Data Analysis Report

Overview

The dataset consists of customer-related information, including demographics, banking behaviors, and feedback on satisfaction. The primary objective of this analysis was to understand customer retention factors and predict the likelihood of customers exiting the bank. Using machine learning techniques, I trained a Random Forest Classifier, which achieved strong predictive performance, indicating a well-structured dataset with clear correlations between features and the target variable.

The dataset included 14 features after preprocessing. Key transformations included one-hot encoding of categorical variables (e.g., 'Card Type' and 'Geography'), and scaling of numerical variables for model optimization. Splitting the data into training, validation, and testing sets ensured unbiased evaluation of the model's effectiveness.

Key Findings

- **Feature Importance**: Variables such as 'Age', 'Balance', and 'Satisfaction Score' had significant impacts on customer retention prediction.
- **Model Performance**: The Random Forest Classifier achieved a high F1-score and accuracy on both validation and testing sets, showing strong generalization capabilities.
- **Customer Insights**:
    - Older customers were more likely to exit.
    - Customers with higher balances and lower satisfaction scores showed a higher likelihood of exiting.
    - Active membership reduced the likelihood of exiting, emphasizing the importance of customer engagement.

Recommendations

1. Enhance customer satisfaction programs targeting lower-scoring customers.
2. Focus retention strategies on high-balance, high-value customers.
3. Promote active membership benefits to improve engagement.