***CUSTOMER CHURN PREDICTION USING MACHINE LEARNING***

**What is Customer Churn?**

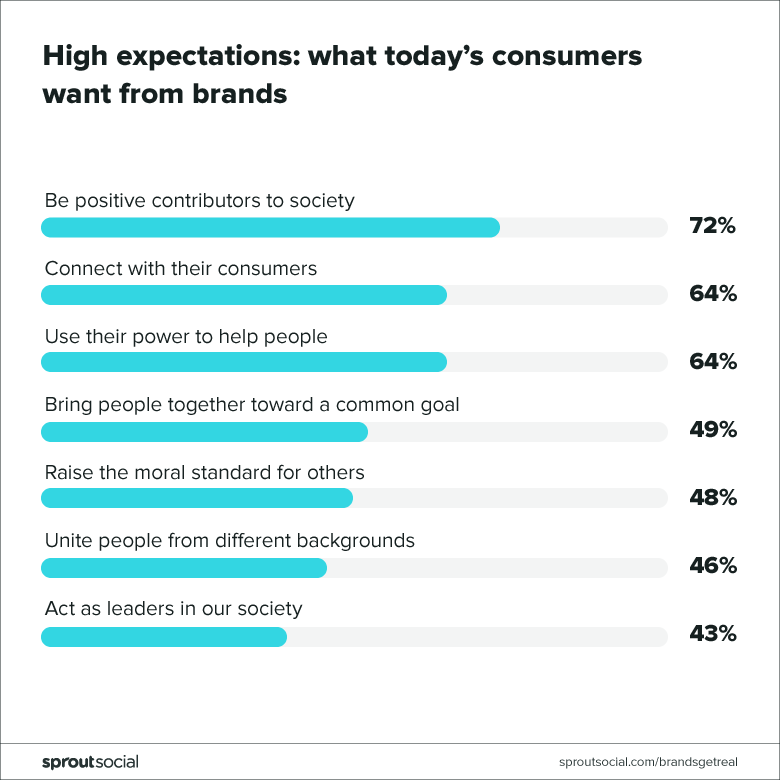
Customer Churn or Customer attrition is the phenomenon where customers of a business no longer purchase or interact with the business. Customer Churn rate is the calculation of percentage of customers who are not likely to make another purchase from a business.

Importanceof Customer Churn!

It is an important parameter for the organization because acquiring a new customer could cost almost 7 times more than retaining an existing customer. Customer Churn is an roadblock for an exponentially growing organization and a retention strategy should be decided in order to avoid an increase in customer churn rates. The most essential step toward predicting customer churn is to start awarding existing customers for constant purchases and support.

Why do customers leaving the brand that once they loved the most?

Because that’s exactly what could happen with just one bad customer experience. As per U.S. reports, even people love a company, 59% will walk away after several bad experiences, 17% will walk away after just one bad experience, 32% would stop doing businesses with a brand they loved after one bad experience.



How does customer churn affect business?

Customer churn rate impacts both retention and acquisition. A high churn rate indicates that customers are not happy with the product or service. It can lead to decrease in sales and overall revenue. Customer churn rate affects several financial metrics of a business.

1.Monthly recurring revenue

2.Net negative MRR churn

3.Customer lifetime value

4.Customer acquisition cost

How do reduce customer churn rate?

1.Analyze why churn is happening

2.Improve the user onboarding process

3.Make yourself indispensable

4.Remind customers the value you provide

5.Increase customer engagement

6.Provide additional services

**Predicting customer churn rate with machine learning**

Data science professionals, like any other machine learning developers, require data to begin with. Researchers determine what data they must collect based on the purpose. Following that, selected data is prepped, preprocessed, and translated into a format appropriate for creating machine learning models. Another important aspect of the task is determining the optimal methods for training machines, fine-tuning the models, and picking the best performers. Once a model with the highest accuracy is chosen, it may be put into production.

Steps to be followed when building machine learning model

* Problem Definition
* Data Analysis
* EDA concluding remarks
* Pre-processing Pipeline
* Building Machine Learning Models
* Concluding Remarks

**Problem Definition**

Telecom Customer churn is a term used when a customer decides to stop using the services of the business. Businesses do customer churn analysis all the time because it is very helpful for a company if they learn which customers are about to leave. The aim of this project is to train a machine learning model on the available data to train a machine learning model that will predict with a high accuracy which customers are about to churn, which in turn will help the business owner in making useful marketing decisions.

It is important to determine what insights are expected from the analysis. In brief, you must pick what question to ask and, as a result, whether to solve a classification or regression problem. Here, we observe Classification is required, as we need to determine whether customer churns or not. The purpose of classification is to establish which class or category a data piece belongs to. To train an algorithm for Classification problems, data scientists would take historical data with predefined target variables.

The Classification algorithm is a Supervised Learning technique used to identify the category of new observations on basis of training data. In Classification, a program learns from given dataset and then classifies new observation into a number of classes. The output variable of Classification is a category, not a value

Here, the problem statement is Binary Classifier, as the classification problem has only two possible outcomes(i.e., whether customer churned or not)

With the classification, we need to answer several questions like..

Will this customer churn or not?

How does contract rate affecting churn rate?

Will a customer renew their subscription?

Will a user downgrade a pricing plan?

Are there any signs of unusual customer behaviour?

**Data Analysis**

Data analysis involves manipulating, transforming and visualizing data in order to infer meaningful insights from the results.

* The very first step is collecting the data. We need to import the dataset using pandas
* Further we need to clean the data i.e., we need to check if there are any missing values and duplicate values.
* We find there is no null data in any of the columns and we also find that there are no duplicate rows. Hence, the data is clean.
* The features of the dataset have classified as follows:

Contextual features: It describe the customers ID, gender, if they have dependents or not and if they are Senior citizen or not

Services that each customers has signed up for: Phone service, Multiple Lines, Internet service, Online security, Online backup, Device protection, Tech support, Streaming TV, Streaming Movies

Customer account information: Tenure, Contract, Paperless billing, Payment method, Monthly charges, Total charges

* We perform Univariate and Bivariate analysis on the data. Univariate analysis is the analysis of one variable. Bivariate analysis is the analysis of exactly two variables to find out the relationship among them. We perform bivariate analysis between target variable and the other variables

**EDA Concluding Remarks**

* We observe that the customers having monthly charges between 70 to 110 have higher churn distribution i.e., customers who have higher monthly charges are leaving the company sooner.
* Gender and partner are evenly distributed with approximate percentage values.
* The difference in churn is slightly higher in females(means more female have left the company in the last month)
* There is a higher proportion of churn in younger customers
* There is also higher proportion of churn in the customers with no partners, no dependents.

**Analysis for all the services that the customer has signed up for:**

* All these services show significant variations across their values
* Almost 90% of the customers have phone services
* Those who don't have phone service, they can't have multiple lines
* Customers who have fibre optic as an internet service are more likely to churn
* Fiber optic service is more expensive than DSL, that may be one of the reasons why customers churn
* Customers with no online security, no online backup, no device protection, no tech support are more likely to churn.
* Streaming TV and movies service is not exactly predictive for churn as its almost evenly distributed for yes and no.

**Insights for the customer account information**

* The shorter the contract, higher the churn rate
* Two year contract, will have the lowest churn rate i.e., if they have more contract years then they will not leave the company, churn rate will be reduced (Customers with a Lower Tenure, are more likely to churn than Customers with a Higher Tenure)
* It explains the motivation for companies to have long-term relationship with their customers then churn rate will be reduced
* Churn rate is highest for the customers opting paperless billing
* Customers who pay through electronic check are more likely to churn.

We need to plot boxplots to check if there are any Outliers in the dataset. An outlier is a data point that differs significantly from other observations. Outliers affect the mean value of the data but have little effect on the median or mode of a given set of data. We plotted box plot to show distributions of numeric data values. We find that there are no outliers in the dataset.

**Pre-processing Pipeline**

* Pre-processing pipeline involves missing data, inconsistent data. We already checked and found that there is no missing data and there are no outlier**s**
* **Feature Engineering** is the pre-processing step of machine learning, which is used to transform raw data into features that can be used for creating a predictive model using Machine learning. Feature Engineering aims to improve the performance of models

**Need for Feature Engineering:** In ML, the performance of the model depends on data pre-processing and data handling. If we create a model without pre-processing, it may not give good accuracy. So we need to build a model with feature engineering, then accuracy of the model will be enhanced.

Having too much data isn’t always good. So, we perform feature engineering to find out which features contribute the most i.e., we check the *correlation. Correlation* to Yes churn and No churn were found. We also checked the correlation between features by plotting heat map. We found that tenure and contract have higher correlation to each other. So, when building the models, we can drop any of them and build the models.

Feature extraction aims at reducing the number of variables (attributes) by leaving the ones that represent the most discriminative information. Feature extraction helps to reduce the data dimensionality (dimensions are columns with attributes in a dataset) and exclude irrelevant information.

During feature selection, specialists revise previously extracted features and define a subgroup of them that’s most correlated with customer churn. As a result of feature selection, specialists have a dataset with only relevant features.

**Building Machine Learning Models**

This project’s major purpose is to create a churn prediction model. Typically, specialists train a large number of models, adjust, assess, and test them in order to identify the one that predicts prospective churners with the appropriate degree of accuracy on training data.

Classic machine learning methods, such as logistic regression, decision trees, random forest, and others, are often used to forecast customer attrition. Data scientists typically use the performance of a baseline model as a criterion to assess the prediction accuracy of more complicated algorithms.

We need to split the data into train and test in a certain ratio. We usually split the data in the ratio 80:20 where 80% of the data will go to the training set and 20% go to the testing set.

After splitting the dataset, we will build the models.

**Logistic regression:** It is a binary classification problem technique. It predicts the likelihood of an event by measuring the relationship between a dependent variable and one or more independent variables (features). The decision for the value of the threshold value is majorly affected by the values of ‘precision’ and ‘recall’. We build the model with an accuracy of 77%

**K Neighbors Classifier:** By default, the KNeighborsClassifier looks for the 5 nearest neighbors. We must explicitly tell the classifier to use Euclidean distance for determining the proximity between neighboring points, then votes for the most frequent label(either Yes or No in this case). We build the model with an accuracy of 69%

**Decision Tree**: It is a type of supervised learning algorithm (with a predefined target variable.) While mostly used in classification tasks, it can handle numeric data as well. This algorithm splits a data sample into two or more homogeneous sets based on the most significant differentiator in input variables to make a prediction. With each split, a part of a tree is being generated. As a result, a tree with decision nodes and leaf nodes (which are decisions or classifications) is developed. A tree starts from a root node – the best predictor. Prediction results of decision trees can be easily interpreted and visualized. Even people without an analytical or data science background can understand how a certain output appeared. Compared to other algorithms, decision trees require less data preparation, which is also an advantage. However, they may be unstable if any small changes were made in data. In other words, variations in data may lead to radically different trees being generated. To address this issue, data scientists use decision trees in a group (AKA ensemble). We build the model with an accuracy of 73%.

We determined feature importance i.e., the features that are contributing more to build the decision tree model. We dropped the unnecessary features and build the model to check if model performance improves. To our conclusion, we found that the accuracy of the model reduced. Hence, feature importance is not fitting for this model.

We also performed Hyperparameter tuning. Decision tree have three main hyperparameters, they are max\_depth, max\_features, class\_weight. We tuned the model with the best parameters then the performance of the model is increased by 4%

**Random Forest:** It is SupervisedML algorithm used widely in classification and regression problems. It is made up of a collection of decision trees on different samples and takes their majority vote for classification. We build this model with an accuracy of 78% for predicting if the customer has churned or not in the last one month.

**Gaussian NB:** Naïve Bayes is a probabilistic machine learning algorithm used for many classification functions and is based on the Bayes theorem. Gaussian Naïve Bayes is a variant of Naïve Bayes that follows Gaussian Normal distribution. In the context of ML, naive bayes classifiers are known to be highly expressive, scalable and reasonably accurate, but their performance deteriorates rapidly with the growth of training set. We build this model with an accuracy of 75%.

**Ada Boost Classifier:** Adaptive Boosting is a technique in ML used as ensemble method. The most common algorithm used with AdaBoost is decision trees with only 1 split. It builds a model and gives equal weights to all the data points. It then assigns higher weights to points that are wrongly classified. By building this model, we got an accuracy of 79%.

**Deployment:**

The last stage of the churn prediction project workflow is deployment. The chosen model or models must be put into production. Data scientists must monitor a model's accuracy and improve it as needed.

"Predicting customer attrition using machine learning and artificial intelligence is an ongoing process that never stops. When customer-facing teams provide input or new data becomes available, we evaluate model performance and update features as needed to enhance accuracy.

We find the model that performs better (i.e., with higher accuracy) and save that model in pickle format. This is the process of deployment.

**Conclusion**

Churn rate is like a health indicator for subscription-based companies. The ability to identify customers that aren’t happy with provided solutions allows businesses to learn about product or pricing plan weak points, operation issues, as well as customer preferences and expectations to proactively reduce reasons for churn.

Companies with a large customer base and numerous offerings would benefit from customer segmentation. The number and choice of ML models may also depend on segmentation results. Data scientists also need to monitor deployed models, and revise and adapt features to maintain the desired level of prediction accuracy.

Dataset source: https://github.com/dsrscientist/DSData/blob/master/Telecom\_customer\_churn.csv

Github link:

https://github.com/Kolaparthi-Likhitha/Data\_trained/blob/main/Evaluation\_Projects/Telecom\_Customer\_Churn.ipynb