

# CLASSIFYING BANK CHURN

Wairimu Kimani



# PROJECT OVERVIEW

## INTRODUCTION

- ❖ Customer churn is a challenge especially in the banking sector due to several issues like competition. Customer churn presents a problem because retaining customers is cheaper than acquiring new ones. In this case, we study the churn of bank customer. By identifying the type customers that are most likely to cancel the bank account, a bank can design campaigns to ensure they retain them. Customer churn refers to the loss of customers for any reason at all. To prevent customers from churning, we can use data analysis to identify potential features and design strategies to retain customers from leaving. It is also noted that an already loyal customer is less likely to churn and easier to retain than try gaining a new customer. Customer churn poses a challenge for banks, as it leads to mistrust and profit loss. To address this, a data-driven solution using classification model is being explored. This predictive model aims to revolutionize customer retention by enhancing transparency, efficiency, and informed decision-making.

# CHALLENGES

- Pricing strategies across different banks
- Competition
- Compliance requirements

# SOLUTIONS

- Create flexible pricing models
- Create special and unique products and services to set the company apart
- Establish strong framework for regulatory compliance to anticipate and respond to changes ensuring smooth adherence and minimizing operational distractions

# PROBLEM STATEMENT

- ❖ Banks have a hard time to retain customers due to challenges such as different pricing strategies. Retaining customers is very crucial because it is cheaper than trying to acquire new ones. Their concern is trying to identify customers who are about to churn. The current lack of a model that can classify customers can result to losses. To address the challenges , I aim to develop a model that can predict the likelihood of user churn in the near future. This will help the bank come up with specific targeted strategies to minimize customer churn and hence enhance their relationships with a customer hence continue making profits.

# OBJECTIVES

- ❖ To identify the top 3 features that have the most impact on customer churn
- ❖ To identify the customer behaviors that indicate a user is about to churn
- ❖ To build a predictive model that forecast the likelihood of customer churn in the bank
- ❖ To derive actionable recommendations for reducing churn based on insights from data analysis

# DATA UNDERSTANDING

## DATA SOURCE

The dataset is a csv file-  
[https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn?source=post\\_page-----fa6e2324c245-----](https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn?source=post_page-----fa6e2324c245-----)

=

## DATA COLUMNS

Credit Score—can have an effect on customer churn, since a customer with a higher credit score is less likely to leave the bank.

Geography—a customer's location.

Gender—Gender of customer

Age—Age of customer

Tenure—refers to the number of years that the customer has been a client of the bank.

## DATA COLUMNS

Balance—Amount of money in the bank for a customer

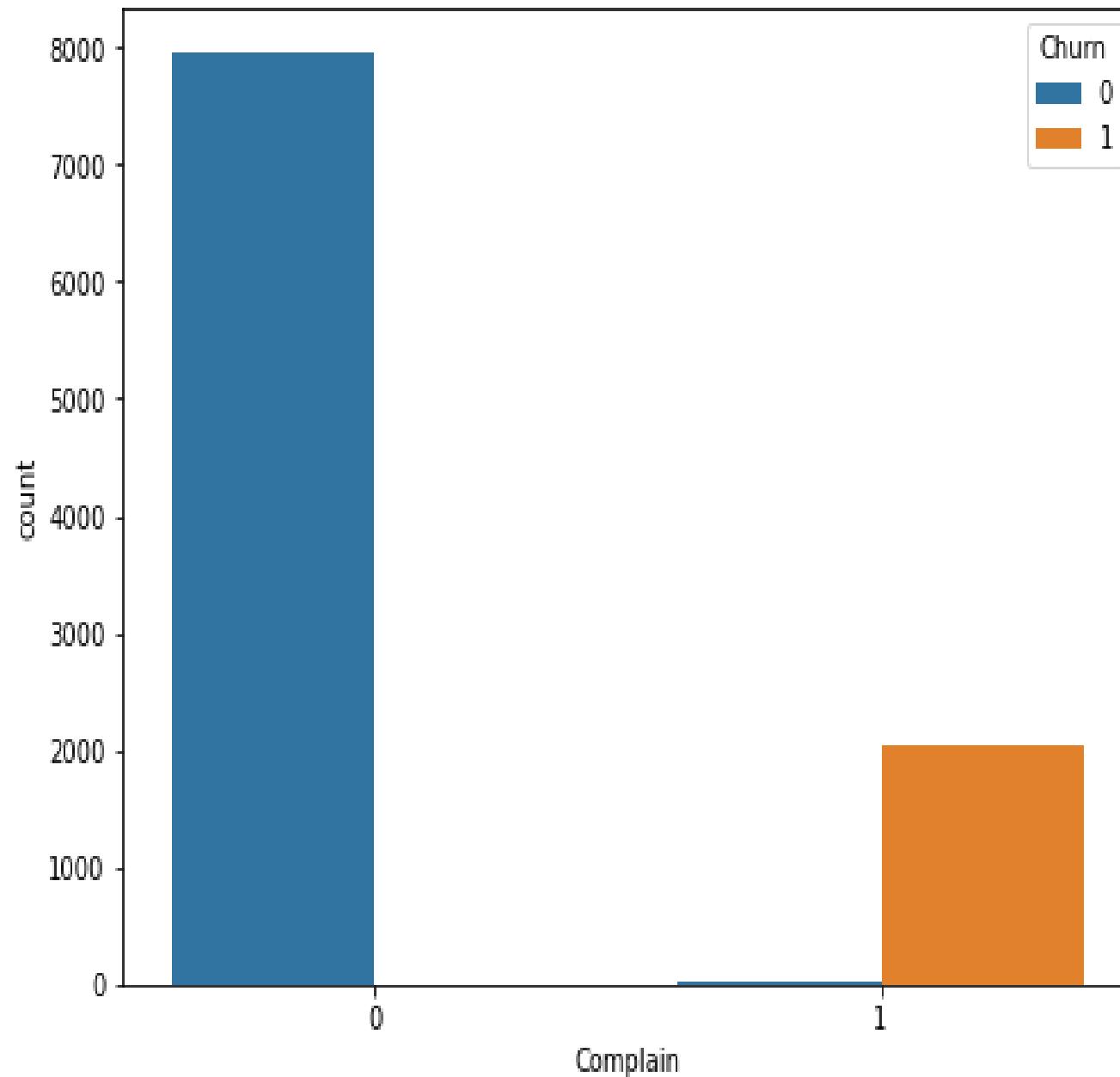
NumOfProducts—refers to the number of products that a customer has purchased through the bank..

IsActiveMember—1 is for active member, 0 is for not active member

Exited—whether or not the customer left the bank.

Complain—customer has complaint or not.

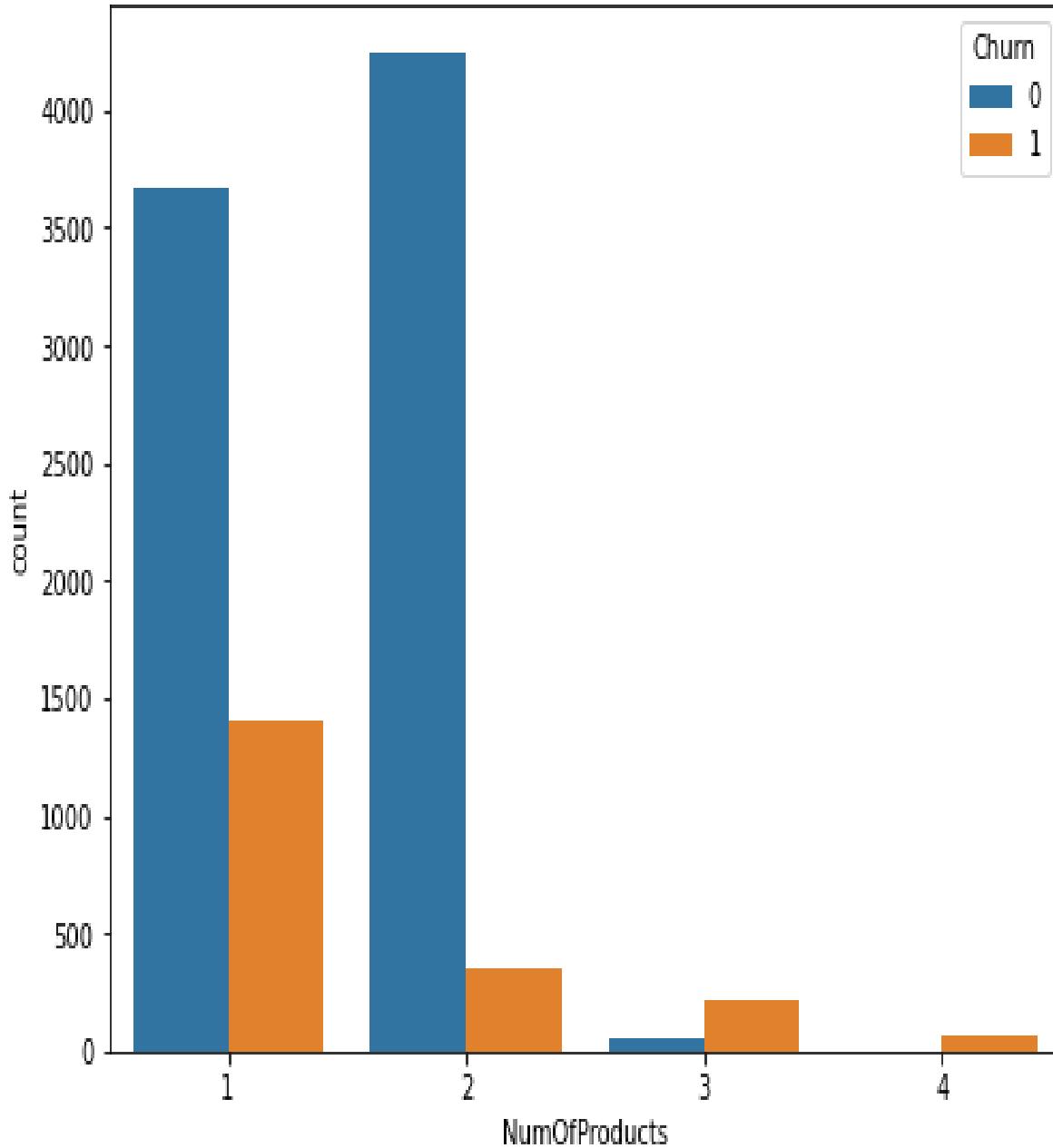
Churn by Complain



**DATA ANALYSIS- COMPLAINS**

From customers who complain, 99 % of them churn. This indicates that when a customer starts to display behaviors such as complaining, it is an indication that they are about to churn. When a customer complains, banks should listen and implement the changes in order to retain customers

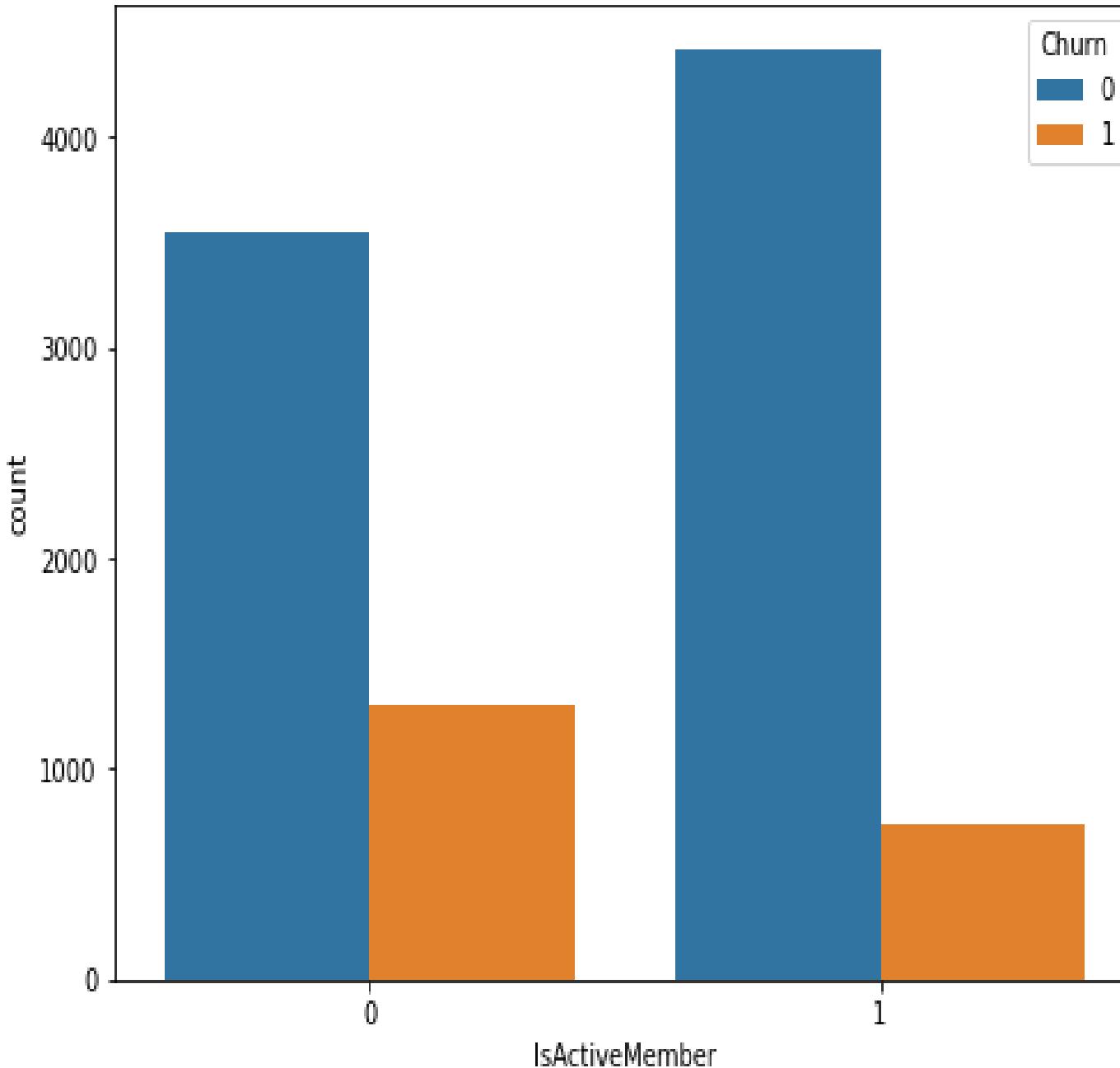
Churn by NumOfProducts



## DATA ANALYSIS- NUMBER OF PRODUCTS

When it comes to the number of products used, the more the number of products used, the less they are likely to churn. Customer who only use one product have a very high rate of churning. This behavior of customers helps us to identify customers who are at a high risk of churning. Banks should come up with strategies to ensure their clients use most of their products.

Churn by IsActiveMember



## DATA ANALYSIS - CUSTOMER ACTIVITY

Customers who are not active members have a higher chance of churning . This indicates that the banks should pay attention to customers who are not active and try to strategize.

# MODELLING

Recall: Recall calculates the proportion of correctly identified churned customers among all actual churned customers. This is particularly critical for a bank, as failing to identify actual churned customers could result in significant revenue loss.

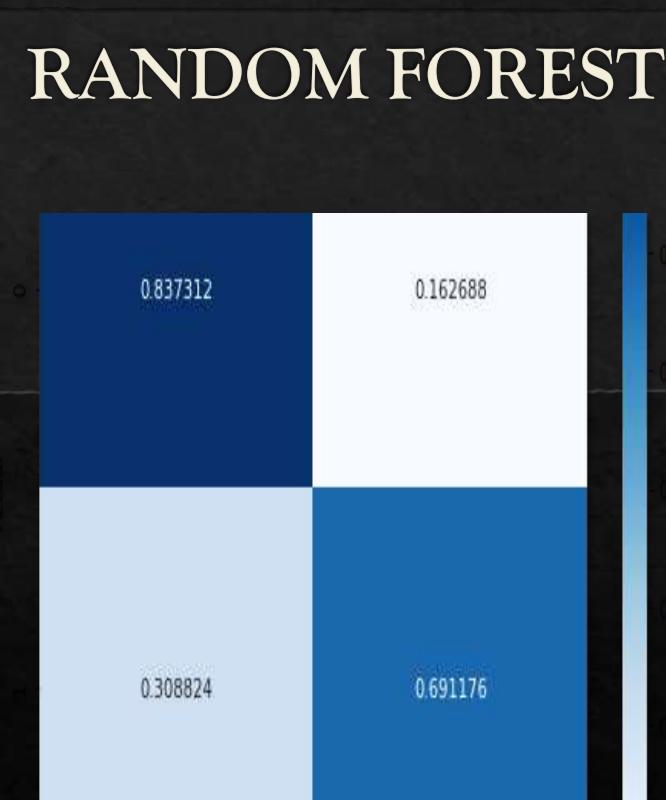
DECISION TREE



Recall Score : 67%

Accuracy of the model: 74%

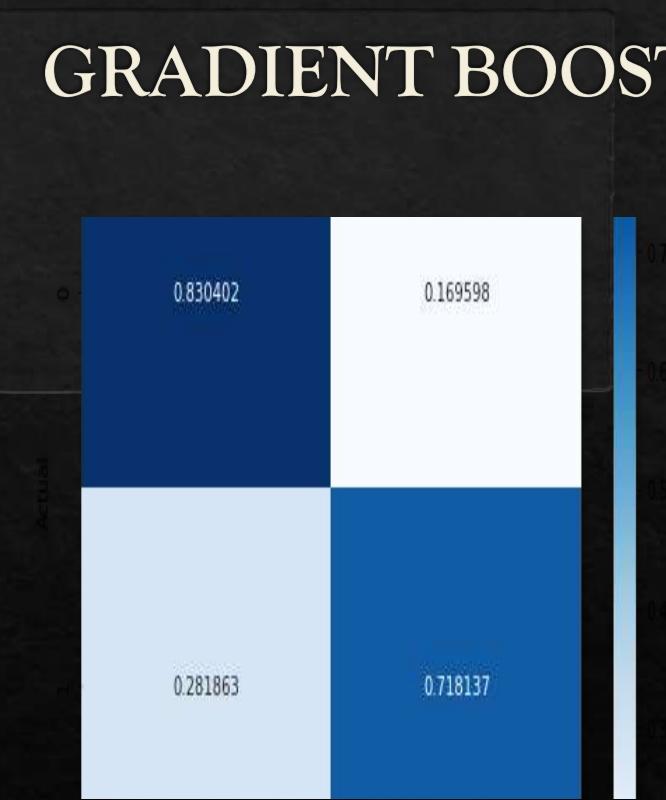
RANDOM FOREST



Recall Score: 70%

Accuracy of the model: 81%

GRADIENT BOOST



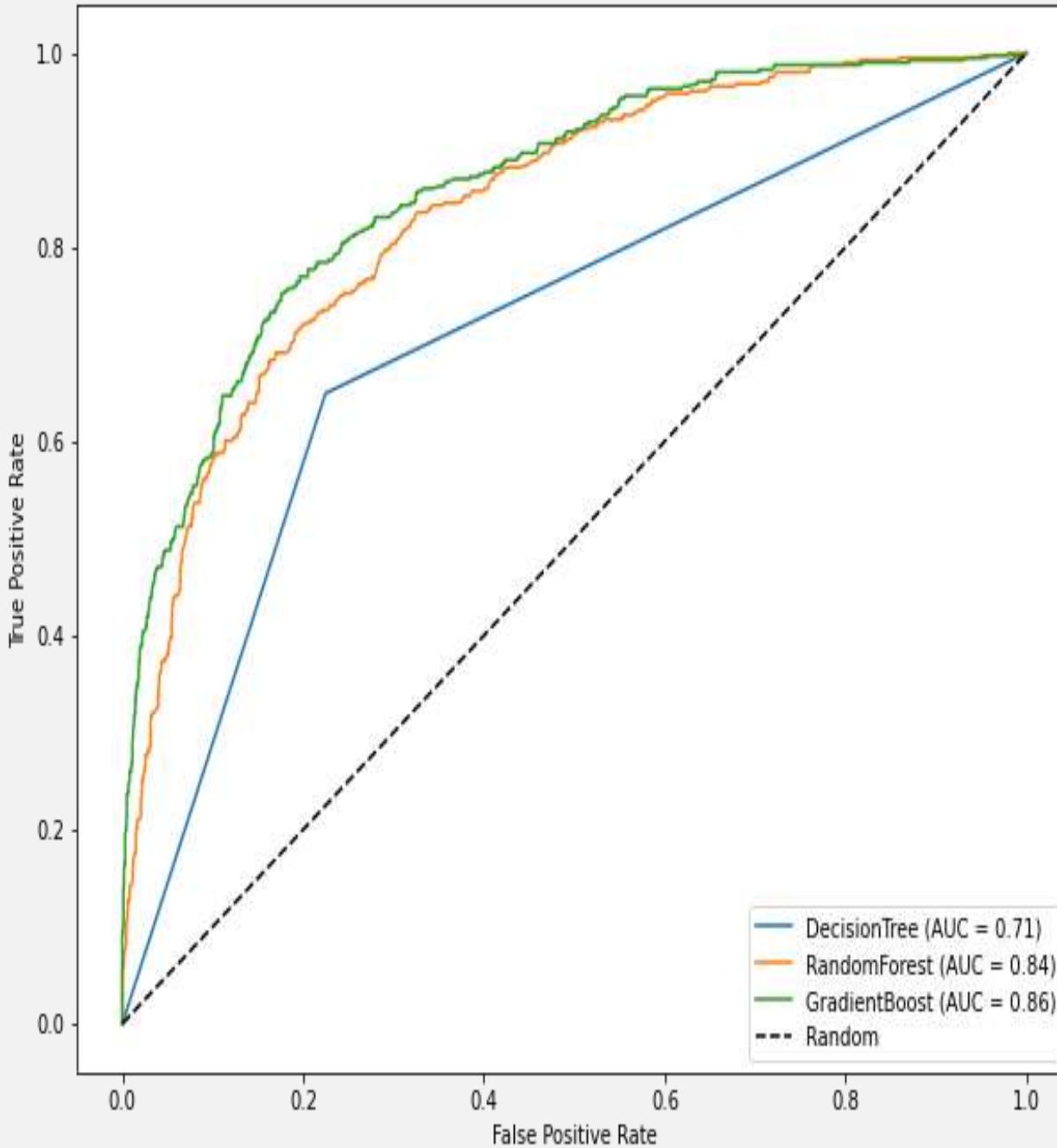
Recall Score : 72%

Accuracy of the model: 81%

# EVALUATION

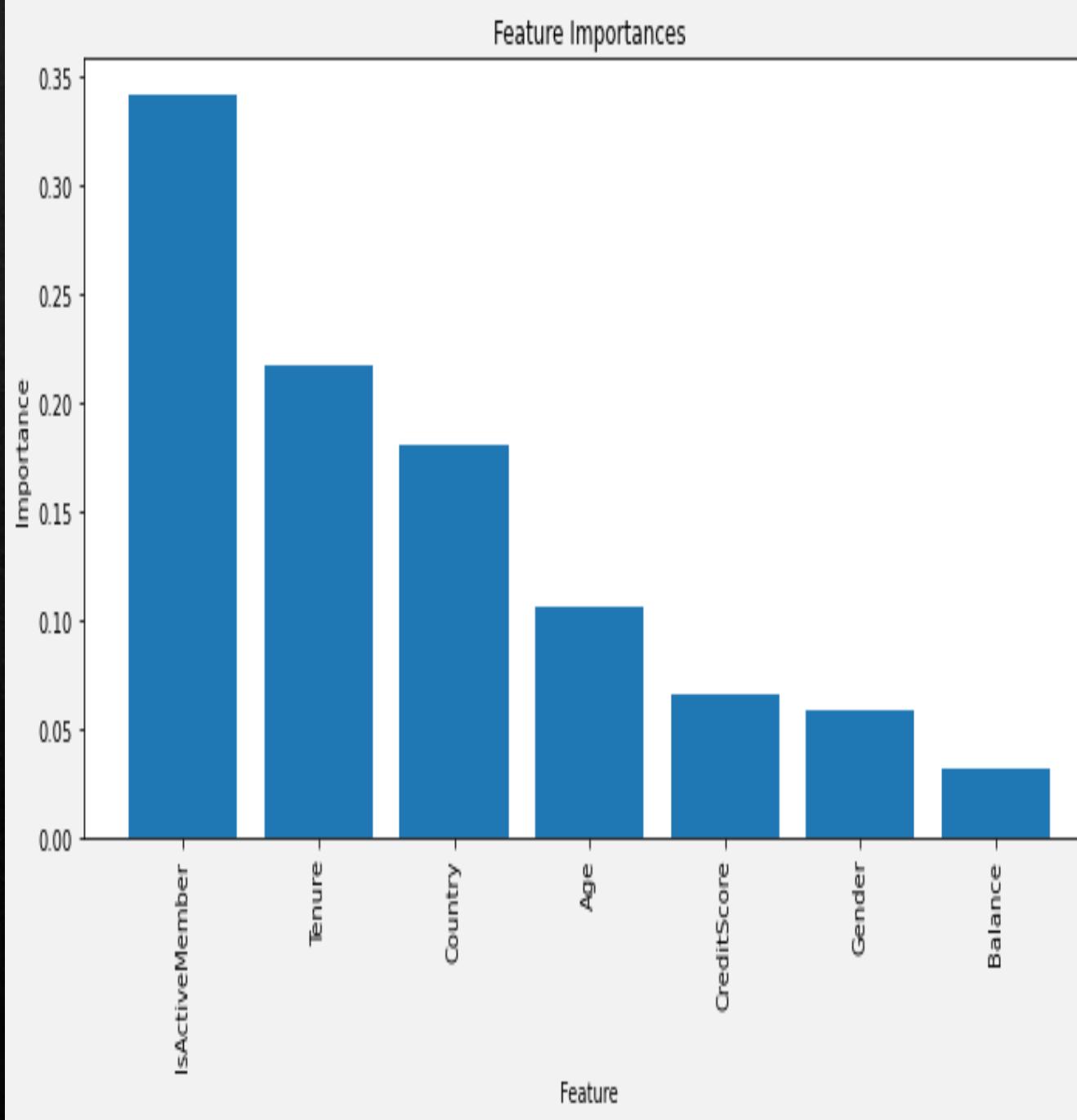
- ❖ Recall: Recall calculates the proportion of correctly identified churned customers among all actual churned customers. This is particularly critical for a bank, as failing to identify actual churned customers could result in significant revenue loss.
- ❖ Maximizing recall helps to capture the majority of the customers who churned. This can now help the bank to take pro active measures .

ROC Curves for Custom Decision Trees



## COMPARING MODEL PERFORMANCE

Based on our ROC curve we can conclude that gradient boost model is our best model since it has an AUC of 0.86 respectively and it is hugging the top left side of our graph. It also has a high recall when predicting a customer will churn as compared to the other models



## FEATURE IMPORTANCE

The most important features:

- Whether a customer is active
- The tenure of the customer
- The country

The bank should pay attention to these factors as they have a lot of importance when it comes to churning

# CONCLUSION

In conclusion, the customer churn prediction model aims to provide valuable insights into customer behavior in the context of the banking industry. Given the financial implications of customer churn, it's crucial to refine the model to accurately identify customers at risk of leaving. Throughout the iterative process, various changes were applied to enhance model performance. These include adjusting hyperparameters, building multiple models, and addressing imbalanced data to improve the model's predictive ability. The rationale behind these alterations was to iteratively fine-tune the model for better accuracy and recall. These results indicate that the gradient boost model performed well in terms of accuracy and recall.

# RECOMMENDATIONS

- ❖ To provide more targeted re-engagement campaigns for at-risk customers, these are clients who complain, are not active and use few bank products
- ❖ Create more focused customer education content to increase customer lifetime value- these include creating different pricing strategies to target customer.
- ❖ Implement retention strategies like personalized onboarding and interactive walkthroughs to encourage product and service adoption
- ❖ Segmentation: Segment users based on their session data, feature usage data, feedback, and more, and tailor their experiences to help them get the most out of their service
- ❖ Offering loyalty programs: Offering incentives and rewards for customers who stay with the provider for a longer period of time can reduce churn

# NEXT STEPS

- ❖ Further Data Collection: collecting additional data to enhance the model's predictive power.
- ❖ Feedback Loop: Continuously monitor the model's performance and incorporate feedback from the business domain experts to iteratively improve it.