



**Department of Computer Science  
Practical Business Analytics Project (COMM053)**

## **Group: D-Walk**

### **Group Members**

<b>Mohamed Ali</b>	6675010
<b>Ranjeet Singh Yadav</b>	6656616
<b>Pavan Siddihally Venugopala Reddy</b>	6644984
<b>Evans Gichuki</b>	6677196
<b>Mallavarapu Arun showri Chowdary</b>	6679047
<b>Hitesh Sai Marisetty</b>	6673633

**Title: Do our customers have the intention of leaving us in the future?**

## Table of Contents

<b>Problem definition</b>	<b>3</b>
Why is Churn prediction Necessary?	4
The dataset	4
<b>Business Analytics Tasks</b>	<b>5</b>
Our approaches	5
<b>Expectation</b>	<b>5</b>
<b>Project Plan</b>	<b>6</b>
<b>References</b>	<b>7</b>

# Problem definition

*Churn rate - the annual percentage rate at which customers stop subscribing to a service or employees leave a job.*

Starting and running a business is itself a tedious task. Factors such as inflation, new regulations by the government, shortage of staff, e.t.c are likely to affect some parts of the business. This would then reflect on the decrease of quality in the product.

With the increasing number of competitors, clients might have to juggle between core competitors, whereby the winner company will be the one that will have met the customer needs in the most efficient way. If smaller businesses can't compete against their larger counterparts, they need to find ways to pivot and stay in business. To do that takes a keen business sense (Adams, 2021)

Of course, gaining customer interest plays an imperative role for the company to be in the market. Ultimately, customers are the one who are going to use the services and provide opportunities for the company to grow more and more. We are able to say that the profit made by the company is directly proportional to the customer satisfaction. Globally, the average value of a lost customer is \$243. (Customer Retention Statistics, 2021)

It seems that it is not an easy task to gain customer interest. The organisation has to take care of various necessary actions to fulfil the needs of the average customer. It is possible to see whether the customers are happy with their services by taking regular surveys and analysing this data. Sometimes the organisations have to make changes in their services so that they can level with the market.

As a team we are going to use a dataset called "Predict Customer Churn" which is available in Kaggle to find out about customers with the intention of leaving the organisations in the near future. Our main focus will be using machine learning algorithms to get as accurate values as possible. The dataset that we are going to use consists of 21 columns and around 7,000 rows - we are going to pass each column and row through various analytic algorithms to be as accurate as possible.

## Why is Churn prediction Necessary?

- 1) Predicting churn rates allows the company to target that individual in an attempt to prevent them from discontinuing their subscription with the provider.
- 2) Predicting churn rates helps the business gain a better understanding of future expected revenue.
- 3) And, since the cost of acquiring a new customer is higher than keeping an existing one, it is cheaper and cost effective to do everything in your power to keep those existing customers.(How Churn Prediction Can Improve Your Business - Baremetrics, 2021)

## The dataset

Each row represents a customer, each column contains customer's attributes described on the column Metadata. The dataset contains 7043 rows and 21 columns. The "Churn" column is our target. Our work is mainly based on one dataset

(<https://www.kaggle.com/blatchar/telco-customer-churn>):

1. PredictCustomerChurn. This contains 1 csv/ excel spreadsheet file as follows
2. **customerID** - This is the customer id
3. **Gender** - Customers gender(male or female)
4. **SeniorCitizen** - Whether the customer is a senior citizen or not(1/0)
5. **Partner** - customer is married or not(yes/no)
6. **Dependents** - customer has dependents(yes/no)
7. **Tenure** - how long one has been a customer in months (int)
8. **PhoneService** - whether client has a phone service(yes/no)
9. **MultipleLines** - whether the customer has many lines (yes/no/other)
10. **InternetService** - internet service type(dsl,fibre,none)
11. **OnlineSecurity** - internet security (yes,no, none)
12. **OnlineBackup** - customer has online backup or not (yes, no, no internet service)
13. **DeviceProtection** - device protection or not (yes, no, no internet service)
14. **TechSupport** - has tech support or not (yes, no, no internet service)
15. **StreamingTV** - has streaming TV or not (yes, no, no internet service)
16. **StreamingMovies** - streaming movies or not (yes, no, no internet service)
17. **Contract** - contract term of the customer (month-to-month, one year, two year)
18. **PaperlessBilling** - paperless billing or not (yes, no)
19. **PaymentMethod** - customer's payment method (electronic check, mailed check, bank transfer, credit card)
20. **MonthlyCharges** - amount charged to the customer monthly
21. **TotalCharges** - total amount charged to the customer

## Business Analytics Tasks

### Our approaches

We will be using “Predict customer churn dataset”. The best approach would be using supervised learning as we have an idea about the output.

1. Initially the data needs to be pre-processed.
2. As we have many columns with string as a data type, we need to convert those columns with numeric datatype.
3. We also have to handle non-values and replace them with either mean or medium of the equivalent columns.
4. The unwanted columns need to be dropped from the dataset.
5. After the data is pre-processed then we need to scale the data, for this we are planning to use a scale function algorithm.
6. After scaling we have to divide the dataset into two parts as training and testing datasets. For dividing the dataset we are planning to use the “” algorithm. It will divide the data as 70% for training and the remaining 30% for the testing dataset.
7. Finally, we are going to use several classifier algorithms and use the algorithm with the best accuracy to create and train the model.

## Expectation

After the use of analytical algorithms we will be able to predict the tendency of customers to leave the organisations in near future. We will also be able to build confusion matrices and also be able to measure the classification by determining the recall and precision values, along with metrics like MAE and RMSE that are used for evaluating the model.

# Project Plan

Project Plan					
	Week-1	Week-2	Week-3	Week-4	Week-5
<b>Mohamed Ali</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using linear and multi-Linear regression	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report
<b>Ranjeet Singh Yadav</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using Naive Bayes and Logistic regression	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report
<b>Pavan Siddihally Venugopala Reddy</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using linear and multi-Linear regression	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report
<b>Evans Gichuki</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using logistic Regression and Naive Bayes	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report
<b>Arun</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using K-nearest neighbors	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report
<b>Hitesh</b>	Planning, Data Preparation	Pre-processing of data	Training and testing using K-nearest neighbors	Model Development, interpretation and Evaluation	Data Interpretation Result collection and project report

## References

Adams, R., 2021. *10 Reasons Why 7 Out of 10 Businesses Fail Within 10 Years*. [online] Entrepreneur. Available at: <<https://www.entrepreneur.com/article/299522>> [Accessed 1 March 2021].

Baremetrics. 2021. *How Churn Prediction Can Improve Your Business - Baremetrics*. [online] Available at: <<https://baremetrics.com/academy/churn-prediction-can-improve-business>> [Accessed 1 March 2021].

Neil Patel. 2021. *Customer Retention Statistics*. [online] Available at: <<https://neilpatel.com/blog/retaining-customers/>> [Accessed 1 March 2021].