CUSTOMER CHURN PREDICTION

TEAM MEMBERS

kesavan.K

DeepakKumar.K

HariDass.K.S

Dhanasekaran.S.A

PROBLEM DEFINTION AND DESIGN THINKING

PROBLEM DEFINTION:

Customer churn analysis and prediction in telecom sector is an issue now a days because it's very important for telecommunication industries to analyze behaviors of various customer to predict which customers are about to leave the subscription from telecom company. So machine learning techniques and algorithm plays an important role for companies in today's commercial conditions because gaining a new customer's cost is more than retaining the existing ones. This project focuses on various machine learning techniques for predicting customer churn through which we can build the classification models such as Logistic Regression, Random Forest and lazy learning and also compare the performance of these models.

DESIGN THINKING:
1.Analysis Objectives:

identified churn factors that are essential in determining the root causes of churn. By knowing the significant churn factors from customers' data, Customer Relationship Management (CRM) can

improve productivity,recommend relevant promotions to the group of likely churn customers based on similar behavior patterns, and excessively
improve marketing.

2.Data collection:

The customer who cease a product or service for a given period is referred as churner. In a telecommunication company the individual who has opted service from a firm is referred to as Churn. The individual who probably intends to depart from the firm in near future was predicted by the churn model. Many industries build a model like a churn as a common application for data mining technique. Mobile telephone organizations present across the globe are almost on the verge of building their own churn model. Furthermore, to retain the customers, churn results can be efficiently utilized for various other goals.Churn Management approach is actually the first step in building a model.

3.Visualization Strategy:

Due to data imbalance churn rate is low in the overall dataset. In order to fix it analysis is performed on certain important features column like age on network(AON), incoming calls usage, outgoing calls usage, operator wise calls usage, recharge amount, recharge count, average revenue per user and 2G and 3G. These columns seem to have outliers at the top percentile which is treated using outliers treatment. The outlier treatment is to cap the outliers at the 99th percentile for the above mentioned features column which derives some mandatory features. Remove the columns of date to perform sample logistic regression on the available data.

4.Predictive Modeling:

Now that the class is well balanced it is splited into train (70%) and test (30%) dataset.

Apply PCA on the training dataset for dimensionality reduction and feature selection. Draw the screeplot for the PCA components and pick the right number of PC components to build the model and chose 60 PCA components for model building using the following PCA algorithm.

Step 1: : Consider a Data with n-dimensions .

Step 2: : Subtract the mean - from each of the data dimensions.

Step 3: Calculate the covariance matrix

.**Step 4:** Calculate the eigen values and eigenvectors of the covariance matrix

Step 5: Reduce dimensionality and form feature vector

Step 6: : Deriving the new data Final Data = Row Feature Vector x Row Zero Mean Data

Build models like Random Forest , KNN and Logistic Regression .Stratified k-fold cross validation method is used to select the best model by estimating their performance