ANN FOR CUSTOMER CHURN PREDICTION

TABLE OF CONTENTS

Introduction & Problem Definition

XYZ is a mid-size telecommunications company offering a variety of services for mobile, internet and TV services. The company, despite their growth through the years, has noticed customer churning, customer's cancelling their services and turning to competitor's. The company realised that it is more cost effective to retain the existing customers rather than acquire more. Before implementing the strategies, the company would like to predict the ones who might churn, and personalize strategies towards them. This report seeks to cover a model that might help the company in predicting customers who might churn.

The dataset used contains data of 10,000 customers, with the following columns below

- CustomerID: Unique identifier for each customer.
- Surname: Last name of the customer.
- Credit Score: The credit score of the customer.
- Geography: Geographical country in Europe.
- Gender: Male or Female.
- Age: Age of the customer.
- Tenure: Number of years the customer has been with the company.
- Number of Products: How many bank products the customer has availed.
- HasCrCard: Whether the customer has a credit card with the bank.
- IsActiveMember: Is the customer an active member?
- EstimatedSalary: Estimated salary of the customer.
- Exited / Churn: Target variable indicating whether the customer churned (1) or stayed (0).

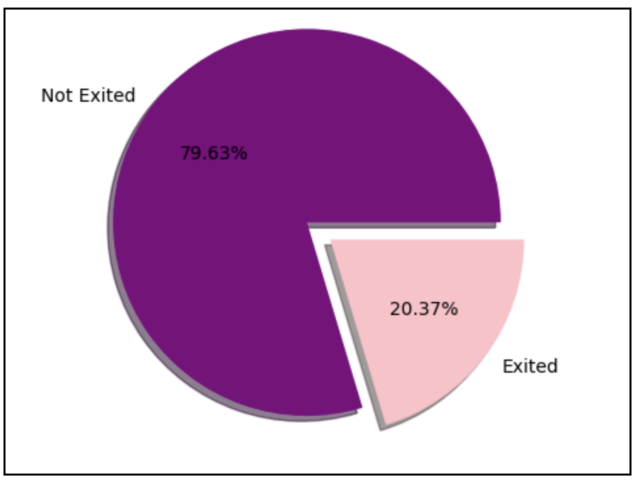
Python and keras will be used to build an ANN that predicts the customer churned and then the model will be evaluated using accuracy, precision, recall and F1-Score.

Exploratory Analysis

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   RowNumber        10000 non-null  int64  
 1   CustomerId       10000 non-null  int64  
 2   Surname          10000 non-null  object 
 3   CreditScore      10000 non-null  int64  
 4   Geography        10000 non-null  object 
 5   Gender           10000 non-null  object 
 6   Age              10000 non-null  int64  
 7   Tenure           10000 non-null  int64  
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64  
 10  HasCrCard        10000 non-null  int64  
 11  IsActiveMember   10000 non-null  int64  
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64  
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```

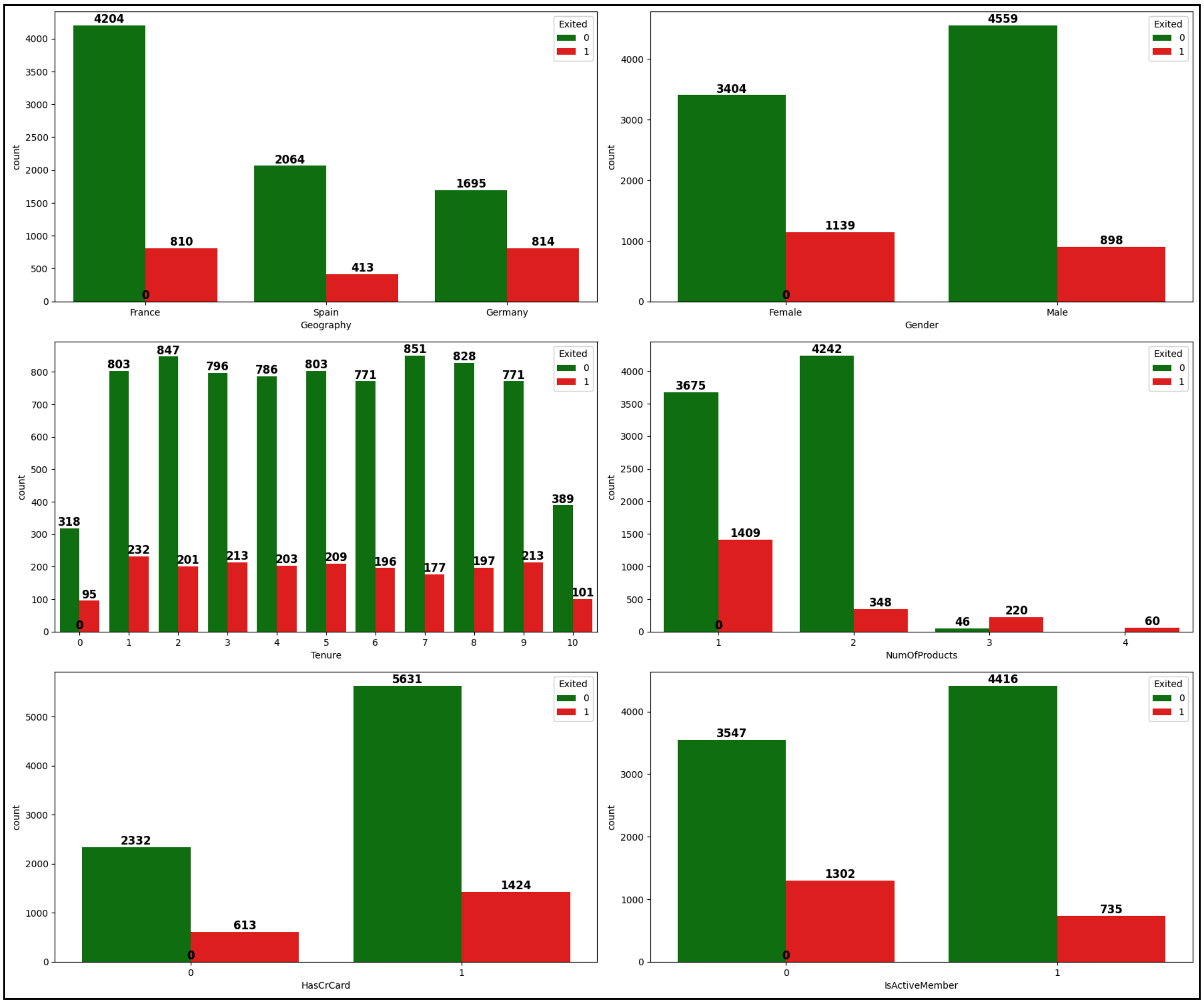
	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	10000.00000	1.000000e+04	10000.000000	10000.000000	10000.000000	10000.000000	10000.000000	10000.00000	10000.000000	10000.000000	10000.000000
mean	5000.50000	1.569094e+07	650.528800	38.921800	5.012800	76485.889288	1.530200	0.70550	0.515100	100090.239881	0.203700
std	2886.89568	7.193619e+04	96.653299	10.487806	2.892174	62397.405202	0.581654	0.45584	0.499797	57510.492818	0.402769
min	1.00000	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.00000	0.000000	11.580000	0.000000
25%	2500.75000	1.562853e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.00000	0.000000	51002.110000	0.000000
50%	5000.50000	1.569074e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.00000	1.000000	100193.915000	0.000000
75%	7500.25000	1.575323e+07	718.000000	44.000000	7.000000	127644.240000	2.000000	1.00000	1.000000	149388.247500	0.000000
max	10000.00000	1.581569e+07	850.000000	92.000000	10.000000	250898.090000	4.000000	1.00000	1.000000	199992.480000	1.000000

The data set contains 10,000 non-null records. Gender and geography being the 2 categorical variables and rest numeric.



The above pie chart shows the class distribution. We can see that not-exited class is the majority and dominates the dataset.

Analysis of bar plots:

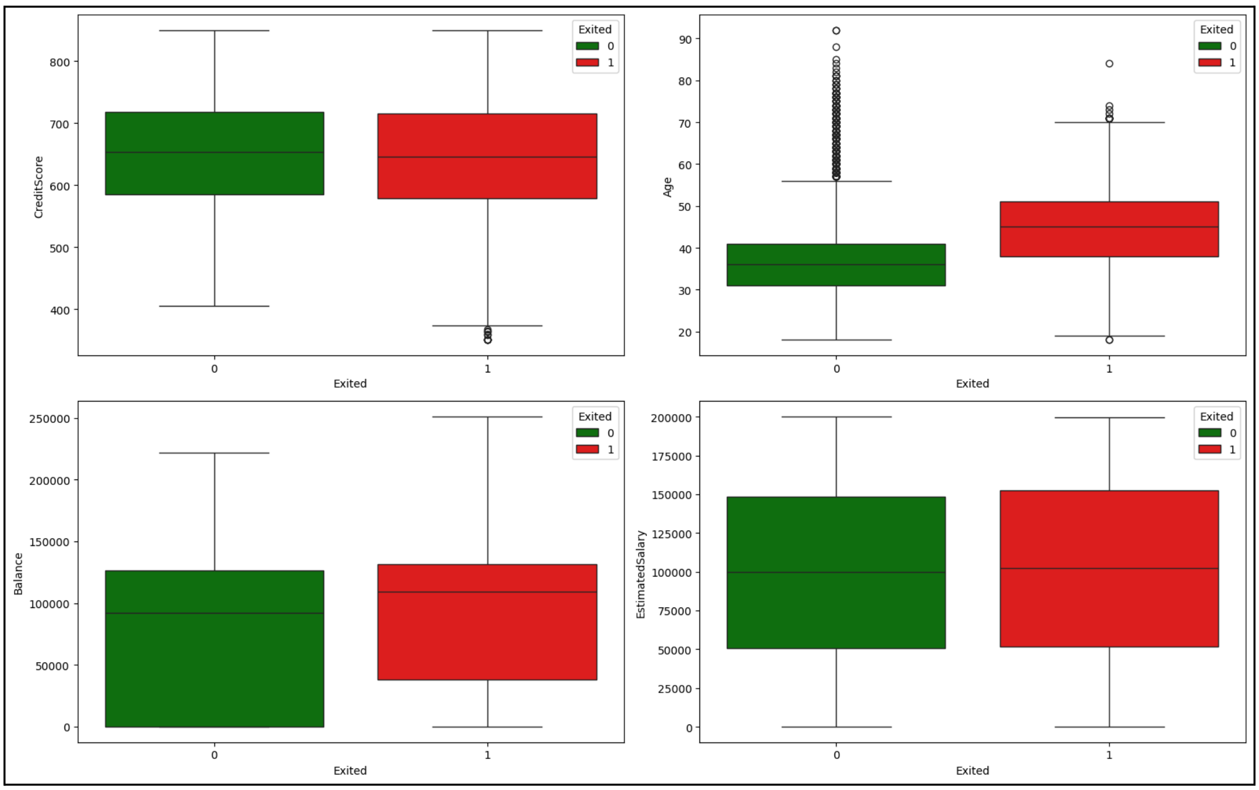


- The first plot - Geography: A visual representation of the distribution of excited and not excited customers across Geography-Spain, France, Germany. France has the most customers but also the highest number of retained customers. Germany has the highest churn rate compared to other countries (almost equal churn vs. retention). Spain has a lower churn rate than Germany but still shows noticeable churn. Customers in Germany are more likely to churn than in other countries. It may indicate dissatisfaction or external factors affecting German customers.

- Second Plot - Gender: More males are in the dataset yet the proportion of churned females is higher than churned males. Gender plays a role in churn, and female customers might be at a higher risk of leaving
- Third plot - tenure: shows us tenure across churning. The plot shows us that there are more customers spread through 1 to 9 years of tenure, but there is approximately similar churning through these years. If the data provided the date of churning, it would have been easier to detect the year with maximum churning and might help the company determine the policy changes or competitor''s actions which might have prompted the customers to churn. Newer customers (0-3 years) and long-term customers (10 years) are more likely to churn. Possible dissatisfaction at the start or completion of a typical banking lifecycle.
- The Fourth - plot no of products: A common insight is that the customer who have used only one product from the company have churned the most. The customers who have used 3 to 4 products, despite less in numbers have mostly churned as compared to using 1 or 2 products. From this we can determine that customers using only one or two products of the company have mostly retained. Also the fact that customers who have used 3 or more products have mostly churned could signify that the company has only one to two good products that sell well or well utilized. Figuring out whether these two products are the same for all customers requires additional data. The distribution of churning across products can help the company focus on those products that retain the customers.
- The Fifth plot - having credit card: Majority of customers have a credit card (1). Churn exists in both groups, but having a credit card does not strongly affect churn.

- The Sixth plot - Active: Inactive members have almost twice the churn rate. Inactive customers are at high churn risk. Increasing engagement (e.g., personalized offers, customer interactions) may help retention.

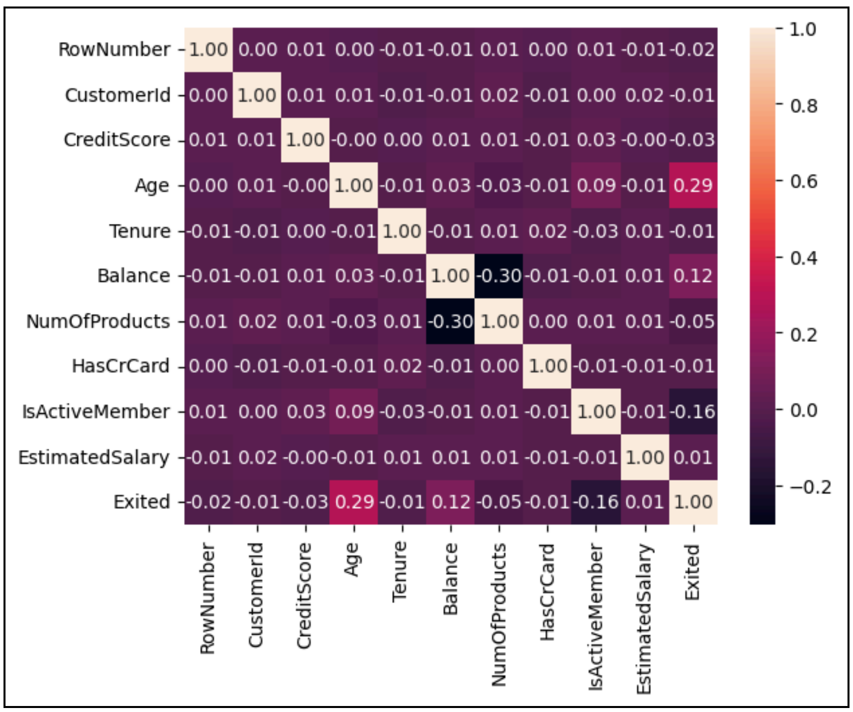
Analysis of Box PLot for Continuous Numeric Independent Variables



- Box Plot 1 - credit score: The median of exited and not exited customers is almost the same, but with not- exited customers having slightly higher median credit score. The plot shows that credit score might not be a strong predictor of customer churning.
- Box Plot 2 - Age and Exited: The median age of customers who have exited is 45 as compared to not exited customers being 36. The company should target between the age group 39 to 52, in order to retain and avoid the customers churning. The company could

try providing additional benefits for the customers belonging to this age group who comprise established working professionals.

- Box Plot 3 - Balance score and Exited: An interesting insight here is that customers who have exited have a higher median as compared to not exited customers. The XYZ company should focus more on customers with higher balance. This can be linked with the 39 to 52 age group as this is the period of career growth for most working customers and who are mostly cautious, preparing for retirement. Older customers are more likely to leave.
- Box Plot 4 - Estimated Salary score and Exited : The exited customers have almost same median of estimated salary as compared to the not exited.



The correlation matrix shows us that there is not strong co-relationship between any of the variables. Some variables have a weak positive or negative correlation.

Data Preprocessing

```
df.drop(columns=["RowNumber","CustomerId","Surname"], axis=1, inplace=True)
df.columns
```

Dropping columns that do not add information or importance to build the model and then checking if the necessary columns are included.

```
df['Geography'] = df['Geography'].map({'France' : 0, 'Germany' : 1, 'Spain' : 2})
df['Gender'] = df['Gender'].map({'Male' : 0, 'Female' : 1})
df.head()
```

Encoding categorical variables geography and gender. Used mapping instead of one hot encoding for better model performance.

```
[ ] X = df.drop('Exited', axis = 1)
    y = df.Exited

    # splitting data into training set and test set

    from sklearn.model_selection import train_test_split

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20)


    from sklearn.preprocessing import StandardScaler

    sc = StandardScaler()
    X_train = sc.fit_transform(X_train)
    X_test = sc.transform(X_test)
```

Splitting the dataset into a training and testing set with 80:20 split and then standardising the independent variables. This prevents one variable from dominating the model training. Helps converge faster and.

Model Architecture

```
import keras
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dropout
from keras.layers import BatchNormalization

# initializing ann
model = Sequential()

# adding the first input layer and the first hidden layer
model.add(Dense(6, kernel_initializer = 'he_normal', activation = 'relu', input_shape = (10, )))

# adding batch normalization and dropout layer
model.add(Dropout(rate = 0.1))
model.add(BatchNormalization())

# adding the third hidden layer
model.add(Dense(6, kernel_initializer = 'he_normal', activation = 'relu'))

# adding batch normalization and dropout layer
model.add(Dropout(rate = 0.1))
model.add(BatchNormalization())

# adding the output layer
model.add(Dense(1, kernel_initializer = 'glorot_normal', activation = 'sigmoid'))

# compiling the model
model.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['accuracy'])

# fitting the model to the training set

model_history = model.fit(X_train, y_train, validation_data = (X_test, y_test), epochs = 100)
```

- The ann model is initialised using keras sequential into the variable called model.
- The first layer with 10 inputs and the first hidden layer with 6 neurons are initialised. The kernel initializer is he-normal. ReLU activation (max(0, x)) sets negative values to zero, which can cause a vanishing gradient issue, he-normal initialization scales weights to prevent neuron activations from shrinking too much. The activation is relu, Customer churn depends on complex relationships between factors like age, credit score, balance, and salary. ReLU allows the model to learn non-linear patterns more effectively, with less computational cost.
- A 0.1 dropout rate is added, 10% of neurons are dropped randomly during training which prevents overfitting and makes the model better at generalizing.

- Batch normalization is added to stabilize the outputs of neurons which become inputs for the next layer of neurons. This speeds up the convergence and reduces internal covariate shift and speeds up the convergence.
- The third layer with 6 neurons is initialized with he-normal and activation relu.
- Dropout of 0.1 and batch normalization is also added after the third layer.
- The final layer is added with 1 neuron with glorot_normal initialization and sigmoid activation for probability(values between 0-1).
- The model is then compiled with adam optimizer. Binary cross entropy is used since the output is binary classification using sigmoid. Accuracy is used to keep track of the model performance.

```
model.summary()
```

Model: "sequential_5"

Layer (type)	Output Shape	Param #
dense_15 (Dense)	(None, 6)	66
dropout_10 (Dropout)	(None, 6)	0
batch_normalization_10 (BatchNormalization)	(None, 6)	24
dense_16 (Dense)	(None, 6)	42
dropout_11 (Dropout)	(None, 6)	0
batch_normalization_11 (BatchNormalization)	(None, 6)	24
dense_17 (Dense)	(None, 1)	7

Total params: 443 (1.73 KB)
Trainable params: 139 (556.00 B)
Non-trainable params: 24 (96.00 B)
Optimizer params: 280 (1.10 KB)

$$Parameters = (Input\ Neurons \times Output\ Neurons) + Biases$$

Dense (6 neurons, input 10): $(10\times6) + 6 = 66$

Dense (6 neurons, input 6): $(6\times6) + 6 = 42$

Output Layer (1 neuron, input 6): $(6\times1) + 1 = 7$

Extra non-trainable, learnable parameters from batch normalization: 24

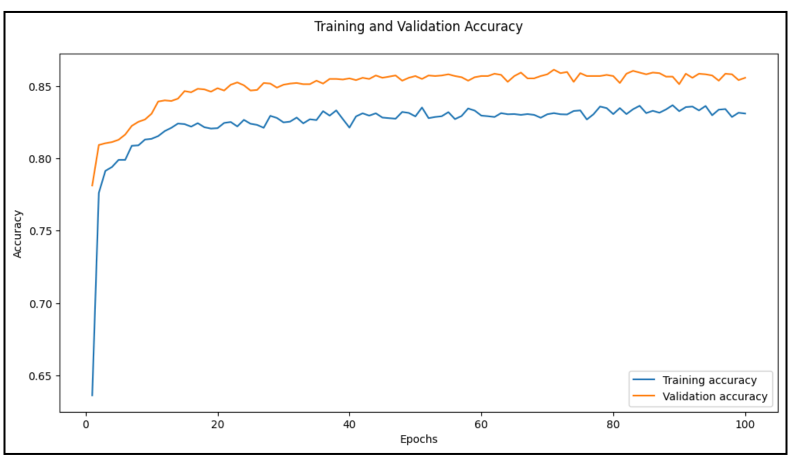
Total trainable parameters: $66 + 42 + 7 + 24 = 139$

Optimizer parameters: 280

Total parameters: $139 + 280 = 443$

Evaluation

Loss and accuracy plots



Both training and validation loss decrease over epochs and are consistent after the first initial 20 epochs. The validation loss is below the training loss. Similarly, the training and validation loss increase and stabilize at 20 epochs, with the validation accuracy higher than training accuracy.

This shows that the model is not overfitting and generalises well.

Threshold 0.5

```
from sklearn.metrics import confusion_matrix, accuracy_score
y_pred = model.predict(X_test)
y_pred = (y_pred > 0.5)
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)

from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

```
79/79 ━━━━━━━━━━━━━━━━━━━━ 0s 2ms/step
[[1965   46]
 [ 315  174]]
              precision    recall  f1-score   support

           0       0.86      0.98      0.92      2011
           1       0.79      0.36      0.49       489

    accuracy                           0.86      2500
   macro avg       0.83      0.67      0.70      2500
weighted avg       0.85      0.86      0.83      2500
```

At 0.5 threshold, the accuracy is 86%. The precision for class 1 the churners is 79% whereas the recall and f1 score slightly low at 36% and 49% respectively. High precision means fewer false negatives. But the low recall and f1 score mean that the model might have a higher false positive and miss actual churners. The company XYZ needs to predict customers who might churn, therefore it requires a higher recall and f1-score.

Threshold 0.4

```
from sklearn.metrics import confusion_matrix, accuracy_score
y_pred = model.predict(X_test)
y_pred = (y_pred > 0.4)
cm = confusion_matrix(y_test, y_pred)
print(cm)
accuracy_score(y_test, y_pred)

from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

```
79/79 ━━━━━━━━━━━━━━━━━━━━ 0s 3ms/step
[[1889  122]
 [ 259  230]]
              precision    recall  f1-score   support

           0       0.88      0.94      0.91      2011
           1       0.65      0.47      0.55       489

    accuracy                           0.85      2500
   macro avg       0.77      0.70      0.73      2500
weighted avg       0.84      0.85      0.84      2500
```

At 0.4 threshold the accuracy is 85% with slightly improved recall and f1-score at 47% and 55%. This threshold better recognised customers who might churn, but the precision for class 1 decreases to 65%, therefore it incorrectly labels unchurned customers as churners, more true negative. The company can use this threshold to detect churned or customers who might churn, but might face the trade off between precision and recall.

Feature importance

```
feature_importance = pd.DataFrame({'Feature': X.columns,
                                   'Importance': model.layers[0].weights[0].numpy().flatten()[:len(X.columns)]})
feature_importance = feature_importance.sort_values(by='Importance', ascending=False)
print("\nFeature Importance:")
print(feature_importance)
```

```
Feature Importance:
           Feature  Importance
8   IsActiveMember    0.783193
3              Age    0.175544
7        HasCrCard    0.152912
2           Gender    0.138447
6    NumOfProducts    0.061200
5          Balance   -0.010147
0      CreditScore   -0.059520
1        Geography   -0.115704
4           Tenure   -0.137227
9  EstimatedSalary   -0.217031
```

The feature importance analysis of the customer churn prediction model reveals that IsActiveMember is the most influential factor, indicating that active customers are less likely to churn. Age and HasCrCard also play significant roles, suggesting that middle-aged customers and those with credit cards exhibit distinct churn behaviors. Gender and NumOfProducts have moderate influence. Interestingly, Balance, CreditScore, and Geography have relatively lower impact, suggesting that financial stability and location are not primary churn determinants. Tenure shows a slight negative influence, implying that long-term customers may still leave due to evolving needs or dissatisfaction. EstimatedSalary has the least importance, indicating that income level alone does not drive customer retention. These insights suggest that focusing on customer engagement strategies, rather than purely financial metrics, could be more effective in reducing churn.

Discussion & Business Recommendation

Using the ANN model XYZ company can predict the customers who might churn, then they can target these customers with personalized offers and benefits to retain them. The model showed that active engagement was a crucial variable that retained most of the customers, therefore customer engagement should be their priority. The company can do this through loyalty programs, exclusive offers and discounts. Have robust customer service, have check-ins or targeted messaging for customers who have gone inactive for long periods of time.

Since age is also a strong predictor of customer churning, the company can further drill down and target middle aged customers, especially females, which the data has shown have churned the most. The company can have stable and cost-effective plans for them.Customers who had credit cards churned less, this can be due to the customers being able to link their credit card with the plans, setting up automatic payment. The company can give benefits and offers such as discounts or cash-backs to the customers who link their credit card. Promoting long-subscription plans will lock the customers in better.The company can also optimize their product offerings through innovation. They can also bundle subscription offers so that customers who use only one product will not opt for the bundle subscription. Providing upgrade incentives and personalised upselling campaigns.

Customers who have had long tenures have also churned, meaning that the customers were dissatisfied and hence churned even after using the company's service for a span of 1 to 9 years. The company can look into getting the feedback and implementing changes based on the customer feedback. The company can also offer special discounts for long tenured customers. Credit, salary and geography have less impact on customer churn, the company can still focus on local regional trends and, country employment and inflation to help retain customers through plans that parallel the situation.