Customer Churn Analysis

* Problem Statement:-

Customer churn is when a company’s customers stop doing business with that company. Businesses are very keen on measuring churn because keeping an existing customer is far less expensive than acquiring a new customer. New business involves working leads through a sales funnel, using marketing and sales budgets to gain additional customers. Existing customers will often have a higher volume of service consumption and can generate additional customer referrals.

Customer retention can be achieved with good customer service and products. But the most effective way for a company to prevent attrition of customers is to truly know them. The vast volumes of data collected about customers can be used to build churn prediction models. Knowing who is most likely to defect means that a company can prioritize focused marketing efforts on that subset of their customer base.

Preventing customer churn is critically important to the telecommunications sector, as the barriers to entry for switching services are so low.

* DATA ANALYSIS:-

Customer Churn Analysis data contain total 7043 rows and 21 columns, data have total 20 features naming-

Customer Id :- It Shows the Unique Id of each Customer .Contain Object Data.

Gender :- Gender of the Customer .Contain Object Data

Senior Citizen:- Shows whether the customer is Senior Citizen or not and contain the numerical data, having binary numbers 0 and 1

Partner :- The customer is in partnership with company or not and contain object data having two outcomes yes and no.

Dependents :- This column shows the Customer is financially dependent or not and have the object data having two outcomes with yes and no.

Tenure :- Tenure of the service company providing to the service. Contain the numerical data.

Phone Service :- Company provides the phone service to the customer and data is in object type having and yes and no.

MultipleLine :- Having object data and shows if the company provides multiple lines.

Internet Service :- Shows the data in object type is that customer getting Internet services in DSL, Fiber Optic and No internet service.

Online Security :- Shows if the customer getting InternetServices alongwith is he getting Online Security.Contain the object data having binary classification.

,Online Backup:- This column contain the data regarding online backup. Is that customer have the facility of Online Backup and the data is in object type.

Device Protection:- Customer have the facility of Device Protection and contain categorical object data.

Tech Support:- Column shows the information about the customers who are getting technical support.

Streaming Tv:- Information about the StreamingTv data.

Streaming Movies:- Data shows the Streaming Movies facility provided to the consumers.

Contract:- Inforamtion about the customers who have contract with the company on the month to month,one year and two year basis.

Paperless Billings:- Customers with and without paperless billings.

Payment Method:- Electronic check,Mailed check,Bank transfer (automatic)

Credit card (automatic) These are the payment methods used by specifc

customers.

Monthly Charges:- Monthly charges for the service.

Total Charges:- Total charges of the service.

Target Varaible:-

Churn:- It is the target variable contain the binary classification having the object data which is No and Yes and it shows the customers who left within the last month.

This is the overall dataset information about the data.

* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents
* EXPLORATORY DATA ANALYSIS:-

In the Exploratory data analysis I have checked for the data shape which is 7043 rows and 21 columns.

Data Does not have any null values.

Columns containing object data.

Target data variable count which is Yes-1869 ,No-5174 , here yes is for the customer count who had left the service and No. for the retained customers which are taking the services.

And the data count is imbalanced.

Customer Id is the just unique id it is not useful , So I have dropped that column.

* Data Visualization :-

In the visualization process I have plotted the countplot for every column to get idea about realation between target variable.

While comparing with gender data, churn exist with both. Customers who are partner stay longer with the services.

* A lot of customers choose the Fiber optic service and it's also evident that the customers who use Fiber optic have high churn rate, this might suggest a dissatisfaction with this type of internet service.
* Customers having DSL service are majority in number and have less churn rate compared to Fibre optic service.
* Customers without dependents are more likely to churn
* Customers that doesn't have partners are more likely to churn
* About 42% of the senior citizens churn.
* Most customers churn in the absence of online security.
* Customers with no TechSupport are most likely to migrate to another service provider.
* Very small fraction of customers don't have a phone service and out of that, about 25% Customers are more likely to churn.
* Customers with higher Monthly Charges are more likely to churn.
* Customers with higher Total Charges are more likely to churn.
* Customers who stayed with the company for longer time are more less likely to churn now.
* Data Preprocessing:-

While Checking for the Co-relation between features and variables , Contract and tenure are highly correlated with each other and has multicolinearity . So I decided to drop the one of them and after checking the corealtion with target variable, Contract is correlated than the tenure and along with it I also have dropped the gender column it’s also not useful to predict the churn.

After that I have separated features in X variable and Target in y. I have used ExtraTreeClassifier to analyse which are the important features.

SeniorCitizen','PhoneService','Dependents','StreamingTV','StreamingMovies and these are the features which are less important. Excluding this features I trained my model.

To balance the target variable count I used SMOTE oversampling method. And I have with final Count 3877 and 2907.

* Building Machine Learning Models:-

As we have the binary predication So After preprocessing I build total 7 models with various accuracy scores ,classification report and confusion matrix.

|  |  |
| --- | --- |
| Model Name | Accuracy Scores |
| Logistic Regression | 80% |
| Decision Tree Classifier | 72% |
| Random Forest Classifier | 76% |
| Support Vector Classifier | 73% |
| Ada Boost Classifier | 80% |
| Bagging Classifier | 77% |
| XGB Classifier | 79% |
| KNeighborsClassifier | 73% |

Logistic Regression and Adaboost are giving same and high accuracy as compared to the another models . After checking for cross validation scores and applying hyperparameter tuning Adaboost is giving the highest accuracy and area under the curve for the Adabosst classifier is 84%.

**Confusion Matrix:-**

|  |  |
| --- | --- |
| 918  True Positive Rate | 121  False Positive Rate |
| 166  False Negative Rate | 204  True Negative Rate |

From the confusion matrix also false rates are low. So I saved Adaboost Classifier as final model.

* Conclusions:-

I have achieved about 80% accuracy on the test.

Customer churn is definitely bad to a firm ’s profitability. Various strategies can be implemented to eliminate customer churn. The best way to avoid customer churn is for a company to truly know its customers. This includes identifying customers who are at risk of churning and working to improve their satisfaction. Improving customer service is, of course, at the top of the priority for tackling this issue. Building customer loyalty through relevant experiences and specialized service is another strategy to reduce customer churn. Some firms survey customers who have already churned to understand their reasons for leaving in order to adopt a proactive approach to avoiding future customer churn.

Reducing churn is more important than ever, especially while the telecom industry is growing its competitive advantage. However, many operators have not yet taken the important steps to build a strong analytical foundation for success. The companies that move quickly towards these changes could be best positioned for success in the future.