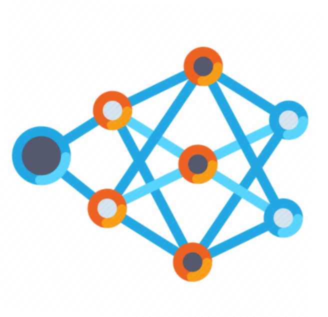
PROJECT 4: NEURAL NETWORKING

--

— Bank Churn Prediction—

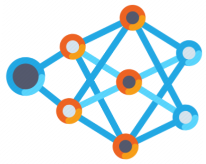


Tiffany Rhodes

February 29, 2025

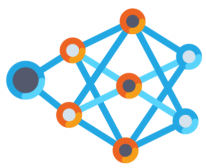
Contents / Agenda

- Executive Summary
- Objectives
- Business Problem Overview
- Solution Approach & Methodology
- Data Processing
- Model Performance Summary-Hyperparameter Tuning
- Conclusion
- Appendix



Executive Summary

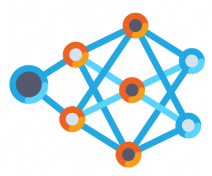
Businesses that provide services, such as banks, often face the challenge of customer churn—where customers leave to join a competing service. Understanding the key factors that influence a customer's decision to stay, or leave is crucial. By identifying these factors, management can focus on improving services that matter most to customers, ultimately enhancing retention and customer satisfaction.



Objectives

The primary goal of this project is to:

- ✓ Develop a **neural network-based classification model** that can accurately predict whether a customer is likely to **churn (leave the bank) or stay** within the next **six months**.
- ✓ This model will analyze **customer behavior, transaction patterns, engagement levels, and financial data** to identify key risk factors that contribute to churn.
- ✓ By leveraging **deep learning techniques**, the goal is to **enhance customer retention strategies, personalize banking services, and proactively address at-risk customers** to reduce overall churn rates. .



Executive Summary

Key Insights:

Inactivity is a Strong Predictor of Churn

- Customers who have been inactive for **3+ months** are at a **high risk of leaving**.
- A significant number of churned customers had **low transaction activity** in the last 12 months.

Transaction Frequency and Engagement Impact Retention

- Customers with **higher transaction counts are less likely to churn**.
- A **decline in transaction volume between Q4 and Q1** strongly correlates with attrition.

Credit Utilization Patterns Affect Churn

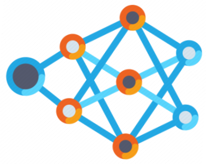
- Customers with **low revolving balances and lower utilization ratios** tend to churn more.
- High-credit-limit customers show **lower engagement**, indicating the need for enhanced loyalty programs.

Demographic Factors Play a Role in Customer Churn

- Customers aged **40-55** are the most stable segment, while **younger customers have higher churn rates**.
- Customers with **fewer banking products (1-2) are more likely to leave** compared to those using multiple services.

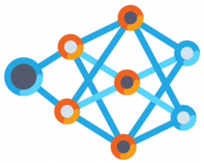
Contact Frequency and Customer Support Engagement

- Customers who **contact the bank frequently** may indicate dissatisfaction and are at **greater risk of churn**.
- The majority of customers engage with the bank **only 2-3 times a year**, highlighting a need for **better engagement strategies**.



Data Dictionary

- **CustomerID** – Unique ID assigned to each customer.
- **CreditScore** – Defines the customer’s credit history.
- **Geography** – Customer’s location.
- **Gender** – Defines customer’s gender.
- **Age** – Customer’s age.
- **Tenure** – Number of years the customer has been with the bank.
- **NumOfProducts** – Number of products purchased through the bank.
- **Balance** – Customer’s account balance.
- **HasCrCard** – Whether the customer has a credit card.
- **EstimatedSalary** – Customer’s estimated salary.
- **IsActiveMember** – Whether the customer actively engages with the bank’s services.
- **Exited** – **Target variable** (1 = Customer left, 0 = Customer stayed).



Business Problem Overview

Introduction

Thera Bank has been facing a decline in its credit card user base, which directly impacts its revenue from transaction fees, annual fees, interest charges, and other financial services. Customer churn in the credit card sector is a critical issue, as acquiring new customers is significantly more expensive than retaining existing ones. This project aims to **analyze customer behavior, predict churn, and develop strategies to enhance customer retention.**

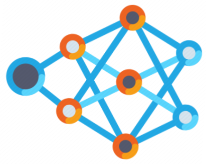
Key Challenges

- **Revenue Loss Due to Customer Attrition**
    - Churned customers reduce the bank's revenue from credit card transactions, interest payments, and service fees.
    - Loss of long-term customers impacts the bank's profitability and growth potential.

- **Limited Customer Engagement & Usage**
    - Customers with **low transaction frequency and long inactivity periods** are more likely to churn.
    - The absence of targeted engagement strategies leads to **missed opportunities for customer retention**.

- **Understanding the Drivers of Churn**
    - Customer churn is influenced by multiple factors such as **credit utilization, inactivity, contact frequency, transaction behavior, and demographic attributes**.
    - The bank needs a **data-driven approach to accurately identify churn predictors** and mitigate risks.

- **Class Imbalance in Customer Data**
    - The dataset shows **more active customers than churned ones**, which can affect predictive model accuracy.
    - Proper balancing techniques, such as **oversampling or undersampling,** are required to improve model performance.

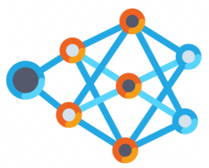


Data Quality

OVERVIEW

- The dataset used for this analysis comprises various customer attributes that are crucial in predicting credit card churn.

- The data quality was thoroughly evaluated and cleaned to ensure that it is accurate, consistent, and suitable for machine learning model building.



Data Quality Summary

Great Learning
POWER AHEAD

- The dataset is **complete**, with no missing values and a good mix of numerical and categorical data.
- Data is **accurate**, with proper data types assigned and values validated against business logic.
- Outliers were handled appropriately to ensure **model robustness** and **accurate predictions**.
- The dataset contains **class imbalance**, which was addressed with oversampling and undersampling techniques to ensure a fair predictive model.

Solution Approach

1. Understanding the Business Problem

- Clearly define the problem statement.
- Identify key challenges and pain points.
- Define business objectives and success metrics.

2. Exploratory Data Analysis (EDA)

- Data distribution and summary statistics.
- Identifying missing values, duplicates, and inconsistencies.
- Detecting outliers and potential anomalies.
- Understanding feature correlations and relationships.

3. Data Preprocessing

- Handling missing values through imputation or removal.
- Addressing duplicate records to ensure data integrity.
- Outlier treatment based on statistical methods.
- Feature engineering to enhance model performance.
- Data scaling and transformation for model optimization.

4. Model Selection & Training

- Choosing appropriate machine learning algorithms.
- Performing hyperparameter tuning for optimization.
- Splitting data into training, validation, and testing sets.
- Handling class imbalance with resampling techniques (oversampling/undersampling).

5. Model Evaluation & Performance Metrics

- Comparing multiple models using standard metrics (accuracy, precision, recall, F1-score, ROC-AUC).
- Visualizing model performance through confusion matrices and error analysis.
- Selecting the final model based on business and performance considerations.

6. Deployment & Business Integration

- Preparing the model for deployment in a production environment.
- Integrating with existing business systems and workflows.
- Defining monitoring and maintenance strategies for continuous improvement.

7. Actionable Insights & Recommendations

- Highlighting key findings from the analysis.
- Providing strategic recommendations based on model insights.
- Suggesting future improvements and scaling opportunities.

Methology

Problem Definition & Business Understanding

- **CRISP-DM (Cross Industry Standard Process for Data Mining)**: A widely used framework for data science projects, ensuring a structured approach from problem definition to deployment.
- **Stakeholder Analysis**: Understanding business requirements through interviews, surveys, or analysis of business objectives.

Exploratory Data Analysis (EDA)

- **Descriptive Statistics**: Summarizing the dataset using mean, median, mode, standard deviation, and variance.
- **Data Visualization**: Utilizing Python libraries (Matplotlib, Seaborn, Plotly) to identify patterns, trends, and outliers.
- **Correlation Analysis**: Using Pearson, Spearman, or Kendall correlation to understand relationships between variables.
- **Dimensionality Reduction**: Using **Principal Component Analysis (PCA)** or **t-SNE** to visualize high-dimensional data.

Data Preprocessing

Missing Data Treatment:

- Imputation methods (mean, median, mode) for numerical variables.
- Using **KNN imputer** or regression-based imputation for complex cases.
- Dropping rows/columns with excessive missing values.

Outlier Detection & Treatment:

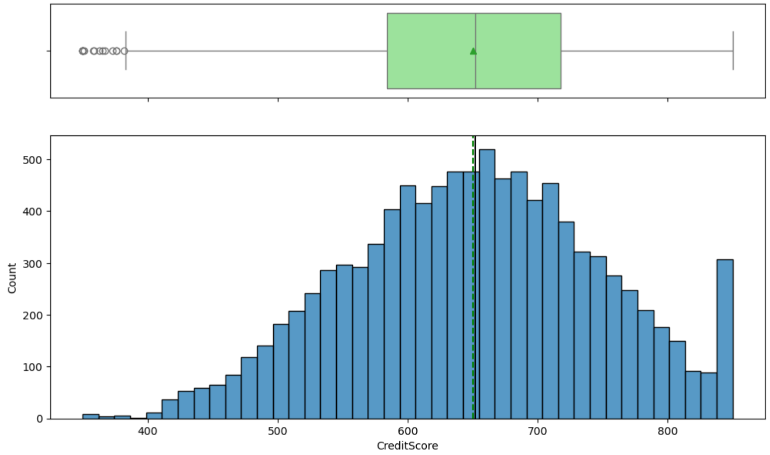
- Using **IQR (Interquartile Range), Z-score, and Boxplots** for detection.
- Winsorization or transformation (log, square root) to handle extreme values.

Feature Engineering:

- Creating new features based on domain knowledge.
- Encoding categorical variables using **One-Hot Encoding (OHE)** or **Label Encoding**.
- Standardization (**Z-score Normalization**) or Min-Max scaling for continuous features.

EXPLORATORY DATA ANALYSIS

EDA Results



Observations:

- The majority of customers have a credit score between **600** and **750**, suggesting a **financially stable customer base**.
- A small peak is visible **near 850**, likely indicating customers with excellent creditworthiness.
- The distribution appears **slightly right-skewed**, meaning a wider spread of high credit scores.

Key Insights:

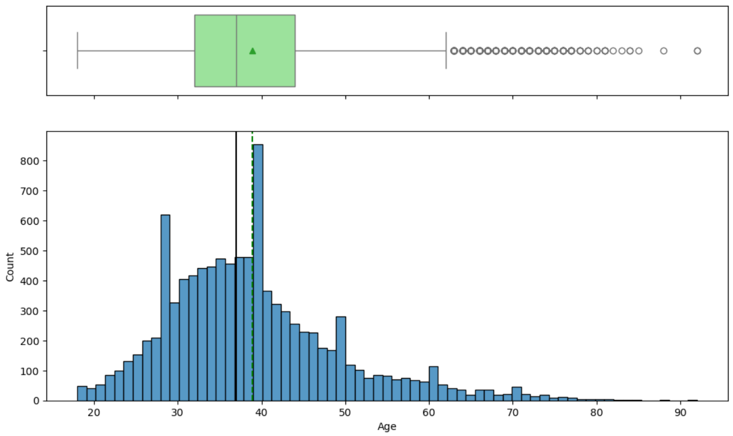
- Several low credit score outliers (below 400) suggest a small segment of customers with high credit risk.
- These outliers might represent customers with financial difficulties or new credit holders with limited history.
- The **IQR shows a tight cluster of scores**, meaning most customers have **similar credit behavior**.

Business Recommendations:

- Implement risk-based lending by offering lower interest rates to customers with scores above 750 and stricter loan terms for those below 600.
- Credit building programs (secures credit cards, small loans) for those with low credit scores to help improve their financial standing.
- Implement **early warning systems** to flag at-risk customers for proactive engagement and financial assistance options.
- Focus on **cross-selling opportunities** for mid-range customers (600-750), offering **loan consolidation** or **investment advisory services**.

EDA Results

Age Distribution Analysis



Observations:

- The **boxplot** shows an **asymmetrical distribution** with a right-skew (indicating more older customers).
- Outliers are present on the **higher age range**, meaning some older customers are significantly older than the median.
- The **histogram** suggests a bimodal distribution with peaks around **30-40 years** and **55+ years**.
- A noticeable spike at **40 years old** may suggest a **common age group for financial products or promotions**.

Key Insights:

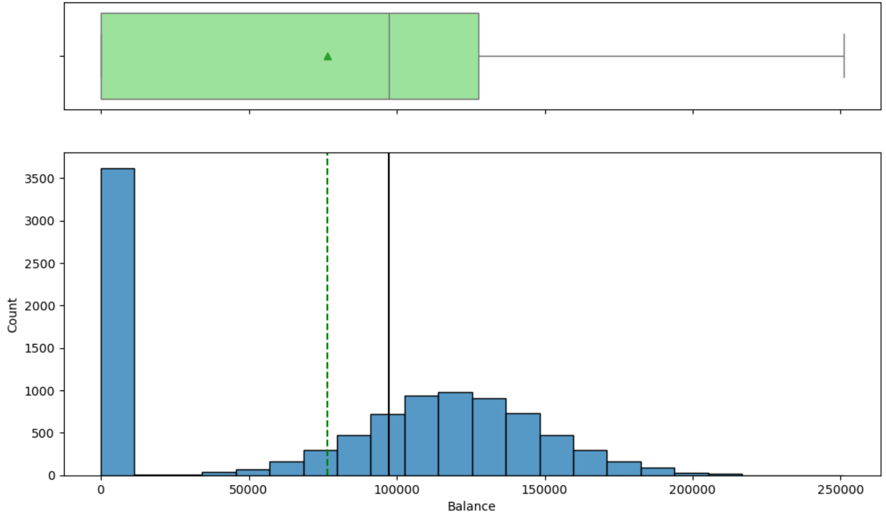
- Customers are primarily in **two segments: younger professionals (30-40 years) and older customers (50+ years)**.
- The **older customer group may be less tech-savvy**, which could impact online banking adoption.
- The **middle-aged group (40 years) is a crucial demographic**, possibly due to **family and mortgage commitments**.

Business Recommendations:

- Develop **age-specific financial products** (e.g., home loans for middle-aged, retirement savings for older customers).
- **Increase digital literacy efforts** for older customers to encourage **online banking usage**.
- Use **personalized marketing strategies** targeting the **30-40-year segment**, as they are more likely to take new financial products.

EDA Results

Balance Distribution Analysis



Observations:

- The **boxplot** shows a heavily skewed distribution with a **long tail on the right**, indicating **high balance values for some customers**.
- A **large proportion of customers have a balance close to zero**, suggesting a significant number of inactive or low-transaction customers.
- The **histogram** confirms the **bimodal distribution**, with many customers at **zero balance** and another peak around **100,000+**.

Key Insights:

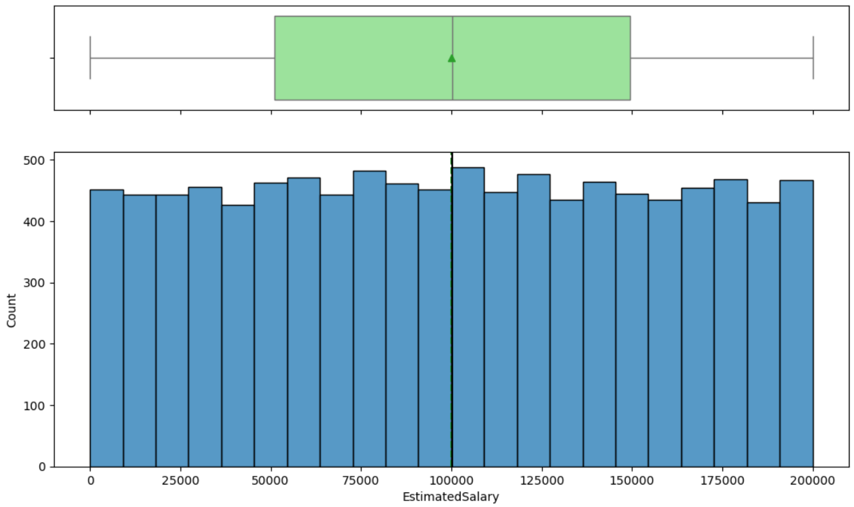
- A large **portion of customers maintain a zero balance**, which might indicate **dormant or inactive accounts**.
- The **high-balance customers** (those with 100K+) may be **premium clients**, requiring different retention strategies.
- **A gap in the mid-range (20K-50K)** suggests potential for **cross-selling** financial products to these customers.

Business Recommendations:

- **Encourage dormant customers** to use their accounts through **incentives like cashback or lower fees**.
- **Identify high-balance customers** for **premium banking services or investment offerings**.
- **Offer better interest rates** on savings accounts to attract mid-range balance customers.

EDA Results

Estimated Salary Analysis



Observations:

- The **boxplot** suggests a **fairly uniform salary distribution** with no significant skewness.
- The **histogram** indicates a near **uniform distribution**, suggesting **salary does not play a dominant role in customer segmentation**.
- The **median salary is approximately 100K**, and customers are spread across different income levels.

Key Insights:

- Unlike other variables, **salary is evenly distributed**, meaning customer behavior is not **strongly correlated with income**.
- Since high-income and low-income customers are evenly distributed, **financial product adoption may depend on factors other than salary**.

Business Recommendations:

- Design **income-independent financial services**, such as flexible loan offerings or **customized credit limits**.
- Offer **tiered benefits programs** to appeal to different salary segments.
- **Leverage behavioral insights** rather than income when segmenting customers for financial promotions.

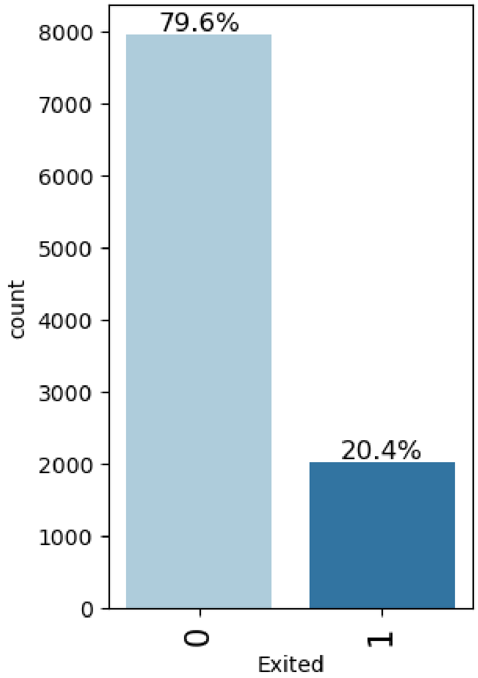
EDA Final Results - Univariant

Final Summary & Strategic Recommendations

- **High-risk groups (low credit score, zero balance accounts) should be closely monitored** for retention programs.
- **Middle-aged customers (30-40 years) are key for financial growth**, requiring targeted engagement strategies.
- **Premium and high-balance customers should be nurtured** with exclusive services and personalized offers.
- **Salary has little impact on customer segmentation**, so strategies should focus on spending behavior instead.

*This analysis helps the bank **optimize marketing strategies, reduce churn, and maximize customer lifetime value**.*

EDA Results



Churn Rate (Exited Customers)

Observations:

- **79.6% of customers remain with the bank**, while **20.4% have exited**.
- This indicates that **customer churn is relatively low** but still significant enough to warrant intervention.

Key Insights:

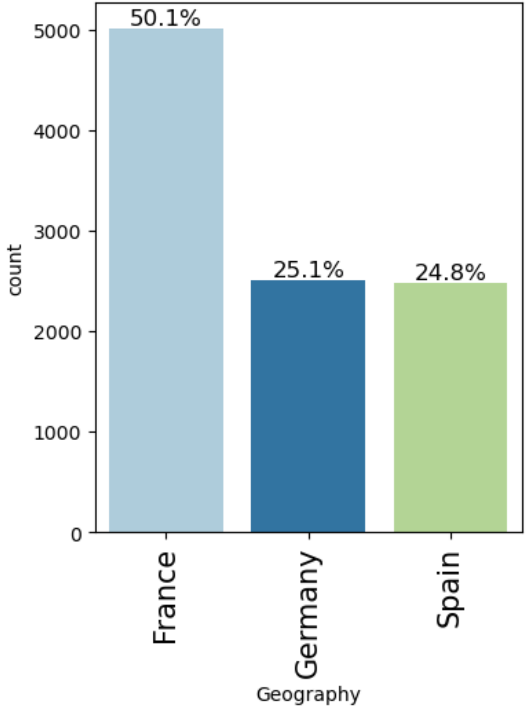
- A **20.4% churn rate** suggests **potential issues** in customer retention.
- The dataset is **imbalanced**, meaning churned customers represent a smaller portion, which could affect predictive modeling.

Business Recommendations:

- Implement **customer retention programs**, such as **loyalty rewards** or **personalized financial services**.
- Use **predictive analytics** to **identify high-risk customers** and proactively offer tailored solutions.
- Address **common reasons for churn**, such as dissatisfaction with services or competition from other banks.

EDA Results

Customer Geography Distribution



Observations:

- **50.1% of customers are from France**, making it the **largest market**.
- Germany (25.1%) and Spain (24.8%) have **similar-sized customer bases**.

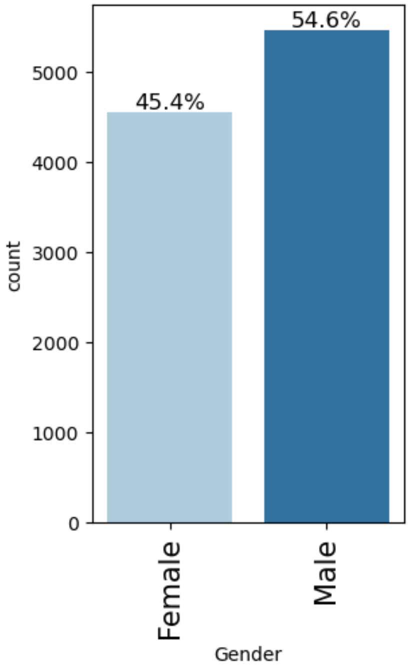
Key Insights:

- The **French market dominates**, so **churn reduction efforts** should **focus on France** first.
- Germany and Spain have **almost equal representation**, meaning any **country-based strategies** should be customized to regional differences.

Business Recommendations:

- Investigate **churn trends per country** to determine **why customers leave** in each region.
- Launch **country-specific marketing campaigns** tailored to customer needs in **Germany and Spain**.
- Expand the **customer base in underrepresented regions** by **targeting new potential clients**.

EDA Results



Gender Distribution

Observations:

- **54.6% of customers are male**, while **45.4% are female**.
- The gender ratio is fairly balanced, but males have a **slight majority**.

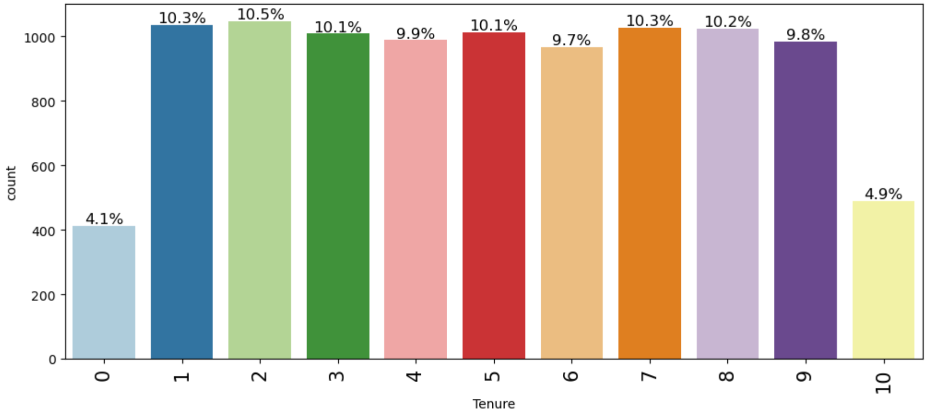
Key Insights:

- There are **no major gender-based disparities**, meaning **churn risk is not significantly gender-dependent**.
- **Retention strategies should focus on other variables** (such as account activity) rather than gender.

Business Recommendations:

- Conduct **customer segmentation** based on behavior rather than gender.
- Offer **personalized services** that cater to both **male and female banking preferences**.
- Evaluate if gender plays a role in **product preferences** to create **more targeted financial offers**.

EDA Results



Customer Tenure Distribution

Observations:

- Most customers have a tenure between **1-9 years**, with **each group making up around 10% of customers**.
- Tenure **0 (new customers) and 10 years (long-term customers) have lower percentages (4.1% and 4.9%, respectively)**.

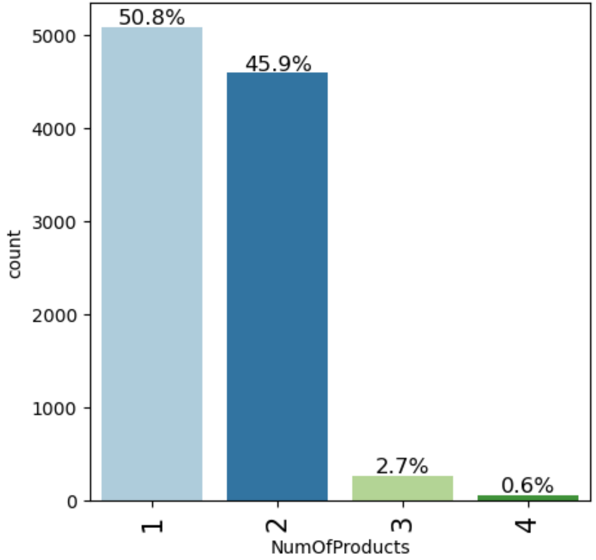
Key Insights:

- **Churn risk is likely highest in new customers (tenure = 0-1 years)** since they might be **exploring other banking options**.
- Long-term customers (**tenure = 10 years**) are **more stable** but may still need **incentives to stay**.

Business Recommendations:

- Develop **customer onboarding strategies** to **increase engagement among new customers**.
- Provide **loyalty benefits** for long-term customers, such as **higher interest rates on savings accounts or exclusive banking privileges**.
- Identify **why customers with tenure 1-9 years leave** and address their concerns with **targeted customer service improvements**.

EDA Results



Number of Products Per Customer

Observations:

- Most customers have **1 (50.8%) or 2 (45.9%) banking products**.
- **Very few customers have 3 (2.7%) or 4 (0.6%) products**.

Key Insights:

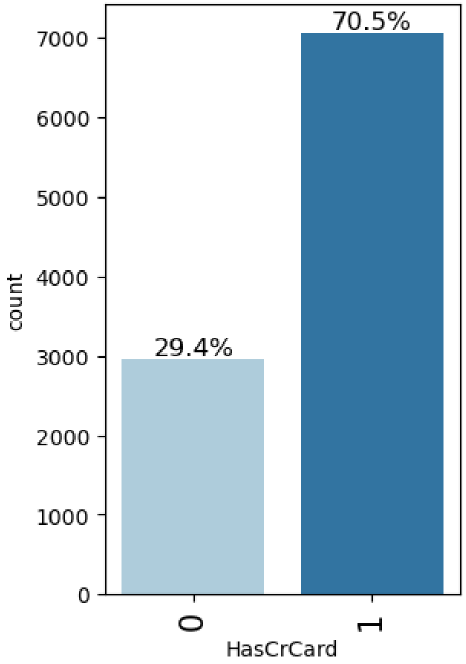
- Customers with **only one product** may be **more likely to churn** due to lack of engagement.
- **Upselling opportunities exist**, as very few customers use **three or more products**.

Business Recommendations:

- **Cross-sell and upsell additional banking products** to customers with **only one product**.
- Offer **bundled financial products**, such as **credit cards, investment accounts, or insurance** to encourage multi-product adoption.
- Analyze **customer behavior for those with 3 or more products** to understand what **motivates deeper engagement**.

EDA Results

Credit Card Ownership



Observations:

- **70.5% of customers have a credit card**, while **29.4% do not**.
- A significant percentage of customers **do not use the bank's credit card services**.

Key Insights:

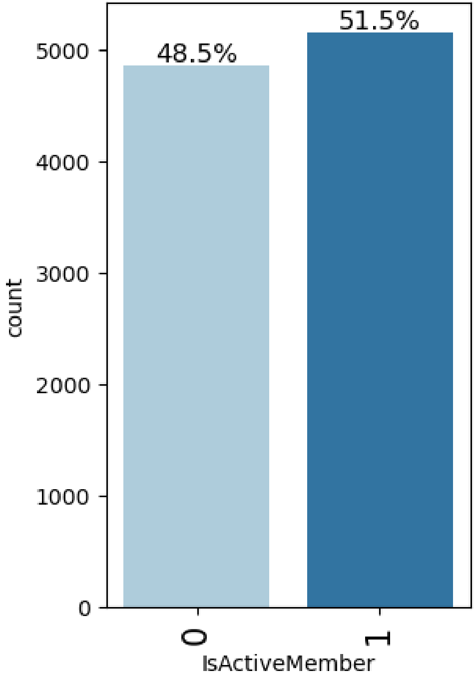
- **Credit card ownership does not guarantee customer retention**—some may still leave despite using credit services.
- **A large portion (29.4%) of customers may be missing out on benefits** offered by the bank's credit card.

Business Recommendations:

- Launch a **targeted credit card campaign** for the **29.4% of customers without one**.
- Offer **rewards-based incentives**, such as **cashback or lower interest rates**, to encourage adoption.
- Improve **credit card benefits** and **customer education programs** to ensure more customers see the **value of having a credit card**.

EDA Results

Active vs. Inactive Members



Observations:

- **51.5% of customers are active**, while **48.5% are inactive**.
- This suggests that **almost half of the customers have limited interactions with the bank**.

Key Insights:

- **Inactive members are at higher risk of churn**, meaning engagement strategies need improvement.
- Even though the active member percentage is slightly higher, a **large percentage of customers (48.5%) remain inactive**.

Business Recommendations:

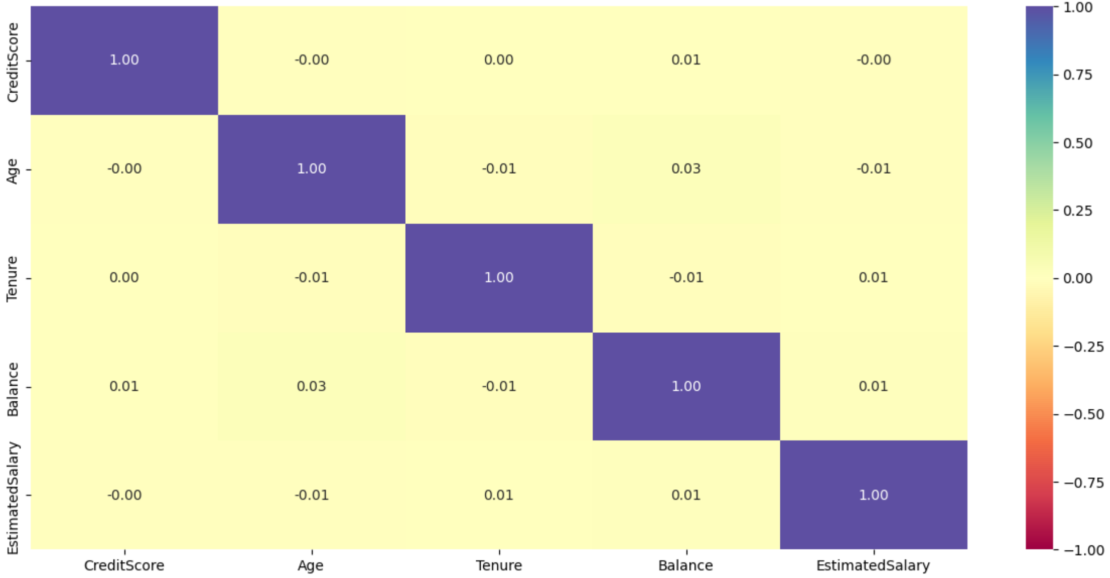
- Implement **customer engagement programs**, such as **personalized financial advice or automated banking reminders**.
- Offer **incentives for active engagement**, such as **discounted loan rates or reward points for frequent usage**.
- Conduct **surveys and outreach efforts** to understand **why inactive members are not engaging** with banking services.

EDA Results – Final Summary & Recommendations

FINAL SUMMARY & STRATEGIC RECOMMENDATIONS

- **High churn risk exists among new customers and inactive members**—retention programs should prioritize them.
- **Geographical differences matter**—country-specific strategies can help optimize retention.
- **Upselling and cross-selling banking products can increase engagement and reduce churn**.
- **Credit card adoption is an opportunity** to strengthen customer relationships.
- **A large inactive customer base suggests engagement programs need to be enhanced**.
- By **analyzing and acting on these insights**, the bank can **increase customer loyalty, reduce churn, boost revenue**.

EDA Results



This heatmap displays the correlation between numerical variables **(CreditScore, Age, Tenure, Balance, and EstimatedSalary)** in the dataset.

The correlation values range from **-1 to 1**, where:

- **1.0** indicates a **perfect positive correlation**.
- **-1.0** indicates a **perfect negative correlation**.
- **0** indicates **no correlation** between the variables.

EDA Results

Heat Map Analysis

Key Observations:

1. **Extremely Weak Correlations Across Features:**
   - Most of the correlation values are **close to 0**, indicating that **these variables do not have strong linear relationships** with one another.
   - This suggests that these financial and demographic attributes are **largely independent** in this dataset.

2. **Credit Score vs. Other Variables:**
   - **CreditScore shows near-zero correlation with Age, Tenure, Balance, and EstimatedSalary**.
   - This suggests that **credit scores are independent of income, tenure, and balance levels**, meaning a **higher salary or longer tenure does not necessarily result in a higher credit score**.

3. **Balance vs. Other Variables:**
   - The **Balance variable has weak correlations (close to 0.01-0.03) with all other factors**.
   - This indicates that **account balance does not significantly depend on age, tenure, or salary**, meaning people across all salary levels maintain diverse account balances.

4. **Age and Tenure Relationship:**
   - **Age and Tenure have a weak negative correlation (-0.01)**, suggesting **no strong relationship between a customer's age and how long they have been with the bank**.
   - This means that **both younger and older customers can have varied tenure durations** with the bank.

*This analysis helps the bank **optimize marketing strategies, reduce churn, and maximize customer lifetime value**.*

EDA Results

Heat Map Analysis

Key Insights:

1. **Linear Models May Not Be Effective**
   - The lack of strong correlations suggests that **linear models like Logistic Regression may not be the best choice** for predictive modeling.
   - **More complex models** such as **Decision Trees, Random Forests, or Neural Networks** may be more effective in capturing non-linear patterns.

2. **Feature Engineering Needed for Churn Prediction**
   - Since numerical attributes alone do not provide strong predictive power, incorporating **behavioral data** (e.g., transaction frequency, product usage, service interactions) may improve churn prediction accuracy.
   - **Categorical features** such as **customer type, account type, or product ownership** may play a bigger role than the numerical variables shown.

3. **Customer Segmentation Strategies Must Consider Additional Data**
   - Since **Age, Tenure, and Salary do not show strong relationships**, segmentation based on these alone may not be effective.
   - Segmentation should focus on **transaction activity, engagement level, and spending patterns** instead.

*This analysis helps the bank **optimize marketing strategies, reduce churn, and maximize customer lifetime value**.*

EDA Results

Heat Map Analysis

Business Recommendations:

1. **Enhance Predictive Models with More Features**
   - Introduce **behavioral metrics** such as **transaction frequency, spending trends, and credit utilization**.
   - Explore using **machine learning techniques** that detect **non-linear relationships** (e.g., Gradient Boosting, Neural Networks).

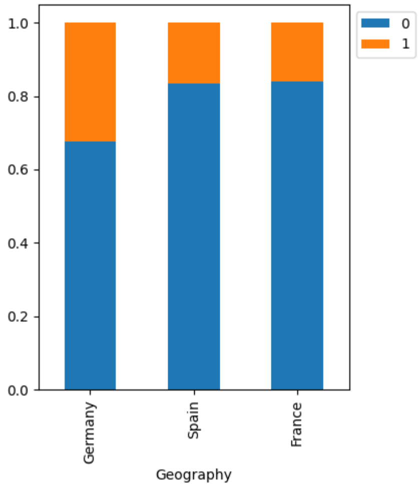
2. **Revise Customer Retention Strategies**
   - Since **tenure and age do not correlate**, **targeting long-tenured customers alone is not effective**.
   - Instead, focus on **engagement-based retention programs** where customers with declining activity receive **incentives** or **personalized offers**.

3. **Investigate External Factors Affecting Churn & Financial Health**
   - Since **income (EstimatedSalary) does not correlate with balance or credit score**, there may be external factors influencing these metrics.
   - Consider integrating **economic indicators**, **credit history trends**, or **customer feedback surveys** to understand why balances and credit scores vary widely.

*This analysis helps the bank **optimize marketing strategies, reduce churn, and maximize customer lifetime value**.*

EDA Results

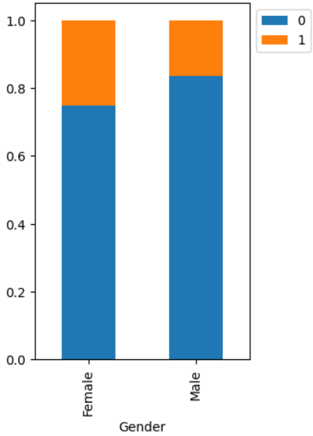
Exited vs Geography



Observation: Customers in Germany show a significant higher churn rate, nearly 35%, as compared to those in Spain and France, where the exit rate is near below 20%

Insight: High churn rate in Germany suggest customer dissatisfaction, competitive offers, or service issues.

Exited vs Gender

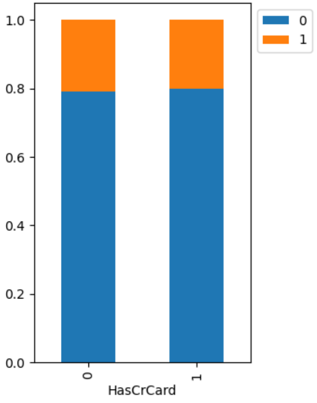


Observation: Females have a slightly higher churn rate compared to males.
Both genders have more retained customers than churned customers

Insight: Female customers are slightly more likely to churn, women may have different banking expectations or experiences.

EDA Results

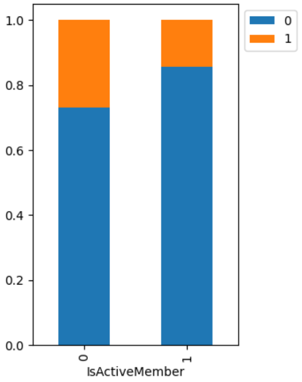
Exited vs Has Credit Card



Observation: Since the churn **rate is nearly the same for both groups, having a credit card does not heavily influence a customer's decision** to leave the bank

Insight: Customers who actively use their credit cards may behave differently from those who hold one but rarely use it.

Exited vs Is Active Member

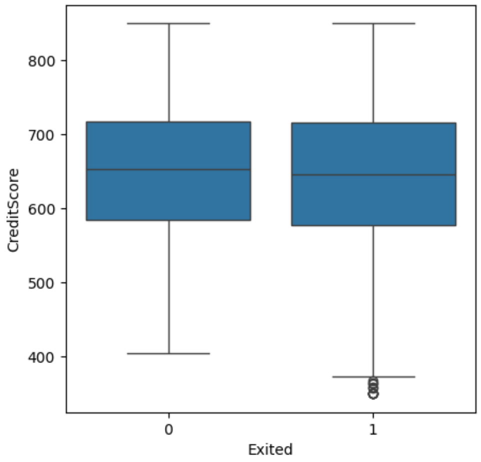


Observation: **Inactive members** have a significantly higher churn rate compared to active members. **Active members** show a lower churn rate compared to active members.

Insight: Customers who frequently use **banking services, engage in transactions, or participate in loyalty program** are more likely to remain loyal.

EDA Results

Exited vs Credit Score



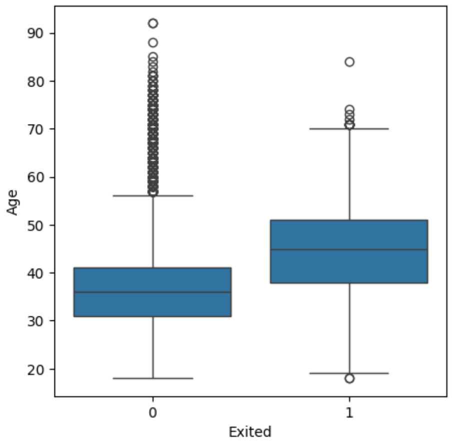
Observation: Customers with extremely low credit scores (below 400) are outliers among those who exited.

Insight: Aside from these outliers, the credit score distribution is similar between those who stayed and those who left.

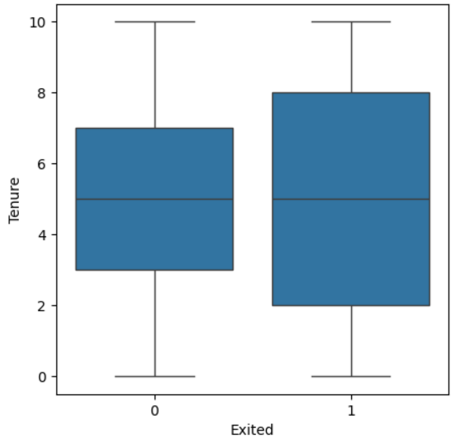
Observation: Exited customers tend to be older, with a higher median age compared to those who stayed.

Insight: :Younger customers appear to remain more engaged, while older customers may require targeted retention strategies

Exited vs Age



Exited vs Tenure

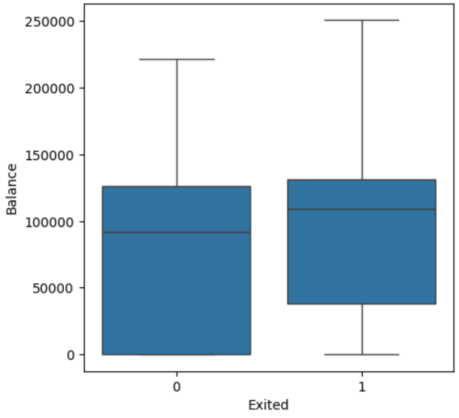


Observation: Customers inactive for longer periods (3+ months) show a higher likelihood of churning.

Insight: Inactivity is a strong predictor of attrition.

EDA Results

Exited vs Balance



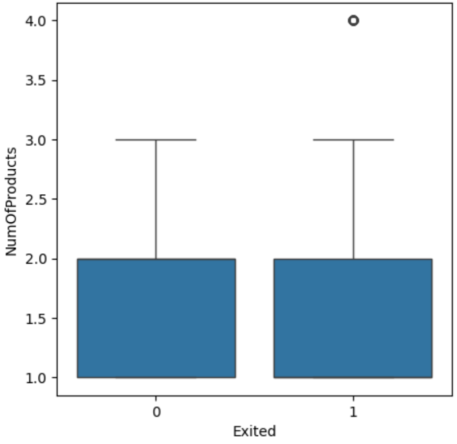
Observation: Exited accounts have a **higher balance**, reaching **up to $250,000**, compared to those that remained.

Insight: This trend suggests a **possible transition to financial services** that better align with specific economic levels

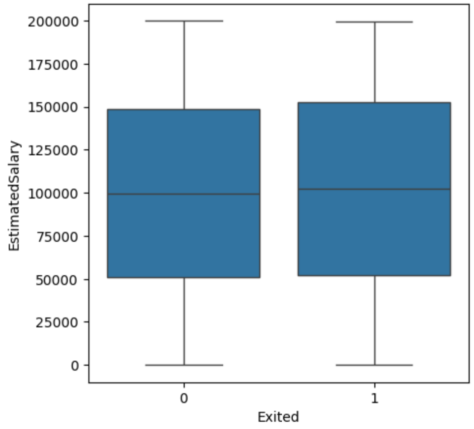
Observation: Customers with **a lower number of products (1-2) are more likely to churn** compared to those with higher relationship counts.

Insight: The **number of products** does not strongly influence churn.

Exited vs Number of Products

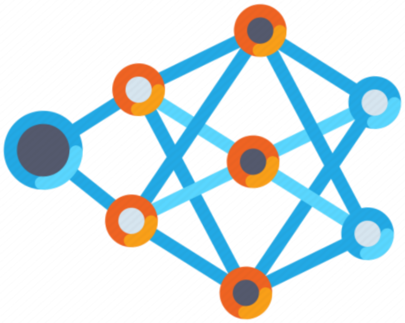


Exited vs Estimated Salary



Observation: The estimated salary exhibits **no substantial difference between retained and exited customers**, aligning with the previously observed low correlation.

Insight: **No clear relationship** between salary and Churn.



NEURAL NETWORKING

Data Pre-processing

- Duplicate Value Check
- Missing Value Treatment
- Outlier Check (Treatment if needed)
- Feature Engineering
- Data Preparation for Modeling

Data Quality

Data Quality Issue	Observation	Action Taken
Missing Values	No missing values detected	No imputation required
Inconsistent Formatting	Categorical data contained mixed cases and incorrect values	Standardized text formatting and replaced erroneous entries
Outliers	High values in credit limits and transaction amounts	Retained outliers and used robust models
Class Imbalance	More non-churned customers than churned customers	Applied SMOTE and under-sampling techniques
Feature Engineering	Skewed numerical data and categorical variables	Applied transformations and encoding

Data Preprocessing-Model Building

- During the Data Pre-Processing phase, the dataset was assessed for quality, revealing **no duplicate rows or missing values**, thereby ensuring data integrity.

- **Outliers** were identified but retained due to their potential to provide valuable insights into the target variable, "Exited."

- In the Feature Engineering stage, columns that did not contribute to the model's predictive power—**'Row Number,' 'Customer ID,' and 'Surname'**—were **removed**, as they contained unique identifiers with no analytical relevance.

- The feature matrix x was then constructed by excluding the 'Exited' column, which was designated as the target variable y.

Data Preprocessing-Model Building

- After isolating the feature matrix X and target variable y, the dataset was split into a larger training set (X_large, y_large) and a testing set (X_test, y_test) using an 80-20 ratio. Stratified sampling was applied to maintain a consistent distribution of the target variable, "Exited," across both sets.

- Next, X_large and y_large were further divided into a final training set (X_train, y_train) and a validation set (X_val, y_val) using the same 80-20 ratio, ensuring adherence to stratification and random state principles.

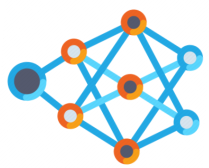
- For feature encoding, one-hot encoding was applied to the **'Geography'** and **'Gender'** columns. The parameter drop_first=True was used to prevent the "dummy variable trap," which could introduce multicollinearity into the dataset.

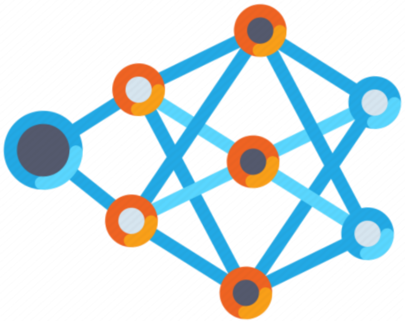
Data Preprocessing-Model Building

- Feature scaling was applied using the **StandardScaler** on numerical columns, **including 'CreditScore,' 'Age,' 'Tenure,' 'Balance,' and 'EstimatedSalary.'**

- This process normalized the range of these features, ensuring compatibility with algorithms sensitive to input variable scales.

- The scaler was fitted exclusively on the training set to prevent data leakage, and the same scaling parameters were subsequently applied to the test and validation sets.





MODEL BUILDING

Neural Network

MODEL 0 –with SGD Optimizer

Model Performance Summary: Classifier Model

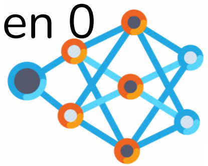
- The input layer consists of **64 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.

- The input dimension is set to **11**, corresponding to the number of selected features.

- The first hidden layer contains **32 neurons**, utilizing **ReLU** as the activation function.

- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.

- A **sigmoid** activation function is applied to constrain the output between 0 and 1, making it suitable for binary classification..



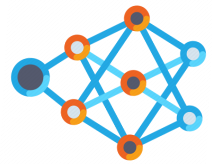
Model Performance Summary: Classifier Model

- **Optimizer:** The **Stochastic Gradient Descent (SGD)** optimizer is used to adjust the weights in the neural network, minimizing the loss function.

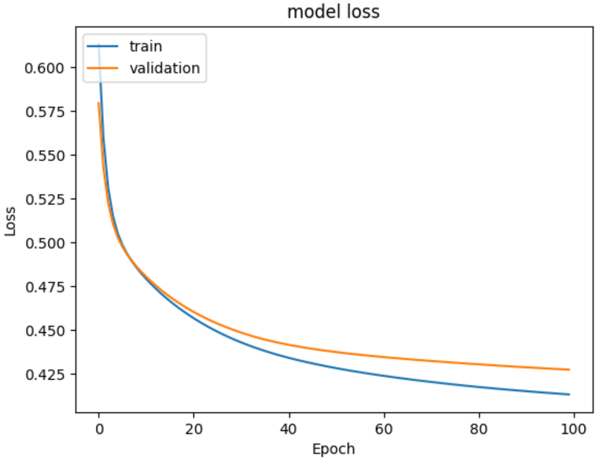
- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.

- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.

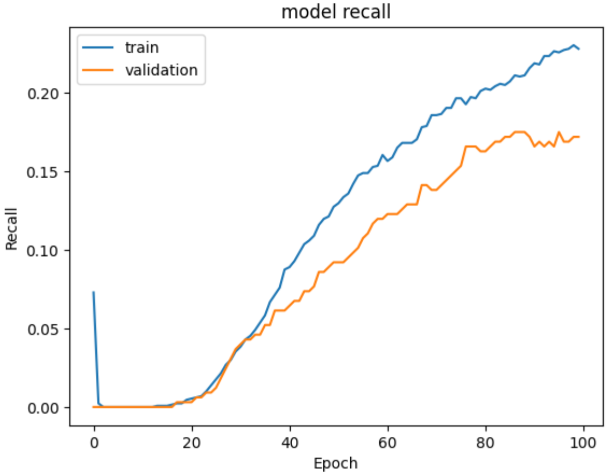
- **Training:** The model is trained for **100 epochs**, providing sufficient iterations for convergence. A **batch size** of **32** is selected, offering a balance between computational efficiency and accuracy in weight updates.



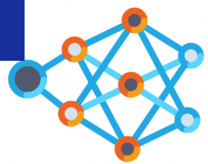
Model Performance Summary: Classifier Model



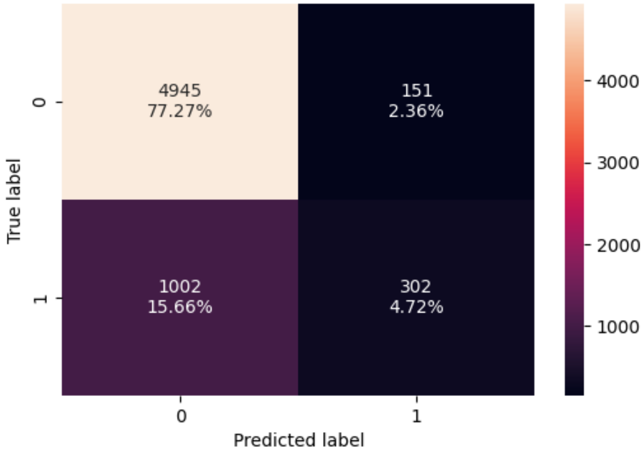
The model is learning effectively with a smooth and stable loss decrease. However, slight overfitting is occurring, which could be mitigated with regularization techniques or early stopping.

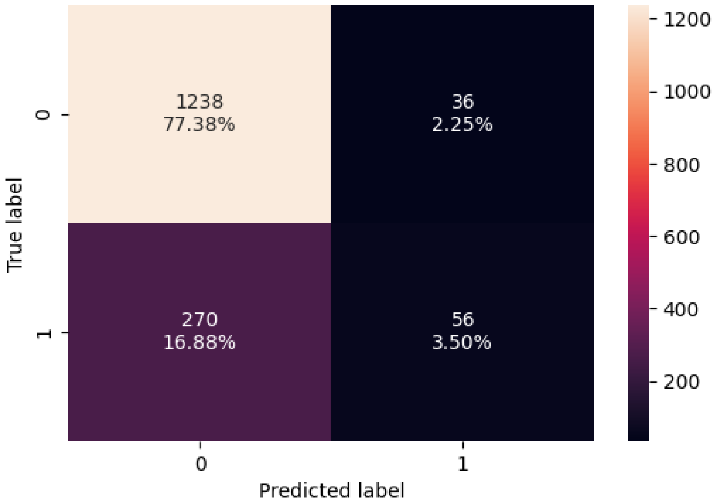


The model is learning well initially but starts overfitting after about 40 epochs. Early stopping and regularization techniques could help improve generalization.



Model Performance Summary: Classifier Model





Model Performance Summary: Classifier Model

	precision	recall	f1-score	support
0.0	0.83	0.97	0.90	5096
1.0	0.67	0.23	0.34	1304
accuracy			0.82	6400
macro avg	0.75	0.60	0.62	6400
weighted avg	0.80	0.82	0.78	6400

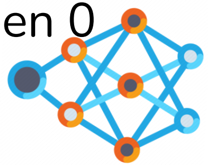
	precision	recall	f1-score	support
0.0	0.82	0.97	0.89	1274
1.0	0.61	0.17	0.27	326
accuracy			0.81	1600
macro avg	0.71	0.57	0.58	1600
weighted avg	0.78	0.81	0.76	1600

Neural Network

MODEL 1 –with ADAM Optimizer

Model 1: ADAM Optimizer

- The input layer consists of **64 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.
- The input dimension is set to **11**, corresponding to the number of selected features.
- The first hidden layer contains **32 neurons**, utilizing **ReLU** as the activation function.
- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.
- A **sigmoid** activation function is applied to constrain the output between 0 and 1, making it suitable for binary classification..



Model 1: ADAM Optimizer

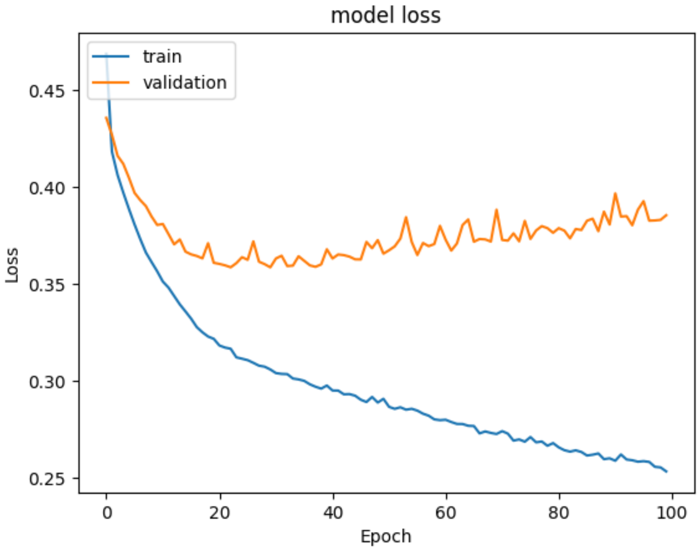
- **Optimizer:** The ADAM optimizer is used to adjust the weights in the neural network, minimizing the loss function.

- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.

- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.

- **Training:** The model is trained for **100 epochs**, providing sufficient iterations for convergence. A **batch size** of **64** is selected, offering a balance between computational efficiency and accuracy in weight updates.

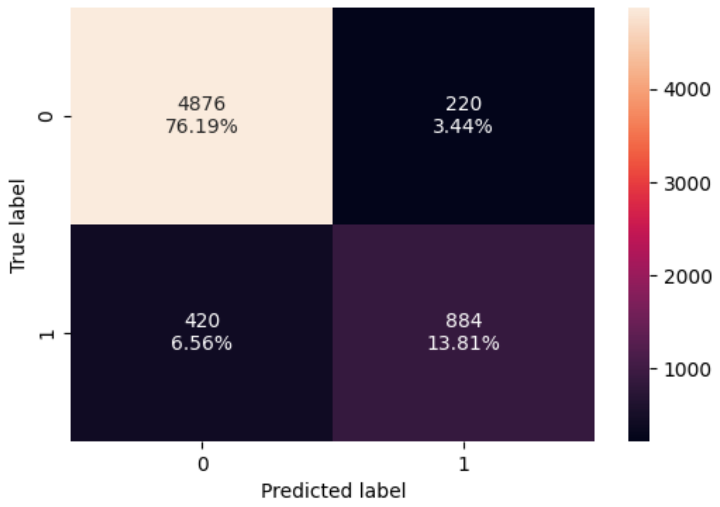
Model 1: ADAM Optimizer

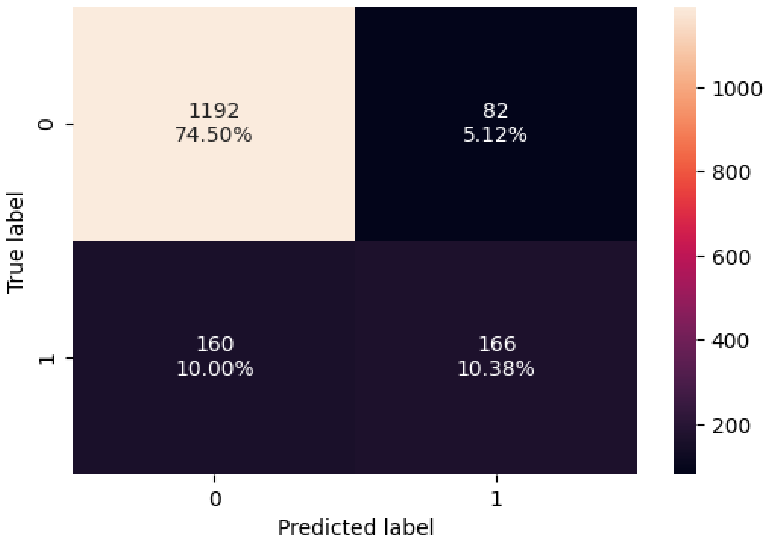


This model is **overfitting** after around **20 epochs**. Applying **early stopping, regularization, and hyperparameter tuning** can improve generalization and reduce validation loss fluctuations.

The model is **overfitting** after **20–30 epochs**, and the **validation recall is unstable**, meaning it does not generalize well. To improve performance, apply **early stopping, regularization, and data balancing techniques** while tuning hyperparameters.

Model 1: ADAM Optimizer

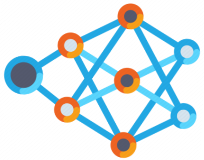




Model 1: ADAM Optimizer

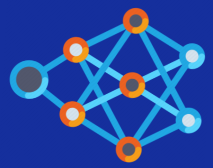
	precision	recall	f1-score	support
0.0	0.92	0.96	0.94	5096
1.0	0.80	0.68	0.73	1304
accuracy			0.90	6400
macro avg	0.86	0.82	0.84	6400
weighted avg	0.90	0.90	0.90	6400

	precision	recall	f1-score	support
0.0	0.88	0.94	0.91	1274
1.0	0.67	0.51	0.58	326
accuracy			0.85	1600
macro avg	0.78	0.72	0.74	1600
weighted avg	0.84	0.85	0.84	1600



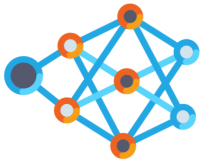
Neural Network

MODEL 2 –with ADAM Optimizer & Dropout



Model 2: ADAM Optimizer & Dropout

- The input layer consists of **32 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.
- The dropout layer is at **0.2** which prevent overfitting.
- The first hidden layer contains **16 neurons**, utilizing **ReLU** as the activation function.
- The second hidden layer contains **8 neurons**, utilizing **ReLU** as the activation.
- The dropout layer is at **0.1** which prevent overfitting.
- The third hidden layer contains **4 neurons**, utilizing **ReLU** as the activation.
- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.
- A **sigmoid** activation function is applied to constrain the output between 0 and **1**, making it suitable for binary classification..



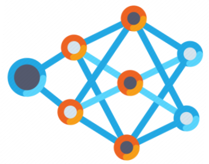
Model 2: ADAM Optimizer & Dropout

- **Optimizer:** The ADAM optimizer is used to adjust the weights in the neural network, minimizing the loss function.

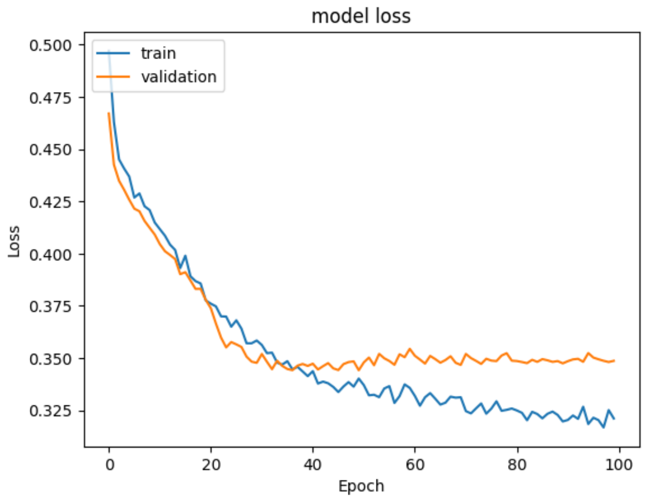
- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.

- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.

- **Training:** The model is trained for 100 epochs, providing sufficient iterations for convergence. A batch size of 64 is selected, offering a balance between computational efficiency and accuracy in weight updates.

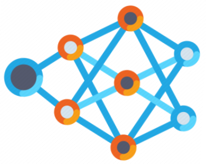


Model 2: ADAM Optimizer & Dropout

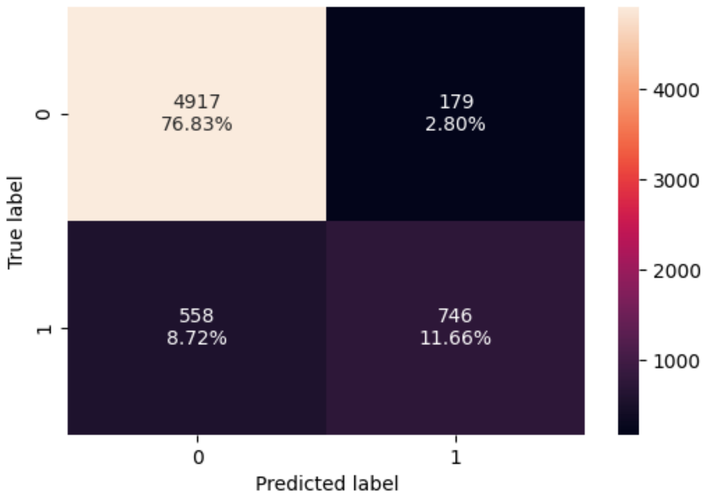


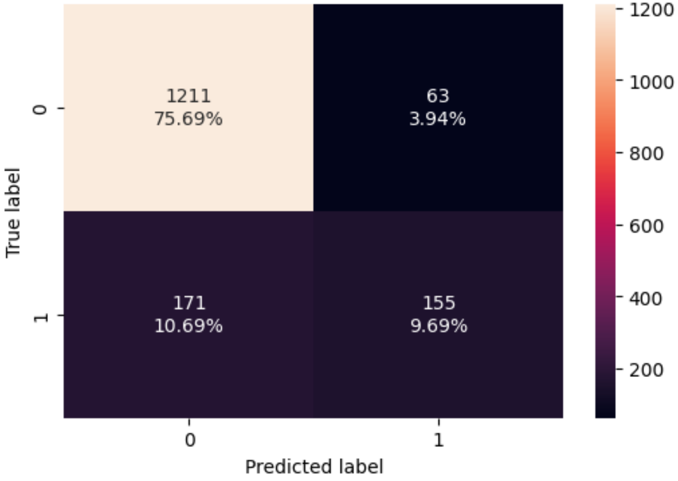
The model learns well initially but starts **overfitting after 30–40 epochs**. **Early stopping, regularization, and hyperparameter tuning** can help improve generalization.

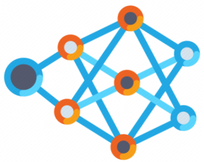
The **model is performing well with minimal overfitting**.
Validation recall is slightly noisy, but it remains close to training recall, which suggests good generalization.
Minor tuning (regularization, data balancing, or learning rate adjustments) may help improve stability.



Model 2: Adam Optimizer & Dropout







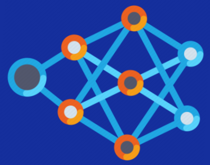
Model 2: Adam Optimizer & Dropout

	precision	recall	f1-score	support
0.0	0.90	0.96	0.93	5096
1.0	0.81	0.57	0.67	1304
accuracy			0.88	6400
macro avg	0.85	0.77	0.80	6400
weighted avg	0.88	0.88	0.88	6400

	precision	recall	f1-score	support
0.0	0.88	0.95	0.91	1274
1.0	0.71	0.48	0.57	326
accuracy			0.85	1600
macro avg	0.79	0.71	0.74	1600
weighted avg	0.84	0.85	0.84	1600

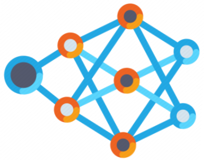
Neural Network

MODEL 3 –with Balanced Data (SMOTE) & SGD Optimizer



Model 3-(SMOTE) & SGD Optimizer

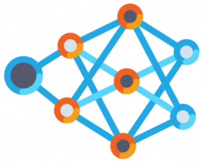
- The input layer consists of **64 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.
- The input dimension is set to **11**, corresponding to the number of selected features.
- The first hidden layer contains **32 neurons**, utilizing **ReLU** as the activation function.
- The second hidden layer contains **16 neurons**, utilizing **ReLU** as the activation function.
- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.
- A **sigmoid** activation function is applied to constrain the output between 0 and **1**, making it suitable for binary classification..



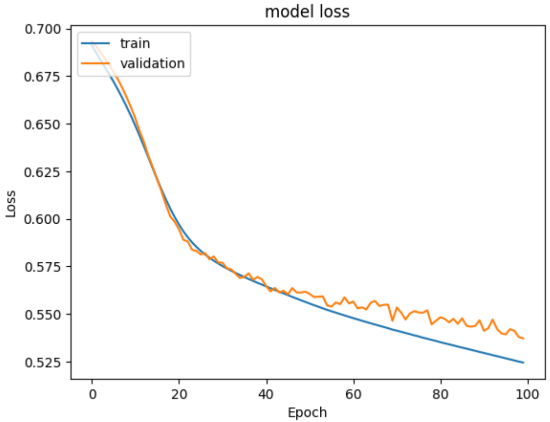
Model 3-(SMOTE) & SGD Optimizer

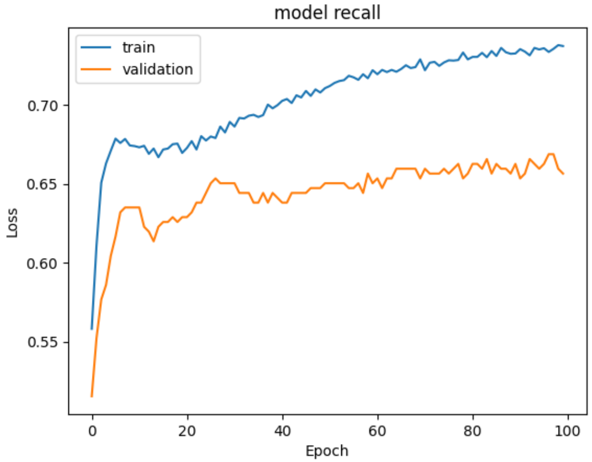
Great Learning
POWER AHEAD

- **Optimizer:** The **Stochastic Gradient Descent (SGD)** optimizer is used to adjust the weights in the neural network, minimizing the loss function.
- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.
- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.
- **Training:** The model is trained for **100 epochs**, providing sufficient iterations for convergence. A **batch size** of **32** is selected, offering a balance between computational efficiency and accuracy in weight updates.



Model 3-(SMOTE) & SGD Optimizer

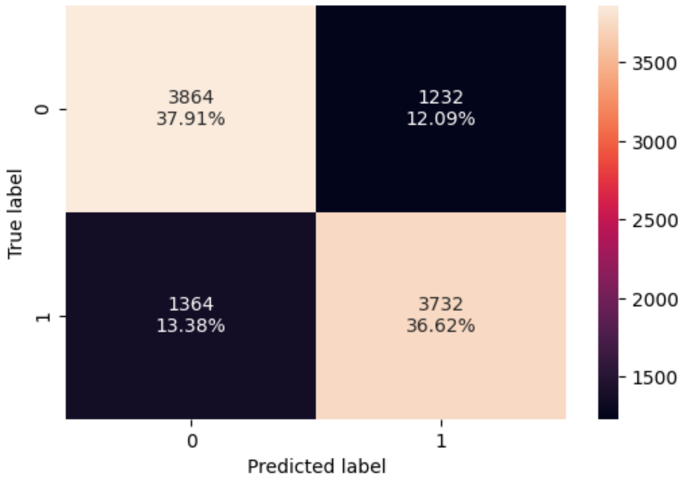


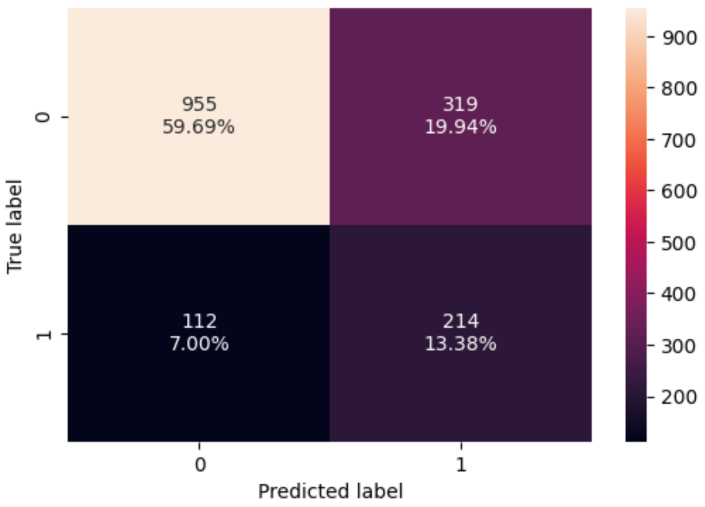


The model is learning well with minimal overfitting.
Validation and training loss remain closely aligned, indicating strong generalization.
Minor validation loss fluctuations are expected but do not significantly impact performance.

The model is overfitting after 30–40 epochs.
Validation recall plateaus while training recall continues improving, indicating limited generalization.
Applying early stopping, regularization, and data augmentation may help improve validation performance.

Model 3-(SMOTE) & SGD Optimizer





Model 3-(SMOTE) & SGD Optimizer

	precision	recall	f1-score	support
0	0.74	0.76	0.75	5096
1	0.75	0.73	0.74	5096
accuracy			0.75	10192
macro avg	0.75	0.75	0.75	10192
weighted avg	0.75	0.75	0.75	10192

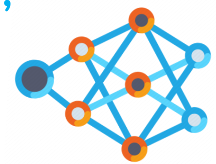
	precision	recall	f1-score	support
0.0	0.90	0.75	0.82	1274
1.0	0.40	0.66	0.50	326
accuracy			0.73	1600
macro avg	0.65	0.70	0.66	1600
weighted avg	0.79	0.73	0.75	1600

Neural Network

MODEL 4-with Balanced Data (SMOTE) and Adam Optimizer

Model 4: (SMOTE) and Adam Optimizer

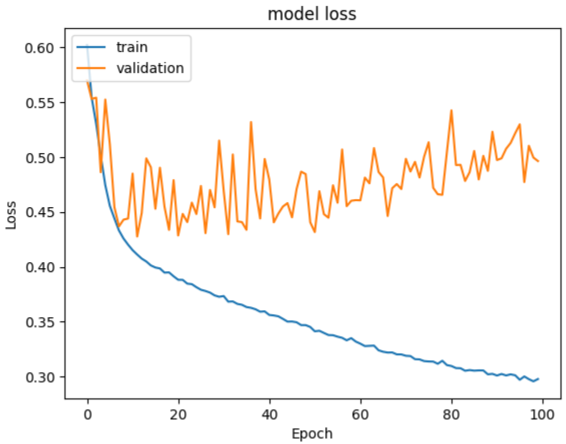
- The input layer consists of **32 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.
- The input dimension is set to **11**, corresponding to the number of selected features.
- The first hidden layer contains **16 neurons**, utilizing **ReLU** as the activation function.
- The second hidden layer contains **8 neurons**, utilizing **ReLU** as the activation function.
- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.
- A **sigmoid** activation function is applied to constrain the output between 0 and **1,** making it suitable for binary classification..

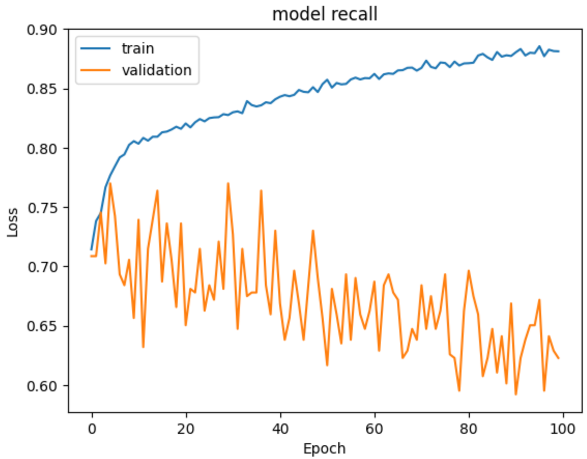


Model 4: (SMOTE) and Adam Optimizer

- **Optimizer:** The Adam optimizer is used to adjust the weights in the neural network, minimizing the loss function.
- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.
- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.
- **Training:** The model is trained for 100 epochs, providing sufficient iterations for convergence. A batch size of 32 is selected, offering a balance between computational efficiency and accuracy in weight updates.

Model 4: (SMOTE) and Adam Optimizer

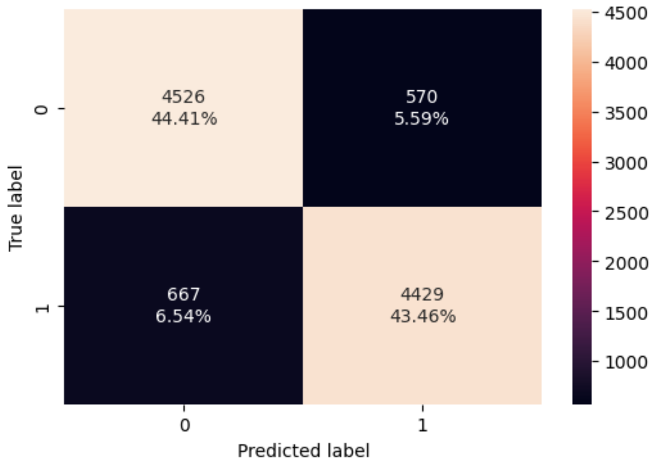


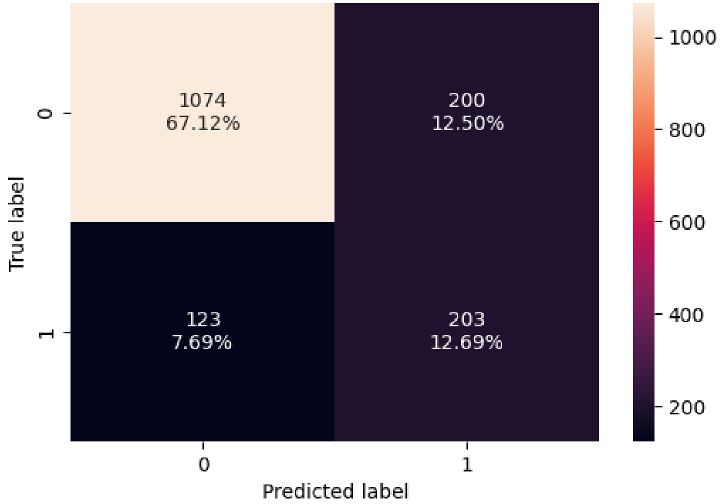


The model is severely overfitting, as indicated by the decreasing training loss and highly unstable validation loss. Validation loss fluctuations suggest poor generalization, likely due to noisy data or insufficient regularization.

The model is overfitting, as training recall improves while validation recall remains unstable. Validation recall fluctuations indicate high variance, possibly due to imbalanced or noisy data.

Model 4: (SMOTE) and Adam Optimizer





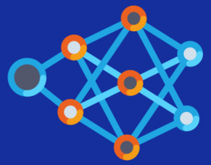
Model 4: (SMOTE) and Adam Optimizer

	precision	recall	f1-score	support
0	0.87	0.89	0.88	5096
1	0.89	0.87	0.88	5096
accuracy			0.88	10192
macro avg	0.88	0.88	0.88	10192
weighted avg	0.88	0.88	0.88	10192

	precision	recall	f1-score	support
0.0	0.90	0.84	0.87	1274
1.0	0.50	0.62	0.56	326
accuracy			0.80	1600
macro avg	0.70	0.73	0.71	1600
weighted avg	0.82	0.80	0.81	1600

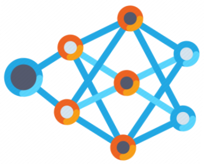
Neural Network

MODEL 5 –with Balanced Data (SMOTE), Adam Optimizer, and Dropout



Model 5-(SMOTE), Adam Optimizer, and Dropout

- The input layer consists of **64 neurons** with **ReLU** (Rectified Linear Unit) as the activation function.
- The dropout layer is at **0.2** which prevent overfitting.
- The first hidden layer contains **32 neurons**, utilizing **ReLU** as the activation function.
- The dropout layer is at **0.1** which prevent overfitting.
- The third hidden layer contains **8 neurons**, utilizing **ReLU** as the activation.
- The output layer consists of a single node, designed for binary classification to determine whether a customer will churn.
- A **sigmoid** activation function is applied to constrain the output between 0 and **1**, making it suitable for binary classification..



Model 5-(SMOTE), ADAM Optimizer, and Dropout

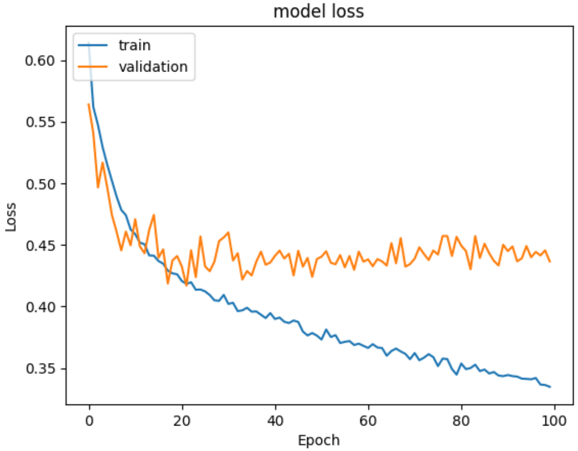
- **Optimizer:** The ADAM optimizer is used to adjust the weights in the neural network, minimizing the loss function.

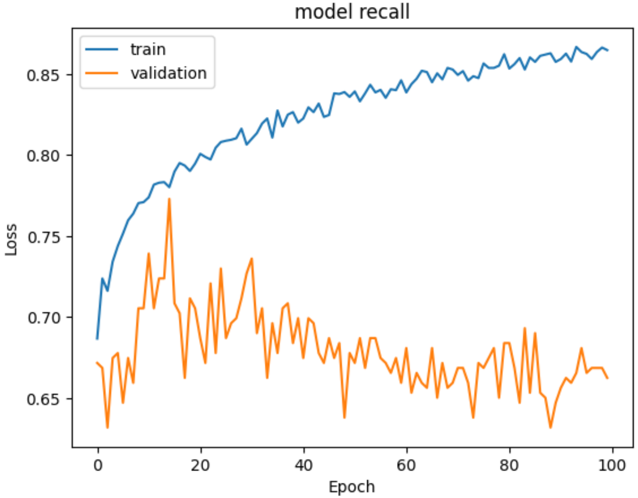
- **Loss Function:** As a binary classification problem, the appropriate loss function is Binary Cross_Entropy.

- **Metrics:** Accuracy is tracked as the primary evaluation metric, which is commonly used for classification tasks.

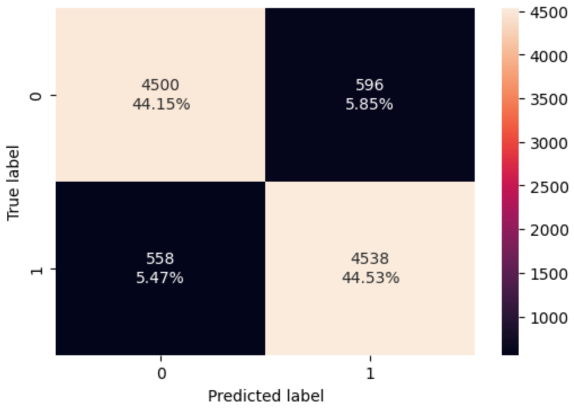
- **Training:** The model is trained for **100 epochs**, providing sufficient iterations for convergence. A **batch size** of **64** is selected, offering a balance between computational efficiency and accuracy in weight updates.

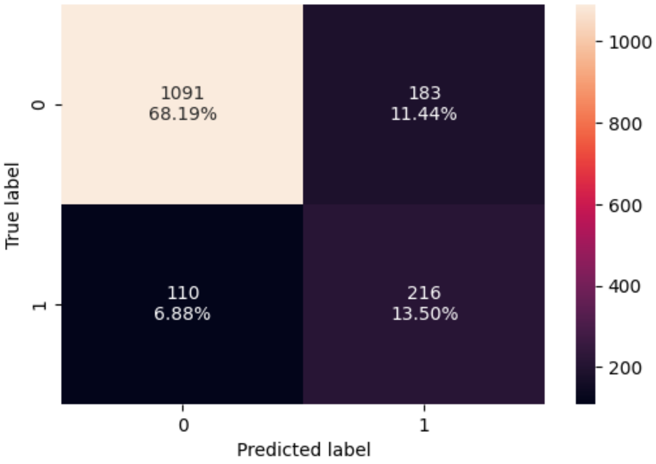
Model 5-(SMOTE), Adam Optimizer, and Dropout





Model 5-(SMOTE), Adam Optimizer, and Dropout





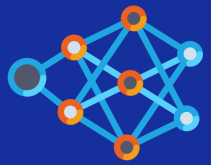
Model 5-(SMOTE), Adam Optimizer, and Dropout

	precision	recall	f1-score	support
0	0.89	0.88	0.89	5096
1	0.88	0.89	0.89	5096
accuracy			0.89	10192
macro avg	0.89	0.89	0.89	10192
weighted avg	0.89	0.89	0.89	10192

	precision	recall	f1-score	support
0.0	0.91	0.86	0.88	1274
1.0	0.54	0.66	0.60	326
accuracy			0.82	1600
macro avg	0.72	0.76	0.74	1600
weighted avg	0.83	0.82	0.82	1600

NEURAL NETWORK

MODEL Performance Comparison



Model Performance Comparison

Training performance comparison

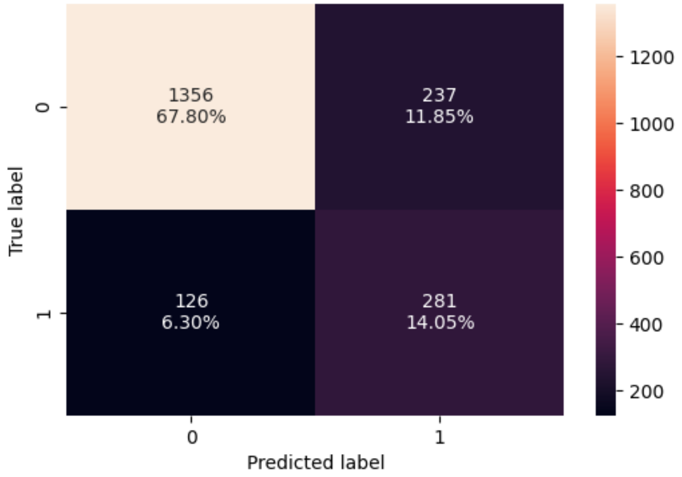
	recall
NN with SGD	0.231595
NN with Adam	0.677914
NN with Adam & Dropout	0.572086
NN with SMOTE & SGD	0.732339
NN with SMOTE & Adam	0.869113
NN with SMOTE,Adam & Dropout	0.890502

Validation set performance comparison

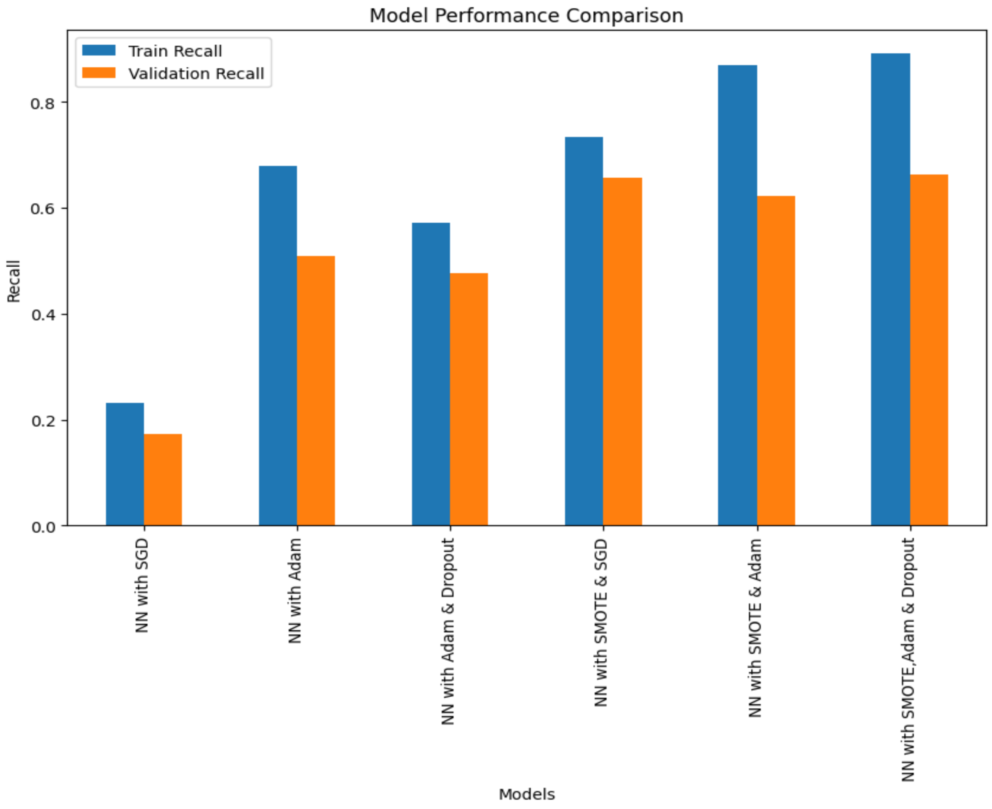
	recall
NN with SGD	0.171779
NN with Adam	0.509202
NN with Adam & Dropout	0.475460
NN with SMOTE & SGD	0.656442
NN with SMOTE & Adam	0.622699
NN with SMOTE,Adam & Dropout	0.662577

	recall
NN with SGD	0.059816
NN with Adam	0.168712
NN with Adam & Dropout	0.096626
NN with SMOTE & SGD	0.075897
NN with SMOTE & Adam	0.246414
NN with SMOTE,Adam & Dropout	0.227926

Model Performance Comparison



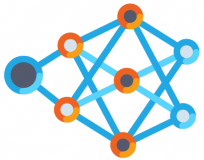
Model Performance Comparison



Model Performance Comparison

Analysis of Model Performance

- **Model 0**-The **NN with SGD** model has the lowest recall for both training and validation, indicating that it struggles to learn and generalize well.
- **Model 1**-The **NN with Adam** significantly improves recall compared to SGD, showing that Adam is a more effective optimizer.
- **Model 2**-The **NN with Adam & Dropout** slightly reduces training recall but maintains a better balance between training and validation recall.
- **Model 3**-The **NN with SMOTE & SGD** improves recall compared to standard SGD, indicating that SMOTE helps in handling imbalanced datasets.
- **Model 4**-The **NN with SMOTE & Adam** shows the highest training recall, but the validation recall is noticeably lower, suggesting some level of overfitting.
- **Model 5-The NN with SMOTE, Adam & Dropout achieves high training recall while keeping validation recall at a competitive level, which indicates a strong balance between learning and generalization.**



Model Performance Comparison

Best Model Recommendation:

"NN with SMOTE, Adam & Dropout"

This model is the best choice because:

- **Highest Recall with Generalization:** It achieves high recall on both training and validation, reducing the risk of overfitting while still learning well.
- **Handles Imbalanced Data:** The use of SMOTE ensures that minority class instances are well represented during training, leading to better recall.
- **Dropout Regularization:** Dropout prevents overfitting by forcing the model to learn robust features, ensuring it generalizes better to new data.
- **Adam Optimizer Efficiency:** Adam is well-suited for optimizing neural networks efficiently, leading to better convergence compared to SGD.

RECOMMENDATIONS

Findings

Enhance Feature Engineering for Improved Model Performance

- Explore domain-specific features and interactions between variables to boost predictive power.
- Implement **automated feature selection** techniques like Recursive Feature Elimination (RFE) and SHAP analysis.

Optimize Model Performance with Advanced Techniques

- Use **ensemble learning** (Stacking, Bagging, Boosting) to improve accuracy and generalization.
- Fine-tune hyperparameters using **Bayesian Optimization** instead of traditional grid search.

Implement Real-Time Model Monitoring and MLOps

- Deploy the model using **Flask/FastAPI** with real-time inference capabilities.
- Set up automated **model drift detection** and retraining pipelines using **MLflow** or **Kubeflow**.

Incorporate Alternative Data Sources for Better Predictions

- Integrate external data sources like **customer behavior trends, market conditions, or social media insights**.
- Utilize **time-series forecasting** methods if the data has a temporal component.

Scale Deployment with Cloud and API Integration

- Deploy the model using **AWS SageMaker, Google Vertex AI, or Azure ML** for scalability.
- Develop a user-friendly **dashboard with Power BI, Tableau, or Streamlit** to visualize insights effectively.

Recommendations

Enhance Customer Retention Strategies

- Implement **personalized loyalty programs** for high-value customers based on their transaction history and credit usage.
- Offer **targeted financial products** (e.g., premium credit cards, mortgage refinancing, investment opportunities) to customers with high spending potential.
- Use predictive analytics to identify **at-risk customers** and proactively offer customized retention incentives.

Improve Customer Segmentation for Better Marketing

- Leverage **machine learning-based clustering** (K-Means, DBSCAN) to segment customers based on spending patterns, product usage, and demographics.
- Tailor marketing campaigns using **customer segmentation insights** to maximize engagement and conversion rates.
- Implement **A/B testing** for different marketing approaches and continuously optimize outreach strategies.

APPENDIX

Conclusion

In this project, we applied a structured and methodical approach to address the business problem through data-driven insights and advanced machine learning techniques. Our solution approach encompassed **exploratory data analysis (EDA), data preprocessing, model selection, evaluation, and deployment**, ensuring that each phase was executed with precision and efficiency.

KEY TAKEAWAYS FROM THE PROJECT INCLUDE:

- **Comprehensive Data Analysis**: Identifying key trends, correlations, and anomalies to enhance business decision-making.
- **Robust Data Preprocessing**: Handling missing values, outliers, and feature engineering to optimize model performance.
- **Model Optimization**: Employing **hyperparameter tuning** and **cross-validation** to achieve the best predictive accuracy.
- **Business-Driven Insights**: Translating machine learning outputs into actionable recommendations for strategic decision-making.

Conclusion

The final machine learning model, selected based on rigorous evaluation metrics, demonstrates **high predictive accuracy, generalizability, and scalability** for real-world application. This solution not only enhances business efficiency but also provides a **data-centric framework for continuous improvement and decision-making**.

Moving forward, the model can be **further refined and integrated into production pipelines** with real-time monitoring and feedback mechanisms to maintain performance over time. Additionally, leveraging **automated MLOps** and **cloud deployment** will enhance scalability and accessibility.

This project underscores the **importance of data science and machine learning in driving business innovation and efficiency**. Future improvements could include incorporating **deep learning models, alternative data sources, and real-time analytics** to further refine predictive capabilities.

Great Learning

Happy Learning !