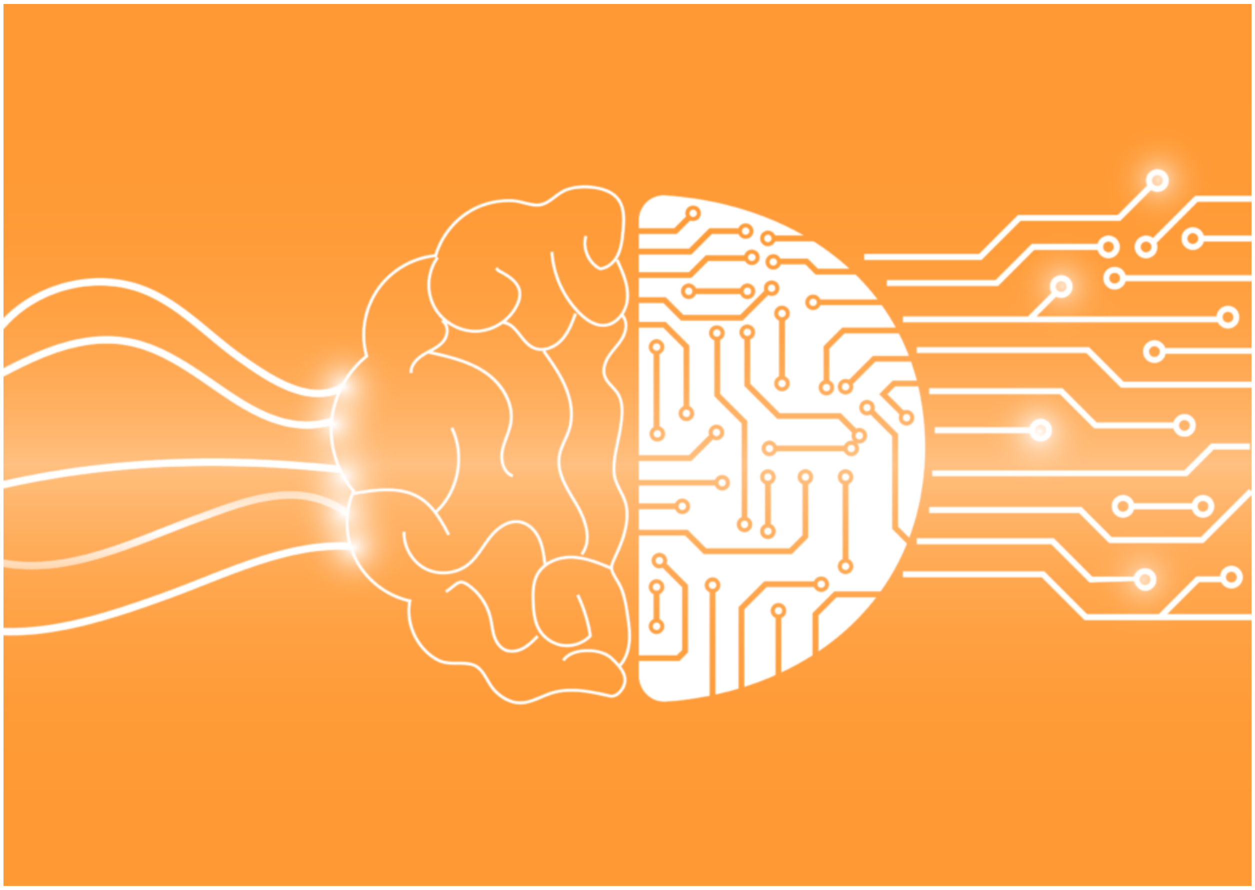
Churn for Bank Customers Analysis



JPMorganS

Jinwoo Oh

Minjae Kwon

Pranav Reddy

Srihith Duggi

Introduction

Background

As advancements in technology continue to reshape various industries, it comes as no surprise that the banking sector has also been significantly influenced. The rapid growth of financial technology, better known as fintech, is captivating a substantial portion of bank customers. Interestingly, a study has revealed that a remarkable 74% of Americans are willing to leave their traditional banks in favor of safer and more cost-effective digital fintech solutions (Francis et al., 2012). Delving deeper into this statistic, it turns out that 39% of these Americans are drawn to the prospect of low-rate credit cards, while around 20% are keen on utilizing budgeting apps. It's important to note that a customer's decision to leave a bank can be influenced by various factors, such as their gender, age, salary, or even the number of products they have purchased through the bank.

Objective

In this report, we acknowledge the well-known fact that acquiring new clients is often more complex and stressful than retaining existing ones. Understanding the factors that influence a client's decision to leave a bank can be immensely valuable. By pinpointing the reasons behind customer attrition, banks can develop targeted strategies, such as loyalty programs or retention marketing techniques, to encourage clients to remain with the institution. Gaining insight into these driving forces not only fosters customer loyalty but also enables banks to adapt and evolve in a rapidly changing financial landscape.

In this report, we chose to utilize the "Exited" column as the label for our predictive model, which indicates whether a customer has remained with the bank or not. The remaining variables were employed as features to inform our analysis. During the initial stages, we visualized the data to discern which factors held greater importance and which were less relevant. Subsequently, we trained our dataset using various classifiers and, based on performance, selected the most suitable model.

Data Processing

Data Sources

The "Churn for Bank Customers" dataset, which contains information on bank customers and their churn status, has been sourced from Kaggle. Designed to assist companies in better understanding customer churn—a major issue faced by banks and other businesses—this dataset offers valuable insights for tackling the problem.

Dataset: https://www.kaggle.com/code/kmalit/bank-customer-churn-prediction/input

There are 10,000 rows and 14 variables in this dataset. The 14 columns include the followings: RowNumber, CustomerId, Surname, CreditScore, Geography, Gender, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary, and Exited. The dataset consists of numerical, categorical, and binary types of data.

Data Type	Features
Numerical	CreditScore, Age, Tenure, Balance, NumofProducts, EstimatedSalary
Categorical	RowNumber, Customerid, Surname, Geography, Gender
Binary	HasCrCard, IsActiveMeber, Exited

Feature Explanation

Feature	Key
CreditScore	Customer's credit score
Gender	Customer's gender
Age	Customer's age
Tenure	The number of years that the customer has been a client of the bank

Balance	How much left in the customer’s bank account
NumOfProducts	The number of products that a customer has purchased through the bank
HasCrCard	Denotes whether or not a customer has a credit card
IsActiveMember	If the customers active or not
EstimatedSalary	Customer’s salary
Exited	Whether or not the customer left the bank

Feature Transformation

In terms of machine learning, labels are a specific type of variable or attribute containing essential information about a dataset. These labels are provided to the model to facilitate predictions based on observed patterns within the dataset. By employing the "Exited" column as the label, we can train a machine learning model to identify patterns and correlations related to customer churn within the data. Consequently, the model can be utilized to predict the likelihood of other customers churning, empowering the bank to take proactive measures to retain their clientele and minimize churn.

We explore the use of one-hot encoding; it operates by converting each category within a categorical variable into a separate binary variable. We applied one-hot encoding to the gender column, which comprises two categories: male and female. By doing so, we transformed this column into two distinct binary columns—one for males and one for females. The values in these new columns indicate whether an observation corresponds to a male or female customer. This method enables a smooth and user-friendly integration of categorical data into machine learning models, ultimately improving their ability to make accurate predictions.

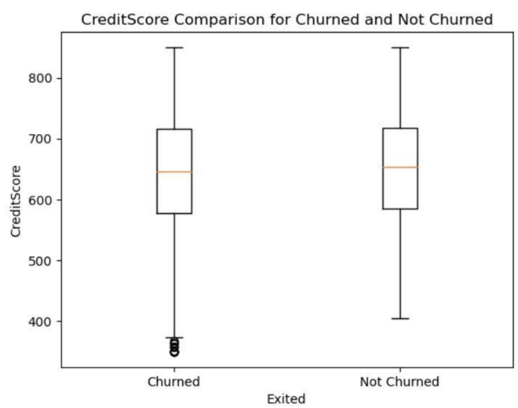
Methodology

Data Visualization

Data visualizations are important because they enable us to communicate complex information in a simple and concise way. They allow us to easily identify patterns, trends, and outliers, which can inform decision-making and reveal insights that might be hidden in raw data. Since the goal of our project is to predict the likelihood of churn, we will use box plots to represent the frequency of each attribute of the refined dataset in 2 categories, churned and not churned.

Credit Score Comparison

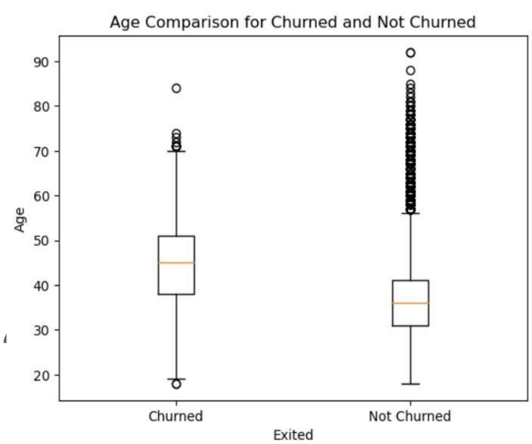
Hypothesis: People with lower credit scores are likely to churn as it may be difficult for them to obtain credit in the future given their financial history. This could result in them switching to a competing bank.



Observation: The credit score comparison box plots for both churned and not churned are similar implying that credit score is not a significant indicator of churn. Although the lower extremes for the plots differ, the upper and lower quartiles, medians as well as the upper extremes in the plots are far too similar for credit score to be a strong indicator of churn.

Age Comparisons

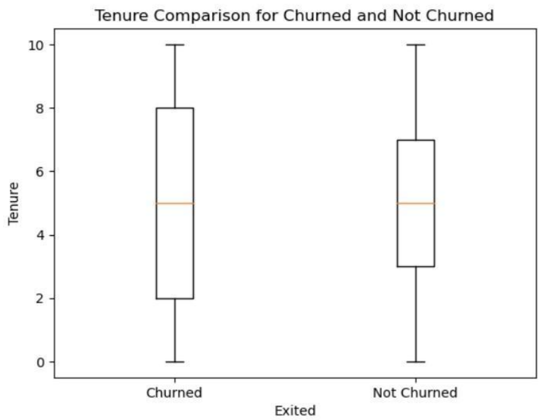
Hypothesis: Younger people are likely to churn as they are likely to have lower incomes resulting in them being more susceptible to debt compared to older individuals. This could lead to younger people switching to a different bank for better prices.



Observation: The box plots indicate a different correlation between churn and age than what the hypothesis suggests. The upper quartile of the not churned plot is at the same level as the lower quartile of the churned plot. This indicates that the percentage of the frequency of younger ages in the not churned category is far greater compared to the churned category.

Tenure Comparisons

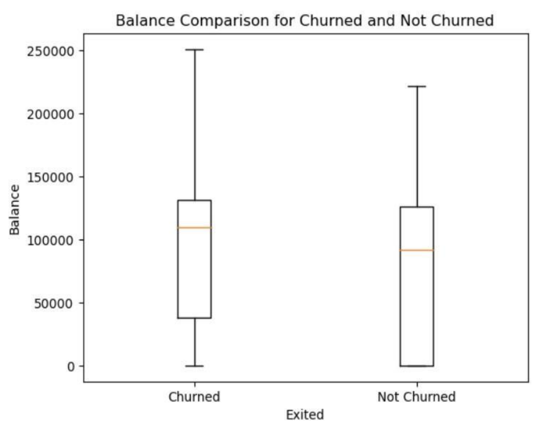
Hypothesis: People with a longer tenure with the bank are less to churn as they would be satisfied and comfortable with the banking services as well as establish a good credit history



Observation: For both the plots in tenure comparison the upper and lower extremes as well as the medians are the same. Also the interquartile range of the churned plot is within the interquartile range of the churned plot. This suggests that tenure is not a very strong indicator of churn.

Account Balance Comparisons

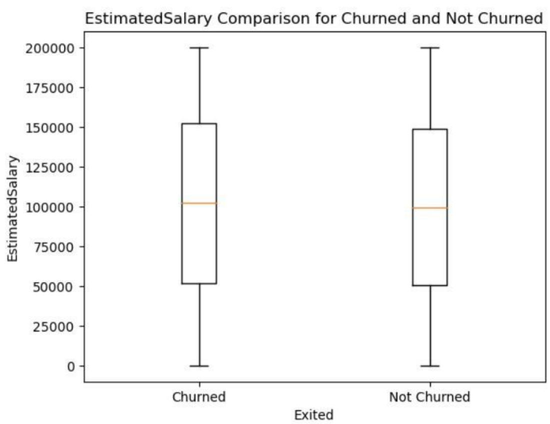
Hypothesis: People with higher account balances are less likely to churn as they are beneficial to the bank and so they would receive special promotions and incentives to stay at the bank. People with lower balances are subject to financial instability indicating that may switch to a better pricing option for them



Observation: The not churned box plot in this comparison shows there is a greater percentage of lower account balances in the not churned category compared to the churned category.

Estimated Salary Comparisons

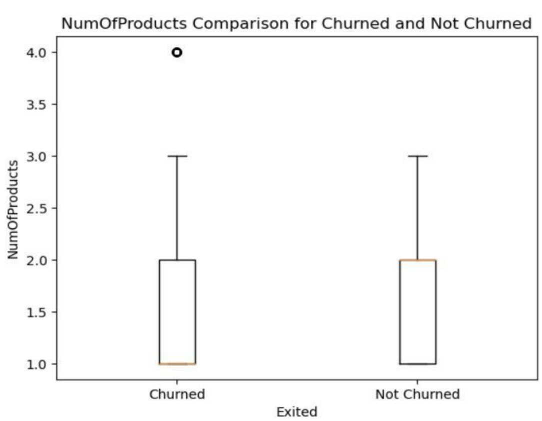
Hypothesis: People with higher salaries are less likely to churn. This hypothesis suggests that there is a negative correlation between estimated salary and churn rate, meaning that as estimated salary decreases, the likelihood of churn increases. This could be because individuals with lower salaries may have fewer resources to cope with unexpected expenses or financial setbacks, and may be more likely to switch to a different service provider if they find a better deal.



Observation: The estimated salary plots for both churned and not churned are once again too similar for salary to be a strong indicator. It may have an influence on churn, however, it is difficult to determine based on these plots.

Number of Products comparison

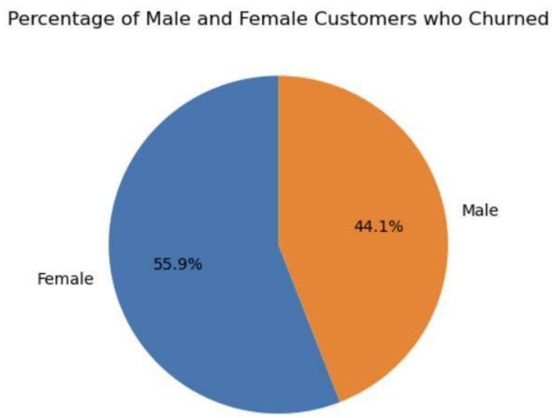
Hypothesis: People that have purchased a higher number of products through the bank are less likely to churn as it would imply satisfaction with the bank's products and services



Observation: Although both the plots have similar extremes and interquartile ranges, they have different median values. The not churned plot’s median value is at 2.0 while the churned plot’s median value is at 1.0. This implies that there is a greater percentage of people that churned had a lower number of products purchased through the bank.

Percentage of Male and Female clients that churned

Hypothesis: There should be a small difference between the percentages of men and women who churned.



Observation: There is an almost even split with women consisting of 55.9% of the churned population and men consisting of 44.1%

Model Training

We randomly selected about 500 rows from the model, which is the test.csv file. We tried the five different classifiers in order to evaluate our model's performance: Logistic, MLP, Decision Tree, Gradient Boosting, and Random Forest Classifiers. As a result, the accuracy of the Random Forest Classifier was the greatest as 0.8641. Therefore, we saved the Random Forest Classifier as the best model.

Classifier	Accuracy
Logistic	0.8021978021978022
MLP	0.8211788211788211
Decision Tree	0.7842157842157842
Gradient Boosting	0.8631368631368631
Random Forest	**0.8641358641358642**

```python
from sklearn.metrics import accuracy_score

# Load the test dataset
test_df = pd.read_csv("test.csv")

# Drop unnecessary columns
test_df.drop(["RowNumber", "CustomerId", "Surname", "Geography"], axis=1, inplace=True)

# Perform one-hot encoding on "Gender" feature
test_df = pd.get_dummies(test_df, columns=["Gender"])

# Extract the input features and actual labels
test_inputs = test_df.drop("Exited", axis=1)
y_actual = test_df["Exited"]

test_inputs.head()

# Predict using Logistic Regression Classifier
y_predicted_lr = lr_classifier.predict(test_inputs.to_numpy())
lr_accuracy_score = accuracy_score(y_predicted_lr, y_actual)

# Predict using MLP Classifier
y_predicted_mlp = mlp_classifier.predict(test_inputs.to_numpy())
mlp_accuracy_score = accuracy_score(y_predicted_mlp, y_actual)

# Predict using Decision Tree Classifier
y_predicted_dt = dt_classifier.predict(test_inputs.to_numpy())
dt_accuracy_score = accuracy_score(y_predicted_dt, y_actual)

# Predict using Gradient Boosting Classifier
y_predicted_grad_boost = grad_boost.predict(test_inputs.to_numpy())
grad_boost_accuracy_score = accuracy_score(y_predicted_grad_boost, y_actual)

# Predict using Random Forest Classifier
y_predicted_random_forest = random_forest.predict(test_inputs.to_numpy())
random_forest_accuracy_score = accuracy_score(y_predicted_random_forest, y_actual)


print (f"Accuracy of the Logistic Classifier = {lr_accuracy_score}")
print (f"Accuracy of the MLP Classifier = {mlp_accuracy_score}")
print(f"Accuracy of the Decision Tree Classifier = {dt_accuracy_score}")
print(f"Accuracy of the Gradient Boosting Classifier = {grad_boost_accuracy_score}")
print(f"Accuracy of the Random Forest Classifier = {random_forest_accuracy_score}")
```

Code: Evaluating our model's Performance

Next, with the best model, we tried to predict for a single instance starting with importing 'pickle' to our Jupyter Notebook. Then, select the random row from the csv file. We input CreditScore, Gender, Age, Tenure, Balance, NumOfProduct, HasCrCard, IsActiveMember, and EstimatedSalary to predict if this user is likely to quit the service or not. Our model prediction and information on the csv file were the same: the user likely to churn.

```python
import pickle

file_to_write = open("churn_best_model.saved","wb")
pickle.dump(random_forest,file_to_write)
file_to_write.close()
```

Import pickle with Random Forest

Result

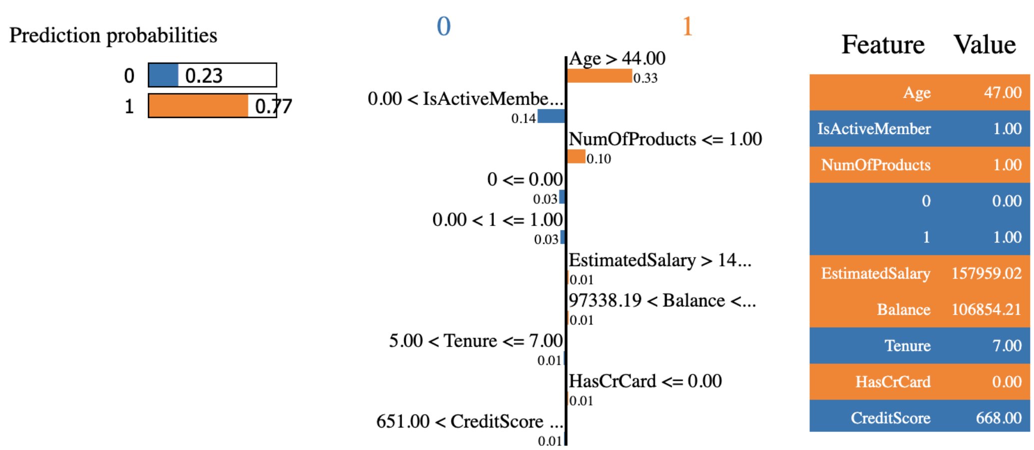
```
import pickle
import numpy as np

# Load the trained model
model_file = open("churn_best_model.saved", "rb")
random_forest = pickle.load(model_file)
model_file.close()

# Prepare a sample input
CreditScore = 668
Gender = "Male"
Age = 47
Tenure = 7
Balance = 106854.21
NumOfProducts = 1
HasCrCard = 0
IsActiveMember = 1
EstimatedSalary = 157959.02

# Perform one-hot encoding for the "Gender" feature
gender_transformed = {"Female": [1, 0], "Male": [0, 1]}  # Update with the mapping used during training
Gender = gender_transformed[Gender]

# Create a numpy array of the input data, including the one-hot encoded "Gender" feature
input_data = np.array([[CreditScore, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary] + Gender])

# Predict using the trained decision tree classifier
y_predicted = random_forest.predict(input_data)

# Interpret the prediction result
if y_predicted[0] == 1:
    print("The person is likely to churn")
elif y_predicted[0] == 0:
    print("The person is not likely to churn")
else:
    print("Invalid prediction")
```

How We Tested a Single Instance; The Result was Correct

We confirmed our model is showing high accuracy by testing a random sample. However, we explained a trained model's prediction using LIME as well. We used the same sample we did above; the model told us the customer is likely to churn. The final graph worked well. By adjusting some values from the instance, we could find out lots of things. We found Age, IsActiveMember, NumofProducts, and Tenure had significant impact on the result. Additionally, there was a difference between the male and female. The graph proves female customers are more likely to leave the bank service. In the case of other features such as CreditScore, Balance, EstimatedSalary had much less impact on the result than we expected.

```
import numpy as np
from lime import lime_tabular

CreditScore = 668
Gender = "Male"
Age = 47
Tenure = 7
Balance = 106854.21
NumOfProducts = 1
HasCrCard = 0
IsActiveMember = 1
EstimatedSalary = 157959.02

# Perform one-hot encoding for the "Gender" feature
gender_transformed = {"Female": [1, 0], "Male": [0, 1]}  # Update with the mapping used during training
Gender = gender_transformed[Gender]

input_data = np.array([CreditScore, Age, Tenure, Balance, NumOfProducts, HasCrCard, IsActiveMember, EstimatedSalary] +Gender)


exp = explainer.explain_instance(
    data_row=input_data,
    predict_fn=random_forest.predict_proba
)

exp.show_in_notebook(show_table=True)
```

Code: Generating LIME Graph



LIME Graph

Other than testing the accuracy and LIME graph, we tested the model's bias and fairness. The result of the code below shows the accuracy is low for both Class 1 (Churn). The accuracy is about 10% lower for male in Class 1. This might indicate that there is a probability of bias in this model.

	Male	**Female**
Class 1 (Churn)	0.36470588235294116	0.4574468085106383
Class 0 (Not Churn)	0.9705263157894737	0.9510086455331412

Demo

We created the demo version of the model. Here is the link to a sample video of the demo: https://youtu.be/_GhC2Xi92qM. In the demo the user is able to control all features that we used in the model above. After adjusting all figures, the user can click the "Predict Churn" button to get the result. We used the IPython display tool to make this demo. So, you can download the code and run it in the jupyter notebook.

```python
import ipywidgets as widgets
from IPython.display import display

# Define input widgets
credit_score = widgets.IntSlider(min=300, max=850, description="Credit Score")
age = widgets.IntSlider(min=18, max=100, description="Age")
tenure = widgets.IntSlider(min=0, max=10, description="Tenure")
balance = widgets.FloatSlider(min=0, max=250000, description="Balance")
num_of_products = widgets.IntSlider(min=1, max=4, description="NumOfProducts")
has_cr_card = widgets.Checkbox(value=True, description="Has Credit Card")
is_active_member = widgets.Checkbox(value=True, description="Is Active Member")
estimated_salary = widgets.FloatSlider(min=0, max=200000, description="Est. Salary")
gender = widgets.Dropdown(options=["Male", "Female"], description="Gender")

# Define a button to trigger the prediction
predict_button = widgets.Button(description="Predict Churn")

# Function to handle the button click event
def on_button_click(button):
    # one-hot encoding for the "Gender" feature
    gender_transformed = {"Female": [1, 0], "Male": [0, 1]}
    Gender = gender_transformed[gender.value]

    # Preprocess the input data
    user_input = [[
        credit_score.value,
        Gender[0],
        Gender[1],
        age.value,
        tenure.value,
        balance.value,
        num_of_products.value,
        int(has_cr_card.value),
        int(is_active_member.value),
        estimated_salary.value
    ]]

    user_input_df = pd.DataFrame(user_input, columns=["CreditScore", "Female", "Male", "Age", "Tenure", "Balance", "NumOfProducts", "HasCrCard", "IsAc

    # Make the prediction
    prediction = random_forest.predict(user_input_df)

    # Display the result
    if prediction[0] == 1:
        print("The customer is likely to churn.")
    else:
        print("The customer is not likely to churn.")

# Assign the button click event handler
predict_button.on_click(on_button_click)

# Display the input widgets and the button
display(credit_score, age, tenure, balance, num_of_products, has_cr_card, is_active_member, estimated_salary, gender, predict_button)
```

Code: Demo

Conclusion

Our project and the model focused on whether the customer is likely to churn or not rather than why the customer is churn. The selected model, Random Forest Classifier, shows accuracy of 86.4%, which is reasonable. In the process of bias and fairness tests, it seems like there is a possible bias for male customers who churn. Therefore, we need further investigation with the bias because we cannot simply be sure that this model is biased with this data only. Moreover, we need to find a way to raise the accuracy for class 1 (churn) because the accuracy of class 1 is significantly low. We could guess that this issue is because the percentage of people who churned on train and test data sets is extremely lower than the percentage of people who didn't churn. However, further investigation is needed as well because we cannot make a decision with this single result and a guess.

This model tells you whether the user will churn or not, but it does not reveal the factors that affect the churn. If the bank wants to identify what makes the user churn after realizing that many of their customers are expected to churn using our current model, we should find another dataset with much more features that can affect the customer churn. In this process, our model can be used in the evaluation of the new models that can find the main features that make the user churn.

Reference

Francis, B., Hasan, I., Huang, Y., & Sharma, Z. (2012). Do banks value innovation? evidence from US firms. *Financial Management*, *41*(1), 159–185. https://doi.org/10.1111/j.1755-053x.2012.01181.x

Code

Github repository: https://github.com/starJin2003/Churn_for_Bank_Customers_Analysis.git

All information/code used for the project is in the github. Read the Readme file please.