Credit Card Customer Churn Prediction

Anusha S Rao
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
ctckm7@gmail.com

Moyank Giri
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
moyank110@gmail.com

Moulya T
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
moulyat9353@gmail.com

Ganesh Tarun S R
Computer Science Engineering
PES University,BSK – 3 Campus
Bengalure, India
ganeshtarunsr101@gmail.com

I. PROBLEM STATEMENT

The primary trouble this paper intends to address is to investigate and give you a predictive version for predicting Credit Card Customer Churn so as to aid corporations tormented by patron churn such that they could hit upon Churning Customers as early as feasible and might take essential movement so as to maintain the Churning Customers

II. INTRODUCTION

Banks fall below the Service Sector have a big huge style of customers interacting with them on daily basis. Banks gives many services which consist of debit cards, credit score cards, ATM, net banking, UPI transactions etc. Credit card client churn isn't an uncommon vicinity problem faced via way of means of any economic group wherein the clients start to depart the credit score card carrier furnished via way of means of the economic group due to multiple reasons and factors. Predicting the motives of the churn is critical as it permits the banks in identifying which client is possibly to churn and what measures can be taken to meet the desires of the client. Also, this enables to enhance the quality of service for clients leading to the retention of the clients. This prediction is crucial for the banks due to the fact that the clients are prioritizing quality of service furnished and the primary impediment for each financial institution is the opposition among them. Customers have all of the liberty to pick any of the offerings provided through any banks primarily based totally on how different factors consisting of interest rates ,customer-friendliness of the the financial institute is in terms of vicinity, technology, new upgradation of financial institution etc. and diverse other factors. It will become the duty of the banks to maintain their clients via diverse agendas for which predicting the reasons for churn will become the root. The churning of the clients isn't simply bounded to financial institution however is proliferated in nearly all of the provider sectors. The important cause for this may be the freedom possessed by the clients or the opposition among the service providers. Churning can be viewed in many ways such as it might be the number of customers dropped, ratio or percentage of customers lost in comparison with total customers a bank has and so on. Churn can be calculated on a period of a particular quarter or annually .

For Banks or any service sector , customer churn rate prediction is critical to long-term success rate. Prediction of customer churn which is accurate drives many facets. Therefore, drastic improvements in terms of total revenue of the bank can be achieved by slight improvements in accuracy of prediction . Another considerable challenge is the data we obtain to predict the model which may have missing data, inconsistent data or may be the feature selection which accounts for the timely data. The way we train and test the model also plays important role in contributing to the accuracy and precision of the model. Initially many credit card customer churn prediction models have existed, these models may not address the growth in complexity of the present world or may not be accurate enough to meet the needs of present requirements due to advancement in technology which may have led to different problem scenario. Present day models make use of automation algorithms such as random forest classifiers, SVM (support vector machine), Gradient Boosting and many more can be implemented to create extremely accurate models.

The dataset used is taken from Kaggle website. Thus, summarizing all the above points, the main aim of the project is to use effective methods to build an accurate model which can have multiple benefits to a bank or any such organization which suffers from customer churn. According to some statistics provided "A bank can increase its profits by up to 85 % by improving the retention rate by up to 5 % ". It is also much more cost effective to retain a customer than attracting new customers.

III. KEYWORDS

Churn Prediction; Logistic Regression (LR); Decision Tree (DT); Random Forest Classifier (RF); Support Vector Machine (SVM); SMOTE;

Oversampling; Under sampling; XGBoost; KNN; Hybrid approach; Naive Bayes Classifier.

IV. RELATED WORK (LITERATURE SURVEY)

Hybrid data mining models approach [1]: This paper describes about working on UCI data .Variations in the data was normalized . Using unsupervised methods like K-means and rough k-means algorithms the data is divided into different clusters. The performance is measured in terms of precision, sensitivity, specificity, accuracy, and misclassification error. The insights from this paper is hybrid system works well when compared to single classifier model. Higher accuracy and lower misclassification error over using a single model was the result from Hybrid Model. Of all the hybrid models SVM combined with rough k-means clustering algorithm works well with better accuracy.

One Class support vector machine (OCSVM) based undersampling [2]: This paper reflects the work on Automobile Insurance fraud dataset and Credit card customer churn dataset. Decision Tree (DT), SVM, LR, Group Method of Data Handling (GMDH), Probabilistic Neural Network (PNN) approaches are used for the classification. Undersampling with the Radial Basis kernel for the dataset with respect to DT yielded significant results. The paper recommends use of DT and yields "if-then" rules a in comparison with other classifiers . It also demonstrated OCSVM based undersampling. Proposed Under-sampling methodology reduced the complexity of building the system and at the same time, yielded significantly accurate results. Also, the paper gives the insight of preferring DT over SVM as there is no statistically significant difference between the two.

Data mining techniques to predict the churn: This paper works on 97% non-churned and 3% churned dataset and have used SMOTE and also undersampling ,oversampling to balance it. The paper is inclusive of LR, RF, SVM as the constituents of the model. Through experimentations the best results were concluded when the unbalanced original data is SMOTED, RF was implemented and for combination of undersampling and oversampling [3].

In Guangli Nie and team's churn prediction of credit card in China's banking industry, the paper focuses on the execution and the understanding of the model rather than building a new one [4]. The paper proposes the development of a criterion measure called misclassification cost. logistic regression and Decision tree classification model are being proposed in the paper . Some selected variables shows that the demographic information makes little contribution to the churn prediction. This idea can be implemented in our model also. The test results shows that LR performs better than the DT. The concept of Multicollinearity has not been discussed here.

Machine Learning approach to resolve gap of churn and non-churn customers: The paper infers on accuracy levels that can be achieved by classifiers. A novel approach KNN is proposed for grouping the data into training and testing sets depending on weighted scales along with XGBooster algorithm [5], aiming for high accuracy in model. The experiment concludes that XGBoost gives the best result in terms of accuracy, sensitivity and specificity. This brings the model designer to the conclusion that XGBoost can be used while building the model for accurate results and better forecasts.

Telecommunication sector's customer churn can be used as backing to our scenario. Hybrid approach [6] discussed in the paper focuses on hybrid methodology rather than the non-hybrid ones to increase the accuracy of the classifiers. Algorithms such as LOLIMOT and C5.0 are proposed in the paper. Approaches such as ANN and ICA is proposed for building a better schema. The paper concludes with a conclusion that the number of features and their subsets majorly affect the prediction accuracy. Thus, better feature selection leads to better model development is one of the major inferences.

This paper describes a study on bank customers churn in India [7]. It speaks about converting raw customer data into meaningful and useful data that suits modelling. It deals with two techniques of classifications namely CART and C 5.0. While CART yielded 95.01% classification rate on training data and 91.22 per cent on test data, C5.0 yielded 69.3% and 68.4% classification rate on training data and on test data respectively. The results obtained on Churned class by CART is quite high but C 5.0 showed not very significant results in predicting churned customers. Also, the paper also infers on future predictions of churns by formulating intervention strategies.

Rule extraction from SVM is another proposed method [8]. This paper speaks about Naïve Bayes Tree (NBTree) resulting in the SVM + NBTree hybrid. The data set consists of 93.11% non-churned and 6.89% churned customers. Using the original unbalanced data only the observation proposed hybrid SVM + NBTree yielded the best sensitivity compared to other classifiers. The paper recommends that it is better to support vectors and use case-sp (sensitivity-68.33%, specificity-74.38%, accuracy- 75.18%) to generate rules

Deep Learning model [9] is also one of the proposed solutions. Techniques such as LSTM can be employed for time series data prediction. Bi-LSTM allows sequence time step information in both forward and backward direction. Customers transaction details are created as features and are passed to the model. For each feature Recency,

frequency, and monetary features are extracted and the model is allowed to learn from the pattern. The paper gives the insight on improved churn prediction. The idea of Down sampling and up sampling is being implemented for raw data set using REHC and SMOTE. Results shows better performance when compared with other deep learning models.

V. PROPOSED SOLUTION

Initially, in order to comply with the requirements of numerical columns for model training and prediction, we initially convert the categorical variables into one hot encoded vector such that each numerical value represents one unique value for each attribute. For example, the dataset had a column for gender which was split into columns namely 'Gender_F' and 'Gender_M' which for a male customer would have '01' whereas for a female customer would have '10' respectively.

To further help our cause to provide reliable and accurate predictions, we make use of the Min Max Scaling Algorithm in order to scale every numerical column into a common range.

Upon using this scaling algorithm all the numerical column were bought into a range between 0 and 1. This was mainly done because some of our implemented algorithms made use of distance measures and hence scaling ensures that all features contribute approximately proportionately to the final distance measure.

80% - 20% was the division made in the data set to train and test the model parts respectively. This was done so that we have ample amount of data for both training and testing the model. The 80-20 is most commonly used split amount and this is optimal split amount such that there is very less chance for either over-fitting or the under-fitting.

Classification techniques like Decision tree, Random Forest Classifier, Naïve Bayes Classifier etc yield accurate results when worked upon individually or when combined together. From the literature survey made we can infer those different classifiers have yielded different accuracy. This shows that the data set we use also has the impact on the classifiers we tend to implement.

Our Approach intends to make use of a voting classifier model for prediction which intends to provide predictions based on multiple models rather than just using a single model.

The main advantage of using this is that this model in general can give high accuracies while utilizing simpler models for its prediction. Making use of simpler models also helps in increasing the speed at which predictions are made while improving upon accuracy of the prediction if used only a single model. It is "vote" based method where results from all models are taken and integrated to produce one better result . Our voting model approach consists of several models consisting of:

1. AdaBoost
2. Gradient Boosting Classifier
3. Random Forest Classifier(RF)
4. Extra Trees Classifier
5. Decision Tree Classifier(DT)
6. Support Vector Machine(SVM)
7. Bagging Classifier

Here are few advantages that can be seen with each of the above models mentioned above

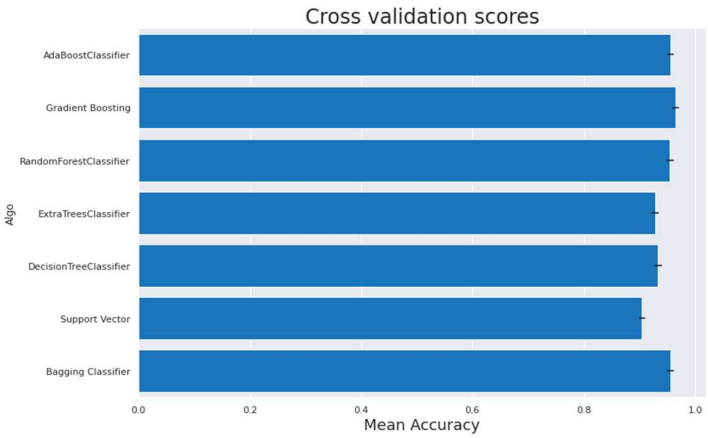
XGBoost: gradient boosting is will mechanically give estimates of feature importance from a trained prognosticative model. XGboost is one in every of the implementations of gradient boosting concept, it uses a lot of regularized model formalization to manage over-fitting, which provides it higher performance. XGB consists of variety of hyper-parameters that can be tuned, has an in-built capability to handle missing values. It provides numerous intuitive features, akin to parallelization, distributed computing, cache optimization and more. once tree is constructed, it retrieves feature importance scores for each attribute. The feature importance contributes a score that indicates how much valuable every feature was within the construction of the boosted decision trees within the model.

Decision Tree: This algorithmic program is one among the supervised learning algorithms which may be used for each classification and regression problems. during this algorithm a training model is formed that predicts the category or value of target variable by learning decision rules gathered from training data. It uses a flow diagram kind of tree structure which shows the predictions from a sequence of feature-based splits. The approach for construction decision tree is sometimes top-down during which a variable is chosen at every step that splits the set of items. In decision tree there are 2 varieties of nodes particularly decision Node and Leaf Node. Leaf Nodes are the outputs of decision nodes and don't contain branches whereas decision nodes may be used to build any decisions and take branch.

SVM: This is supervised machine learning algorithmic rule that may be used for each classification or regression challenges. SVM has method called the kernel trick. The SVM kernel function takes low dimensional input space and transforms it to a higher dimensional space. SVM offers sensible accuracy and perform quicker prediction. They conjointly use less memory as a result of using a subset of training points within the decision phase. SVM works well high dimensions. It constructs a hyperplane in an iterative manner, that is employed to reduce an error. The core plan of

SVM is to seek out maximum marginal hyperplane that best divides the dataset into classes.
Random forest Classifier: This is a supervised machine learning rule which may be used for each regression and classification. it's an ensemble technique which mixes the results of small decision trees referred to as estimators and combines the predictions of those estimators to provide a lot of correct prediction. These are typically used as black box models as they produce predictions across a large range of data with very little configuration. There are 2 ways for random forest to confirm that the behavior of every individual tree isn't correlated with behavior of different trees within the model. they're bagging and feature randomness. sacking is a process of permitting every individual tree to arbitrarily sample from the dataset with replacement. In Feature randomness each tree in an exceedingly random forest will choose solely from a random subset of features.

For Evaluating the models, we make use of multiple evaluation metrices such as precision, recall, f1-score, accuracy etc. Below is a graph showing the cross-validation score means of every model used:



As it can be seen from the above all models used have accuracies above 90%

Below is a table showing the models cross validation score means and cross validation errors

	CrossVal_Score_Means	CrossValerrors	Algo
0	0.955932	0.005984	AdaBoostClassifier
1	0.964449	0.006362	Gradient Boosting
2	0.954698	0.006984	RandomForestClassifier
3	0.928157	0.006437	ExtraTreesClassifier
4	0.933219	0.006555	DecisionTreeClassifier
5	0.904457	0.005608	Support Vector
6	0.955809	0.006502	Bagging Classifier

We also make use of confusion matrix to display results of classification for various models. Below, we can see the confusion matrix for various algorithms:

```
AdaBoostClassifier
[[ 255   54]
[  34 1683]]

Gradient Boosting
[[ 261   48]
[  19 1698]]

RandomForestClassifier
[[ 236   73]
[  24 1693]]

ExtraTreesClassifier
[[ 183  126]
[  16 1701]]

DecisionTreeClassifier
[[ 240   69]
[  65 1652]]

Support Vector
[[ 160  149]
[  37 1680]]

Bagging Classifier
[[ 271   38]
[  45 1672]]
```

All the above models were combined into a single voting classifier so that we can have highly generalized classification while maintaining high accuracies. Upon combining the model, we get the following result:

	CrossVal_Score_Means	CrossValerrors	Algo
0	0.960006	0.006239	Ensemble Model

This shows that the model is able to perform very well on the data and also is generalizing effectively.

VI. RESULTS

The above analytics of the dataset used provided with various useful insights with respect to the data itself such as

1. There are about 16 % of customers who have churned in the dataset.
2. The proportion of gender count is almost equally distributed
    - 57.2% female existing customers, 52.1% attrited female customers
    - 42.9% existing male customers, 47.9% attrited customers

compared to proportion of existing and attributed customer count (83.9% and 16.1%) which is highly imbalanced.

3. Most of the attributes in the dataset are not normally distributed and are skewed
4. There are few columns which are also multi modal
5. There are various instances where we can observe that there is many strong correlations among various attributes in the dataset
6. For the proportion of churned customers by gender there are 4.2% more female than male who have churned.
7. Customers who have churned are highly educated - A high proportion of education level of attrited customer is Graduate level (30.9 %)
8. A high proportion of marital status of customers who have churned is Married (46.3%), followed by Single (38.9%) compared to Divorced (7.4%) and Unknown (7.4%) status - Marital status of the attributed customers are highly clustered in Married status and Single.
9. Also, the portion of income category of churned customer, it is more around "Less than 40K" income.
10. Among all the models built, we see that, The Gradient Boosting Classifier in our case gave the highest accuracy of 96.69%.
11. The least accurate model obtained was the Support Vector Machine with an accuracy of 90.82%. Hence, it can be concluded that all models are able to predict quite well.

VII. CONCLUSIONS

The project helped us understand potential problem that any bank faces that is, customer churn
We saw that there are various factors which contribute to customer churn. Analytics of these factors and attributes helped in gaining insights over what factors essentially has contributed highly over a churning customer and what have not contributed. Using these insights, we came across, one could essentially understand why a particular customer is churning and what can be done to help improve their predicament. These insights also help to predict customer churning at a early stage and hence a company can use this early warning to reduce their churn rate.
After all these predictions are done it is the responsibility of the bank to take steps on how to avoid the churn .A few ways of these could be to have constant interaction with the customers , provide them with more facilities than their peer competitors and many more because any one who leaves the service will have a reason behind doing so .The motive of bank should not be just increasing their profits but also the benefits of the customers.

This project was done as part of Elective Course: Data Analytics and was done under the guidance of Dr. Gowri Srinivasa. This project was led to its current state by the collective effort of every team member who equally contributed time and effort for the completion of the project.

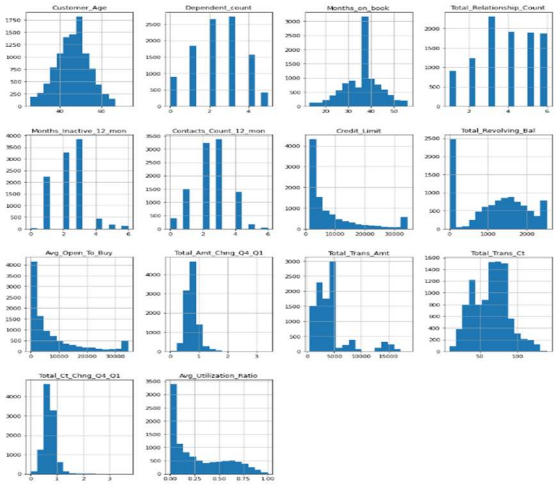
VIII. REFERENCES

[1] R. Rajamohamed ,J. Manoaran , ' *Improved credit card churn prediction based on rough clustering and supervised learning technique*',2018 https://link.springer.com/article/10.1007/s10586-017-0933-1
[2]G. Ganesh Sundarkumar, Vadlamani Ravi, V. Siddeshwar,' *One-Class Support Vector Machine based undersampling: Application to Churn prediction and Insurance Fraud detection* ',2015 https://ieeexplore.ieee.org/abstract/document/7435726
[3] Dudyala Anil Kumar and V. Ravi *, 'Predicting credit card customer churn in banks using data mining'*, 2008 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1086.1056&rep=rep1&type=pdf
[4] Guangli Nie , Wei Rowe , Lingling Zhang , Yingjie Tian , Yong Shi , ' *Credit card churn forecasting by logistic regression and decision tree* ',2011 https://www.researchgate.net/profile/Xingsen-Li/publication/224759968_The_Analysis_on_the_Customers_Churn_of_Charge_Email_Based_on_Data_Mining_Take_One_Internet_Company_for_Example/links/5891e87a92851cda256a0358/The-Analysis-on-the-Customers-Churn-of-Charge-Email-Based-on-Data-Mining-Take-One-Internet-Company-for-Example.pdf
[5] Hemlata Dalmia, Ch V S S Nikil , Sandeep Kumar ,' *Churning of Bank Customers Using Supervised Learning* ',2020 www.researchgate.net/profile/Sandeep-Kumar-249/publication/340855263_Churning_of_Bank_Customers_Using_Supervised_Learning/links/5f54b5ea92851c250b96c697/Churning-of-Bank-Customers-Using-Supervised-Learning.pdf
[6] Elham Jamalian, Rahim Foukerdi,' *A Hybrid Data Mining Method for Customer Churn Prediction* ',2018 https://etasr.com/index.php/ETASR/article/view/2108/pdf
[7] Dr. U. Devi Prasad, S. Madhavi,' *PREDICTION OF CHURN BEHAVIOR OF BANK CUSTOMERS USING DATA MINING TOOLS* ',2012 https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.4094&rep=rep1&type=pdf#page=100

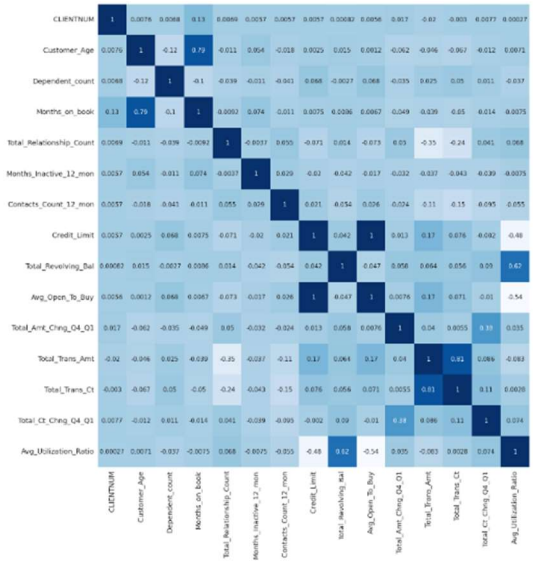
[8] M.A.H. Farquad1, V. Ravi1,S. Bapi Raju,' *Data Mining Using Rules Extracted from SVM: An Application to Churn Prediction in Bank Credit Cards*',2009
https://link.springer.com/chapter/10.1007/978-3-642-10646-0_47
[9] Mr. M. John Britto, Dr. R. Gobinath,' *Improved Churn Prediction Model In Banking Industry And Comparison Of Deep Learning Algorithms* ',2021
http://www.journal-aquaticscience.com/article_133719_d883304ae722b3c710a2722f1f0d3e7a.pdf

IX. APPENDIX

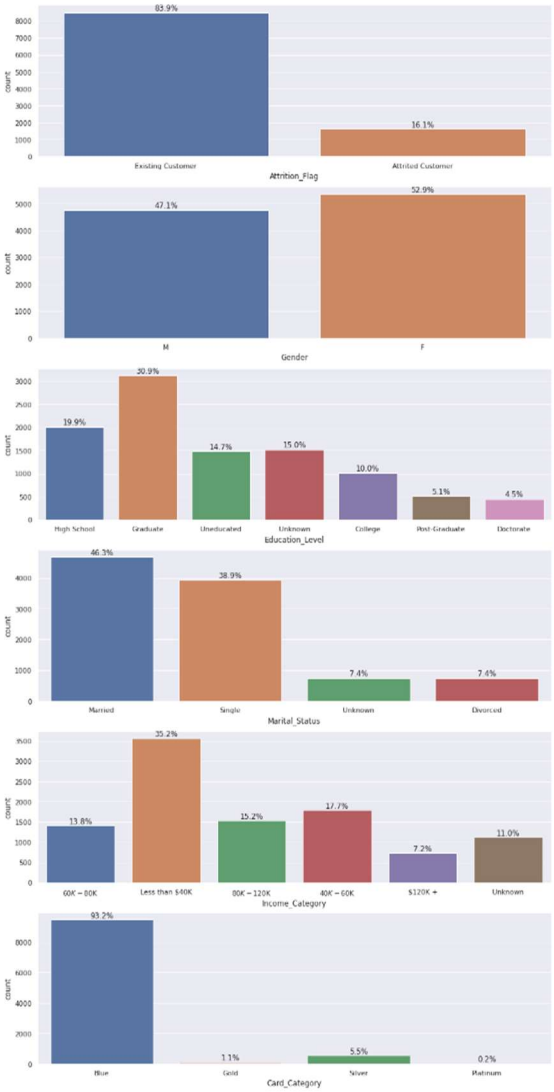
Distribution of every attribute in the dataset



Correlation Analysis of attributes



Categorical Columns Analytics



Classification Report for models used:

AdaBoostClassifier

	precision	recall	f1-score	support
Attrited Customer	0.88	0.83	0.85	309
Existing Customer	0.97	0.98	0.97	1717
accuracy			0.96	2026
macro avg	0.93	0.90	0.91	2026
weighted avg	0.96	0.96	0.96	2026

Accuracy 0.9565646594274433

Gradient Boosting

	precision	recall	f1-score	support
Attrited Customer	0.93	0.84	0.89	309
Existing Customer	0.97	0.99	0.98	1717
accuracy			0.97	2026
macro avg	0.95	0.92	0.93	2026
weighted avg	0.97	0.97	0.97	2026

Accuracy 0.9669299111549852

RandomForestClassifier

	precision	recall	f1-score	support
Attrited Customer	0.91	0.76	0.83	309
Existing Customer	0.96	0.99	0.97	1717
accuracy			0.95	2026
macro avg	0.93	0.87	0.90	2026
weighted avg	0.95	0.95	0.95	2026

Accuracy 0.9521224086870681

ExtraTreesClassifier

	precision	recall	f1-score	support
Attrited Customer	0.92	0.59	0.72	309
Existing Customer	0.93	0.99	0.96	1717
accuracy			0.93	2026
macro avg	0.93	0.79	0.84	2026
weighted avg	0.93	0.93	0.92	2026

Accuracy 0.9299111549851925

DecisionTreeClassifier

	precision	recall	f1-score	suppo
Attrited Customer	0.79	0.78	0.78	3
Existing Customer	0.96	0.96	0.96	17
accuracy			0.93	20
macro avg	0.87	0.87	0.87	20
weighted avg	0.93	0.93	0.93	20

Accuracy 0.9338598223099703

Support Vector

	precision	recall	f1-score	support
Attrited Customer	0.81	0.52	0.63	309
Existing Customer	0.92	0.98	0.95	1717
accuracy			0.91	2026
macro avg	0.87	0.75	0.79	2026
weighted avg	0.90	0.91	0.90	2026

Accuracy 0.9081934846989141

Bagging Classifier

	precision	recall	f1-score	support
Attrited Customer	0.86	0.88	0.87	309
Existing Customer	0.98	0.97	0.98	1717
accuracy			0.96	2026
macro avg	0.92	0.93	0.92	2026
weighted avg	0.96	0.96	0.96	2026

Accuracy 0.9590325765054294
