GITHUB LINK:

PROJECT TITLE: Predicting customer churn using machine learning to uncover hidden patterns

1.**Problem Statement**

The problem statement of customer churn prediction in the telecom industry using ML is to develop accurate and reliable models that can identify customers who are at risk of leaving the telecom service provider. The goal is to predict customer churn before it occurs and take proactive measures to retain those customers, thereby reducing the customer churn rate and improving customer retention. The challenge is to develop models that can effectively analyze and make sense of the vast amounts of customer data generated by telecom service providers. These models should be able to identify patterns and trends in the data that indicate customer churn risk factors and provide actionable insights to service providers. The problem statement also involves selecting appropriate ML techniques that can handle the large and complex datasets and account for the variability in customer behavior. The models should also be transparent and explainable to build customer trust and regulatory compliance. In summary, the problem statement of customer churn prediction in the telecom industry using ML is to develop accurate and reliable models that can effectively analyze large and complex customer datasets to predict customer churn risk and provide actionable insights to reduce customer churn rates and improve customer retention. The problem being addressed in a churn prediction project is to identify customers who are likely to stop using a product or service in the near future. Churn can be a costly problem for businesses, as it means losing revenue and potentially damaging their reputation. Therefore, predicting and preventing churn is an important goal for businesses to retain their customers and maintain a profitable customer base. The churn prediction project typically involves analyzing customer data such as usage patterns, demographics, and past behavior to identify factors that are strongly associated with churn. Machine learning algorithms such as Support Vector Machines (SVM) and Random Forest can be trained on this data to predict which customers are likely to churn in the future. The output of the churn prediction project is a model that can be used to predict churn for new customers based on their data. This can be integrated into the company's existing customer relationship management (CRM) system or other relevant tools to allow the business to take proactive measures to retain customers at risk of churning. In summary, the problem definition for a churn prediction project is to use machine learning algorithms to analyze customer data and predict which customers are at risk of churning, allowing the business to take appropriate measures to retain their customers and maintain a profitable customer base.

2.Abstract

In the highly competitive e-commerce industry, customer churn represents a major challenge to profitability and sustainability. This study aims to develop a robust predictive model for customer churn using a publicly available e-commerce dataset. The research leverages various machine learning algorithms, including Logistic Regression, Random Forest, XGBoost, and LightGBM, to compare performance. We address class imbalance with SMOTE and utilize SHAP and LIME for model interpretability. Our results demonstrate the effectiveness of the Random Forest model, achieving a ROC AUC of 0.9850. This study provides valuable insights into the factors driving customer churn, offering actionable recommendations for businesses to reduce churn rates and enhance customer retention strategies. Businesses must compete fiercely to win over new consumers from suppliers. Since it directly affects a company's revenue, client retention is a hot topic for analysis, and early detection of client churn enables businesses to take proactive measures to keep customers. As a result, all firms could practice a variety of approaches to identify their clients early on through client retention initiatives. Consequently, this study aims to advise on

the optimum machine-learning strategy for early client churn prediction. The data included in this investigation includes all customer data going back about nine months before the churn. The goal is to predict existing customers' responses to keep them. The study has tested algorithms like stochastic gradient booster, random forest, logistics regression, and k-nearest neighbors methods. The accuracy of the aforementioned algorithms are 83.9%, 82.6%, 82.9% and 78.1% respectively. We have acquired the most effective results by examining these algorithms and discussing the best among the four from different perspectives.

3. System Requirement

- Hardware:
  a. Minimum 4GB RAM (8GB recommended)
  b. Any processor (Intel i3/i5 or AMD equivalent)
  c. System : Pentium Dual Core.
  d.Hard Disk : 120 GB.
  e. Monitor : 15”LED
  f. Input Devices : Keyboard, Mouse
  g. Ram : 1GB.
- Software:
  a. Python
  b. Operating system : Windows 7.

c. Coding Language : python

d. Toolkit : Jupiter Notebook ,Google Colab

e. DATABASE : EXCEL

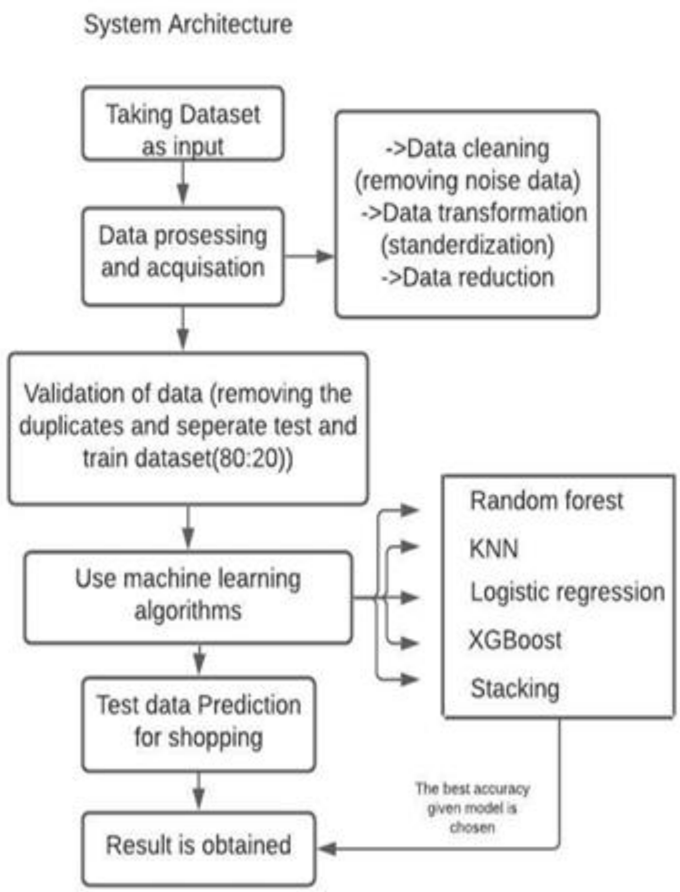
f. Libraries: pandas,numpy,matplotlib,seaborn.scikit-learn,gradio,plotly

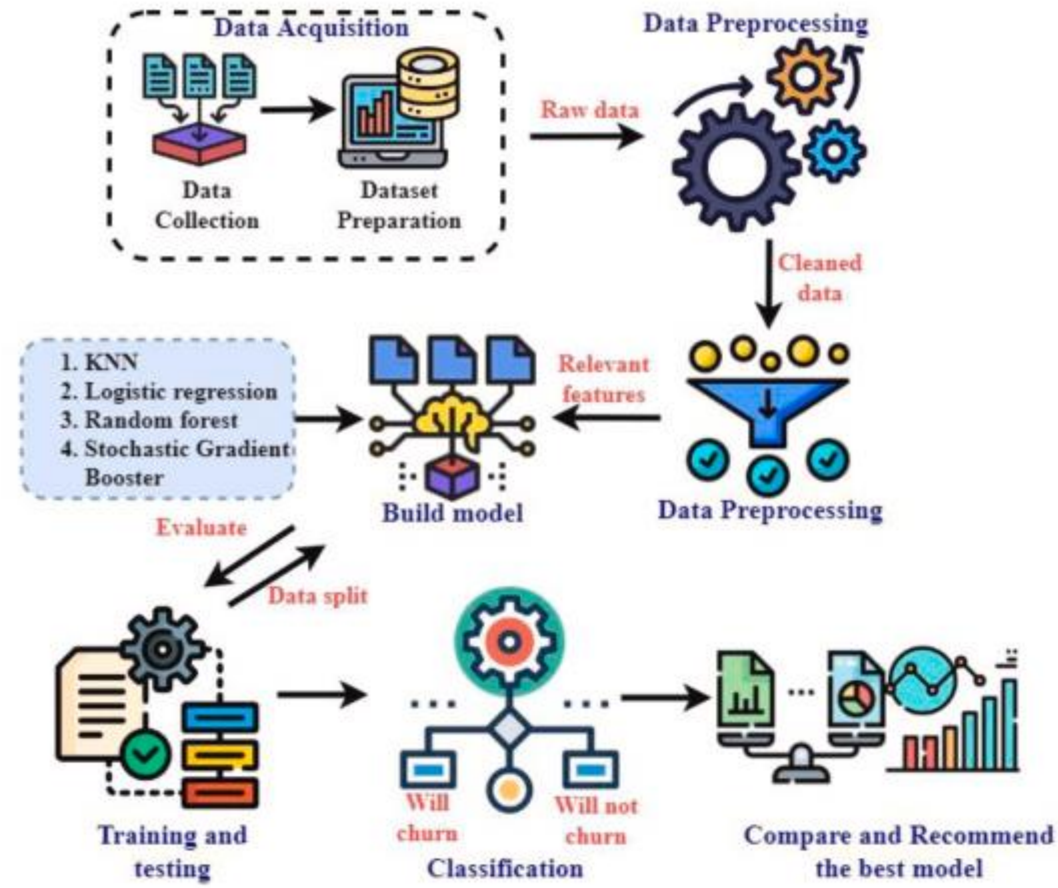
4. Objectives

To get and keep loyal customers for every business organization is a big challenge. Correct prediction about a customer is going to churn or not and then successfully convincing him to stay with that company can increase the revenue of that company.Therefore,predicting customer churn, i.e.if a customer is about to leave for a better service, that is an important part for analyzing the customer behavior. The churn model is a representation of various calculations that are built on existing historical data. The customer churn can be defined in other ways also, like low switching cost, deregulation motivates a customer to replace the sector. The churn is also classified into two: voluntary and involuntary churn Voluntary churn is defined as the termination of services by the customer itself, whereas involuntary churn is defined as the termination of services by the bank for fraud, non-payment services. The customer churn is very risky if it is not managed carefully, it may bring a company to its knees for poor maintenance services. Cost of customer churn also includes loss of revenue, profit. Previous case study has shown that the cost of maintaining a new customer is higher than the cost of maintaining the old one There are various banks who are suffering with this customer churn problem. So the most defined way to deal with this problem is developing a predictive model that can reliably and easily identify the possible churner. In the recent past the most frequently used technique is data mining to develop such models with satisfactory results Day by day when this churn prediction problem gets importance, much more research efforts are generated towards improving churn prediction rates.

5.Flowchart of the project workflow

System analysis is the process of gathering and interpreting facts, diagnosing problems and using the information to recommend improvements on the system. System analysis is a problem-solving activity that requires intensive communication between the system users and system developers. System analysis or study is an important phase of any system development process. The system is viewed as a whole, the inputs are identified and the system is subjected to close study to identify the problem areas. 11 The solutions are given as a proposal. The proposal is reviewed on user request and suitable changes are made. This loop ends as soon as the user is satisfied with the proposal. They are many steps involved in this process and has importance to each and every step here.





6.Dataset Description

Data preparation is one of the most important steps in building a reliable machine learning model. In my experience, this phase often takes the majority of the time because preparing the data involves a lot of detailed work. For this project, the data preparation process included cleaning the data, addressing any missing values and converting categorical variables into a format that machine learning models can understand. Without thorough data preparation even the best algorithms will struggle to make accurate predictions. First, I loaded the dataset and took a quick look at the first few rows to get a sense of the data. This dataset contains information like customer tenure, the type of internet service they use, contract types, and whether they eventually churned (the target variable). After loading the data, I checked for missing values since missing data can introduce bias or errors. Depending on the dataset, I sometimes choose to fill missing values with an appropriate measure (like the mean or median) or as in this case simply drop rows where critical information is missing. Finally, because machine learning models require numerical inputs, I needed to convert the categorical variables into numerical values. I used one-hot encoding which creates new binary columns for each category in a feature.

- Type: CSV dataset
- Size: the raw data contains 7043 rows (customers) and 21 columns (features).
- Nature: Structured tabular data
- Attributes:

    a. Demographics: Age, gender, Senior citizen, Parental Education

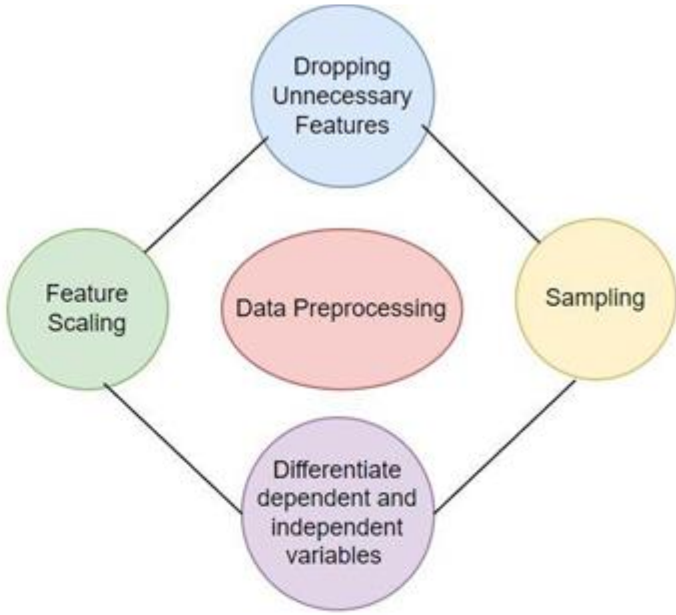
    b. Academics: Grades (G1,G2), Study time

    c. Behavior: Absences

Sample dataset (df.dtypes)

gender	object
SeniorCitizen	int64
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64

7. Data Preprocessing



Data are acquired online after preprocessing. The data can be vibration, voltage, current, acoustic emission, sound signals, etc. Corresponding to the object, different preprocessing approaches are used, such as filtering (high-, low-, and band-pass), wavelet transform, and averaging. Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

Steps involved are: 1] Data Cleaning

2] Data Integration

3] Data Reduction

4] Data Transformation

DATA CLEANING :

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time. Essentially, garbage data in is garbage analysis out. Various Steps involved are: 1.Remove Duplicate or irrelevant observations: Remove unwanted observations from your dataset, including duplicate observations or irrelevant observations. 2.Fix structural errors: Structural errors are when you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization. 3.Filter Unwanted Outliers: If an outlier proves to be irrelevant for analysis or is a mistake, consider removing it. 4.Handle Missing Data: Missing data can be handled as- 1.Remove missing values from dataset if there are very small number. 2.Replace the missing data with mean, median or mode observations. 5.Validate the data so that data can be free from all the cases mentioned above.

DATA INTEGRATION

Data Integration is a data preprocessing technique that involves combining data from multiple heterogeneous data sources into a coherent data store and provide a unified view of the data. These sources may include multiple data cubes, databases, or flat files. The data integration approaches are formally defined as triple where, G stand for the global schema, S stands for the heterogeneous source of schema, M stands for mapping between the queries of source and global schema. There are mainly 2 major approaches for data integration – one is the "tight coupling approach” and another is the “loose coupling approach”. Tight Coupling: • Here, a data warehouse is treated as an information retrieval component. • In this coupling, data is combined from different sources into a single physical location through the process of ETL – Extraction, Transformation, and Loading. Loose Coupling: • Here, an interface is provided that takes the query from the user, transforms it in a way the source database can understand, and then sends the query directly to the source databases to obtain the result. • And the data only remains in the actual source databases.

DATA REDUCTION

Data Reduction or Dimensionality reduction refers to techniques for reducing the number of input variables in training data. When dealing with high dimensional data, it is often useful to reduce the dimensionality by projecting the data to a lower dimensional subspace which captures the “essence” of the data. Data reduction techniques are used to obtain a reduced representation of the dataset that is much smaller in volume by maintaining the integrity of the original data. By reducing the data, the efficiency of the data mining process is improved, which produces the same analytical results. The most used method in this project is Dimensionality Reduction. Dimensionality Reduction: Whenever we encounter weakly important data, we use the attribute required for our analysis. Dimensionality reduction eliminates the attributes from the data set under consideration, thereby reducing the volume of original data. It reduces data size as it eliminates outdated or redundant features. Here are three methods of dimensionality reduction. 15 i. Wavelet Transform: In the wavelet transform, suppose a data vector A is transformed into a numerically different data vector A' such that both A and A' vectors are of the same length. Then how it is useful in reducing data because the data obtained from the wavelet transform can be truncated. The compressed data is obtained by retaining the smallest fragment of the strongest wavelet coefficients. Wavelet transform can be

applied to data cubes, sparse data, or skewed data. ii. iii. Principal Component Analysis: Suppose we have a data set to be analyzed that has tuples with n attributes. The principal component analysis identifies k independent tuples with n attributes that can represent the data set. In this way, the original data can be cast on a much smaller space, and dimensionality reduction can be achieved. Principal component analysis can be applied to sparse and skewed data. Attribute Subset Selection: The large data set has many attributes, some of which are irrelevant to data mining or some are redundant. The core attribute subset selection reduces the data volume and dimensionality. The attribute subset selection reduces the volume of data by eliminating redundant and irrelevant attributes.The attribute subset selection ensures that we get a good subset of original attributes even after eliminating the unwanted attributes. The resulting probability of data distribution is as close as possible to the original data distribution using all the attributes.

DATA TRANSFORMATION

Data transformation is the process of converting raw data into a format or structure that would be more suitable for model building and also data discovery in general. It is an imperative step in feature engineering that facilitates discovering insights. This article will cover techniques of numeric data transformation: log transformation, clipping methods, and data scaling. The scaling transformation techniques are: Min Max Scaler — normalization 16 MinMaxScaler() is usually applied when the dataset is not distorted. It normalizes the data into a range between 0 and 1 based on the formula: x' = (x — min(x)) / (max(x) — min(x))

Eg: sklearn.preprocessing import MinMaxScaler.fit(train_data) Standard Scaler — standardization We use standardization when the dataset conforms to normal distribution. StandardScaler() converts the numbers into the standard form of mean = 0 and variance = 1 based on z-score formula: x' = (x — mean) / standard deviation.

Eg: sklearn.preprocessing import StandardScaler.fit(train_data)
Robust Scaler RobustScaler() is more suitable for dataset with skewed distributions and outliers because it transforms the data based on median and quantile:

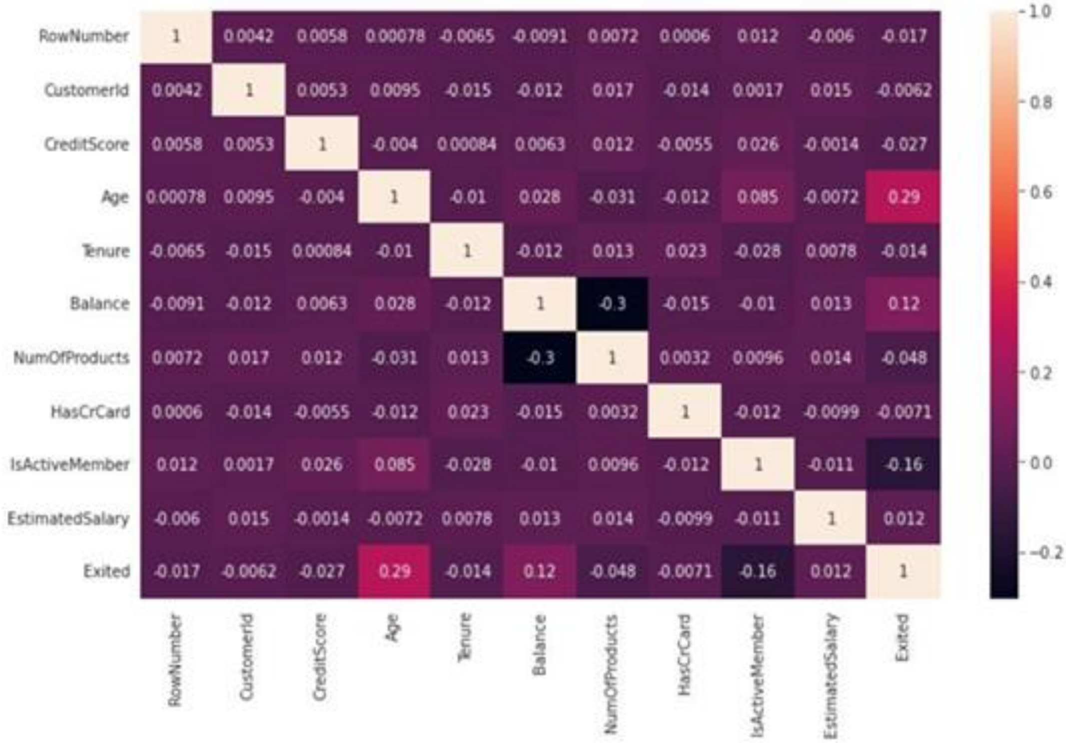
x = (x — median) / inter-quartile range.

Eg: sklearn.preprocessing import RobustScaler.fit(train_data)

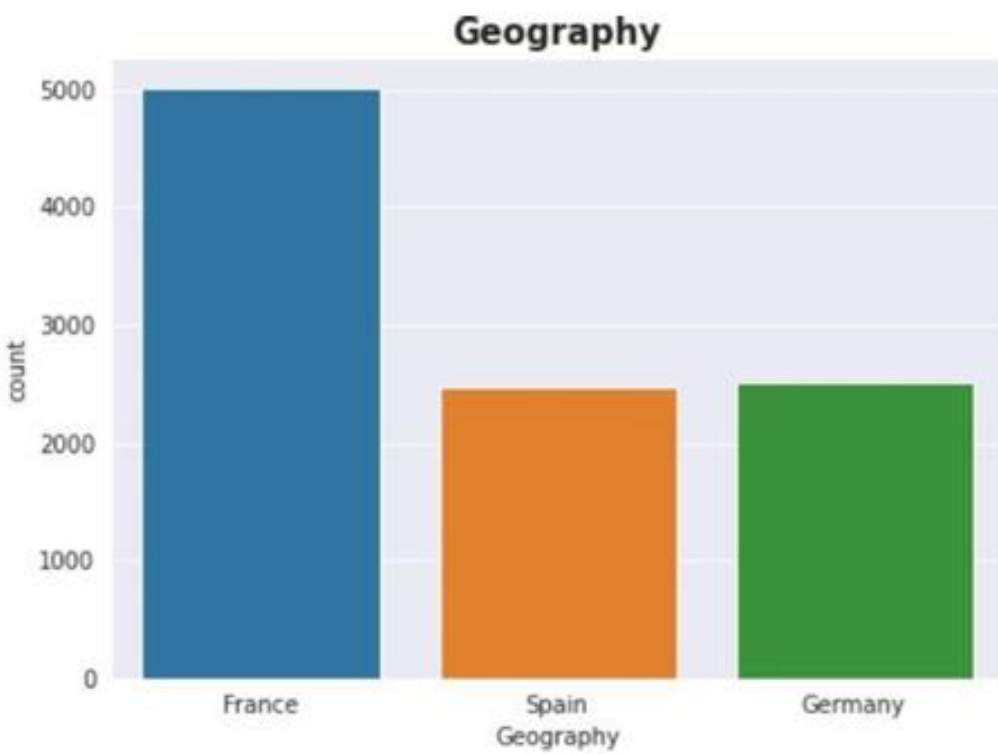
Feature	Description
CustomerID	Unique identifier for each customer
Tenure	Number of years the customer has been with the company
PreferredLoginDevice	Customer's preferred login device (mobile, desktop, etc.)
CityTier	Tier of the customer's city (market regions)
WarehouseToHome	Distance from the warehouse to the customer's home
SatisfactionScore	Customer's satisfaction score (1-5)
Churn	Whether the customer has churned (1) or not (0)

8.Exploratory Data Analysis (EDA)

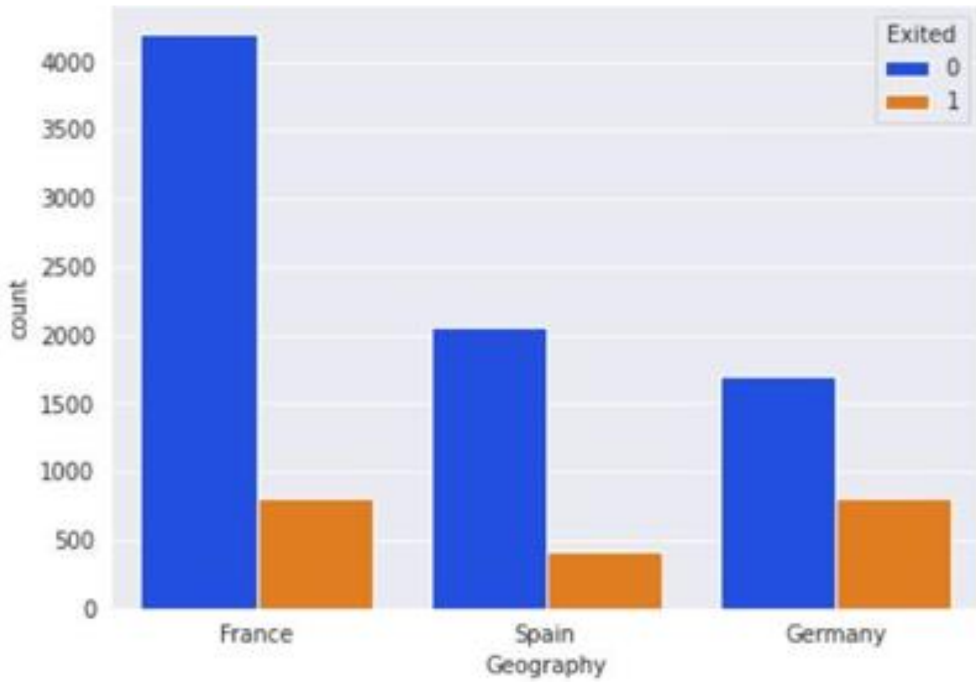
In the EDA stage, we investigated the information through different factual and visual strategies. We analyzed the dissemination of factors, checked for lost values, and recognized exceptions. We too performed highlight designing to extricate significant highlights and expelled insignificant highlights. Additionally, we analyzed the relationship between factors to maintain a strategic distance from multicollinearity issues. We made a few box plots to visualize the relationship between factors and their affect on the target variable. We recognized the most critical factors that influence client churn utilizing univariate and bivariate investigation. By and large, the EDA stage made a difference us to pick up experiences into the information and get it the relationship between factors, which is basic for building an compelling churn forecast demonstrate. Another critical component of EDA is data transformation, which entails changing data into a more acceptable format for analysis. Visualization is an effective technique in EDA, allowing analysts to visually analyse data and detect patterns or trends. Scatter plots, histograms, and box plots are common visualizations used in EDA to help understand variable distributions and interactions.

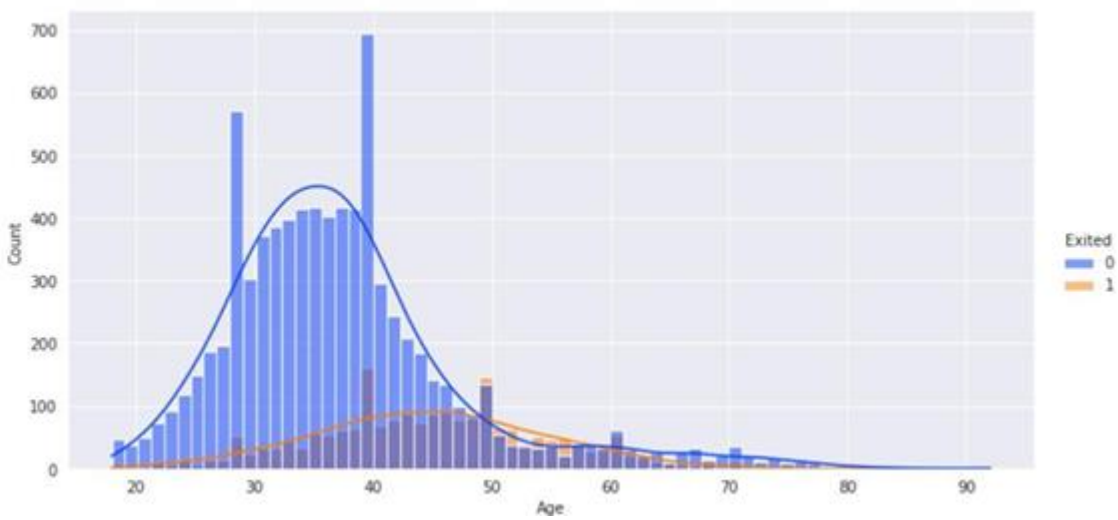


RowNumber
CustomerId
CreditScore
Age
Tenure
Balance
NumOfProducts
HasCrCard
IsActiveMember
EstimatedSalary
Exited



Geography
count
France
Spain
Germany
Geography





we can see the churn rate according to age. Here the blue line represents not churners and orange line represents churner. In this figure we can find out that the age between 40 to 50 are more churn than others.

represents the number of products that used by the customers. We can see in this figure that the count of using one product customer is higher.

Using two products of customer is also high. A very little group of customers are using three products and the number of using four products customer is very less.

represents the churn rate according to using number of total products. Customers who are using 4 products their churning rate is 100%. Using 3 products customer's churning rate is higher than not churn.

9.Feature Engineering

For getting better accuracy, we need a suitable dataset. To make the dataset suitable for this project I have dropped some features. Earlier the dataset has 14 features. I have dropped 3 features. These three features are RowNumber, CustomerId and Surname. After dropping these three features the data set now have 11 features and 10000 columns. In Figure 3.16 we can see the 1st 5 rows of these new preprocessed dataset.

	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	619	0	1	42	2	0.00	1	1	1	101348.88	1
1	608	1	1	41	1	83807.86	1	0	1	112542.58	0
2	502	0	1	42	8	159660.80	3	1	0	113931.57	1
3	699	0	1	39	1	0.00	2	0	0	93826.63	0
4	850	1	1	43	2	125510.82	1	1	1	79084.10	0

Dependent variable usually depends on another variable. Mostly they depend on independent variable. On the other hand, independent variable does not depend on other variables. In this project the exits feature is dependent variable. Exits feature represents churner. This feature is dependent variable because it depends on the other features. A customer will be churned or not churned it totally depends on the other independent features. Here I have divided the independent and dependent variables and put the independent variables in ‘x’ variable and in ‘y’ variable put the independent variables. To improve model performance, several feature engineering techniques were applied, including one-hot encoding for categorical features and the creation of new interaction terms. All numerical features were standardized using Standard Scaling to mitigate unit discrepancies.

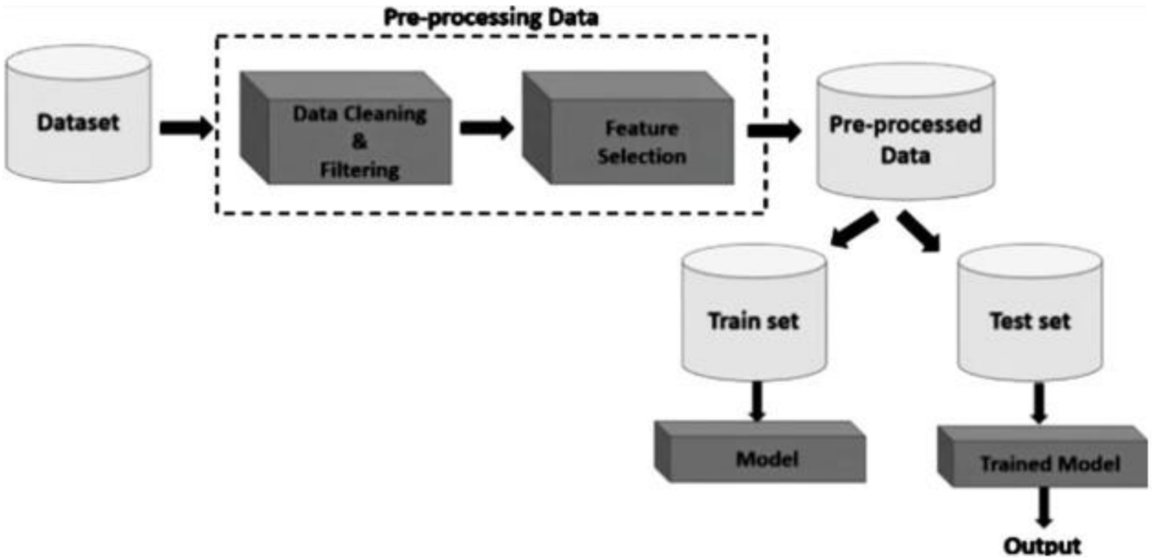
The dataset exhibited a significant class imbalance, with approximately 83% of customers not churning and 17% churning. To address this issue, the Synthetic Minority Over sampling Technique (SMOTE) was applied to the training dataset. This technique artificially oversamples the minority class (churned customers) to achieve a balanced class distribution, which mitigates the risk of bias in predictive models (Zimal et al., 2023).

Including choice is an critical errand to discover the client is chur or not. Highlight determination can be done on two perspectives some time recently applying the classification calculations. Data's are plotted utilizing visualization methods and following is utilized to discover the esteem utilizing "lasso coefficient". Three highlights are (residency, month to month charges, add up to charges) chosen after applying the two diverse sorts of include determination strategies. Select the most important highlights that are likely to contribute to foreseeing churn. Strategies like relationship investigation, include significance from models, or space information can be accommodating. Furthermore, feature engineering may entail altering existing features to make them more suited for modeling, such as converting continuous variables to categorical ones or scaling features to a standard range

Feature scaling is also known as normalization. Usually feature scaling is performed in the data preprocessing steps. It is an important method. The range of feature data or independent variables are normalized by this method. Feature scaling can be done unit of the values. If we did not perform feature scaling, then machine learning algorithm considers greater values as higher values and tends to weigh smaller values as lower values. Feature scaling is necessary for any machine-learning algorithm. Because it can calculate distance between data. In raw data the range of values varies widely. For this reason, in some machine learning algorithms, without feature scaling the objective function do not work properly. In Neural Network Algorithms, feature scaling gives better error surface shape. Feature scaling prevents the chances from getting stuck in local minima. It can also make the training faster. So that I have used feature scaling in the entire dataset of this project.

10.Model Building

Choosing suitable machine learning calculations for churn forecast. Commonly utilized calculations incorporate calculated relapse, choice trees, irregular timberlands, back vector machines, or slope boosting strategies. We utilized a 5-fold cross-validation strategy to assess the execution of each show based on exactness, exactness, review, and F1 score. Based on the assessment comes about, we chosen the top-performing models for assist examination. By using insights from EDA to understand the dataset's characteristics and the underlying relationships between features, analysts can choose the best model or combination of models to accurately predict customer churn, assisting businesses in implementing targeted retention strategies and reducing customer attrition.



Utilize the preparing information to train the chosen demonstrate. while in this stage, the demonstrate learns the designs in the information that are related with churn. Use the preparing information to fit the chosen machine learning calculation to the designs in the information. Amid preparing, the show learns the connections between the input highlights (e.g., client graph , exchange history) and the target variable (churn). Training for models such as Random Forest and Decision Tree, as well as approaches such as SMOTEEN, entails fitting the selected model to the training data while guided by EDA insights. To test the trained model's performance, a validation set is used. Accuracy, precision, recall, F1 score, are some common churn prediction evaluation criteria.

In the model assessment step, we surveyed the execution of the prepared models utilizing different measurements. To begin with, we calculated the exactness score to decide the extent of accurately classified occasions out of all occasions. Be that as it may, exactness alone is not continuously the best metric for assessing classifier models, particularly when the dataset is imbalanced. Hence, we too calculated the F1-score, which considers both exactness and review, to assess the in general execution of the models.

In this project, model building is done by Deep Learning Hyper Parameter and different famous Machine learning algorithm. Random Forest, Decision Tree, KNN, Logistic Regression these are the Machine Learning algorithm which are used to build the model. I split the dataset into training and testing data. 75% of total data are in training data and rest of the 25% data are in testing data for performance evaluation of the model. In chapter 2 I have already described all these Machine Learning algorithms briefly. In the next section I will show the result and output part briefly.

For the binary classification problem of customer churn, various machine learning algorithms were explored. Logistic Regression served as the baseline model due to its simplicity and interpretability. Additionally, more complex models, including Decision Trees, Random Forests, XGBoost, and LightGBM, were selected to capture non-linear patterns in the data. The evaluation metrics used to compare model performance included Accuracy, Precision, Recall, F1-Score, and ROC AUC. Given the imbalanced nature of the dataset, particular emphasis was placed on Precision, Recall, and ROC AUC to ensure robust performance in identifying churned customers.

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Logistic Regression	0.78	0.95	0.42	0.55	0.8403
Decision Tree	0.92	0.96	0.74	0.78	0.8826
Random Forest	0.95	0.98	0.83	0.86	0.9851
XGBoost	0.96	0.98	0.88	0.89	0.9872
LightGBM	0.93	0.96	0.78	0.80	0.9708

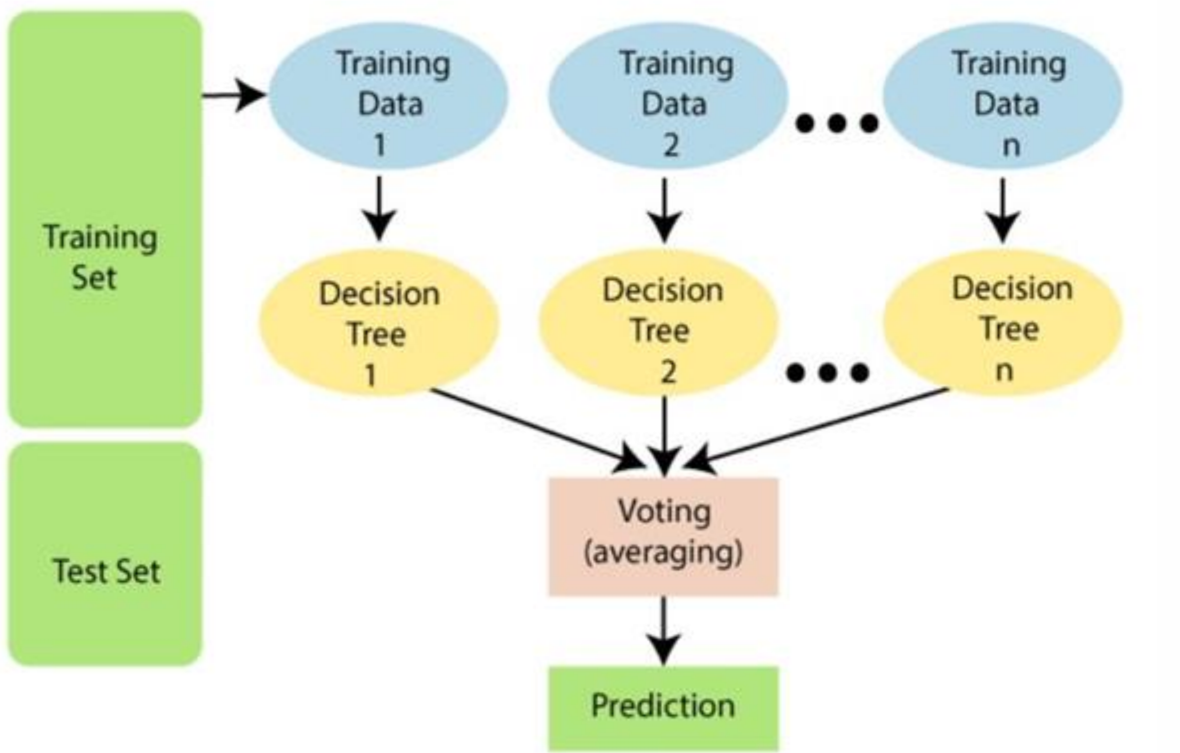
- Training Details:
  - a. 80% training /20% testing split
  - b.train_test_split (random_state=42)

Here the blue line represents not churners and orange line represents churner. In this figure ‘1’ means active member and ‘0’ means not active. We can see that the rate of non-churner of active member is higher than active member. And the churn rate of non-active member is higher than active member.

11. Model Evaluation

a. Random Forest:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier: o There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result. o The predictions from each tree must have very low correlations.



Steps involved in this method are: •

Step-1: Select random K data points from the training set. •

Step-2: Build the decision trees associated with the selected data points (Subsets). •

Step-3: Choose the number N for decision trees that you want to build. •
Step-4: Number of Decision trees used are depends on the hyperparameterisation . •

Step-4: Repeat Step 1 & 2. •

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes. Majority voting is given priority, this is called Row Sampling with replacement. •

Step-6: When the dataset is changed like out of 1000 records 200 records got changed, it does not have more impact on models, as only some random samples will be transferred for the algorithm, but it has less impact on accuracy and error loss.

Advantages

Random Forest is capable of performing both Classification and Regression tasks. o It is capable of handling large datasets with high dimensionality. o It enhances the accuracy of the model and prevents the overfitting issue.

Disadvantages

Although random forest can be used for both classification and regression tasks, it is not more suitable for Regression tasks.

Applications

There are mainly four sectors where Random forest mostly used:

1. Banking: Banking sector mostly uses this algorithm for the identification of loan risk.

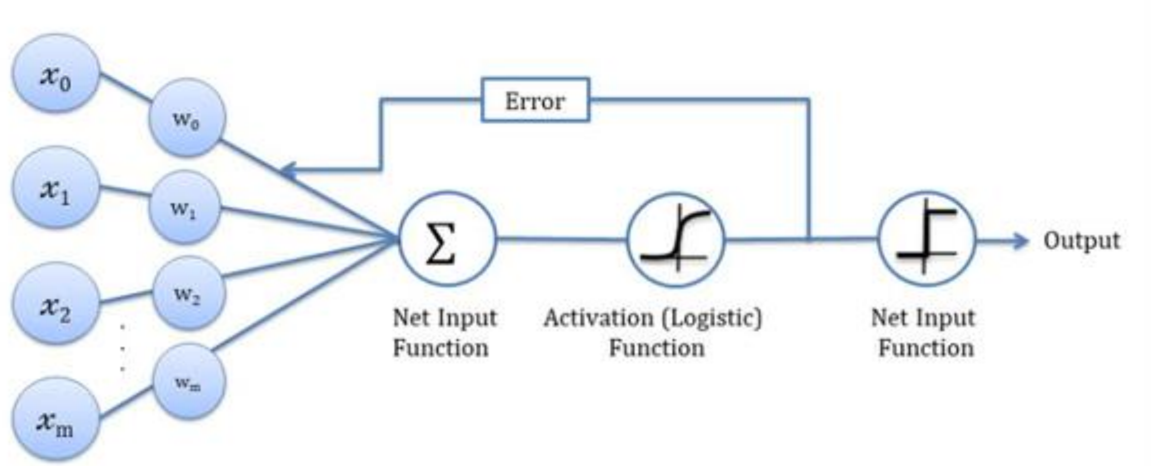
2. Medicine: With the help of this algorithm, disease trends and risks of the disease can be identified.

3. Land Use: We can identify the areas of similar land use by this algorithm.

4. Marketing: Marketing trends can be identified using this algorithm.

b. Logistic Regression:

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.



Steps involved in this method are: •

Step-1: When there are points on the graph then the line is drawn in such a way that has the minimum distance between line and points i.e, R^2 • Step-2: Now when the new data points are added then the minimized line has to change and the line equation will be in form of y=mx+c where m=slope, c=intercept. •

Step-3: But when the outlier comes into model then the line total changes towards the outlier to make it minimize ass it results in high error rate .so we use Logistic Regression here.

• Step-4: Then it starts dividing the points based on y values. There are few conditions that apply is :

i]y>0, the slope is positive

ii]y<0, the slope is negative

Advantages

• Logistic regression is easier to implement, interpret, and very efficient to train.

• It can easily extend to multiple classes (multinomial regression) and a natural probabilistic view of class predictions.

• It is very fast at classifying unknown records.

• It can interpret model coefficients as indicators of feature importance.

• Good accuracy for many simple data sets and it performs well when the dataset is linearly separable.

• It not only provides a measure of how appropriate a predictor(coefficient size)is, but also its direction of association (positive or negative).

Disadvantages

• If the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise, it may lead to overfitting. 29 30

• The major limitation of Logistic Regression is the assumption of linearity between the dependent variable and the independent variables.

• Non-linear problems can’t be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.

• It is tough to obtain complex relationships using logistic regression. More powerful and compact algorithms such as Neural Networks can easily outperform this algorithm.

• It constructs linear boundaries.

<u>Applications</u>

• Credit scoring

• Medicine

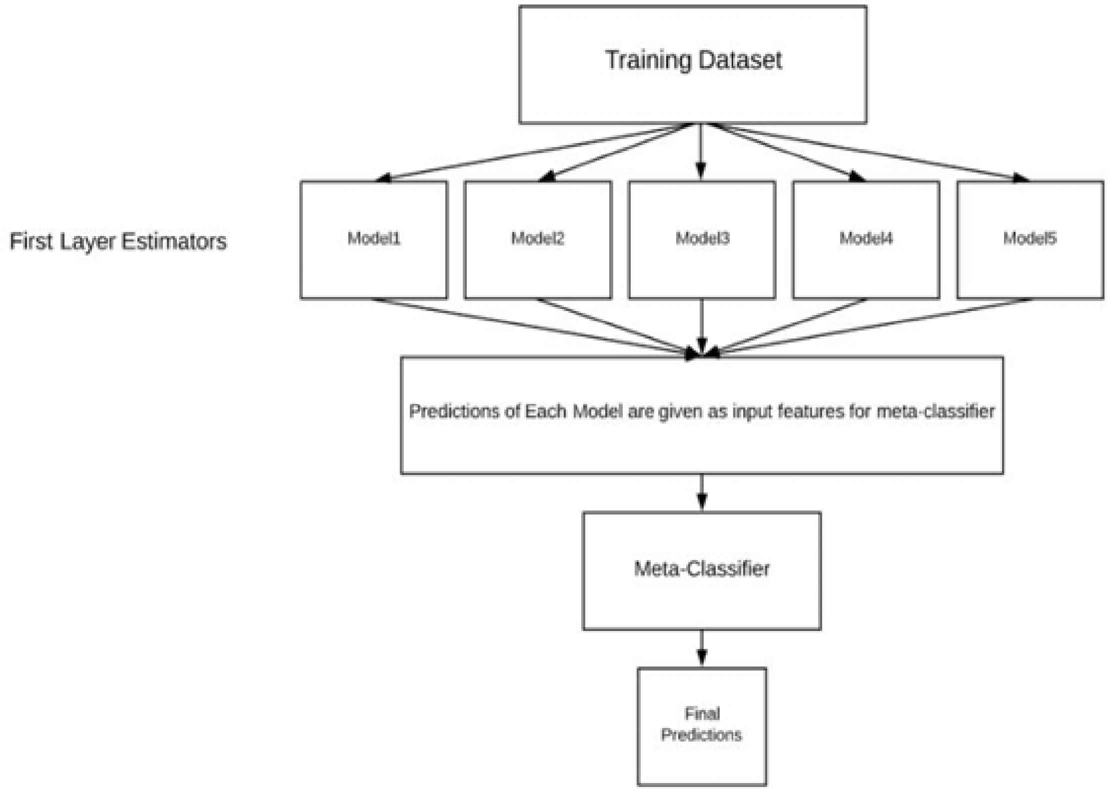
• Text editing

• Hotel Booking

• Gaming

<u>c.stacking</u>

Stacking is a way to ensemble multiple classifications or regression model. There are many ways to ensemble models, the widely known models are Bagging or Boosting. Bagging allows multiple similar models with high variance are averaged to decrease variance. Boosting builds multiple incremental models to decrease the bias, while keeping variance small. Stacking (sometimes called Stacked Generalization) is a different paradigm. The point of stacking is to explore a space of different models for the same problem. The idea is that you can attack a learning problem with different types of models which are capable to learn some part of the problem, but not the whole space of the problem. So, you can build multiple different learners and you use them to build an intermediate prediction, one prediction for each learned model. Then you add a new model which learns from the intermediate predictions the same target. This final model is said to be stacked on the top of the others, hence the name. Thus, you might improve your overall performance, and often you end up with a model which is better than any individual intermediate model. Notice however, that it does not give you any guarantee, as is often the case with any machine learning technique. They can improve the existing accuracy that is shown by individual models. We can get most of the Stacked models by choosing diverse algorithms in the first layer of architecture as different algorithms capture different trends in training data by combining both of the models can give better and accurate results.



Steps involved in this method are:

• We split the training data into K-folds just like K-fold cross-validation.
• A base model is fitted on the K-1 parts and predictions are made for Kth part. • We do for each part of the training data.

• The base model is then fitted on the whole train data set to calculate its performance on the test set.

• We repeat the last 3 steps for other base models.

• Predictions from the train set are used as features for the second level model.

• Second level model is used to make a prediction on the test set.

Advantages

• The benefit of stacking is that it can harness the capabilities of a range of well-performing models on a classification or regression task and make predictions that have better performance than any single model in the ensemble.

• Strengthens different models to combine their predictions.

• Capability to use as many as models it required.

• Majority of all algorithms output is considered as final output.

Disadvantages

• Stacked models can take significantly longer to train than simpler models and require more memory.

• Generating predictions using stacked models will usually be slower and more computationally expensive. This drawback is important to consider if you are planning to deploy a stacked model into production.

Applications

• Ensemble model so applications also depends on the applications of algorithms it uses.

12.Deployment

The model with the highest performance metrics on the test dataset is chosen as the final model for customer churn prediction. The chosen model should strike a compromise between accuracy and generalization, ensuring that it can accurately forecast customer attrition without overfitting to the training data. Deploy the trained model into production. This could involve integrating it into existing systems or applications where churn predictions are needed. Set up monitoring to track model performance over time and retrain periodically if necessary.

Advance bits of knowledge like electronic check medium, tech assistance , and online security can be looked an decided based on churning as well.For client churn forecast, a combination of exploratory information investigation (EDA), the SMOTE-ENN strategy for rectifying course imbalance, and the utilize of choice tree and irregular forest models created extraordinary results. Following broad information examination, pre processing, and model preparing, the random forest and decision tree models come to a 94% prediction accuracy. The EDA stage given profitable experiences into the dataset, counting the dispersion of churned and non-churned customers, major qualities affecting churn, and likely relationships between factors. This data impacted the pre handling methods, which included dealing with lost information, encoding category factors, different categories for it to be and scaling numerical features.

13.Source Code

```
import pandas as pd

from matplotlib

import pyplot as plt i

mport numpy as np

%matplotlib inline

from google.colab

import files uploaded = files.upload()

import pandas as pd

df=pd.read_csv("customer_churn.csv")

print(df.head())

Df.dtypes

Df.TotalCharges.values

pd.to_numeric(df.TotalCharges,errors='coerce').isnull()

df[pd.to_numeric(df.TotalCharges,errors='coerce').isnull()]

Df.shape

df.iloc[488].TotalCharges

df[df.TotalCharges!=' '].shape

df1 = df[df.TotalCharges!=' ']

Df1.shape

Df1.dtypes

Df1.TotalCharges.values
```

```
df1[df1.Churn=='No']

tenure_churn_no = df1[df1.Churn=='No'].tenure

 tenure_churn_yes = df1[df1.Churn=='Yes'].tenure

 plt.xlabel("tenure")

plt.ylabel("Number Of Customers")

 plt.title("Customer Churn Prediction Visualiztion")

blood_sugar_men = [113, 85, 90, 150, 149, 88, 93, 115, 135, 80, 77, 82, 129]

blood_sugar_women = [67, 98, 89, 120, 133, 150, 84, 69, 89, 79, 120, 112, 100]

plt.hist([tenure_churn_yes, tenure_churn_no],

 rwidth=0.95,

color=['green','red'],

label=['Churn=Yes','Churn=No'])

 plt.legend()

mc_churn_no = df1[df1.Churn=='No'].MonthlyCharges

mc_churn_yes = df1[df1.Churn=='Yes'].MonthlyCharges

plt.xlabel("Monthly Charges")

 plt.ylabel("Number Of Customers")

 plt.title("Customer Churn Prediction Visualiztion")

blood_sugar_men = [113, 85, 90, 150, 149, 88, 93, 115, 135, 80, 77, 82, 129]

blood_sugar_women = [67, 98, 89, 120, 133, 150, 84, 69, 89, 79, 120, 112, 100]

plt.hist([mc_churn_yes, mc_churn_no],

rwidth=0.95,

color=['green','red'],
```

```
label=['Churn=Yes','Churn=No'])

plt.legend()

df1['gender'].replace({'Female':1,'Male':0},inplace=True)

df1.gender.unique()

for col in df1:
        print(f'{col}: {df1[col].unique()}')
df2 = pd.get_dummies(data=df1, columns=['InternetService','Contract','PaymentMethod'])
 Df2.columns
df2.sample(5)
Df2.dtypes
X_test.shape
X_train[:10]
```

14.Future Scope

Further research can concentrate on refined data-side preprocessing and exhaustive hyperparameter standardization to improve the model performance. More advancedoptimization methods can be used for hyperparameter optimization. To achieve the most accuracy despite machine intensity, a better hyperparameter optimization method would probably increase the classification accuracy of the models to some extent.

15.Team Members and Roles

1.Tamil Thendral – Involved in collecting the datasets and developing the source code.
2. Divya – Involved in data preparation and data preprocessing
3. Swetha – Involved in model building and model training
4. Subulakshmi – Involved in EDA
5. Vaishalini – Involved in model selection