JOURNAL OF BASIC SCIENCE AND ENGINEERING

# A STUDY ON MACHINE LEARNING TECHNIQUES FOR CUSTOMER CHURN PREDICTION IN TELECOM INDUSTRY

**Mrs.T.Pratheebha[1,*], Dr. S. Santhana Megala[2], Dr.V.Ganesh Babu[3]**

[1]*Research Scholar, School of Computer Studies, RVS College of Arts and Science, Coimbatore & Assistant Professor in Computer Science P.A College of Education, Pollachi,*
*E-mail: prathe21@gmail.com.*
[2]*Assistant Professor, Department of Computer Technology, Hindusthan College of Arts and Science, Coimbatore. E-mail: santhanamegala.s@hicas.ac.in*
[3]*Associate Professor and Head, Department of Computer Science, Government First Grade College, Basavanagudi, Bangalore. E-mail: vgbzone@gmail.com*

**ABSTRACT**

*As the unstable development of the advanced change, which gives clients a more prominent chance to choose from a subscription-based administrations and item approach, it's presently an incredible challenge in foreseeing the behavior of client and to preserve current clients. Since obtaining modern clients is five times more costly than keeping current ones, the issue of client churn, which postures a grave hazard to all businesses, must be tended to. The most objective of this inquire about is to form a novel prescient client churn strategy that will empower the distinguishing proof of planned clients who are most likely to take off, giving earlier notice that will empower therapeutic activity to be taken to keep them as clients. The "XGBOOST," Classifier was appeared to be the foremost viable reply to the client churn issue after we evaluated and inspected the execution of a few tree-based machine learning methods and calculations such as Logistic Regression, Support Vector Machine, ADA Boost, Random Forest and XG Boost classifier. This paper will assist many different companies in understanding the root cause of customer attrition and the actual rate of customer turnover using machine learning, which the business may utilize to reduce the attrition of customer and also build techniques to bring back the customer who has left. Our most significant contribution to the telecom industry is the establishment of a special model for predicting customer churn.*

**Keywords:** Customer Churn Prediction, Support Vector Machine, Logistics Regression, Random Forest, Ada Boost Classifier, XG Boost Classifier.

## 1. INTRODUCTON

With each passing decade, the telecommunications industry's expansion has accelerated. Information is transmitted electronically over distances through telecommunications. The information could be voice conversations, text, images, video, or other types of data. Today, telecommunications is used to connect relatively distant computer systems into networks. Churners are customers who no longer use your product or service for a period of time. In telecommunications companies, people who decide to stop using the company's services are called churn. [1]. Due of their difficulty in acquiring new clients, Companies in the telecommunications sector believe that losing one is very costly. In the telecom sector, keeping loyal consumers always comes at a lower cost than bringing in new ones. So Churn prediction and customer segmentation are the two main parts of customer analytics in the telecom sector. Customer churn management has taken on increasing importance as hiring new clients is no longer a practical approach in a crowded market. To anticipate customer churn, data mining technology is the most widely utilized technique. A show for anticipating client turnover is built utilizing information mining innovation, which too joins choice trees, neural systems, classification regression trees, and other advances. Businesses can create appropriate customer work plans based on this information by using the model to extract information valuable for predicting customer loss from a large number of customer data. As a result, for the purposes of my research paper, the customer prediction model is crucial for identifying primary characteristics of consumers that could lead a telecom consumer to switch providers and anticipate churn in the past.

## 2. RELATED STUDY

There are numerous studies on the focus of customer churn that come at it from different directions and use different datasets, algorithms, and industry-specific statistics. Churn analysis is one of the tool used universally to investigates customer behaviour and spot customers who are about to end their service agreements with a company is churn analysis [2]. In this still developing field, various machine learning strategies have been put forth to address the problem of high customer turnover in numerous industrial sectors. Therefore, the vital aim of our work is to add a novel optimal approach for consumer churn prediction to the simulation method reinforced by typical interpretability and clarification [3].

Suhel Malik, Siddhart Runwal, Yash Shah, Vishal Raut, Prof. S.S. Hire [4] With an attention on machine learning procedures including, random forests, logistic regression, decision trees, and artificial neural networks. It emphases on the difficulties related with forecasting client turnover and examines the contemporary in this area, taking into account recent developments in deep learning and natural language processing. Approximately 7,043 records with 20 attributes were included in the collection. To safeguard consumer privacy, the data was acquired from a telecom company's database and anonymised. With an accuracy of 87% and an AUC-ROC score of 0.91, the model stack of all four classifiers performed well. The most successful model has an AUC-ROC score of 0.91 and an accuracy of 87% for the stack of all four classifiers.

Khulood Ebrah, Selma Elnasir [2], discussed three algorithms Naive Bayes, SVM, and decision trees utilized to forecast churn by comparing two benchmark datasets. The Cell2Cell dataset provides 7,033 observations and 21 attributes, while the observation dataset provides 71,047 observations and 57 attributes. These datasets were acquired via IBM Watson. We assessed the model's efficacy using the Area Under the Curve (AUC) approach. For the IBM dataset, they earned scores of 0.82, 0.87, and 0.77, while for the cell2cell dataset, they received values of 0.98, 0.99, and 0.98. The performance of the model was assessed using the (AUC) Area Under the Curve, and the best AUCs were (0.82, 0.87, 0.78) for IBM dataset and (0.98, 0.99, 0.98) for the cell2cell dataset. In addition, the models that were suggested performed more accurately than previous research on the same datasets.



Vol. 20, No. 1, (2023)
ISSN: 1005-0930

Khulood Ebrah, Selma Elnasir [5], construct a Model Churn Prediction, a variation of learning tactics has been used AdaBoost, random forest, extreme randomized tree, XGBoost, gradient boosting, bagging and stacking, and other gathering learning approaches have all been explored in arrange to discover the most excellent demonstrate for making a Model of Churn Prediction for Users. Moreover, these tactics were associated to more broadly use grouping approaches as logistic regression, k-nearest Neighbour and decision trees. Results Comes about from the ANN and CNN are outflanked from the other strategies on both datasets. Utilizing CNN and ANN, 99 % accuracy on first dataset and 98%, individually on the second dataset, precision was 98% and 99%.

Jitendra Maan, Harsh Maan [6], discussed the purpose of Customer Churn Prediction model. The open source community is where the dataset that can be utilized to calculate customer attrition was found. The information collection incorporates a range of client profiles in expansion to highlights that are broken down over a number of columns and incorporate account level points of interest, membership plans with day/night call charges, and other highlights. The effectiveness of a few tree-based Machine Learning approaches and calculations and found that the Extreme Gradient Boosting "XGBOOST" Classifier was the foremost proficient implies of tending to the issue of client turnover. This would progress the straightforwardness and explainability of the model.

J. Faritha Banu,S. Neelakandan, B.T Geetha, V. Selvalakshmi, A. Umadevi,and Eric Ofori Martinson [7] created an real AICCP-TBM show to categories churned and non-churned clients within the telecom locale. With less computational complexity, the AICCP-TBM demonstrates is planning to improve churn location. Highlight subsets are chosen from pre-processed client information utilizing the CSSO-FS strategy within the proposed AICCP show. Additionally, for churner categorization within the media transmission, the QPSO-FRC approach is connected. A comprehensive comparative precision assessment for the proposed AICCP-TBM procedure was carried out with three datasets. The AICCP-TBM strategy appeared way better execution than the other two strategies, with a most extreme exactness of 97.25%.

V. Kavitha, S. V Mohan Kumar, G. Hemanth Kumar, M. Harish [8], focused to obtain precise numbers and aid in predicting customer attrition, the variety of techniques like Random Forest, XGBoost, and Logistic Regression has been used. Data filtering was used to eliminate all of the null values from the dataset once it was downloaded from Kaggle. In comparison to previous algorithms, we will get greater accuracy by employing Random Forest, XGBoost, and Logistic Regression. Compared to the other procedures, Random Forest RF provided better outcomes and had higher accuracy and performance.

## 3. METHODOLOGY

This study consider about that the creators looked for and examined unused procedures and arrangements to illuminate current troubles in predicting client churn rates based on their own perceptions in a generally comparable environment. Although the suggested solution is conceptually generalized, its actual execution necessitates a special strategy that takes modern technological improvements into account.

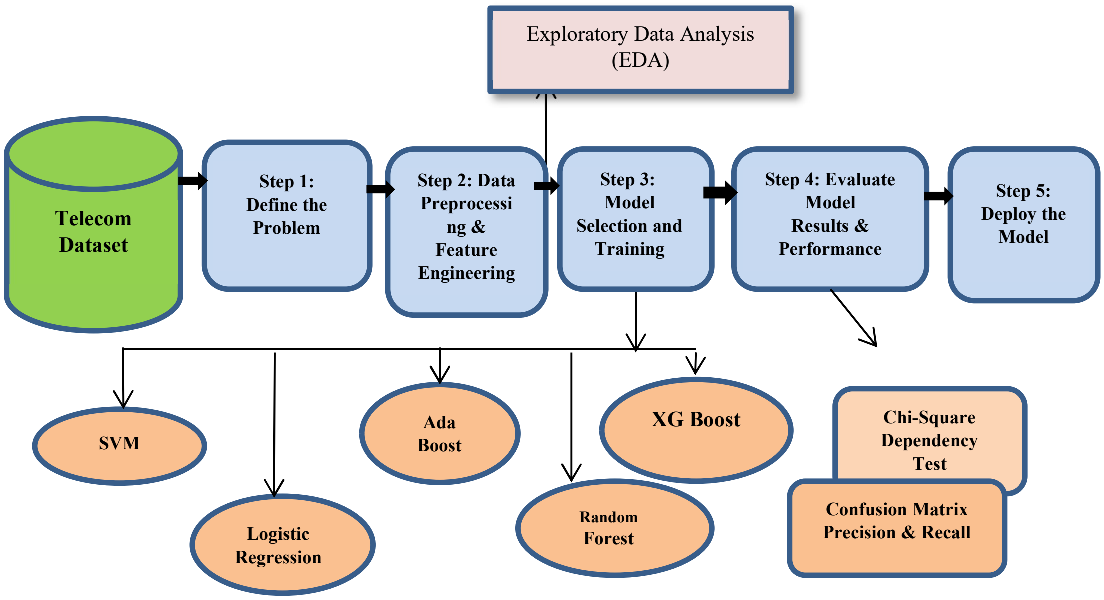



**Fig 1. Architectural Design for Predicting Customer Churn**

### 3.1 Problem Statement

In the Customer Churn Prediction (CCP), every customer is separated into one of two probable behaviors (i) churning or (ii) not churning. Further, customer churn behaviour can be smashed down into two categories, (a) voluntary customer churn, in which a customer decides to stop using the service or even the company and (b) involuntary customer churn, in which the company or service provider opts to terminate the customer's contract. This study emphasizes voluntary customer churns since it is more complicated to forecast these types of customer churns than involuntary ones, which can more easily be filtered out.

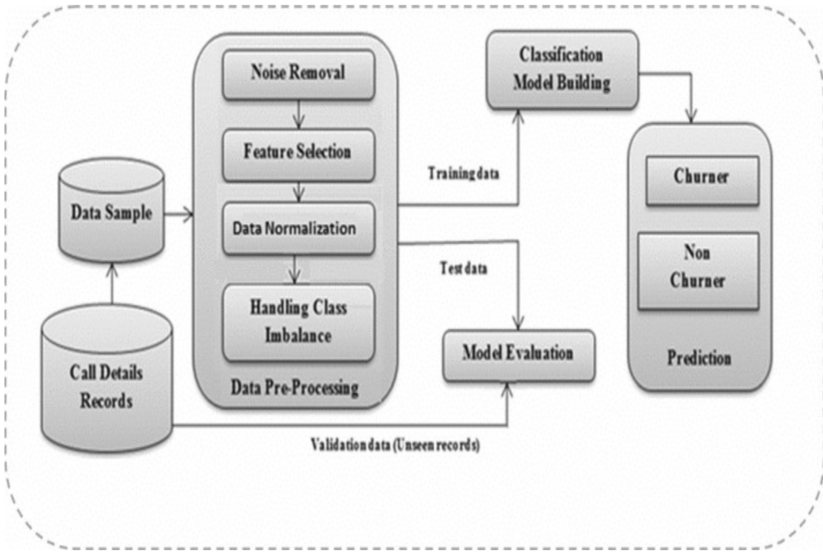



**Fig 2. Process of CCP**

**Dataset:**

In this research, making use of the built-up basic libraries, the data analysis and gathering process starts with the import. The data has certain distinctive qualities, such as traits that facilitate inferences in this paper. The study's raw data comprises 7043 distinct values that reflect consumers and attributes. These values are organized into 21 columns of independent variables that describe the characteristics of the clients. It also has a churn column, which is a response variable that indicates whether or not clients continued to use the business's services in the previous month.

```
telecom_cust.head()
```

| | customerID | gender | SeniorCitizen | Partner | Dependents | tenure | PhoneService | MultipleLines | InternetService | OnlineSecu |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 7590-VHVEG | Female | 0 | Yes | No | 1 | No | No phone service | DSL | |
| 1 | 5575-GNVDE | Male | 0 | No | No | 34 | Yes | No | DSL | |
| 2 | 3668-QPYBK | Male | 0 | No | No | 2 | Yes | No | DSL | |
| 3 | 7795-CFOCW | Male | 0 | No | No | 45 | No | No phone service | DSL | |
| 4 | 9237-HQITU | Female | 0 | No | No | 2 | Yes | No | Fiber optic | |

**Fig 3. Data Columns**

To preserve consumer privacy, the information was taken from the database of a telecom provider and anonymised. Using SQL queries, this paper took the data out of the database and saved it in a CSV file format for later analysis. This paper cleaned the data before beginning the analysis to get rid of any outliers, duplicate records, and missing values. The cleansed dataset was used for exploratory data analysis and to build predict customer churn models.



Vol. 20, No. 1, (2023)
ISSN: 1005-0930

### 3.2. Exploratory Data Analysis

#### A. Preprocessing

One of the primary techniques used in data mining to clean and filter the data is called data pre-processing. By doing so, inconsistencies are eradicated and raw data is transformed into valuable information that can be managed effectively. The data collection must be cleaned up of any null or missing values, and it must also be checked for imbalanced class distributions, an upcoming problem in data mining. Re-sampling techniques, better evaluation requirements, etc. can be employed to deal with the challenge of an uneven data set [9]. In the Data Pre-processing stage of Machine Learning Models, the following tasks are carried out [8].

#### B. Noise Removal

The dataset contains numerous wrong or lost values. As it were the foremost valuable components from our investigation of the whole dataset were recorded. Enhanced precision can be accomplished by utilizing the highlight list, which as it were incorporates valuable highlights. Since undesirable or invalid values may result in less information, making the information useable is vital.

**a. Missing value:** There are no values that are lost within the facts, and it is found among investigation.

**b. Data Type Conversion**: Data sorts are reasonably progressed from Address data sorts (Boolean) to numerical data sorts for effective creation of models. For event, Churn, Vmail Organize, and Intl Organize are changed into the numerical data sort for planning and examination by various classifiers.

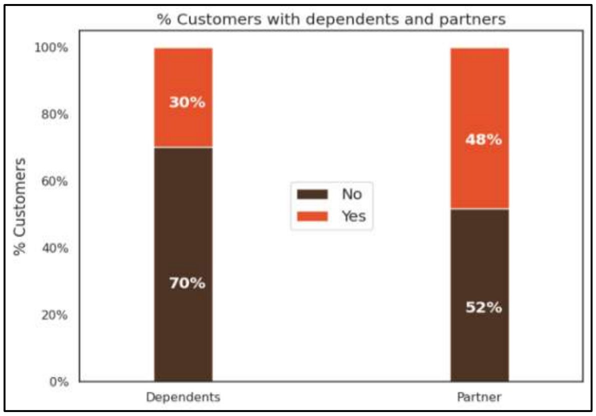
**c. Redundancy of Data:** For the distinctiveness dataset's, we have additionally looked for redundant customer contact information.

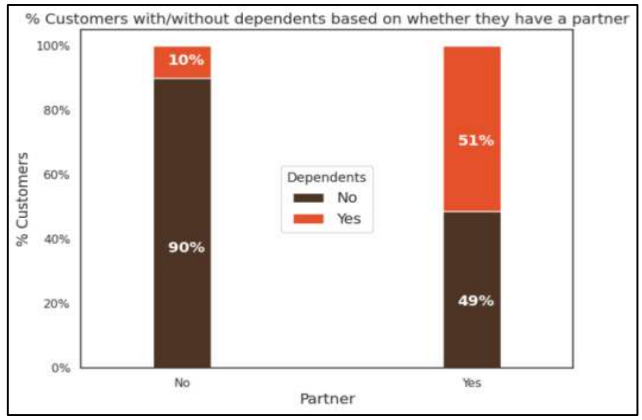
Using a range of statistical and visual methods, we investigated the data throughout the EDA phase. We looked at the variable distribution, looked for any missing data, and found any outliers. In order to identify trends in the data and possibly generate some hypotheses, let's first begin by examining our data collection. Prior to analyzing our data for any significant trends, we will examine the distribution of each individual variable.

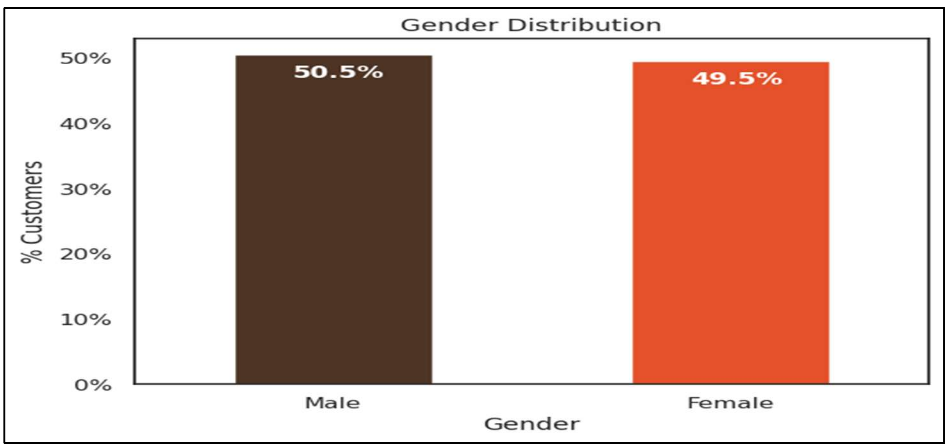


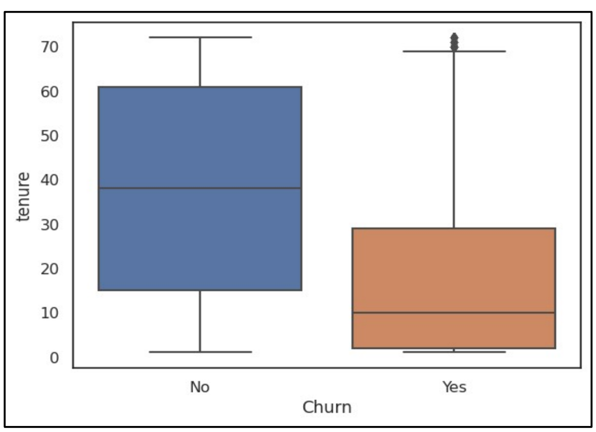
Vol. 20, No. 1, (2023)
ISSN: 1005-0930

**Target Variable Visualization:** First, let's examine the clients' gender, age range, partner status, and level of dependency. In Gender Distribution the clients in our data collection, men make up around half and women the other half. With regard to Not-Churn: Churn customers, the dataset is uneven. Predictions will be biassed in the direction of non-converting clients.



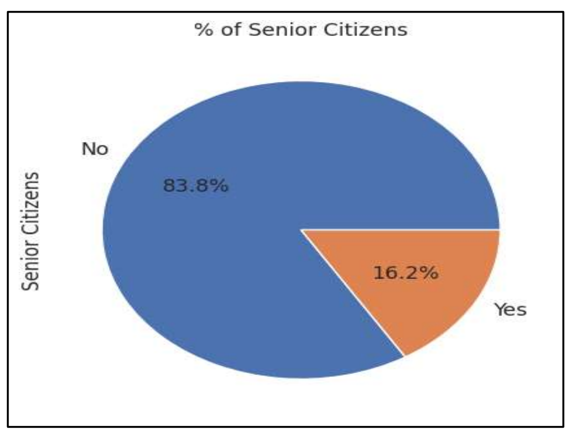



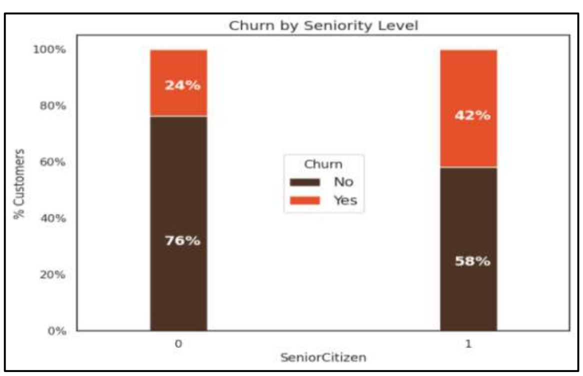

**Fig 4. Target Variable Visualization**

Merely 16% of the clientele at % Senior Citizens are elderly individuals. Consequently, the majority of our clients in the statistics are younger individuals. In The status of a partner and dependent Just 30% of all customers have dependents, compared to about 50% of customers who have a partner. Merely 50% of the clients with partners also have dependents, and the remaining 50% do not have any independents. Furthermore, among the clients without a partner, the majority (80%) do not have any dependents, as would be predicted. This bias will be shown in visualization as well!

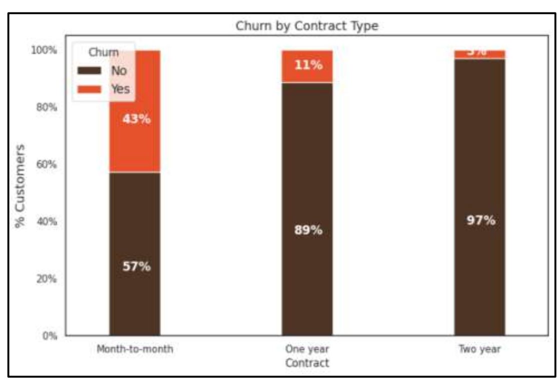
**C. Feature Mining and Assortment:**

The primary goal of choosing attributes is to exclude insignificant variables that are stable throughout time or lack any discernible dispersion. In order to extract relevant characteristics and eliminate unnecessary ones, we also used feature engineering. To prevent multicollinearity problems, we also examined the correlation between the variables. The category and numerical features can be separated, allowing us to apply the proper data preparation methods to each kind of feature independently. This creates it possible for the Machine Learning Algorithms to use all the features in an efficient manner. Treating category and numerical characteristics similarly can lead to biased or erroneous models, therefore it also helps to avoid potential problems [4][9]. Now, let's examine the duration and contract for customer account information. The Numerous telecom companies' clients have just been with them for a month, but many have been there for generally 72 months, agreeing to our investigate. This could be the case due to the certainty that diverse customers' conventions distinction. Consumers' skill to endure with or quit the telecom firm may therefore vary depending on the terms of their contract. 74% of the clients in our data do not churn. The data is obviously biased because we would anticipate that the vast majority of customers would remain loyal. This is crucial to remember because skeweness can result in a lot of false negatives for our modeling. How to prevent skeweness in the data will be covered in the section on modeling. Now that we have examined the churn rate by tenure, seniority, contract type, monthly costs, and total charges, we can understand how these factors affect it.

**Churn vs tenure**



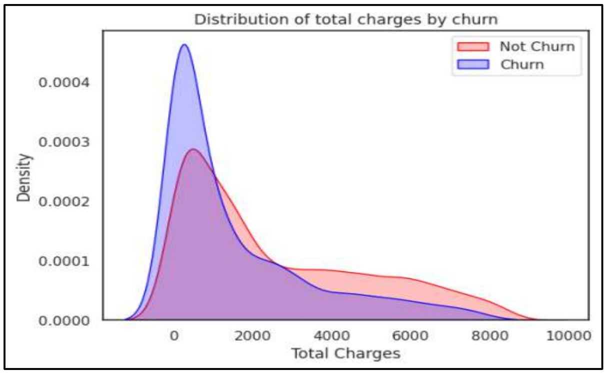

**Churn vs seniority level**






Vol. 20, No. 1, (2023)
ISSN: 1005-0930

**Churn vs total charges** **Churn vs monthly charges**



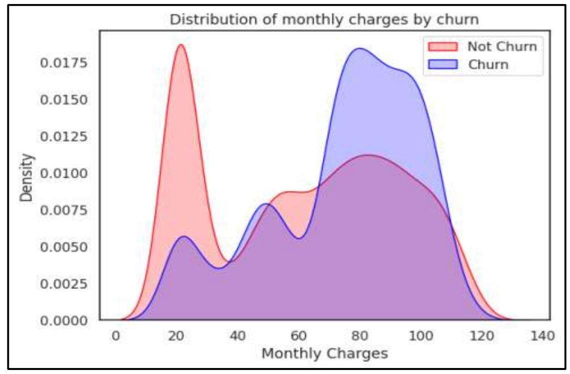




**Fig 5. Churn rate vs tenure| seniority |contract type |monthly costs| total charges**

To determine how each numerical feature relates to customer turnover, numerical features are compared to the target variable (churn). Finding the characteristics that are more crucial for churn prediction is completed easier through the assistance of this contrast. Tenure becomes a significant characteristic if the investigation reveals that consumers with longer tenure are less likely to churn. Likewise, if higher monthly charge clients have a greater turnover rate, Monthly Charges becomes a crucial feature.

**Fitness Function Derivation**

Minimizing error rates improves the precision of classification, which is the aim of the feature selection challenge. The rate of error is considered as a measure of fitness in GSA that is depicted in the following equation and whose reduction is the goal.

$$\text{Err or Rate} = \frac{FP + FN}{TP + TN + FP + FN}$$

Where FN, FP, TN, and TP represent for false negative, false positive, true negative, and true positive, accordingly.

**3.3. Model Evaluation:**

We used a variety of indicators to evaluate the trained models' performance during the model evaluation phase. In order to calculate the percentage of correctly identified cases among all instances, we first calculated the accuracy score. The optimal statistic for assessing classifier models, however, may not always be accuracy, particularly when the dataset is unbalanced. As a result, we also determined the F1-score to measure the complete efficiency of the models. This quantity proceeds into an account of equally precision and recall [4].

**Support Vector Machine**: Another innovative machine learning technique that is quite well-liked right now is Support Vector Machine (SVM). It primarily serves as a classification tool. Assistance with Machine Learning Vector Machines are supervised learning models that analyses data for both regression analysis and classification. They are combined with corresponding learning algorithms.

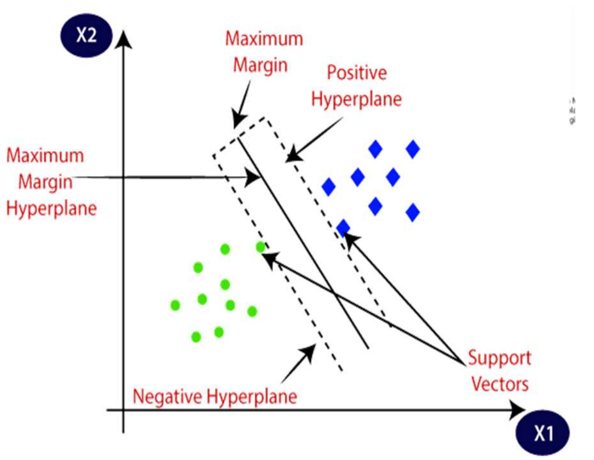



**Fig 6. Structure of Support Vector Machine**

**Data** : Dataset with $p^*$ variables and binary outcome.
**Output**: Ranked list of variables according to their relevance.

Find the optimal values for the tuning parameters of the SVM model;
Train the SVM model;
$p \leftarrow p^*$;
**while** $p \geq 2$ **do**
  $SVM_p \leftarrow$ SVM with the optimized tuning parameters for the $p$ variables and observations in **Data**;
  $w_p \leftarrow$ calculate weight vector of the $SVM_p$ $(w_{p1}, \ldots, w_{pp})$;
  $rank.criteria \leftarrow (w_{p1}^2, \ldots, w_{pp}^2)$;
  $min.rank.criteria \leftarrow$ variable with lowest value in $rank.criteria$ vector;
  Remove $min.rank.criteria$ from **Data**;
  $Rank_p \leftarrow min.rank.criteria$;
  $p \leftarrow p - 1$;
**end**
$Rank_1 \leftarrow$ variable in **Data** $\notin (Rank_2, \ldots, Rank_{p^*})$;
**return** $(Rank_1, \ldots, Rank_{p^*})$

**Fig 7. Pseudo-code of Support Vector Machine**

The knowledge of margin calculation supports SVM. It effectively draws lines between the classes. The margins are drawn to have the shortest possible distance between them and the classes, which reduces classification error. Fig 6 and 7 provides an example of functioning and pseudo-code for SVM [10].

**Logistic Regression:**

The probability of a churn, or the chance that a client would terminate their membership, can be predicted using logistic regression. An algorithm for supervised learning that is utilized in classification is called logistic regression. Only the categorization is done using logistic regression; we establish a threshold based on the limit. Depending on the specifics of the categorization challenge, the value of threshold is changeable. Fig 8 provides a functioning of Logistic Regression.

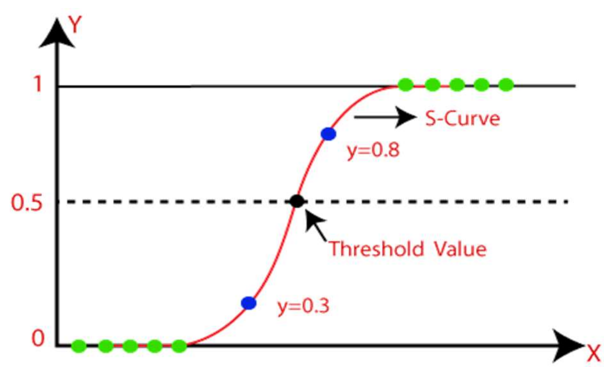



**Fig 8. Structure of Logistic Regression**

From the linear regression equation, the equation logistic regression can be derived. Below are the mathematical processes needed to obtain equations for logistic regression. The straight line equation is expressed as $\frac{y}{1-y}$; 0 for y= 0, and infinity for y=1 follows,

Since y in a logistic regression can only be between 0 and 1, let's divide the previous equation by (1-y)

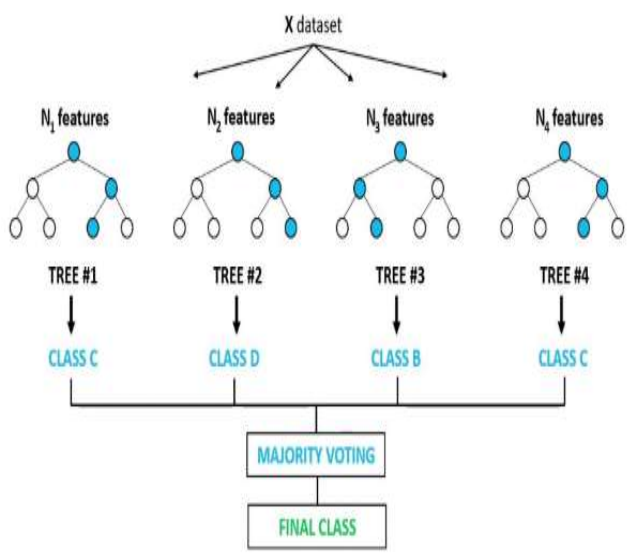
$$y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

However, the range that we require is from -[infinity] to +[infinity]. The result of computing the equation's logarithm is

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

**Random forest:**

Random Forest (RF) is a technique which utilizes outfit learning and DT to categories information. When it is within the arrange of preparing, it creates various trees as well as a forest of decision trees.




.

```
To generate c classifiers:
for i = 1 to c do
  Randomly sample the training data D with replacement to produce D_i
  Create a root node, N_i containing D_i
  Call BuildTree(N_i)
end for

BuildTree(N):
if N contains instances of only one class then
  return
else
  Randomly select x% of the possible splitting features in N
  Select the feature F with the highest information gain to split on
  Create f child nodes of N , N_1,..., N_f , where F has f possible values (F_1, ... , F_f)
  for i = 1 to f do
    Set the contents of N_i to D_i , where D_i is all instances in N that match
    F_i
    Call BuildTree(N_i)
  end for
end if
```

**Fig 9. Structure of Random Forest**

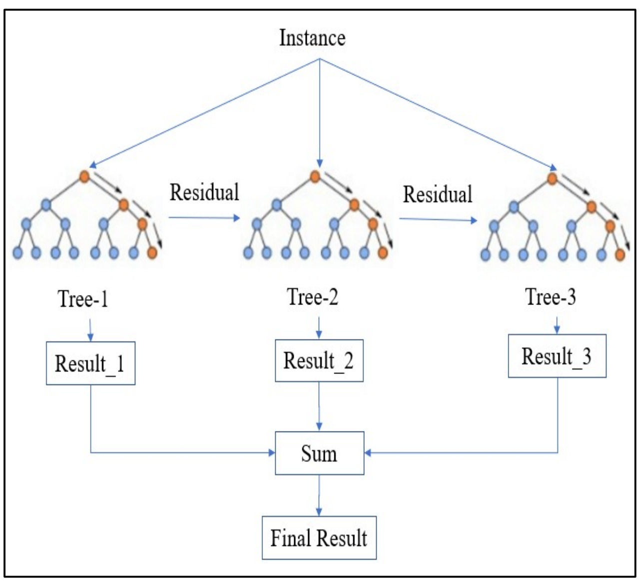
**Fig 10. Pseudo-code of Random Forest**

Each tree, person within the forest, predicts the class label for each occasion all through the testing period. When each tree accurately predicts a class label, majority voting is utilized to create the extreme determination for each test set of data. The label for the class that received the foremost votes is thought to be the one that ought to have been utilized on the test data. This method is rehashed for each piece of data within the collection of facts [11]. We obviously have to understand the impurities of the data set, and we're going to utilize that feature to select the base node that possesses the lowest impurity, or, to put it another way, the lowest Gini index. Fig 9 and 10 shows the working function and pseudo-code for Random Forest Algorithm.

**XGBoost Classifier:**

Gradient boosting is a ensemble machine learning technique based on decision tree. It is appropriate for addressing regression and classification issues, ranking problems, and user-defined prediction problems on unstructured data (pictures, text, etc.)[12]. It has been extensively utilized in various domains to generate innovative results on certain data challenges (such as Kaggle contests).

The extremely efficient, adaptable, and portable Gradient Boosting framework serves as the foundation for Boost's optimisation. Boosting's fundamental principle is to produce an effective classifier with superior classification performance by combining a number of weak classifiers with low accuracy. It can be referred to as the Gradient Boosting Machines if the weak learner for each step is based on the gradient direction of the loss function**.** Fig 11 and 12 explores the Pseudo – code and Architecture of XGBoost Classifier.




**Data:** Dataset and hyperparameters
Initialize $f_0(x)$;
**for** $k = 1, 2, \ldots, M$ **do**
  Calculate $g_k = \frac{\partial L(y,f)}{\partial f}$;
  Calculate $h_k = \frac{\partial^2 L(y,f)}{\partial f^2}$;
  Determine the structure by choosing splits with maximized gain
  $\mathrm{A} = \frac{1}{2}\left[\frac{G_L^2}{H_L} + \frac{G_R^2}{H_R} - \frac{G^2}{H}\right]$;
  Determine the leaf weights $w^* = -\frac{G}{H}$;
  Determine the base learner $\hat{b}(x) = \sum_{j=1}^{T} wI$;
  Add trees $f_k(x) = f_{k-1}(x) + \hat{b}(x)$;
**end**
**Result:** $f(x) = \sum_{k=0}^{M} f_k(x)$

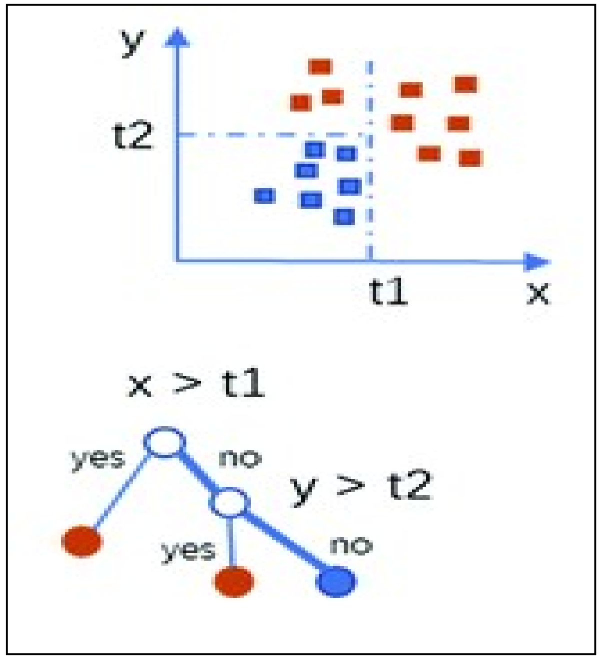
**Fig 11. Structure of XGBoost Classifier.** **Fig 12. Pseudo-code of XGBoost Classifier.**

**AdaBoost:**

The sequential ensemble technique known as "AdaBoost," or adaptive boosting, works by randomly selecting different training subsets from the initial training dataset to produce many weak learners.[13][14] Weights are allocated during each training session and are utilized to understand each hypothesis. The weights represent the relative relevance of each event and are used to compute the hypothesis' error on the dataset. Every iteration results in a new computation of the weights, giving instances that the previous hypothesis mistakenly classified a larger weight. Fig 13 and 14 explores the Pseudo – code and function of XGBoost Classifier.



Vol. 20, No. 1, (2023)
ISSN: 1005-0930




**Input:** Data set $\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_m, y_m)\}$;

Base learning algorithm $\mathcal{L}$;

Number of learning rounds $T$.

**Process:**

$D_1(i) = 1/m$. % Initialize the weight distribution

for $t = 1, \cdots, T$:

$h_t = \mathcal{L}(\mathcal{D}, D_t)$; % Train a weak learner $h_t$ from $\mathcal{D}$ using distribution $D_t$

$\epsilon_t = \Pr_{i \sim D_i}[h_t(\boldsymbol{x}_i \neq y_i)]$; % Measure the error of $h_t$

$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$; % Determine the weight of $h_t$

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t) & \text{if } h_t(\boldsymbol{x}_i) = y_i \\ \exp(\alpha_t) & \text{if } h_t(\boldsymbol{x}_i) \neq y_i \end{cases}$$

$= \frac{D_t(i)\exp(-\alpha_t y_i h_t(\boldsymbol{x}_i))}{Z_t}$ % Update the distribution, where $Z_t$ is

% a normalization factor which enables $D_{t+1}$ be a distribution

end.

**Output:** $H(\boldsymbol{x}) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})\right)$

**Fig 13. Structure of AdaBoost Classifier. Fig 14. Pseudo-code of AdaBoost Classifier.**



Vol. 20, No. 1, (2023)
ISSN: 1005-0930

The algorithm is then able to concentrate on cases that are harder to learn. The algorithm's primary goal is to assign updated weights to the instances that were erroneously classified. Regression uses a real-value mistake instead of instances that are right or wrong, as opposed to categorization. Through the usage of the AdaBoost classifier, the computed error can be classified as either an error or not by comparing it to a predetermined threshold prediction error. The probability (i.e., likelihood) of selecting occasions with bigger deficiencies on earlier learners to prepare another base learner is higher. To create an ensemble prediction of the particular base learner predictions, the weighted normal or middle is utilized [15].

## 4. Experimental Result and Analysis

The process of integrating the predictive model with the telecom company's business processes and putting it into production is known as the "implementation phase" in customer churn prediction. The process entails providing the model to end users, such customer care agents, so they can utilize it to make well-informed decisions and take the necessary steps to keep customers from leaving. When predictive models are deployed and integrated into the organization's database, they can be exploited to predict future events. The telecom provider can then make use of this data to proactively stop customer attrition**.** During the implementation phase, the model's performance is monitored and any necessary adjustments are made to enhance its efficacy and accuracy. It is essential to regularly assess the model's performance to make sure it keeps producing accurate predictions. After going through the EDA we developed some predictive models and compare them, such as Logistic Regression, Random Forest, SVM, ADA Boost and XG Boost.

| S.No | ML Algorithm | Accuracy |
|---|---|---|
| **1.** | Logistic Regression | 0.807583 |
| **2.** | Random Forest | 0.808813 |
| **3.** | Support Vector Machine | 0.820185 |
| **4.** | ADA Boost | 0.815920 |
| **5.** | XG Boost | 0.825970 |

**Table 1: Accuracy score of various classifiers**

Based on the Comparision of above classifiers, in logistic regression, it's crucial to scale the variables so that they're all between 0 and 1. My accuracy increased from 79.7% to 80.7% as a result of this. You'll also see that the variables' relative importance matches what the Random Forest algorithm demonstrates The most significant predictor variables for churn, according to the random forest algorithm, are tenure, monthly contract, and total charges. The random forest's values closely match the logistic regressions and are in line with what we expected from our EDA. It was able to reach up to 82% accuracy using SVM. In order to provide a more precise prediction,



Vol. 20, No. 1, (2023)
ISSN: 1005-0930

we must look further thoroughly at the genuine undesirable and optimistic rates as well as the Area under the Curve (AUC). It's interesting to see that XG Boost raises test data accuracy to approximately 83%. XG Boost is unquestionably superior to the other methods. Table1 examines the accuracy score of various classifiers and fig 13 shows the analysis of the models

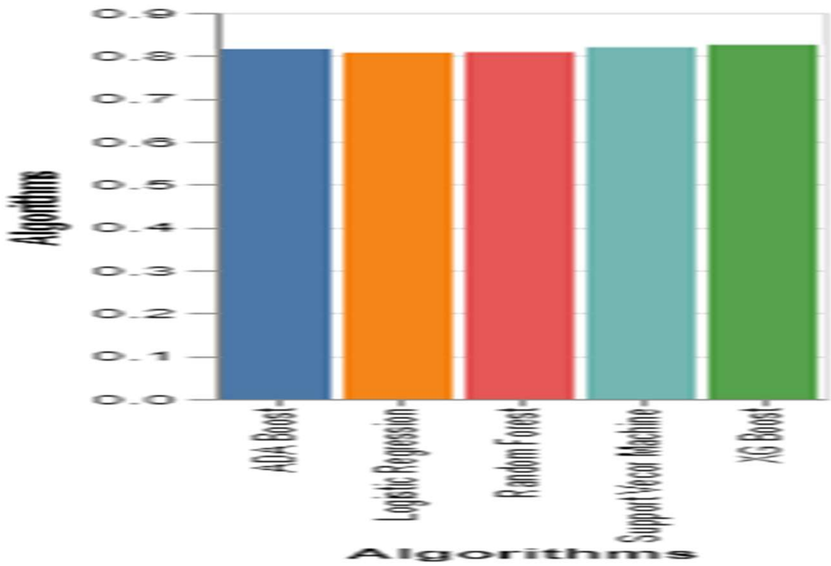



**Fig 15. Accuracy Analysis models**

## 5. Conclusion:

To retain their existing customer base, telecom businesses must comprehend the causes of customer attrition. Data collected from telecom histories can be used to gain this information. This research trains five Machine Learning Models: Random Forest, XG Boost, SVM, Logistic Regression and Ada Boost. Among these, SVM performs ineffectively, whereas XG Boost is the finest model. Logistic regression and Random Forest are normal, and XG Boost is the leading model. The outcomes about of the performance comparison of the previously mentioned classifiers appeared that the XGBOOST Classifier performed the most excellent out of all of them, giving the most prominent F1 score and Precision score compared to the other four models. The most elevated AUC model, XGBoost, is 0.83. This research outlines how vital that feature designing and EDA are making exact machine learning algorithms to foresee client attrition.

**REFERENCES :**

[1] Senthilnayaki B , Swetha M , Nivedha D, "Customer Churn Prediction", *International Advanced Research Journal in Science, Engineering and Technology,* Vol. 8, Issue 6, June 2021.

[2] Khulood Ebrah, Selma Elnasir, *"Churn Prediction Using Machine Learning and Recommendations Plans for Telecoms"*, Journal of Computer and Communications, 2019, 7, 33-53. https://doi.org/10.4236/jcc.2019.711003.

[3] Jitendra Maan , Harsh Maan *," Customer Churn Prediction Model using Explainable Machine learning"*, International Journal of Computer Science Trends and Technology (IJCST) – Volume 11 Issue 1, Jan-Feb 2023.

[4] Suhel Malik , Siddhart Runwal, Yash Shah, Vishal Raut, Prof. S.S. Hire, *"International Research Journal of Modernization in Engineering Technology and Science"*- Volume: 05/Issue:04/April-2023.

[5] Lewlisa Saha , Hrudaya Kumar Tripathy , Tarek Gaber ,Hatem El-Gohary and El-Sayed M. El-kenawy, *"Deep Churn Prediction Method for Telecommunication Industry"* Article - Sustainability 2023, 15, 4543. https://doi.org/10.3390/su15054543.

[6] Jitendra Maan , Harsh Maan, *"Customer Churn Prediction Model using Explainable Machine learning"* International Journal of Computer Science Trends and Technology (IJCST) – Volume 11 Issue 1, Jan-Feb 2023.

[7] J. Faritha Banu ,S. Neelakandan, B.T Geetha, V. Selvalakshmi, A. Umadevi,and Eric Ofori Martinson, *"Artificial Intelligence Based Customer Churn Prediction Model for Business Markets",* Hindawi Computational Intelligence and Neuroscience Volume 2022, Article ID 1703696. https://doi.org/10.1155/2022/1703696.

[8] V. Kavitha, S. V Mohan Kumar, G. Hemanth Kumar, M. Harish, *"Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms"*, International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 05, May-2020, ISSN: 2278-0181.

[9] Praveen Lalwani, Manas Kumar Mishra, Jasroop Singh Chadha,Pratyush Sethi1, *"Customer churn prediction system: a machine learning approach",-* Article in Computing • February 2022, DOI: 10.1007/s00607-021-00908-y.

[10] Ayon Dey, *"Machine Learning Algorithms: A Review"*, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016, 1174-1179, ISSN : 0975-9646.

[11] Md Mamun Ali , Bikash Kumar Paul, Kawsar Ahmed , Francis M. Bui , Julian M. W. Quinn , Mohammad Ali Moni, *"Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison",* Computers in Biology and Medicine, Volume 136, September 2021, 104672.

[12] Kumar, N. and Naik, C. *"Comparative Analysis of Machine Learning Algorithmsfor Their Effectiveness in Churn Prediction in the Telecom Industry".* International Research Journal of Engineering and Technology, 4, 485-489.

[13] Maaz Amjad , Irshad Ahmad , Mahmood Ahmad , Piotr Wróblewski , Paweł Kami´nski and Uzair Amjad *"Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation –Article",* Appl. Sci. 2022, 12, 2126. https://doi.org/10.3390/app12042126 .

[14] Schapire, R.E. Explaining AdaBoost, *"In Empirical Inference"* Schölkopf, B., Luo, Z., Vovk, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.

[15] Seo, D.K, Kim, Y.H,Eo, Y.D,Park, W.Y, Park, H.C, *"Generation of Radiometric, Phenological Normalized Image Based on Random Forest Regression for Change Detection".* Remote Sens. 2017, 9, 1163.