**Churn Prediction Approach Report**

Steps Taken:

1. Imported the datasets (train and test data).

2. Data Exploration and Visualization: to understand the data. It was seen that Age and Balance columns contained outliers.

3. Created Age categories for the age column such as GenX, baby bloomers etc

4. One-Hot Encoding: Some of the values such as Gender, Income, Credit Category, Age Category were in like Male, Female, Income more than 5L etc, used Pandas Get Dummies function to encode them.

5. Used the rename function to avoid the dummy variable trap.

6. Split the Data in Test and Train split: The test ratio was 30% of the dataset, while 70% was training Data.

7. Feature Importance: used Random Forest Classifier to know the features that affected the model prediction more.

8. Standard Scaling: Variables such as Age, Balance were in a very different range as the other variables were in between 0 to 1, scaled the dataset to get them in the same range.

9. Model Selection: Build Model Using Logistic Regression, Gaussian Naïve Bayes, Decision Tree Classifier, Random Forest Classifier, XGB Classifier, Support Vector Machine individually.

Performed Model Selection according to the Model which perform best Prediction Score with Best F1 (Macro) Evaluation Metrics.

And for Model Selection, Selecting the Model with High Prediction Score on Test Data which is Gaussian Naïve Bayes.