**Bank Churn Classification Model**

**About Project**

This project involves analyzing the customer churn rate in order to understand why customers leave and then building a classification model to predict whether a customer will churn out or not so that proper actions can be taken to prevent this churn.

**Business Problem**

Customer retention is as crucial as customer acquisition when it comes to increasing revenue. Also we know, it is much more expensive to sign in a new client than keeping an existing one. So, churn rate needs to be minimized.

To do so it is necessary for banks to know what leads a client towards the decision to leave the bank. Also churn prediction allows companies to develop loyalty programs and retention campaigns to keep as many customers as possible

**Data**

Link to the data : <https://www.kaggle.com/datasets/mathchi/churn-for-bank-customers>

The data is provided by Kaggle and has 10,000 rows and 14 columns as follows :

1. RowNumber—corresponds to the record (row) number and has no effect on the output.
2. CustomerId—contains random values and has no effect on customer leaving the bank.
3. Surname—the surname of a customer has no impact on their decision to leave the bank.
4. CreditScore—can have an effect on customer churn, since a customer with a higher credit score is less likely to leave the bank.
5. Geography—a customer’s location can affect their decision to leave the bank.
6. Gender—it’s interesting to explore whether gender plays a role in a customer leaving the bank.
7. Age—this is certainly relevant, since older customers are less likely to leave their bank than younger ones.
8. Tenure—refers to the number of years that the customer has been a client of the bank. Normally, older clients are more loyal and less likely to leave a bank.
9. Balance—also a very good indicator of customer churn, as people with a higher balance in their accounts are less likely to leave the bank compared to those with lower balances.
10. NumOfProducts—refers to the number of products that a customer has purchased through the bank.
11. HasCrCard—denotes whether or not a customer has a credit card. This column is also relevant, since people with a credit card are less likely to leave the bank.
12. IsActiveMember—active customers are less likely to leave the bank.
13. EstimatedSalary—as with balance, people with lower salaries are more likely to leave the bank compared to those with higher salaries.
14. Exited—whether or not the customer left the bank.

**Approach**

**+ Task1 -> Analyze the customer churn rate for bank**

1. Applied data pre-processing and preparation techniques in order to obtain clean data.
2. Univariate and Bivariate Analysis was done. Here the main interest was to get an understanding as to how the given attributes relate to the 'Exit' status.
3. Inisghts Obtained are as follows:

* 40-50 is the age group for which churn rate is maximum. Also for age group 50-60 customers churned is more than customer retention.The bank may need to review their target market or review the strategy for retention between the different age groups.
* Chances of a non-active member churning out is double the chances of an active member churning out. Bank needs to check with the customers for being inactive , give them some offers plan.
* Majority of the customers that churned are those with credit cards. Given that majority of the customers have credit cards could prove this to be just a coincidence.
* Almost half of the customers are from France followed by Spain and Germany each having 25% customers.Hence churn rate is also maximum for France there but Germany showed the same churn rate despite lower count of customers.
* Male to female ratio of customers is 5:4.Clearly as observed churn probability is more for a female customer.
* Probability of customers having zero balance churning out is maximum.Many people keep 0 balance no matter how high or low their estimated salary is.
* Around one-fourth customers have salary between 150000-175000 and more churn rate for these customers is observed which shows bank is losing it's valuable customers.
* Maximum customers who churned out have used 1 product only (14% out of total 20%). Maybe bank needs to focus on convincing customers to use more of their services and products.¶

**+ Task2 -> Predictive behavior modeling i.e. to classify if a customer is going to churn or not**

1. Splitting data into train and test data in 80:20 ratio
2. Building and training four classification models on the 80% training split: Logistic Regression, Decision Tree, Random Forest and XGBoost that will attach a probability to the churn to make it easier for customer service to target right customer in order to minimize their efforts to prevent churn.
3. Making predictions from the model
4. Testing the performance of the model using performance metrics like Precision, Recall, F-1 score and AUC

**+ Task3 -> Choose the most reliable model**

1. For Decision Tree, Random Forest and XGBoost models, GridSearchCV was used to iterate through relevant parameters and refit the best estimator using 10-fold cross validation. Performance for the latter 2 models was similar.
2. I chose the XGBoost model as it had a high AUC of 0.77, a strong recall of 0.62 and higher precision of 0.69. Also it shows the highest F-1 score of 0.66.