Churn Prediction with Neural Networks: A Data-Driven Strategy for Customer Retention

Project :Bank Churn Prediction
Course: AIML
Date: 06/09/2024

Contents

Executive Summary

Enhancing Customer Retention through Predictive Modeling

Introduction:

Customer churn poses a significant challenge for banks, directly impacting their revenue and profitability. This project leverages advanced machine learning techniques, particularly Neural Networks, to predict and mitigate customer churn. By analyzing comprehensive customer data and employing sophisticated data preprocessing and model improvement strategies, our goal is to accurately identify at-risk customers and provide actionable insights for retention

Executive Summary Actionable Insights

Customer Segmentation

1. **High-Risk Segments**: Identify and prioritize intervention for high-risk customer segments identified by the model, focusing on those with a high likelihood of churn
2. **Behavioral Patterns:** Utilize insights from EDA to understand behavioral patterns leading to churn, enabling targeted engagement strategies

Personalized Retention Strategies

1. **Tailored Offers:** Develop personalized offers and incentives for at-risk customers based on their specific behaviors and preferences
2. **Enhanced Customer Support:** Strengthen customer support for high-risk segments to address their concerns promptly and effectively

Proactive Engagement

1. **Early Intervention:** Implement proactive engagement strategies to address potential issues before they lead to churn, using predictive insights from the model
2. **Customer Feedback**: Regularly gather and analyze customer feedback to identify pain points and improve overall satisfaction

Executive Summary Business Recommendations

Implement Predictive Analytics

1. **Deploy the Model:** Integrate the best-performing churn prediction model into the bank's CRM system to continuously monitor and predict churn risks
2. **Automated Alerts:** Set up automated alerts for customer service teams to take timely action on high-risk customers

Resource Allocation

1. **Efficient Resource Use:** Allocate resources efficiently by focusing retention efforts on customers most likely to churn, optimizing marketing and customer service expenditures
2. **Cross-Functional Collaboration**: Foster collaboration between marketing, customer service, and data science teams to develop cohesive retention strategies

Continuous Model Refinement

1. **Ongoing Monitoring:** Continuously monitor the model's performance and update it with new data to ensure its accuracy and relevance
2. **Incorporate Feedback:** Use feedback from customer interactions and retention efforts to refine and improve the predictive model

Executive Summary

Conclusion :

Implementing these data-driven insights and recommendations will enhance customer retention, reduce churn-related costs, and improve overall customer satisfaction. By proactively addressing churn risks and personalizing engagement strategies, the bank can foster long-term customer loyalty, drive revenue growth, and maintain a competitive edge in the market. This comprehensive approach to churn prediction and mitigation not only safeguards the bank's existing customer base but also positions it for sustainable success in the future.

Business Problem Overview and Solution Approach

Great Learning
POWER AHEAD

Problem Statement

As the analyst tasked with addressing the high churn rate at the bank, my primary challenge is to identify the factors contributing to customer churn and develop a predictive model to accurately forecast which customers are at risk of leaving. By understanding and predicting customer churn, my goal is to recommend targeted retention strategies to reduce churn rates in the next six months, improve customer satisfaction, and optimize resource allocation. This involves a comprehensive analysis of customer data, rigorous model building and evaluation, and the application of actionable insights to drive business decisions.

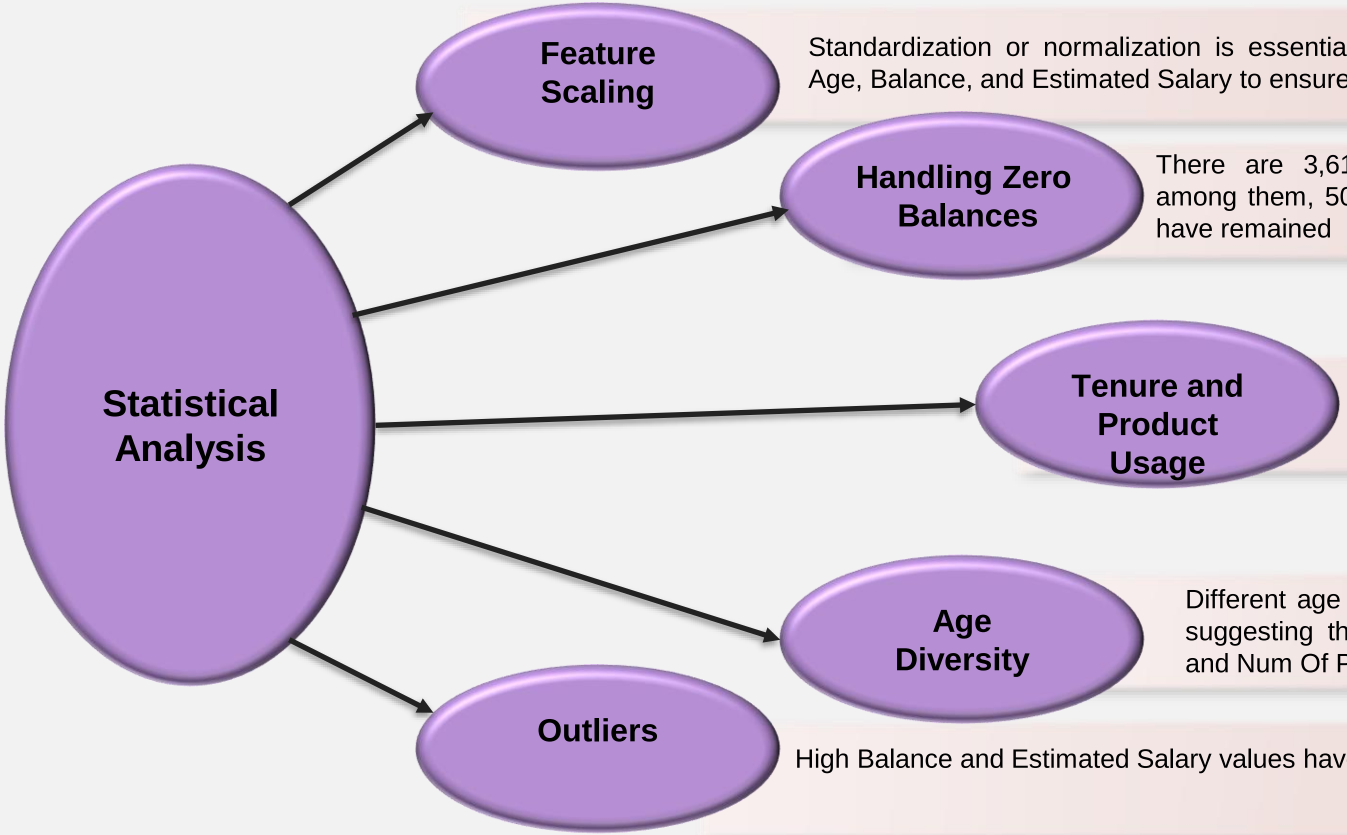
Solution Approach

Data Analysis → Model Development → Implement Solutions → Strategic Execution

Goal

Identify Possible Churners in the next 6 Months

EDA Results Key Findings from Statitical Analysis



Standardization or normalization is essential for features like Credit Score, Age, Balance, and Estimated Salary to ensure optimal model performance

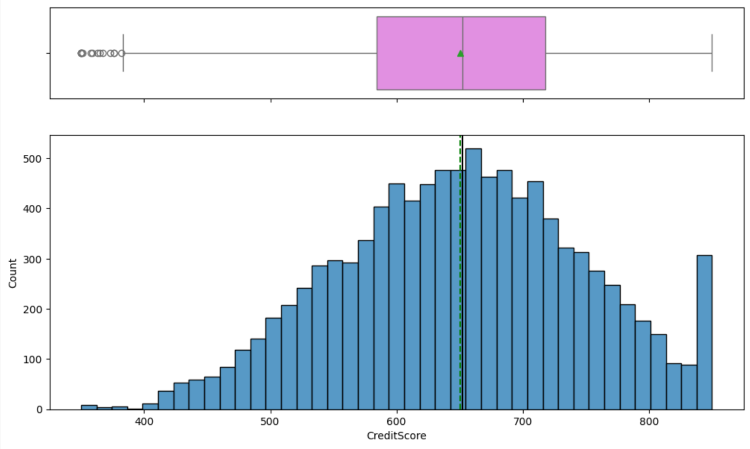
There are 3,617 customers with a 0.0 balance; among them, 500 have exited the bank, while 3,117 have remained

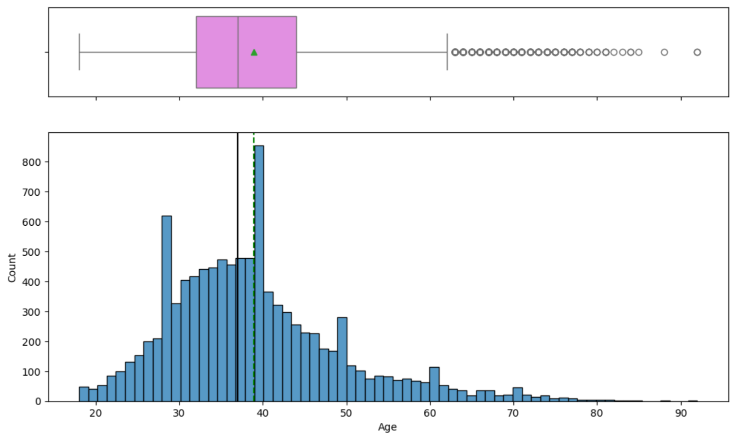
Tenure and Num Of Products are crucial indicators of customer loyalty and engagement

Different age groups exhibit diverse behaviors, suggesting that age interactions with Balance and Num Of Products may be significant

High Balance and Estimated Salary values have been found

EDA – Univariate Analysis





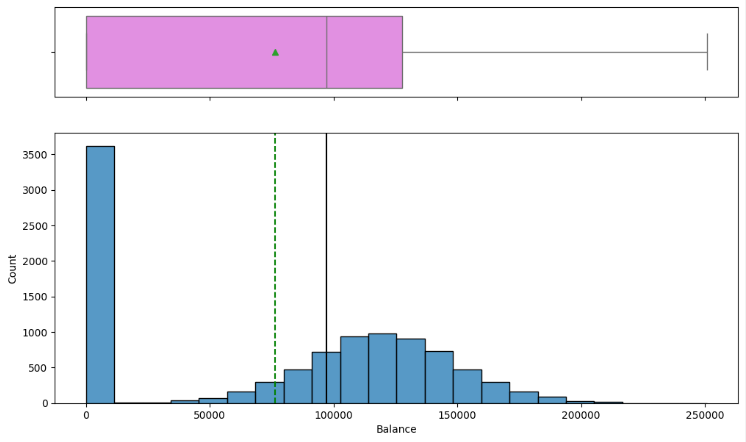
<u>Credit Score</u>

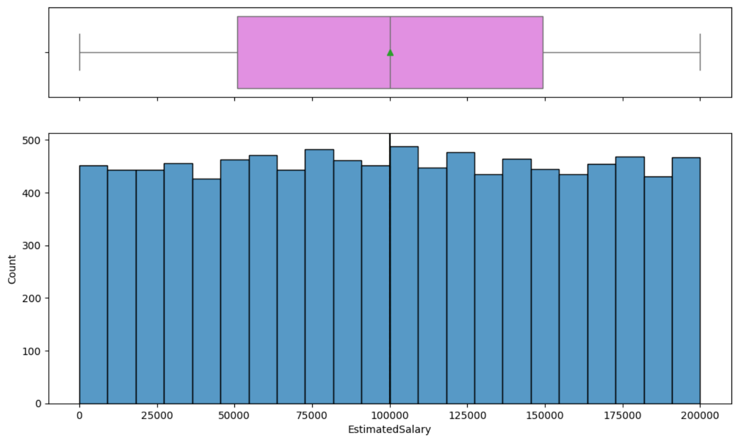
1. Distribution: Credit scores are approximately normally distributed, peaking around 650-700
2. Skewness: Slight left skew, with more customers having higher credit scores near the maximum value of 850
3. Implication: Credit score is a crucial variable; its normal distribution suggests balance in creditworthiness

<u>Age:</u>

1. Distribution: Age shows a right-skewed distribution, with many customers in the 30-40 age range
2. Implication: The concentration of customers in the 30-40 age group suggests age is significant in customer behavior and churn prediction

EDA – Univariate Analysis





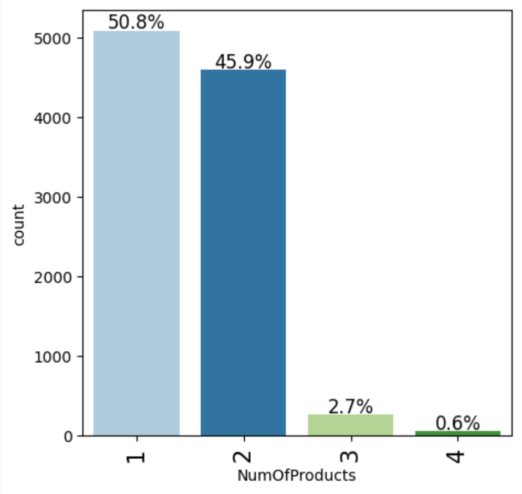
<u>Balance:</u>

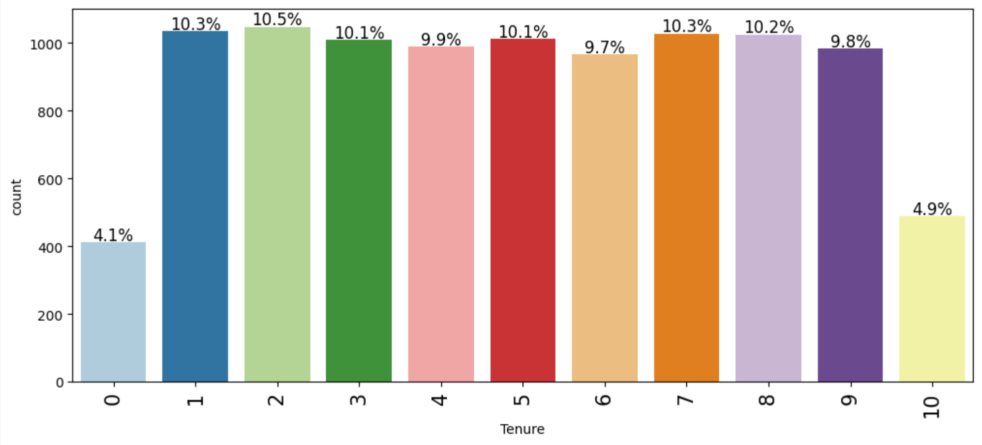
1. Distribution: Balance shows a large number of customers with a zero balance, with the rest of the distribution resembling a normal spread, peaking around 100,000.00
2. Implication: The zero-balance customers might need special consideration as they could represent inactive or low-value customers

<u>Estimated Salary:</u>

1. Distribution: Estimated salary is evenly distributed across the range with a mean of100,090.00
2. Implication: The even distribution suggests income level may not directly correlate with churn, but it should still be included as it might interact with other variables (e.g., balance or number of products)

EDA – Univariate Analysis





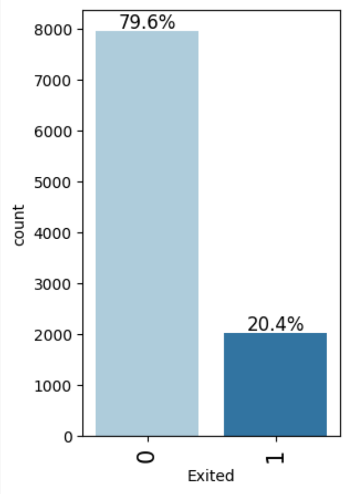
Number Of Products:

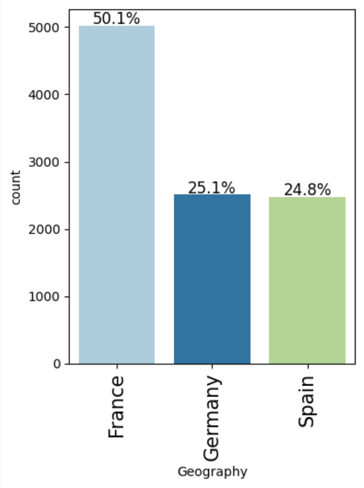
1. Distribution: Most customers have either 1 or 2 products, with very few having 3 or 4 products
2. Implication: The small number of customers with 3 or more products indicates that those with more products might have different behavior patterns. This can be a significant predictor of churn, as customers with fewer products might be more likely to leave

Tenure:

1. Distribution: Tenure appears to be uniformly distributed with a slight decrease at the higher end (10 years).
2. Implication: Since tenure is spread across all values with no specific trend, it might be used as is or binned to understand its impact on churn better

EDA – Univariate Analysis





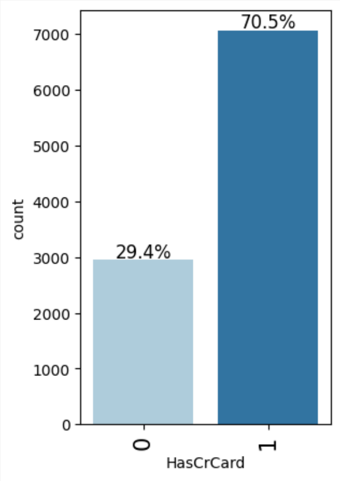
<u>Exited:</u>

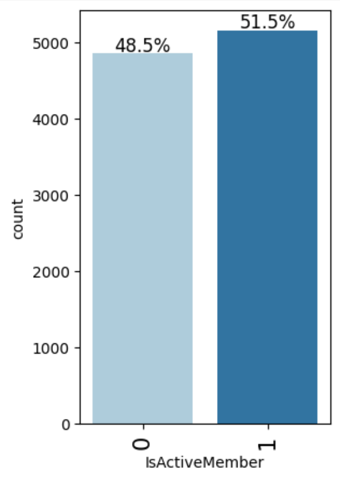
1. There is a clear imbalance, with many more customers not exiting (churning) compared to those who did exit
2. Implication: The imbalance in the target variable (Exited) suggests that class imbalance techniques (e.g., SMOTE) might be necessary to improve model performance. It's important to address this imbalance to ensure the model accurately predicts both classes

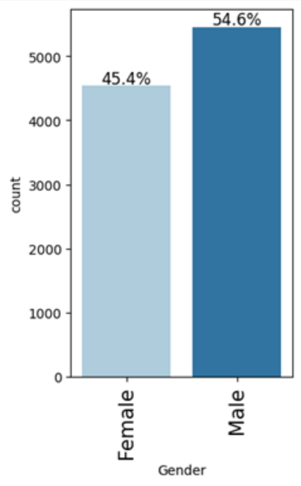
<u>Geogrphy:</u>

1. Distribution: France has the highest number of customers, followed by Spain and Germany.
2. Implication: This geographic distribution should be taken into account, as customers from different regions may have different behaviors and churn rates. It might be useful to include interactions between Geography and other features in the model

EDA – Univariate Analysis







<u>Has Credit Card:</u>

1. Distribution: A significant majority of customers have a credit card
2. Implication: The possession of a credit card might be a factor in customer engagement and satisfaction. Analyzing the impact of having a credit card on churn can provide valuable insights

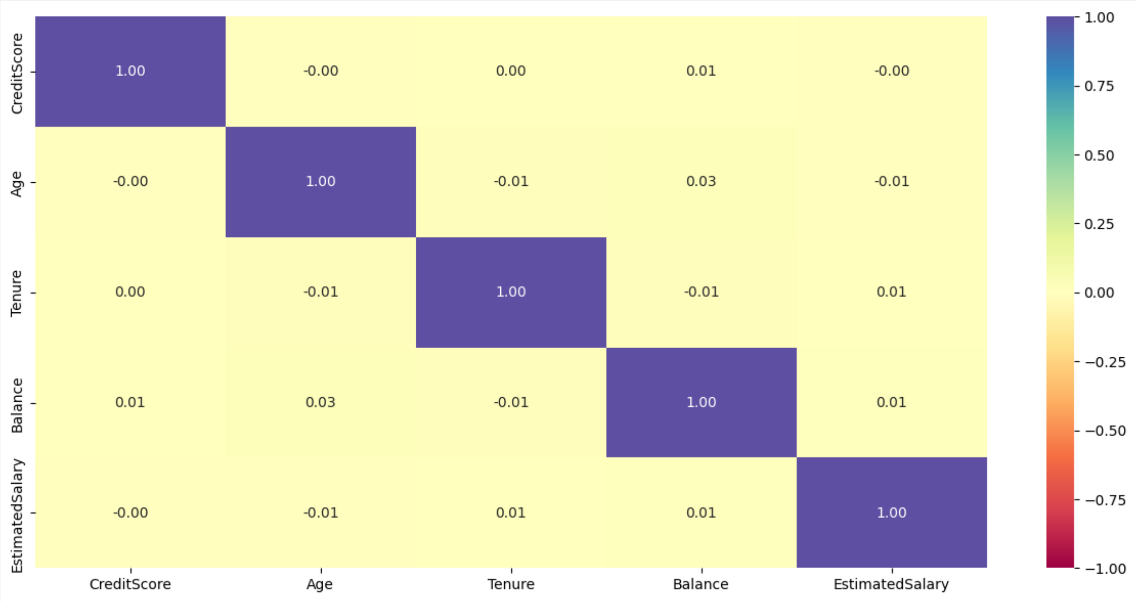
<u>Is Active Member:</u>

1. Distribution: The distribution is nearly even between active and non-active members
2. Implication: Whether a customer is an active member could be a crucial factor in predicting churn. Active members might be less likely to churn compared to inactive ones

<u>Gender:</u>

1. Distribution: The male and female customers is almost equal
2. Implication: Gender balance suggests that any gender-based differences in churn can be observed and modeled

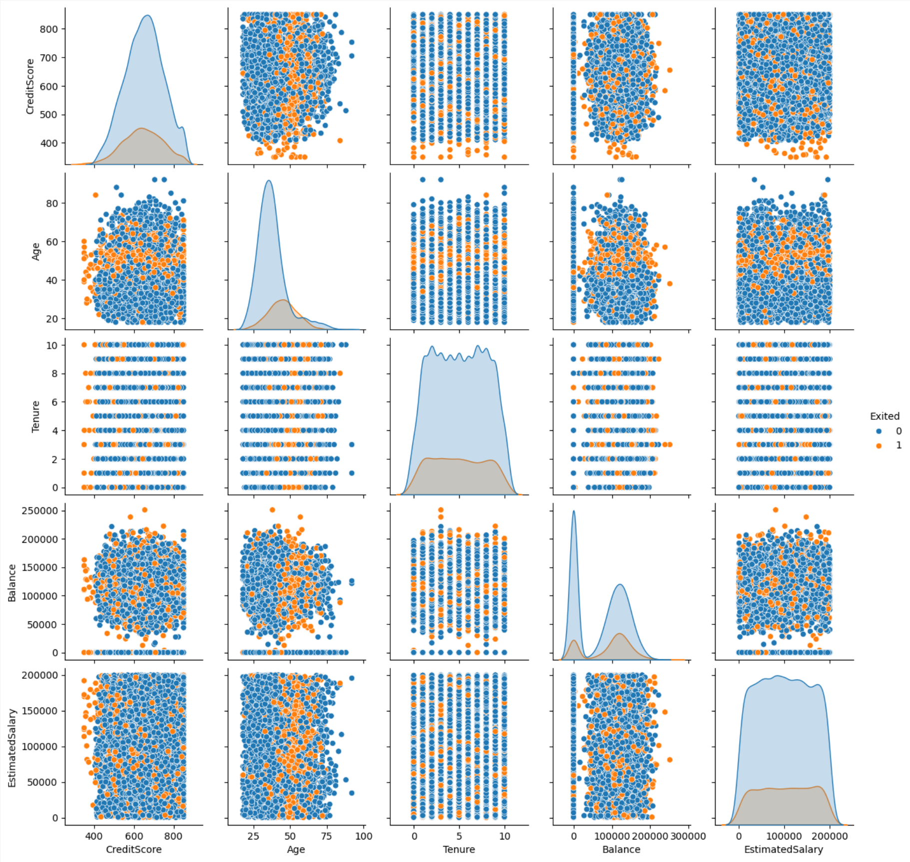
EDA Results of Bivariate Analysis



Key Takeaways:

1. The correlation matrix indicates that the features in the dataset have **very weak linear relationships with each other, suggesting that they are largely independent**
2. This can be advantageous for building a predictive model, as it ensures that each feature adds unique information to the model.

EDA Results of Bivariate Analysis



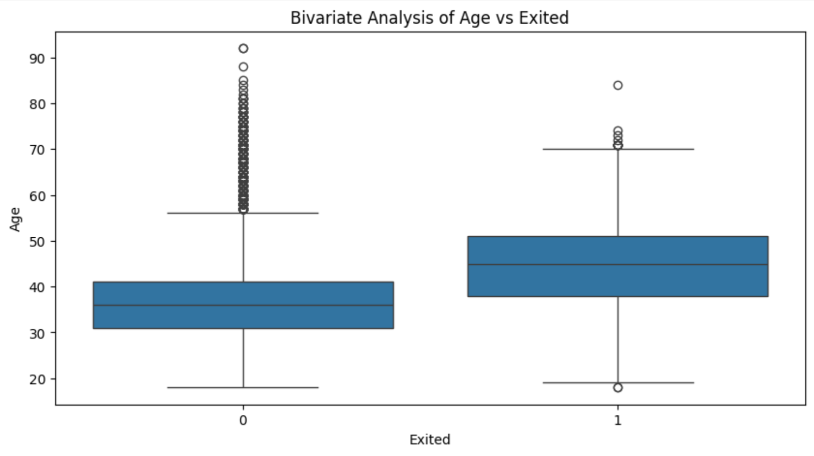
Key Takeaways:

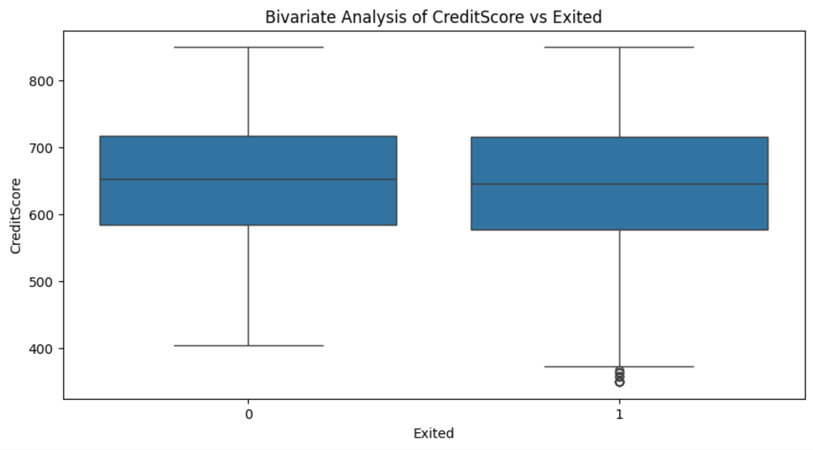
1. The pair plot helps to visually assess the distribution of features and potential interactions

2. **No strong non-linear relationships between features are evident**

3. Some features like age and balance show distinct patterns that might help in predicting churn, but these patterns appear to be linear or weakly correlated rather than non-linear

EDA Results of Bivariate Analysis





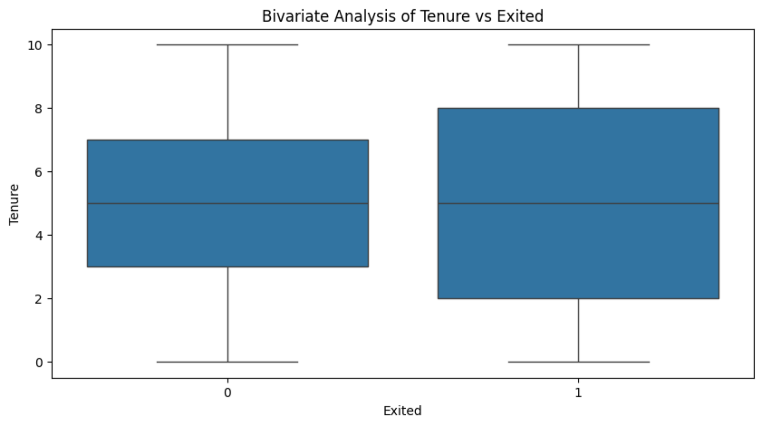
<u>Age vs Exited:</u>

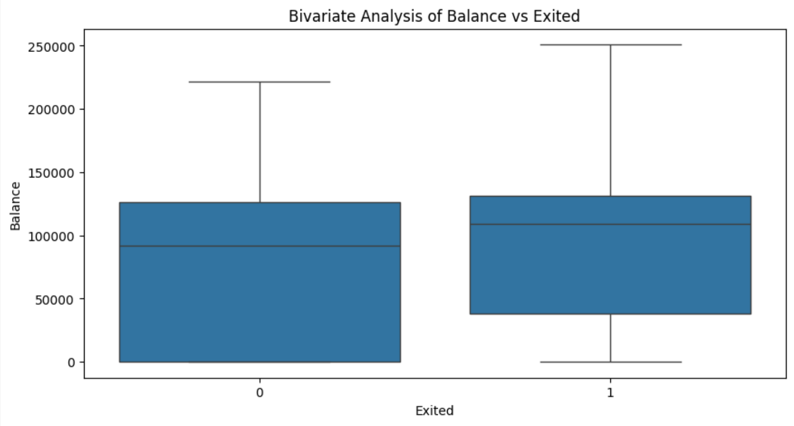
1. Distribution: Analysis of exited customers reveals an age skew
2. Implication: Older customers are more likely to churn, highlighting a need for targeted retention strategies. However, the presence of older customers who did not exit suggests that age alone is not the only factor influencing churn

<u>Credit Score vs Exited:</u>

1. Distribution: Median credit scores are similar for both exited and non-exited customers, around 650-700.The interquartile range (IQR) is almost the same for both groups, indicating a similar spread of credit scores
2. Implication: The presence of lower credit score outliers among the exited group suggests customers with significantly lower credit scores may be more likely to churn. This factor should be considered in churn prediction models

EDA Results of Bivariate Analysis





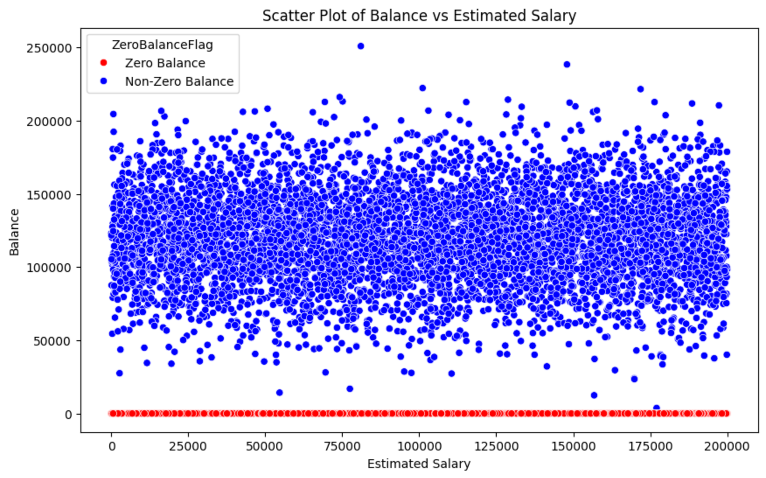
<u>Tenure vs Exited:</u>

1. Distribution: The median tenure for both groups (exited and not exited) is similar, around 5 years
2. Implication: Tenure does not appear to be a strong differentiator between customers who exited and those who did not

<u>Balance vs Exited:</u>

1. High-Balance Churn: Analysis reveals higher balances among exited customers
2. Implication : Balance should be included as a key feature in churn prediction model. Developing targeted retention strategies for high-value customers should be considered

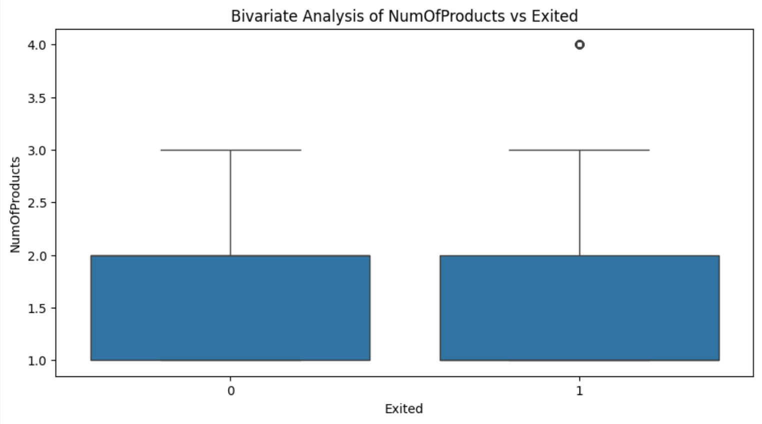
EDA Results of Bivariate Analysis

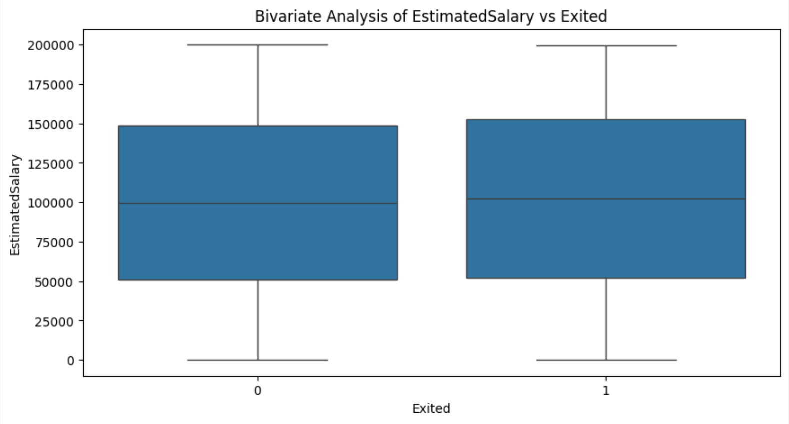


<u>Balance vs Estimated Salary :</u>

1. Zero Balance: There is a noticeable line of points along the x-axis at Balance = 0, indicating a significant number of customers have a zero balance irrespective of their Estimated Salary
2. No Clear Relationship: The spread of points appears to be random, and there is no clear pattern or trend that suggests a strong linear relationship between Balance and Estimated Salary. This is also supported by the calculated correlation coefficient (which would be close to 0 if there is no strong linear relationship) nor is there any non linear relationship

EDA Results of Bivariate Analysis





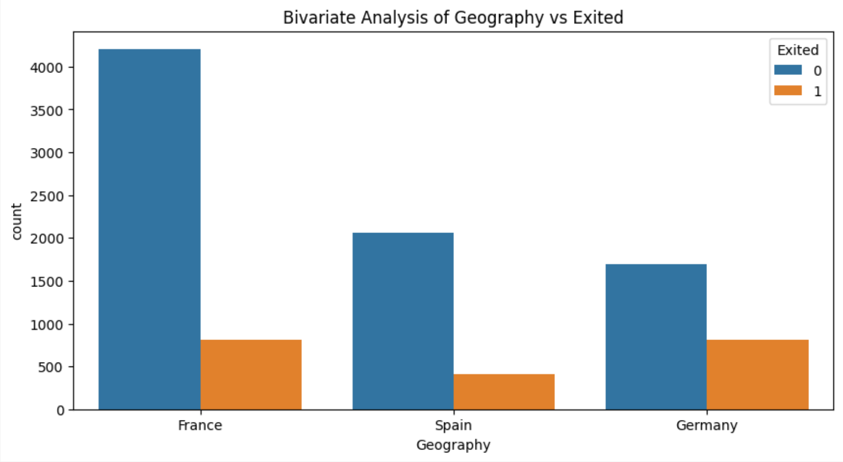
Number of products vs Exited:

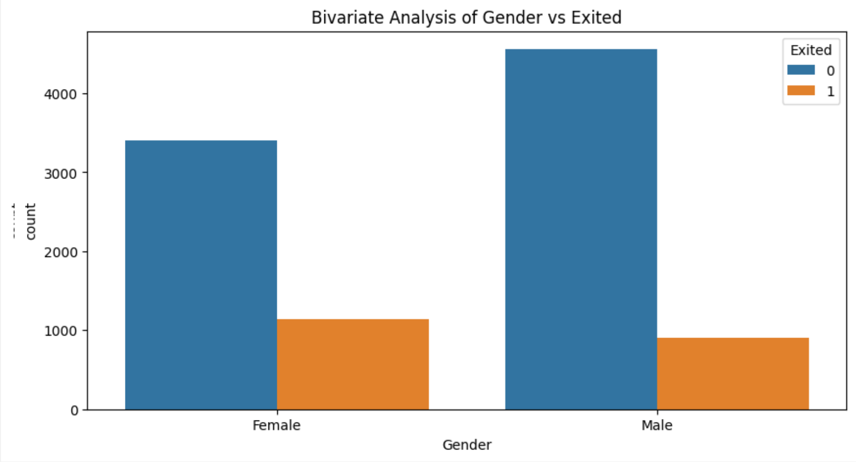
1. Product Usage & Churn: Similar product usage across exited and non-exited customers, but slightly higher for non-churned customers
2. Model Inclusion: Num Of Products included to assess its interaction with other churn factors.
3. Retention Strategy: Encourage product adoption to increase customer engagement

Estimated Salary vs Exited:

1. Salary & Churn: No clear difference in estimated salary between exited and non-exited customers
2. Model Inclusion: Estimated Salary to be included for interaction analysis
3. Retention Focus: Prioritize other factors for targeted retention strategies

EDA Results of Bivariate Analysis





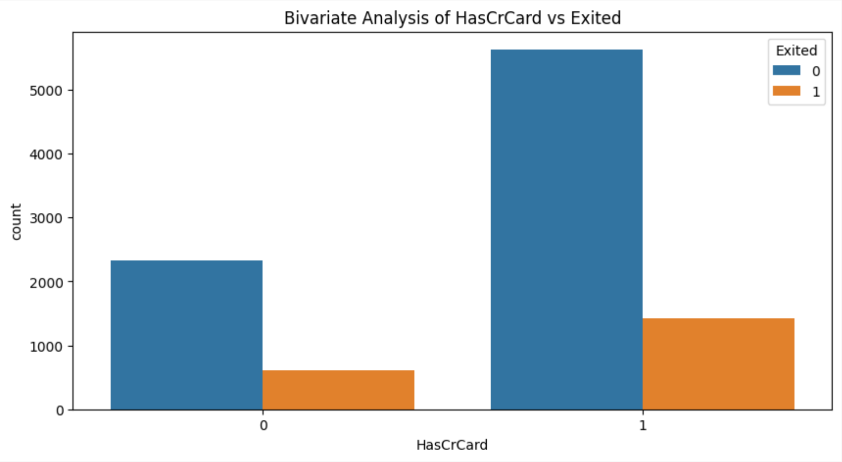
<u>Geography vs Exited:</u>

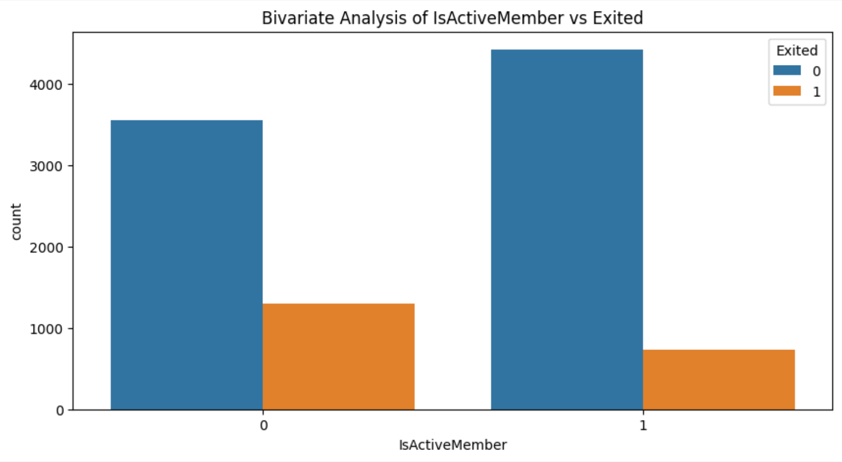
1. Distribution: Churn rates vary significantly by region, with Germany experiencing higher churn than France and Spain
2. Model Update: This suggests that Geography should be included as a key feature for targeted churn prediction
3. Regional Strategies: It might be worth to develop targeted retention efforts specific to each region's needs

<u>Gender vs Exited:</u>

1. Distribution: Higher churn rate observed among females compared to males
2. Model Inclusion: Gender included as a key feature for targeted churn prediction
3. Targeted Strategies: Develop retention efforts specific to address female customer needs

EDA Results of Bivariate Analysis





Has Credit Card vs Exited:

1. Distribution: Customers without a credit card show higher churn rates
2. Model Inclusion: Has Credit Card included as a key feature for churn prediction
3. Targeted Strategies: Retention efforts designed for customers lacking credit cards

Is Active Member vs Exited:

1. Distribution: Active members (Is Active Member = 1) are less likely to exit than inactive members (Is Active Member = 0)is higher than the count of active members who exited
2. Implications: Active membership status strongly indicates lower churn rates. Enhancing engagement and encouraging active membership can reduce churn

Data Preprocessing

Data Overview

- Columns (10000)
- Rows(14)

Data Cleaning → **Handling Zero Values and outliers** → **Feature Scaling** → **Data Split for Modeling**

Data Cleaning

1. Dropped columns with unique values 'Number', 'Customer Id', 'Surname'
2. Ensured there are no missing, duplicate, or unknown values in the data set
3. Applied one-hot encoding on all categorical columns with object data type

Handling Zero Values and outliers

1. Decided to leave the zero balance entries as is, due to the lack of strong correlation with other features and that they appear varied and non-abnormal across the dataset
2. Decided to leave the outliers as they were not abnormal and relatively fewer in number compared to the dataset volume

Feature Scaling

1. Normalized features like Credit Score, Age, Balance, and Estimated Salary that have different ranges and units using standard scalar function

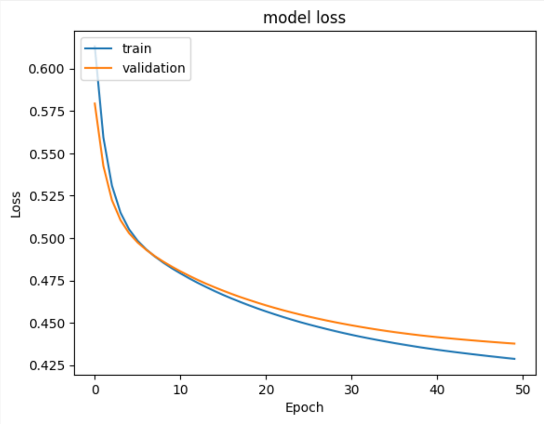
Data Split for Modeling

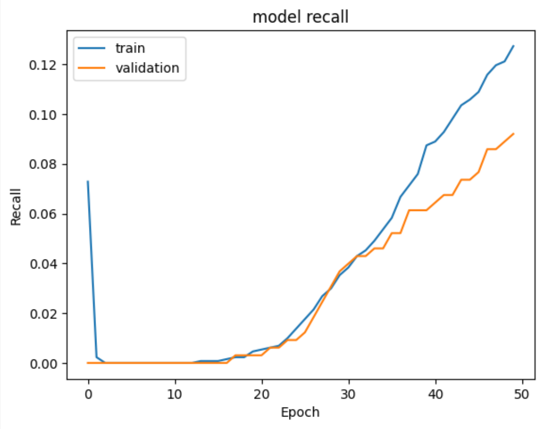
1. Split the data set as follows

X_Training set: (6400, 11)
X_Validation set: (1600, 11)
X_Test set: (2000,11)
y_Training set:6400
y_Validation set:1600
y_Test set:2000

NN Model with SGD Optimizer

Choosing the Metric:
Given the business scenario where the bank wants to prevent customer churn, **Recall** should be a priority metric to ensure most potential churners are identified. F1 Score is also very useful as it balances the need to identify churners (recall) and the need to avoid false positives (precision)





Loss Function Observations:
Training and validation loss decrease steadily over 50 epochs, indicating learning. The small gap between training and validation loss suggests minimal overfitting

Recall Observations:
Recall starts low for both sets but improves gradually, ending at 0.092 on validation, indicating poor identification of churners

Model Performance Metrics (NN with SGD optimizer)

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.82	0.98	0.89
1	0.65	0.13	0.21

Accuracy = 0.81

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.81	0.98	0.89
1	0.59	0.09	0.16

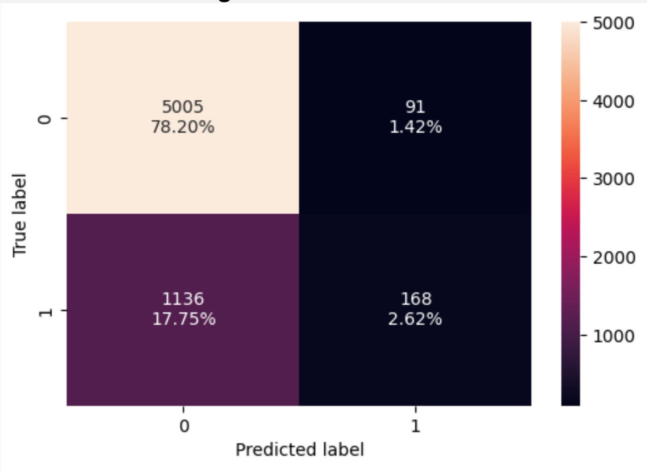
Accuracy = 0.80

Classification Report:

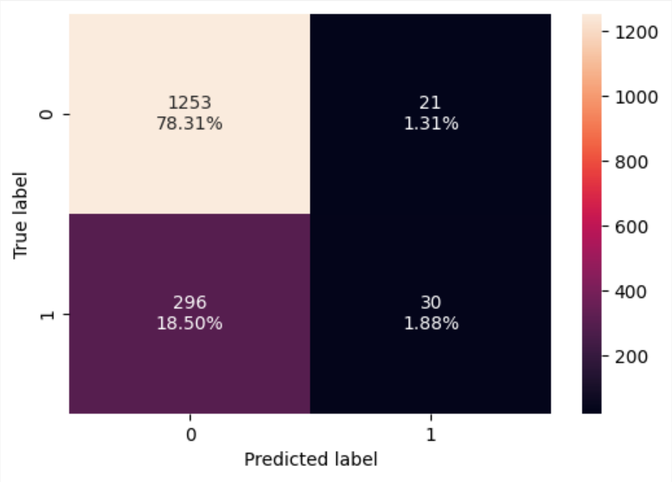
1. The recall for the positive class is very low, indicating that the model misses many actual churners
2. The precision for the positive class (churners) is relatively low, especially for the validation set, which means there are many false positives
3. The overall accuracy is good (around 80% for both training and validation), but this is primarily due to the high accuracy in predicting the non-churners

Confusion Matrix (NN with SGD optimizer)

Training Set Confusion Matrix



Validation Set Confusion Matrix



<u>Summary of Confusion Matrix:</u>

1. For the training set, the model has a high true negative rate (correctly identifying non-churners) but a low true positive rate (correctly identifying churners)
2. Similarly, for the validation set, the true negative rate is high, but the true positive rate remains low

Model Performance Summary (NN with SGD optimizer)

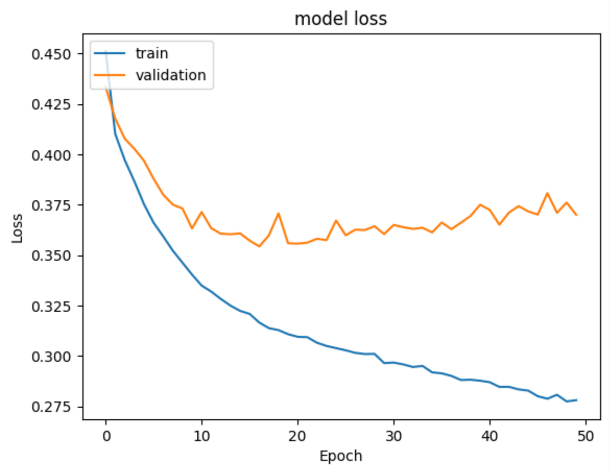
Model Summary: Sequential Model fit with batch size 32 for 50 epochs

Activation Function	Neurons	Parameter #
ReLU	64	768
ReLU	32	2080
Sigmoid	1	33

Summary of Model Performance:

1. The model with SGD optimizer struggles to identify churners accurately, as evidenced by the low recall for the positive class
2. While the overall accuracy is high, this is misleading due to the class imbalance and the model's inability to correctly identify a significant portion of churners
3. The performance indicates a need for further tuning or changing the model to improve recall for the positive class, which is critical for churn prediction tasks

Neural Network Model with Adam Optimizer



Loss Function Observations:

1. Training loss decreases consistently, indicating learning
2. Validation loss decreases initially but fluctuates after 20 epochs, suggesting potential overfitting or need for tuning

Recall Observations:

1. Recall improves over epochs for both sets
2. Validation recall is more volatile, indicating variability in identifying churners

Model Performance Metrics (NN with Adam optimizer)

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.90	0.97	0.93
1	0.82	0.60	0.69

Accuracy = 0.89

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.87	0.95	0.91
1	0.70	0.46	0.55

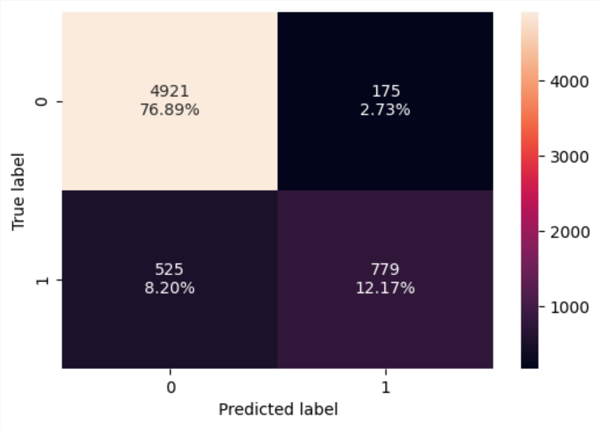
Accuracy = 0.85

Classification Report:

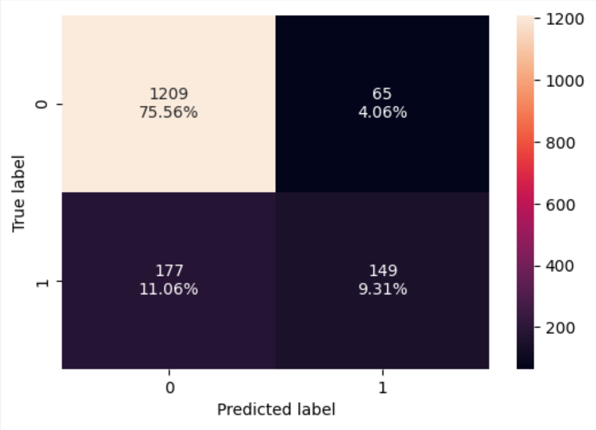
1. The model shows good precision and recall for non-churners (Class 0) but struggles with recall for churners (Class 1)
2. The disparity between training and validation recall for churners suggests overfitting and a need for further model tuning to improve generalization to unseen data

Confusion Matrix (NN with Adam optimizer)

Training Set Confusion Matrix



Validation Set Confusion Matrix



<u>Summary of Confusion Matrix:</u>

1. Strengths: The model is effective at identifying non-churners, as evidenced by the high TN values
2. Weaknesses: The model struggles with identifying churners accurately, with significant FN counts. This implies that the recall for the positive class (churners) is not as high as desired
3. Generalization: The validation set performance is consistent with the training set, suggesting the model generalizes well but still requires improvement in identifying churners accurately

Model Performance Summary (NN with Adam optimizer)

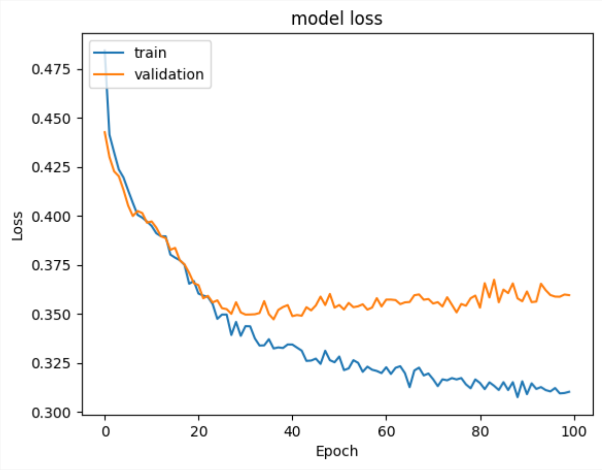
Model Summary: Sequential Model fit with batch size 32 for 50 epochs

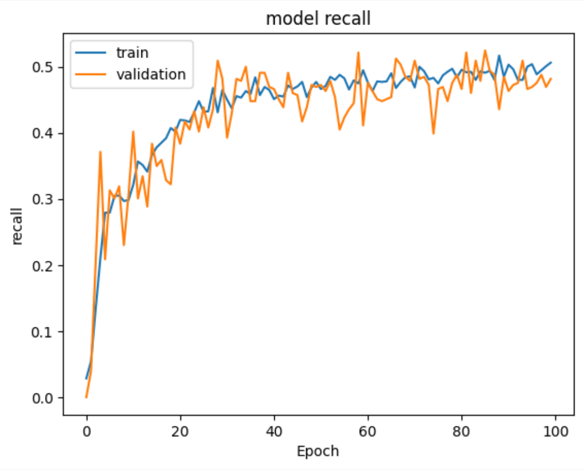
Activation Function	Neurons	Parameter #
ReLU	64	768
ReLU	32	2080
Sigmoid	1	33

Summary of Model Performance:

1. **Improved Performance:** The Adam optimizer improves recall for churners compared to SGD, with better precision and recall, leading to higher F1-scores
2. **Class Imbalance:** Recall for churners is still lower than desired, but high precision indicates correct churn predictions
3. **Generalization:** The model generalizes better than SGD, but validation recall volatility suggests a need for further regularization or dropout adjustments
4. **Potential Overfitting:** Validation loss and recall fluctuations indicate potential overfitting

Neural Network Model with Adam Optimizer and Dropout





<u>Loss Function Observations:</u>

1. Training Loss: Steadily decreases, indicating improved performance
2. Validation Loss: Initially decreases but fluctuates, suggesting some overfitting, though dropout layers help control it

<u>Recall Observations:</u>

1. Training Recall: Improves steadily, indicating better identification of positive cases
2. Validation Recall: Improves but fluctuates, indicating variability in performance

Model Performance Metrics (NN with Adam optimizer and Dropout)

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.90	0.97	0.93
1	0.84	0.56	0.67

Accuracy = 0.89

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.88	0.96	0.92
1	0.73	0.48	0.58

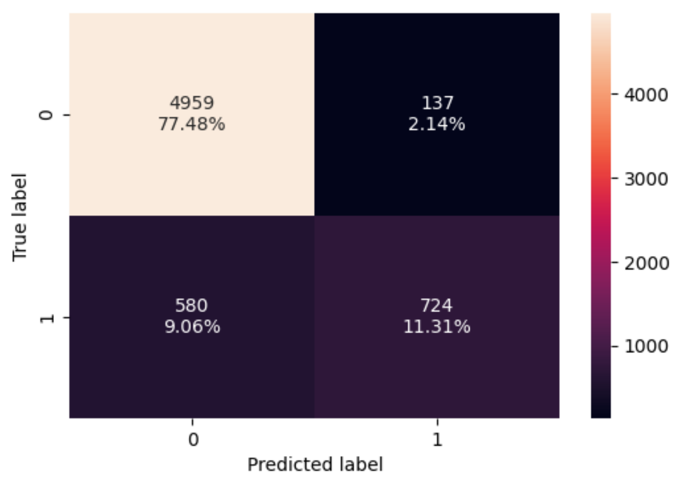
Accuracy = 0.86

Classification Report:

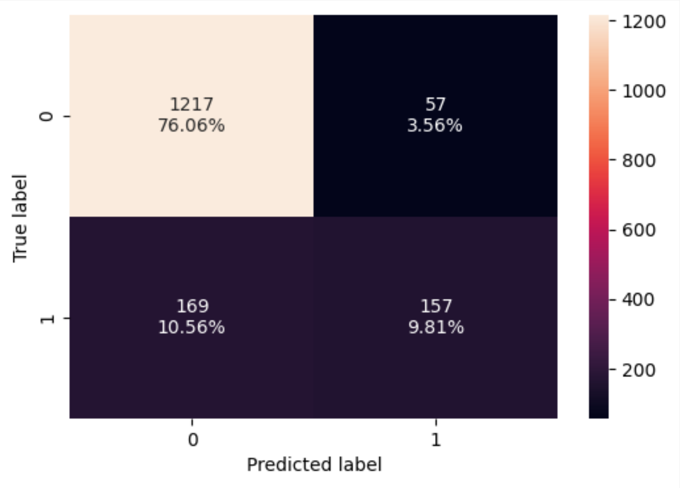
1. The model has high precision for both classes, indicating accurate positive predictions
2. Recall for churners (Class 1) is significantly lower, meaning many churners are missed
3. High accuracy is primarily due to correctly predicting non-churners (Class 0)
4. Slightly lower performance on the validation set suggests minor overfitting, which could be mitigated with further tuning or regularization.

Confusion Matrix (NN with Adam optimizer and Dropout)

Training Set Confusion Matrix



Validation Set Confusion Matrix



<u>Summary of Confusion Matrix:</u>

1. Strengths: Effective at identifying non-churners with a low False Positive Rate
2. Weaknesses: Struggles to identify actual churners and misses many actual churners
3. Generalization: Consistent performance across training and validation sets, but needs improvement in identifying churners accurately

Model Performance Summary (NN with Adam optimizer and Dropout))

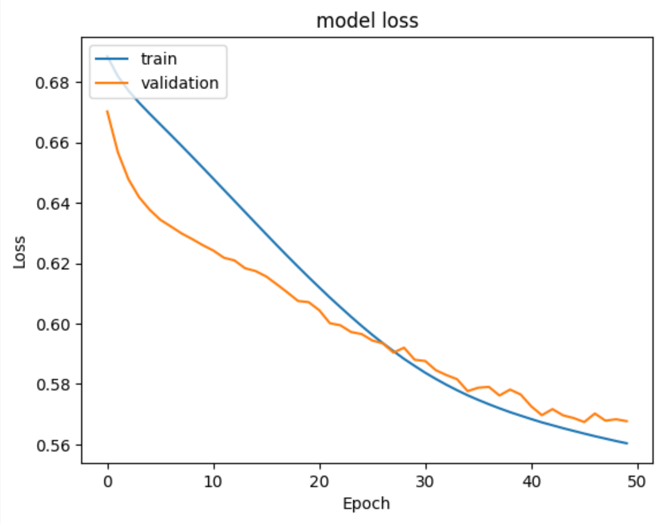
Model Summary: Sequential Model fit with batch size 32 for 100 epochs

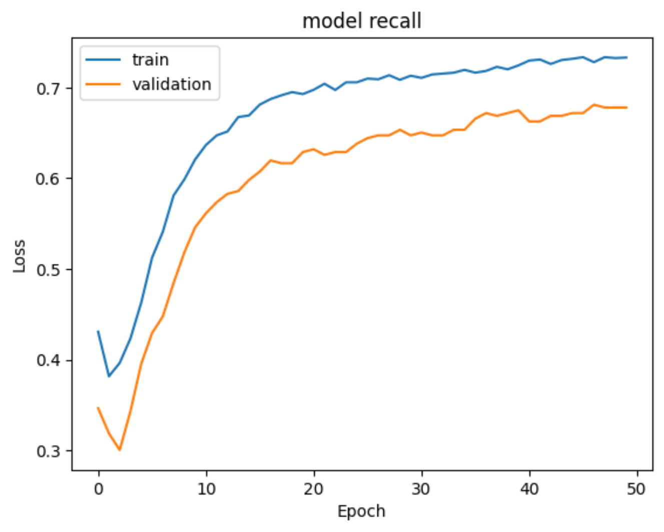
Activation Function	Neurons	Parameter
ReLU	32	384
Dropout with ratio 0.2	0	0
ReLU	32	1056
ReLU	32	1056
Dropout with ratio 0.1	0	0
ReLU	32	1056
Sigmoid	1	33

Summary of Model Performance:

1. **Overall Performance:** The model performs well, with 0.86 accuracy on the validation set
2. **Recall Improvement Needed:** There is room for improvement in recall for the positive class to better identify churn cases
3. **Robustness to Overfitting:** The incorporation of dropout has made the model more robust to overfitting, showing more stable performance across epochs
4. **Critical Focus:** Further improvements are needed to increase recall for the positive class, which is critical in churn prediction

Neural Network Model with SMOTE and SGD Optimizer





Loss Function Observations:

1. Training Loss: Steadily decreases, showing model learning and improvement
2. Validation Loss: Decreases over time, indicating good generalization and minimal overfitting

Recall Observations:

1. Training Recall: Improves significantly, stabilizing around 73%
2. Validation Recall: Improves significantly, stabilizing around 68%, consistent with training recall

Model Performance Metrics (NN with SMOTE and SGD Optimizer

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.73	0.71	0.72
1	0.72	0.73	0.73

Accuracy = 0.89

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.90	0.72	0.80
1	0.38	0.68	0.49

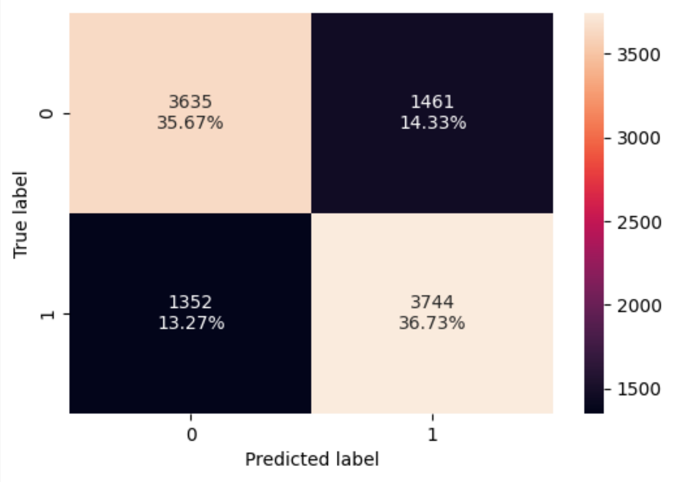
Accuracy = 0.86

Classification Report:

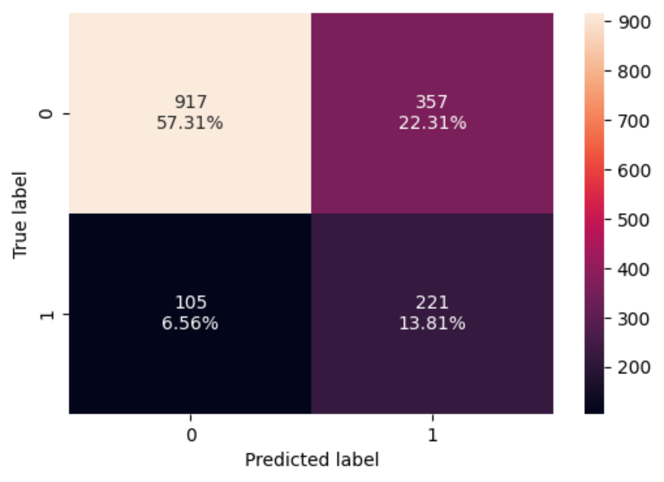
1. Training Data: Balanced precision and recall with an accuracy of 0.72, indicating a good fit due to SMOTE balancing
2. Validation Data: High precision for Class 0 but low precision for Class 1, leading to a higher false positive rate for churn prediction
3. Overall Accuracy: Decent performance at 0.71, but the disparity between precision and recall for Class 1 indicates overfitting and the need for further tuning

Confusion Matrix (NN with SMOTE and SGD Optimizer)

Training Set Confusion Matrix



Validation Set Confusion Matrix



<u>**Summary of Confusion Matrix:**</u>

1. Strengths: The model identifies churners reasonably well but misclassifies a substantial number of non-churners and churners, indicating difficulty in distinguishing between classes
2. Weaknesses: Effective at identifying non-churners with a high True Negative Rate, but struggles with a high False Positive Rate, often incorrectly predicting non-churners as churners
3. The model generalizes well to unseen data, but the high False Positive Rate suggests a need for improved precision in churn predictions

Model Performance Summary (NN with SMOTE and SGD Optimizer)

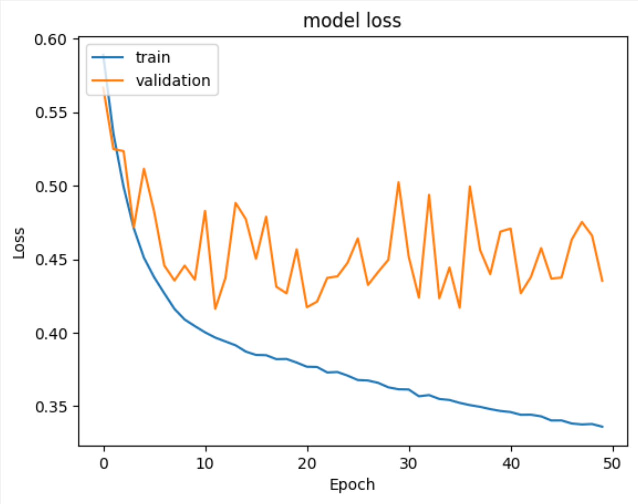
Model Summary: Sequential Model fit with batch size 32 for 50 epochs

Activation Function	Neurons	Parameter
ReLU	32	384
ReLU	16	528
ReLU	16	272
Sigmoid	1	17

Summary of Model Performance:

1. The overall accuracy of 71% on the validation set is decent, with balanced recall
2. The model has improved recall for the positive class, critical for churn prediction
3. The use of SMOTE and SGD has achieved balanced recall for both classes and good generalization
4. There is room for improvement in precision for the churn class

Neural Network Model with SMOTE and Adam Optimizer



Loss Function Observations:

1. Training Loss: Decreases steadily, indicating learning and performance improvement

2. Validation Loss: Decreases initially but fluctuates, suggesting some overfitting

Recall Observations:

1. Training Recall: Steadily improves, indicating better identification of positive cases

2. Validation Recall: Improves but fluctuates more, indicating variability in validation performance

Model Performance Metrics (NN with SMOTE and Adam Optimizer

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.85	0.88	0.86
1	0.87	0.85	0.86

Accuracy = 0.86

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.90	0.85	0.87
1	0.51	0.61	0.56

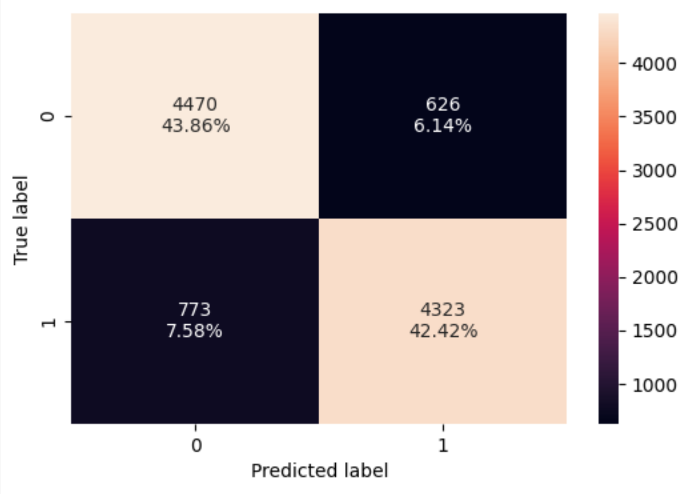
Accuracy = 0.80

Classification Report:

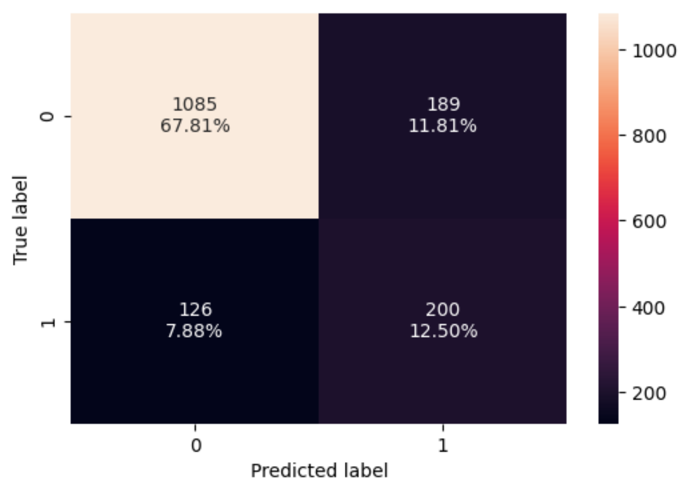
1. Strengths: High Precision and Recall for both classes in the training data (0.85-0.88), Balanced F1-Scores and strong overall accuracy (0.86) in the training set. The model is well-trained on the training data, showing effective learning
2. Weaknesses: Significantly lower Precision for churners (Class 1) in the validation data. Recall for Class 1 is higher than precision, indicating many false positives. Lower F1-Score for Class 1 (0.56) compared to Class 0 (0.87), showing less effectiveness in accurately identifying churners
3. overall accuracy: With 0.80 on the validation set indicates good generalization, but the disparity in precision and recall for the churn class highlights areas for improvement in reducing false positives and improving precision.

Confusion Matrix (NN with SMOTE and Adam Optimizer)

Training Set Confusion Matrix



Validation Set Confusion Matrix



Summary of Confusion Matrix:

1. Strengths: High True Positive and True Negative Rates: The model correctly identifies a large portion of churners and effectively identifies non-churners in both training and validation data
2. Weaknesses: The model has a moderate false positive rate in training data and a high false positive rate in validation data. It also misses some actual churners in both training and validation data
3. Generalization: The model balances precision and recall well on training data with SMOTE but shows overfitting on validation data, indicated by a high false positive rate, requiring further tuning for better precision in churn predictions

Model Performance Summary (NN with SMOTE and Adam Optimizer)

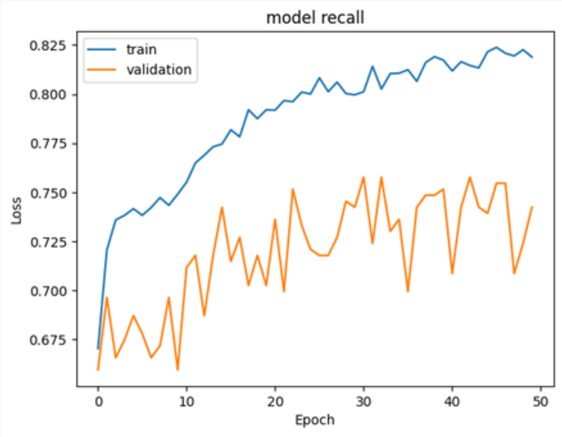
Model Summary: Sequential Model fit with batch size 32 for 50 epochs

Activation Function	Neurons	Parameter
ReLU	32	384
ReLU	16	528
ReLU	16	272
Sigmoid	1	17

Summary of Model Performance:

1. **Good Overall Performance:** The model achieves 0.80 accuracy on the validation set
2. **Improved Detection:** SMOTE and the Adam optimizer enhance the model's ability to detect positive cases compared to previous models
3. **Room for Improvement:** Despite improved recall for the positive class, variability in validation performance suggests further tuning and additional regularization may enhance stability and performance

Neural Network Model with SMOTE , Adam & Dropout



Loss Function Observations:

1. Training Loss: The training loss steadily decreases over the epochs, indicating the model is effectively learning and improving its performance on the training data
2. Validation Loss: The validation loss also decreases but shows more fluctuations compared to the training loss, suggesting variability in performance on the validation set.

Recall Observations:

1. Training Recall: The training recall improves steadily, indicating the model's increasing ability to correctly identify positive cases in the training set
2. Validation Recall: The validation recall shows improvement but fluctuates significantly, indicating variability in the model's performance on the validation data

Model Performance Metrics (NN with SMOTE, Adam & Dropout)

Classification Report on Training Set

Exited	Precision	Recall	F1-Score
0	0.85	0.83	0.84
1	0.83	0.85	0.84

Accuracy = 0.84

Classification Report on Validation Set

Exited	Precision	Recall	F1-Score
0	0.93	0.82	0.87
1	0.51	0.74	0.61

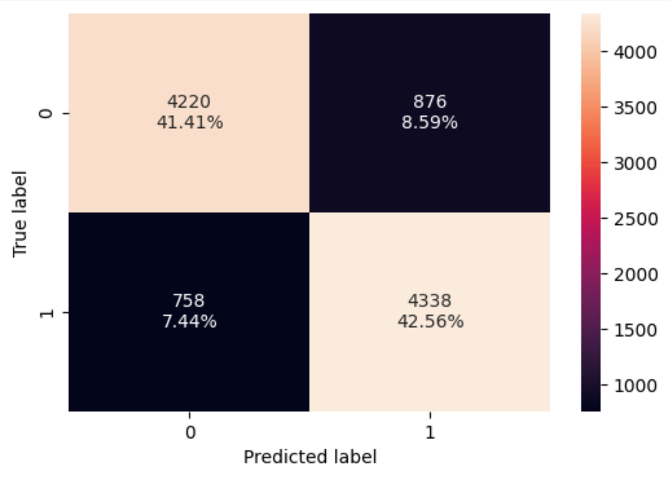
Accuracy = 0.80

Classification Report:

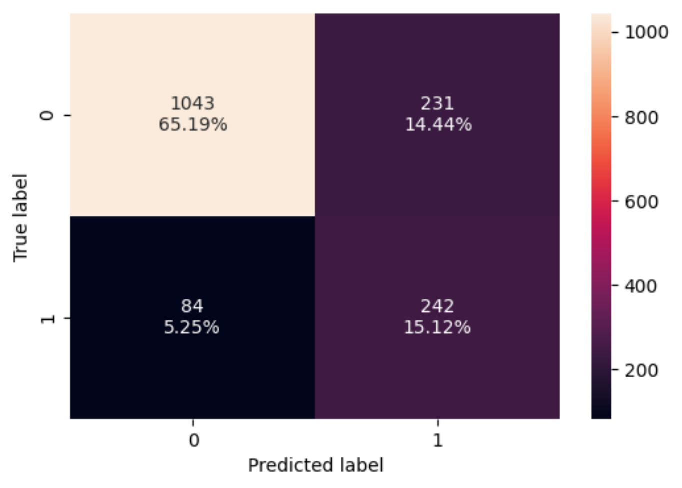
1. Strengths: In both training and validation data, the model has high precision and recall for Class 0, indicating it correctly identifies and predicts non-churners with high accuracy
2. Weaknesses: While the precision for Class 1 is lower in the validation set, indicating some false positives, the recall has improved, showing the model's increased ability to identify actual churners compared to previous models
3. Consistent Performance: The model generalizes well with an accuracy of 0.80, but the lower precision for Class 1 in validation indicates a need to reduce false positives in churn predictions

Confusion Matrix (NN with SMOTE, Adam & Dropout)

Training Set Confusion Matrix



Validation Set Confusion Matrix



<u>**Summary of Confusion Matrix:**</u>

1. High True Positive and Negative Rates: The model correctly identifies a significant portion of churners and non-churners in both training (42.56% churners, 41.41% non-churners) and validation sets (15.12% churners, 65.19% non-churners)

2. Moderate to High False Positive and Negative Rates: The model incorrectly predicts non-churners as churners at moderate rates in the training set (8.59%) and higher rates in the validation set (14.44%). It also misses some actual churners, with moderate false negatives in training (7.44%) and lower false negatives in validation (5.25%)

Model Performance Summary (NN with SMOTE, Adam & Dropout)

Model Summary: Sequential Model fit with batch size 32 for 50 epochs

Activation Function	Neurons	Parameter
ReLU	32	384
Dropout with ratio 0.2	32	0
ReLU	16	528
Dropout with ratio 0.2	16	0
ReLU	8	136
Sigmoid	1	9

Summary of Model Performance:

1. **Improved Recall for Positive Class:** The model, incorporating SMOTE and Adam optimizer with dropout, has significantly improved recall for the positive class (churners), ensuring more churn cases are correctly identified

2. **Consistent Training and Validation Recall:** The training recall stabilizes around 85%, and the validation recall around 74%, showing consistent performance and good generalization from training to validation data.

Model Performance Comparison

Table of Recall values of Class 1 for all Models			
Model Name	Training Set Recall (a)	Validation Set Recall (b)	Difference (a – b)
NN with SGD	0.13	0.09	0.04
NN with Adam	0.60	0.46	0.14
NN with Adam & Dropout	0.56	0.48	0.07
NN with SMOTE & SGD	0.73	0.68	0.05
NN with SMOTE and Adam	0.85	0.61	0.23
NN with SMOTE and Adam & Dropout	0.85	0.74	0.11

Reasons for choosing Model:

1. Conclusion NN with SMOTE, Adam & Dropout is the best model based on these results due to its highest validation recall and balanced performance between training and validation sets

2. Relatively lower difference between recall values across training and validation set

Chosen Model

Rationale for Model Selection

Model NN with SMOTE, Adam & Dropout

Highest Recall on Validation Set

Achieves the highest recall, indicating superior performance in identifying churners

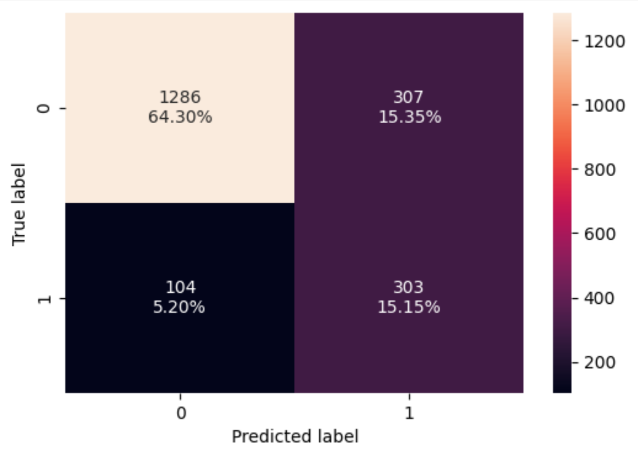
Balanced Metrics

Maintains a good balance between precision and F1-score, ensuring reliable predictions

Effective Generalization

Demonstrates consistent performance across training and validation sets, addressing class imbalance effectively

Model Performance on Test Data



Exited	Precision	Recall	F1-Score
0	0.93	0.81	0.86
1	0.50	0.74	0.60

Model Performance Summary:

1. The model performs well in identifying non-churners with high precision and recall
2. It demonstrates a good balance in identifying churners, with a recall of 74%, indicating it catches most churn cases
3. The precision for predicting churn is lower, meaning there are more false positives
4. Overall, the model shows balanced performance with good generalization capabilities
5. There is room for improvement in reducing false positives for churn predictions but still performs better than the other models

Key Takeaways for the Business

Key Takeaways

High Non-Churner Identification: The model effectively identifies non-churners, ensuring that most loyal customers are accurately classified, which is crucial for maintaining customer satisfaction and retention strategies

Improved Churner Detection: With a recall of 74% for churners, the model successfully identifies most customers likely to churn, providing the bank with valuable insights to target at-risk customers with retention efforts

Need for Precision Enhancement: The lower precision for churn predictions indicates a higher rate of false positives. Improving this precision can reduce unnecessary retention efforts and associated costs, optimizing resource allocation

Balanced Performance and Generalization: The model's balanced performance across training and validation sets suggests it generalizes well to new data. Continuous monitoring and tuning can further enhance its predictive power, making it a reliable tool for churn management

Churn Prediction with Neural Networks: A Data-Driven Strategy for Customer Retention

The End