**A REPORT**

*On Project*

**TELEPHONE CHURNING USING MACHINE LEARNING**

*By*

PRANITA SONAWANE (BE B 61, COMPUTER 2022-23)

ANAGHA KULKARNI (BE B 62, COMPUTER 2022-23)

GAURI PAWAR (BE B 63, COMPUTER 2022-23)

NANDINI MARKE (BE B 64, COMPUTER 2022-23)



**K. K. WAGH INSTITUTE OF ENGINEERING EDUCATION AND RESEARCH, NASHIK, MAHARASHTRA.**

**DEPARTMENT OF COMPUTER ENGINEERING**

**DETAILS OF THE PROJECT**

**Problem Statement :**

Based on the introduction the key challenge is to predict if an individual customer will churn or not. To accomplish that, machine learning models are trained based on 80% of the sample data. The remaining 20% are used to apply the trained models and assess their predictive power with regards to “churn / not churn”. A side question will be, which features actually drive customer churn. That information can be used to identify customer “pain points” and resolve them by providing goodies to make customers stay.

To compare models and select the best for this task, the accuracy is measured. Based on other characteristics of the data, for example the balance between classes (number of “churners” vs. “non-churners” in data set) further metrics are considered if needed.

**Objectives / Motivation Scope**

The main objective of this research is to predict Telephone Churning. Through the use of statistical methods, algorithms are trained to make classifications or Predictions Telephone Churning refers to various customers that whether they will continue to use the service or not.

Provides new approach to concealed patterns in the data.

* Helps to Telecom companies to hold the customers.
* To implement Support Vector Clustering Algorithm that predicts the telephone churning using user’s data.
* Improve business of telecom companies.

**Dataset**

The dataset provides the customers information. It includes over 2,666 records and 20 attributes.

This dataset contains data that stores state, international plan, voice mail plans, total day calls, total day charge, total day minutes, etc.

On this basis whether the customer is likely to leave a service or to cancel a subscription to a service is predicted.

Here target variable is “Churn” column.

**1. Data Pre-processing**

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So for this, we use data pre-processing task.

**2. Exploratory Data Analysis**

* **Missing variables**

There are no attribute having missing values. So no need to treat it.

* **Correlation**

Correlation measures the relationship between two variables. We mentioned that a function has a purpose to predict a value, by converting input (x) to output (f(x)). We can say also say that a function uses the relationship between two variables for prediction.

* **Resampling of Data**

Resampling is the method that consists of drawing repeated samples from the original data samples. The method of Resampling is a nonparametric method of statistical inference.

**3. SVC Algorithm**

Support vector clustering (SVCs) are powerful yet flexible supervised machine learning methods used for classification, regression, and, outliers’ detection. SVCs are very efficient in high dimensional spaces and generally are used in classification problems. SVCs are popular and memory efficient because they use a subset of training points in the decision function.

C-support vector classification whose implementation is based on **l**ibsvm. The module used by scikit-learn is sklearn.svm.SVC. This class handles the multiclass support according to one-vs-one scheme.

The main goal of SVCs is to divide the datasets into number of classes in order to find a maximum marginal hyperplane (MMH) which can be done in the following two steps −

* Support Vector Clustering will first generate hyperplanes iteratively that separates the classes in the best way.
* After that it will choose the hyperplane that segregate the classes correctly.

Some important concepts in SVM are as follows −

* **Support Vectors** − They may be defined as the datapoints which are closest to the hyperplane. Support vectors help in deciding the separating line.
* **Hyperplane** − The decision plane or space that divides set of objects having different classes.
* **Margin** − The gap between two lines on the closet data points of different classes is called margin.

**4. Testing Dataset**

The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. If the test set does contain examples from the training set, it will be difficult to assess whether the algorithm has learned to generalize from the training set or has simply memorized it.

**5. Prediction**

Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome. Accuracy of the model is represented using confusion matrix. The matrix compares the actual target values with those predicted by the machine learning model.

**Conclusion**

After training and testing the dataset SVC algorithm, give the accuracy of 85.45%. Customer churn is always a grievous issue for the Telecom industry as customers do not hesitate to leave if they don’t find what they are looking for. Customer churning is directly related to customer satisfaction. There is no standard model which addresses the issues of global telecom service providers accurately. Keeping all such things into consideration, a research thesis on customer churn prediction based on mobile data usage volumes. Confirmed trends observed through correlations: As Quality increases, Volume increases, accordingly Annoyance and Churn risk decreases. Weekly data volumes with customer’s recent history and necessary attributes like age, gender, tenure, bill, contract, data plan, etc., are pivotal for churn prediction.