting customer churn while simultaneously providing interactive business insights through an easy-to-use web application.

The complete project includes multiple stages such as data preprocessing, feature engineering, exploratory data analysis, Machine Learning model development, model evaluation, and dashboard deployment. Several classification algorithms including Logistic Regression, Decision Tree, and Random Forest were implemented and compared to determine the most suitable prediction model.

Among the implemented algorithms, the Random Forest classifier achieved the highest prediction accuracy of 86.25% and was selected as the final deployment model. The trained model was integrated into a Streamlit dashboard where users can perform real-time customer churn prediction, explore customer information, visualize analytical reports, and receive business recommendations.

The developed dashboard demonstrates how predictive analytics can transform raw banking data into actionable business intelligence that supports strategic decision-making, improves customer retention, and enhances organizational profitability.

Objectives of the Project

The primary objective of this project is to develop a Machine Learning-based Customer Churn Prediction Dashboard that enables banking organizations to identify customers who are likely to discontinue their services. The project combines predictive analytics with interactive visualization to assist organizations in making informed business decisions and improving customer retention strategies.

The specific objectives of this project are as follows:

Develop a Customer Churn Prediction Model

To build an accurate Machine Learning model capable of predicting whether a customer is likely to leave the bank based on demographic and financial information.

Perform Data Cleaning and Preprocessing

To improve the quality of the dataset by removing duplicate records, eliminating unnecessary columns, encoding categorical variables, and standardizing numerical features before model training.

Compare Multiple Machine Learning Algorithms

To implement and evaluate different Machine Learning classification algorithms, including Logistic Regression, Decision Tree, and Random Forest, and select the model with the highest predictive performance.

Develop an Interactive Dashboard

To design and develop a user-friendly Streamlit dashboard that enables users to perform real-time customer churn prediction using customer information.

Visualize Customer Behavior

To create interactive charts and visualizations that help users understand customer characteristics, churn patterns, and business trends.

Generate Business Recommendations

To provide meaningful business recommendations based on prediction results, helping banking organizations improve customer retention and reduce customer attrition.

Support Data-Driven Decision Making

To demonstrate how Machine Learning and Data Analytics can be integrated into Business Intelligence systems for strategic planning and decision-making.

III. TOOLS AND TECHNOLOGIES USED

The successful development of this project required the integration of various tools, programming libraries, and Machine Learning frameworks. Each technology contributed to different stages of data preprocessing, model development, visualization, and deployment.

| Tool / Technology | Purpose |
|-------------------|---|
| Python | Core programming language used for the entire project |
| Pandas | Data loading, cleaning, preprocessing, and analysis |
| NumPy | Numerical computations and array operations |
| Scikit-learn | Machine Learning model development and evaluation |
| Matplotlib | Statistical data visualization |
| Seaborn | Advanced data visualization and heatmaps |
| Plotly | Interactive charts for dashboard visualization |
| Streamlit | Development and deployment of the interactive dashboard |
| Joblib | Saving and loading the trained Machine Learning model |
| Git & GitHub | Version control and project hosting |
| Streamlit Cloud | Online deployment of the dashboard |

Python

Python was used as the primary programming language due to its simplicity, extensive library support, and powerful Machine Learning ecosystem.

Pandas

Pandas was utilized for importing the dataset, cleaning data, removing duplicate records, selecting required features, and performing data manipulation.

NumPy

NumPy supported numerical computations and mathematical operations required during data preprocessing.

Scikit-learn

Scikit-learn was used to implement Machine Learning algorithms, split the dataset into training and testing sets, perform feature scaling, and evaluate model performance.

Matplotlib and Seaborn

These libraries were used to create statistical charts, feature importance plots, confusion matrices, and model comparison graphs.

Plotly

Plotly was used to create interactive visualizations inside the Streamlit dashboard, allowing users to explore customer behavior dynamically.

Streamlit

Streamlit was used to develop the complete web application, enabling real-time customer churn prediction, data visualization, and business recommendation generation.

GitHub and Streamlit Cloud

GitHub was used for version control and project hosting, while Streamlit Cloud was used to deploy the application online for public access.

IV. DATASET DESCRIPTION

The project uses the European Bank Customer Churn Dataset, which contains customer demographic and banking-related information used to predict customer churn.

The dataset consists of customer attributes that describe financial behavior, banking relationship, and account information. Each record represents a unique customer, while the target variable indicates whether the customer has exited the bank.

Dataset Features

| Feature | Description |
|-----------------|---|
| CreditScore | Customer's credit score |
| Geography | Customer's country or region |
| Gender | Male or Female |
| Age | Customer's age |
| Tenure | Number of years with the bank |
| Balance | Current account balance |
| NumOfProducts | Number of banking products used |
| HasCrCard | Credit card ownership status |
| IsActiveMember | Active membership status |
| EstimatedSalary | Estimated annual salary |
| Exited | Target variable indicating customer churn |

Target Variable

The Exited column represents the prediction target.

- 0 → Customer has not left the bank.

- 1 → Customer has left the bank.

Removed Attributes

During preprocessing, the following columns were removed because they do not contribute significantly to customer churn prediction:

- Year
- CustomerId
- Surname

Removing these attributes reduced unnecessary computational complexity and improved the efficiency of the Machine Learning models.

The processed dataset was then used for training and evaluating different classification algorithms to identify the most suitable model for customer churn prediction.

V. DATA UNDERSTANDING

Data understanding is one of the most important stages of any Data Analytics or Machine Learning project. Before building predictive models, it is necessary to understand the dataset, identify important features, examine data types, and analyze how each attribute contributes to the prediction process.

The European Bank Customer Churn Dataset contains demographic, financial, and banking-related information of customers. Each record represents an individual customer, while the target variable indicates whether the customer has continued banking services or has exited the bank.

The dataset consists of both numerical and categorical attributes that influence customer churn. Understanding these features helps in selecting appropriate preprocessing techniques and improving model performance.

The important features used in this project are described below.

Credit Score

Credit Score represents the financial credibility of a customer. It is one of the important factors considered during churn prediction because customers with different credit scores often exhibit different banking behaviors.

Geography

The Geography feature identifies the country or region where the customer belongs. Customer behavior and banking preferences may vary across different geographical locations.

Gender

Gender represents whether the customer is Male or Female. Since Machine Learning algorithms require numerical inputs, this categorical feature was converted into numerical values during preprocessing.

Age

Age indicates the customer's current age. It is one of the most influential variables affecting customer churn, as banking requirements and financial behavior often change with age.

Tenure

Tenure represents the number of years the customer has maintained a relationship with the bank. Customers with longer relationships generally exhibit greater loyalty.

Balance

Balance indicates the amount available in the customer's bank account. It reflects the customer's financial engagement with the bank.

Number of Products

This feature specifies the total number of banking products or services currently being used by the customer.

Has Credit Card

This attribute identifies whether the customer owns a bank-issued credit card.

Is Active Member

The Active Member feature indicates whether the customer actively uses banking services. Active customers generally show lower churn probability.

Estimated Salary

Estimated Salary represents the approximate annual income of the customer and provides additional financial information useful for prediction.

Exited

Exited is the target variable of the dataset.

- 0 indicates that the customer continues banking services.
- 1 indicates that the customer has discontinued banking services.

Since the target variable contains only two possible outcomes, the problem is classified as a Binary Classification Machine Learning Problem.

VI. DATA PREPROCESSING

Real-world datasets often contain unnecessary attributes, duplicate records, and categorical information that cannot be directly processed by Machine Learning algorithms. Therefore, several preprocessing techniques were applied before training the prediction models.

Data Cleaning

Initially, the dataset was examined to identify duplicate records and unnecessary information. Duplicate records were removed to ensure that every customer record represented a unique observation.

Removal of Irrelevant Features

Certain columns were used only for customer identification and did not contribute to churn prediction. These attributes were removed before model development.

The removed columns include:

- Year
- CustomerId
- Surname

Removing these features reduced computational complexity and allowed the model to focus only on meaningful customer attributes.

Label Encoding

Machine Learning algorithms require numerical data; therefore, categorical features were converted into numerical values using Label Encoding.

The following attributes were encoded:

- Gender
- Geography

This conversion enabled the Machine Learning models to process categorical information efficiently.

Feature Selection

The dataset was divided into independent variables and the target variable.

Independent Variables (X)

Credit Score, Geography, Gender, Age, Tenure, Balance, Number of Products, Has Credit Card, Active Member, and Estimated Salary.

Target Variable (Y)

Exited

This separation allowed the Machine Learning algorithms to learn the relationship between customer attributes and churn status.

Train-Test Split

To evaluate model performance fairly, the dataset was divided into two parts:

- Training Dataset: 80%
- Testing Dataset: 20%

The training dataset was used to train the Machine Learning models, while the testing dataset was used to evaluate their predictive performance.

Feature Scaling

The numerical features were standardized using the StandardScaler technique available in the Scikit-learn library. Feature scaling ensures that all numerical variables have a similar range, preventing features with larger values from dominating the learning process.

This preprocessing step significantly improves the stability and performance of Machine Learning algorithms, especially Logistic Regression.

After completing all preprocessing steps, the cleaned and transformed dataset was ready for Machine Learning model training and evaluation.

VII. PROJECT METHODOLOGY

The development of the Bank Customer Churn Prediction Dashboard followed a systematic methodology that integrates Data Analytics, Machine Learning, and Dashboard Development. The project was executed through multiple sequential phases to ensure accurate prediction and effective visualization of customer churn.

The overall methodology consists of the following stages.

Data Collection

The project begins with collecting the European Bank Customer Churn Dataset. The dataset contains customer demographic details, banking information, financial attributes, and customer churn status. It serves as the primary source of data for analysis and prediction.

Data Cleaning

After loading the dataset, data cleaning was performed to improve data quality. Duplicate records were identified and

removed to eliminate redundancy. Additionally, unnecessary columns such as Year, CustomerId, and Surname were removed because they do not contribute to customer churn prediction.

Data Preprocessing

The cleaned dataset was prepared for Machine Learning by converting categorical variables into numerical values using Label Encoding. The features and target variable were separated, and numerical attributes were standardized using the StandardScaler technique. This preprocessing step ensures that all features contribute equally during model training.

Dataset Splitting

The processed dataset was divided into training and testing datasets using an 80:20 ratio. The training dataset was used to train the Machine Learning models, while the testing dataset was reserved for evaluating model performance.

Machine Learning Model Development

Three supervised Machine Learning classification algorithms were implemented and trained using the prepared dataset.

The implemented models include:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier

Each model was trained using identical training data to ensure a fair performance comparison.

Model Evaluation

After training, each model was evaluated using standard classification metrics, including:

- Accuracy Score
- Precision
- Recall
- F1-Score
- Confusion Matrix

The evaluation results were compared to determine the best-performing model for customer churn prediction.

Model Selection

Among the implemented Machine Learning algorithms, the Random Forest Classifier achieved the highest prediction accuracy of 86.25%. Therefore, it was selected as the final prediction model and saved using the Joblib library for deployment.

Dashboard Development

An interactive dashboard was developed using Streamlit to make the Machine Learning model accessible to users. The dashboard allows users to input customer information, predict customer churn, explore customer data, and analyze business insights through interactive visualizations.

Deployment

Finally, the complete application was deployed using Streamlit Cloud, making the dashboard accessible through a web browser without requiring any local software installation.

The deployed dashboard enables users to perform real-time customer churn prediction while simultaneously exploring customer behavior through analytical visualizations.

VIII. PROJECT WORKFLOW

The complete workflow of the Bank Customer Churn Prediction Dashboard follows a structured sequence of analytical and Machine Learning operations.

The workflow consists of the following steps:

- Step 1: Import the customer churn dataset.
- Step 2: Perform data cleaning by removing duplicate records and irrelevant columns.
- Step 3: Encode categorical variables into numerical values.
- Step 4: Separate the dataset into input features and target variable.
- Step 5: Split the dataset into training and testing datasets.
- Step 6: Apply StandardScaler for feature scaling.
- Step 7: Train Logistic Regression, Decision Tree, and Random Forest models.
- Step 8: Evaluate model performance using multiple evaluation metrics.
- Step 9: Select the best-performing model based on prediction accuracy.
- Step 10: Save the trained Machine Learning model using Joblib.
- Step 11: Integrate the trained model into a Streamlit dashboard.
- Step 12: Deploy the dashboard for real-time customer churn prediction.

The systematic workflow ensures accurate prediction, efficient data processing, and effective visualization of customer behavior, making the dashboard a practical business intelligence solution for customer retention analysis.

IX. MACHINE LEARNING MODEL DEVELOPMENT

Machine Learning is the core component of this project. After completing data preprocessing, multiple supervised Machine Learning classification algorithms were implemented to predict whether a customer is likely to leave the bank.

The objective of implementing multiple algorithms was to compare their performance and identify the most accurate model for customer churn prediction.

The following Machine Learning algorithms were implemented during this project.

Logistic Regression

Logistic Regression is a supervised Machine Learning algorithm widely used for binary classification problems. Since customer churn prediction involves only two possible outcomes (Exited or Not Exited), Logistic Regression serves as an effective baseline classification model.

The algorithm estimates the probability of customer churn by analyzing the relationship between customer attributes and the target variable. It is computationally efficient and provides good prediction performance for linearly separable datasets.

Although Logistic Regression produced satisfactory results, its prediction accuracy was lower than the ensemble-based Random Forest classifier.

Decision Tree Classifier

Decision Tree is a rule-based supervised Machine Learning algorithm that predicts outcomes by creating a hierarchical tree structure based on feature values.

The model repeatedly divides the dataset into smaller subsets according to the most informative features until prediction rules are generated.

Decision Trees are simple to understand and interpret because the prediction process follows a sequence of logical decisions. However, they are prone to overfitting when trained on complex datasets.

In this project, the Decision Tree model produced reasonable prediction performance but achieved lower accuracy compared to Logistic Regression and Random Forest.

Random Forest Classifier

Random Forest is an ensemble Machine Learning algorithm that combines multiple Decision Trees to produce more accurate and stable predictions.

Instead of relying on a single tree, Random Forest constructs numerous decision trees during training and determines the final prediction using majority voting.

The ensemble approach minimizes overfitting and significantly improves prediction accuracy by reducing variance.

Because of its robustness and ability to learn complex customer behavior patterns, Random Forest achieved the highest prediction accuracy among all implemented models.

Consequently, Random Forest was selected as the final prediction model for deployment within the Streamlit dashboard.

Model Comparison

To identify the most suitable Machine Learning algorithm, all three models were trained and evaluated using the same training and testing datasets.

The prediction accuracy of each model is summarized below.

| Machine Learning Model | Accuracy |
|--------------------------|----------|
| Logistic Regression | 80.50% |
| Decision Tree Classifier | 77.90% |
| Random Forest Classifier | 86.25% |

The comparison clearly indicates that the Random Forest Classifier outperformed the remaining models and provided the highest prediction accuracy.

Therefore, it was selected as the final Machine Learning model for customer churn prediction.

Model Saving

After selecting the best-performing model, the trained Random Forest model and the StandardScaler object were saved using the Joblib library.

Saving the trained model allows it to be reused without retraining every time the application starts.

The saved model was later integrated into the Streamlit dashboard, enabling users to perform real-time customer churn prediction efficiently.

Overall, the Machine Learning development phase successfully identified the Random Forest Classifier as the most reliable and accurate algorithm for predicting customer churn in the banking sector.

X. MODEL EVALUATION

Model evaluation is an essential phase of every Machine Learning project because it determines how effectively a trained model performs on unseen data. After training the Logistic Regression, Decision Tree, and Random Forest classifiers, each model was evaluated using standard classification metrics.

The evaluation process helped identify the model that provides the most accurate and reliable customer churn predictions.

Accuracy Score

Accuracy Score measures the percentage of correctly predicted observations out of the total number of observations.

It is calculated as:

$$\text{Accuracy} = (\text{Correct Predictions} / \text{Total Predictions}) \times 100$$

The accuracy obtained by each Machine Learning model is shown below.

| Machine Learning Model | Accuracy |
|--------------------------|----------|
| Logistic Regression | 80.50% |
| Decision Tree Classifier | 77.90% |
| Random Forest Classifier | 86.25% |

Among all the implemented algorithms, the Random Forest Classifier achieved the highest prediction accuracy of 86.25%, making it the best-performing model for this project.

Confusion Matrix

A Confusion Matrix provides a detailed summary of prediction results by comparing actual customer classes with predicted customer classes.

It consists of four important components:

True Positive (TP):

Customers who actually churned and were correctly predicted as churned.

True Negative (TN):

Customers who did not churn and were correctly predicted as non-churn customers.

False Positive (FP):

Customers who were predicted to churn but actually remained with the bank.

False Negative (FN):

Customers who actually churned but were predicted as non-churn customers.

The Confusion Matrix provides deeper insight into prediction quality than overall accuracy and helps identify the types of prediction errors made by the model.

Precision

Precision measures how accurately the model predicts churn customers. It answers the following question:

"Out of all customers predicted to churn, how many actually churned?"

A higher Precision value indicates that the model produces fewer false positive predictions.

High Precision helps organizations avoid unnecessary customer retention efforts for customers who are unlikely to leave.

Recall

Recall measures the model's ability to correctly identify customers who are actually at risk of leaving the bank. It answers the following question:

"Out of all customers who actually churned, how many were correctly identified?"

A high Recall value is extremely important in customer churn prediction because missing a potential churn customer may result in revenue loss for the organization.

F1-Score

The F1-Score combines both Precision and Recall into a single performance metric.

It provides a balanced evaluation when both false positives and false negatives are important.

Since customer churn prediction requires both accurate identification and minimal prediction errors, the F1-Score serves as an effective performance indicator.

Performance Comparison

All three Machine Learning algorithms demonstrated satisfactory prediction performance. However, their effectiveness varied based on prediction accuracy and classification capability.

Logistic Regression

The Logistic Regression model produced stable prediction results and served as a reliable baseline model. It performed well on the dataset but was limited in capturing complex relationships between customer attributes.

Decision Tree Classifier

The Decision Tree model generated understandable decision rules but exhibited comparatively lower prediction accuracy. It is also more susceptible to overfitting when compared with ensemble learning methods.

Random Forest Classifier

The Random Forest model produced the highest overall performance among all evaluated algorithms. By combining multiple Decision Trees through ensemble learning, it successfully reduced overfitting and improved prediction accuracy.

Its ability to learn complex customer behavior patterns made it the most suitable algorithm for customer churn prediction.

Selection of Final Model

Based on comprehensive evaluation using Accuracy Score, Precision, Recall, F1-Score, and Confusion Matrix, the Random Forest Classifier was selected as the final prediction model.

The reasons for selecting Random Forest include:

- Highest prediction accuracy (86.25%)
- Better generalization capability
- Reduced overfitting through ensemble learning
- Reliable performance on unseen customer data
- Improved prediction stability compared to other models

The trained Random Forest model was saved using the Joblib library and integrated into the Streamlit dashboard, enabling real-time customer churn prediction.

The evaluation results demonstrate that Random Forest provides an effective and reliable solution for predicting customer churn and can assist banking organizations in identifying high-risk customers before they discontinue banking services.

XI. DASHBOARD DEVELOPMENT

After successfully training and evaluating the Machine Learning model, an interactive web-based dashboard was developed using the Streamlit framework. The primary objective of the dashboard is to provide a simple, user-friendly interface that allows users to perform customer churn prediction and explore analytical insights without requiring programming knowledge.

The dashboard integrates Machine Learning, Data Analytics, and Business Intelligence into a single platform. It enables users to interact with customer data, visualize important trends, and receive business recommendations based on prediction results.

The dashboard is divided into multiple analytical sections, each designed to perform a specific task efficiently.

Home Page

The Home page serves as the introduction to the application. It provides users with a brief overview of the project, its objectives, and the technologies used in the development of the dashboard.

The Home page helps users understand the purpose of the application before exploring its different analytical modules.

Customer Churn Prediction

The Customer Churn Prediction module is the core functionality of the dashboard. Users can enter customer information such as:

- Credit Score
- Geography
- Gender
- Age
- Tenure
- Balance
- Number of Products
- Credit Card Status
- Active Membership
- Estimated Salary

After entering the required details, the trained Random Forest model processes the information and predicts whether the customer is likely to leave the bank.

The prediction result is displayed instantly along with the prediction confidence.

Risk Classification

To improve decision-making, the prediction results are categorized into different risk levels based on prediction probability.

The dashboard classifies customers into:

- High Risk
- Medium Risk
- Low Risk

This classification enables banking organizations to prioritize customer retention efforts according to the severity of churn risk.

Business Recommendation Module

Based on the prediction result, the dashboard automatically generates business recommendations. For customers identified as high-risk, recommendations include:

- Contact the customer immediately.
- Provide personalized offers and discounts.
- Assign a dedicated relationship manager.
- Improve customer engagement.
- Recommend suitable banking services.

For customers predicted as low-risk, the dashboard recommends maintaining regular engagement and exploring cross-selling opportunities.

This module transforms prediction results into actionable business decisions.

XII. DASHBOARD FEATURES

The developed Streamlit dashboard provides several interactive features that improve usability and enhance analytical capabilities.

Interactive Sidebar Filters

The dashboard includes an interactive sidebar that allows users to filter customer information according to their analytical requirements.

The available filters include:

- Select Geography
- Select Gender

These filters dynamically update the visualizations and analytical results, enabling users to focus on specific customer segments.

Exploratory Data Analysis (EDA)

The dashboard contains a dedicated Exploratory Data Analysis section that provides interactive visualizations for understanding customer behavior.

The EDA module includes various charts such as:

- Customer Churn Distribution
- Age Distribution
- Credit Score Distribution
- Balance Distribution
- Geography-wise Customer Analysis
- Gender-wise Customer Analysis
- Active Member Analysis
- Correlation Heatmap
- Feature Importance
- Model Accuracy Comparison

These visualizations help identify important trends and factors influencing customer churn.

Dataset Explorer

The Dataset Explorer enables users to inspect customer records in an organized tabular format. Users can:

- Search customer information.
- View processed records.
- Explore customer attributes.

This module simplifies data exploration and improves the overall usability of the application.

Machine Learning Integration

The dashboard is directly connected to the trained Random Forest Machine Learning model.

Instead of retraining the model every time, the saved model is loaded using the Joblib library, allowing real-time prediction with minimal computational overhead.

This significantly improves dashboard performance and user experience.

User-Friendly Interface

The dashboard was designed with simplicity and usability as primary objectives.

The interface allows users with little or no technical knowledge to perform customer churn prediction and analyze customer behavior efficiently.

Interactive components, organized navigation, and responsive visualizations enhance the overall user experience.

Business Intelligence Support

The dashboard converts raw customer information into meaningful business insights. It assists banking organizations in:

- Identifying high-risk customers.
- Understanding customer behavior.
- Improving customer retention.
- Supporting strategic business decisions.

By integrating Machine Learning with interactive visualization, the dashboard serves as a practical Business Intelligence solution for customer churn analysis.

XIII. BUSINESS INSIGHTS

The analysis of the European Bank Customer Churn Dataset provided several valuable insights regarding customer behavior and the factors influencing customer churn. These insights can help banking organizations develop effective customer retention strategies and improve overall business performance.

Customer Age Influences Churn

The analysis revealed that customer age has a significant impact on churn behavior. Customers belonging to higher age groups were found to have a relatively higher probability of leaving the bank compared to younger customers.

Active Members are Less Likely to Churn

Customers who actively use banking services generally exhibit greater loyalty toward the organization. Active members showed a lower churn rate than inactive customers, highlighting the importance of customer engagement.

Geography Affects Customer Behavior

Customer churn patterns varied across different geographical regions. This indicates that customer behavior is influenced by location, suggesting that banks should implement region-specific retention strategies.

Number of Products Influences Customer Retention

Customers using multiple banking products tend to remain associated with the bank for a longer period. Cross-selling banking services can therefore improve customer loyalty and reduce churn.

Credit Score and Balance are Important Predictors

Credit Score and Account Balance were identified as significant features in predicting customer churn. These financial attributes play an important role in determining customer behavior.

Machine Learning Improves Business Decision Making

The integration of Machine Learning enables banking organizations to identify customers who are at risk of leaving before the actual churn occurs. This allows businesses to implement preventive strategies and improve customer retention.

XIV. BUSINESS RECOMMENDATIONS

Based on the analytical findings and prediction results, several recommendations are proposed for improving customer retention and enhancing banking services.

Identify High-Risk Customers Early

Banks should continuously monitor customers identified as high-risk by the Machine Learning model. Early identification enables organizations to take timely preventive actions.

Improve Customer Engagement

Inactive customers should be encouraged to use banking services through personalized communication, loyalty programs, and promotional campaigns.

Provide Personalized Services

Banks should offer customized financial products, exclusive offers, and personalized assistance to customers with a higher probability of churn.

Strengthen Customer Relationship Management

Assigning dedicated relationship managers to valuable customers can significantly improve customer satisfaction and long-term retention.

Use Data-Driven Decision Making

Business decisions should be supported by analytical insights obtained through Machine Learning and Business Intelligence dashboards instead of relying solely on traditional methods.

Regularly Update Predictive Models

Customer behavior changes over time; therefore, Machine Learning models should be retrained periodically using updated customer data to maintain prediction accuracy.

XV. CONCLUSION

The Bank Customer Churn Prediction Dashboard successfully demonstrates the application of Machine Learning and Data Analytics in solving a real-world business problem within the banking industry.

The project involved multiple stages, including data collection, preprocessing, exploratory data analysis, Machine Learning model development, evaluation, and deployment through an interactive Streamlit dashboard.

Three Machine Learning algorithms—Logistic Regression, Decision Tree Classifier, and Random Forest Classifier—were implemented and compared. Among these models, the Random Forest Classifier achieved the highest prediction accuracy of 86.25%, making it the most suitable algorithm for customer churn prediction.

The developed dashboard provides real-time customer churn prediction, interactive visualizations, business insights, and recommendation support through an easy-to-use interface.

Overall, this project demonstrates how predictive analytics can transform customer data into actionable business intelligence, helping banking organizations improve customer retention, reduce churn, and support data-driven decision-making.

Future Scope

Although the developed system performs effectively, several enhancements can be incorporated in future versions to improve functionality and predictive performance.

The following improvements are recommended:

- Implementation of advanced Machine Learning algorithms such as XGBoost and LightGBM.
- Integration of Deep Learning models for improved prediction accuracy.
- Real-time database connectivity for live customer prediction.
- Development of a mobile-friendly dashboard interface.
- Integration with Customer Relationship Management (CRM) systems.
- Implementation of Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME to improve model interpretability.
- Automated customer notification and retention campaign generation.
- Continuous model retraining using newly collected customer data to maintain prediction accuracy over time.

These future enhancements will further strengthen the applicability of the system in real-world banking environments and contribute to more effective customer retention strategies.

Acknowledgement

I would like to express my sincere gratitude to Unified Mentor Pvt. Ltd. for providing me with the opportunity to work on this Data Analytics and Machine Learning project as a part of my internship.

I am thankful to my mentors and the entire Unified Mentor team for their continuous guidance, valuable suggestions, and encouragement throughout the internship. Their support helped me understand practical applications of Data Analytics, Machine Learning, and Business Intelligence in solving real-world business problems.

I also extend my gratitude to everyone who contributed directly or indirectly to the successful completion of this project.

This internship has significantly improved my technical knowledge and practical skills in Python, Data Preprocessing, Exploratory Data Analysis, Machine Learning, Data Visualization, Dashboard Development using Streamlit, GitHub version control, and project deployment.

The experience gained during this internship has strengthened my confidence in applying Data Analytics concepts to real-world business scenarios and has motivated me to continue learning and exploring advanced technologies in the field of Artificial Intelligence and Machine Learning.

Finally, I sincerely thank Unified Mentor Pvt. Ltd. for providing such an excellent learning platform and an opportunity to enhance my professional and technical skills.

REFERENCES

The following references were used during the development of this project and preparation of this report.

1. Scikit-learn Developers, Scikit-learn: Machine Learning in Python. Available at: <https://scikit-learn.org/>
2. Streamlit Inc., Streamlit Documentation. Available at: <https://streamlit.io/>
3. Pandas Development Team, Pandas Documentation. Available at: <https://pandas.pydata.org/>
4. NumPy Developers, NumPy Documentation. Available at: <https://numpy.org/>
5. Matplotlib Development Team, Matplotlib Documentation. Available at: <https://matplotlib.org/>
6. Seaborn Developers, Seaborn Statistical Data Visualization. Available at: <https://seaborn.pydata.org/>
7. Plotly Technologies Inc., Plotly Python Documentation. Available at: <https://plotly.com/python/>

8. Joblib Documentation. Available at: <https://joblib.readthedocs.io/>
9. European Bank Customer Churn Dataset, Kaggle. Available at: <https://www.kaggle.com/>
10. Géron, A., Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media.
11. Han, J., Kamber, M., & Pei, J., Data Mining: Concepts and Techniques, Morgan Kaufmann.
12. IBM, Customer Churn Analysis and Predictive Analytics. Available at: <https://www.ibm.com/>
13. Gartner Research, Business Intelligence and Analytics. Available at: <https://www.gartner.com/>
14. Harvard Business Review, Customer Retention Strategies. Available at: <https://hbr.org/>
15. Various research articles on Machine Learning applications in customer churn prediction and banking analytics.