College code:4212
Register number:421221243004

CUSTOMER CHURN PREDICTION

PHASE2-DATA ANALYTICS WITH COGNOS: GROUP2

INTRODUCTION:

Predicting the customer churn is a critical business problem in a variety of sectors including telephones, subscriptions services, e-commerce and more. Business may take proactive steps to retain consumers by utilizing churn prediction to identify the customers who are likely to discontinue their goods or services.

In this phase we are going to explain about design and ideology that are going to present to solve this problem.

Dataset link: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

To this problem this dataset is given to us so by using this dataset we are going to solve our problem.

In the phase1 we have defined certain steps to solve the problem step by step now we are going to explain which methodology we are going to use to solve this problem in each step**.**

Clearly define the problem:

We have clearly understood the problem that we are going to understand about the customers who are likely stop the usage of the services in the telecom sector.

Data collection:

The dataset is already given for us:

Dataset link: https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Preparing of the data:

First we have to understand what was the data we are going to analyse for this we have to clean and process the data by using suitable techniques like **dropping the null values, data types, remove the duplicate values, visualize the missing values drop the duplicates, by using the suitable functions like drop, is null etc.…**

Exploratory data analyses:

This was the most important step in this project so we have to represent our data in the **understandable visualization tools like pie chart, bar graph, histogram** to represent the relation between the two attributes in the given coloumn.

Feature selection:

In this step we are going to explain about the features or attributes that are going to select in the dataset and we have to **represent the relationship between the data visualization.**

Model selection:

We are going to solve the problem by using various algorithms **like KNN, svc, random forest, logistic regression and the decision tree classifier,** we can solve by using any of this model and choose the accurate model.

Model training and validation:

We are using various model validation methods like **label encoder, ascending values, dis plot, standardization and scaler.**

Model evaluation:

We are divide our data set by using the **training and testing, predicting the values by using the 2 dimensional planes as a result the accuracy model** will tells us which algorithm do we use.

Result representation:

According to this project can be represented in many ways like using the ROC curve, confusion matrix etc… we are going to use the **confusion matrix** to solve this.

Reporting and visualization:

To inform stakeholders about the **churn prediction and insights provide the periodical reports and dashboards** we are going to do this in a confusion matrix and accuracy of the data.

Business action:

Companies will take the necessary actions after the reports and visualization of the data had been created by us and **provide some special offers to grab the customer attention** to increase their sales.