

Customer churn analysis

Customer churn analysis is the method of measuring the rate at which clients opt out of purchasing more of a company's product or services.

Churn analysis simply tells us what percentage of our customers don't return compared with the percentage who conduct repeat business.

Churn rate tells us what portion of our customers leaves over a period of time.

Customer churn is a measure of loss of customers or loss of revenue to the business during a given period.

Businesses are very keen on measuring churn because keeping an existing customer is far less expensive than acquiring a new customer, customer retention can be achieved with good customer service and products.

There are 7043 rows and 21 columns are present in the data. churn is the target column that holds the objective type of data with the help of this I can say my data has the classification problem input columns are given below.

1. Customer id
2. Gender
3. Senior citizen
4. Partner
5. Dependents
6. Tenure
7. Phone service
8. Multiple lines
9. Internet service
10. Online security
11. Online backup
12. Device protection
13. Tech support
14. Streaming movies
15. Contract
16. Paperless billing
17. Payment method
18. Monthly charges
19. Total charges

Some columns are the objective type and some are int type

Customer id, gender, partner, dependent, phone service, multiple lines, internet service, online security, online backup, device protection tech support, streaming movies, contract, paperless billing payment method, total charges have the object type of data

Senior citizen, tenure has the int data type

Monthly charges has the float data type

With the help of the seaborn library and matplotlib, we are trying to find insight into the data. First, we have to import the seaborn library and matplotlib.

After importing the library we visualize the data with the help of a count plot we are trying to count the value, And find the relation with the target column

When we find the relation between dependent and the churn column we see that those who are not dependent churned more and those who are dependent churned less

Then we find the relation between the internet service and the target column then we visualize that those who have the fiber optic service churned more with the comparison of Dsl and No

Again with the help of count plot we find the relation between partner and churn. We see those who are not partner are churned more and those who are partner churned less. Then we find the relation with dependents and churn column and we see those who are dependents churned less and those who are not dependent churned more. Again with the help of count plot, we find the relation between phone service and churn we see those who have phone service churned more and those who have not phone service churned less. if we find the relation between the online backup and churn we see that those who have no online backup are churned more and those who have no internet service churned less .

When we find the relation between device protection and churn we see that those who have device protection churned less with the comparison of those who have no device protection

After eda process, we build a machine learning model and tuned it and also check the cross validation score then save the model according to the accuracy , we built four classification model 1. Support vector classifier 2. Decision tree classifier 3. K neighbours classifier 4. Random forest classifier and tuned it also then I see support vector classifier give 78 per accuracy and decision tree classifier gives 72 per accuracy when I tuned the decision tree classifier it give the 79 per accuracy knn gives 77 per accuracy when I tuned knn it gives 78 per accuracy rfc gives 79 per accuracy and the cross validation score of rfc is also 79 so I save the random forest classifier.