**CUSTOMER CHURN PREDICTION MODEL**

1. **Business Understanding**
   1. **Business Overview**

Understanding if a customer will cease his or her relationship with the company is a relevant business problem because retained customers form a major chunk of revenue for companies when compared to new targeted customers.

One of the foundational elements of 21st-century sources of growth is data-driven innovation. Huge volumes of data, or "big data," are being produced and used as a result of the convergence of numerous phenomena, including the Data Analytics process. These massive data sets are becoming a key resource for the economy, supporting the growth of new businesses, methods, and goods while giving rise to substantial competitive advantages.

The project takes advantage of big data technologies including Data Analytics to solve real-world problems affecting the organization. This will help the management in better decision-making.

* 1. **Business Objective**

1. To create a model that can provide intuitive action insights and predict NPS
2. To create a model that can provide data, insights, technologies, and dashboarding to intuitively resolve customer issues
   1. **Business Success Criteria**

To create a model that can have a metric of success of 90% and above without overfitting.

* 1. **Assessing the Situation**

1. **Resource Inventory**
2. **Dataset**

We shall source our data from Kaggle. It can be accessed via this link: <https://bit.ly/3Q2Ri1E>

1. **Software**

We shall use Python, Jupyter Notebook, Tableau, and Jira.

1. **Assumption**

We shall assume that the features of the dataset are almost similar to the actual day-to-day data about the problems faced by the company.

1. **Constraints**

Since we are dealing with big data, we may face challenges to deploy the big data frameworks on personal PCs. But with the help of better equipment, we can deploy the model to a production level.

* 1. **Data Mining Goals**

1. To determine the most common type of payment and the relationship with other features
2. To check the distribution of data based on gender and the relationship with the churn rate
3. To determine the relationship between the most common contract and the churn rate
4. To check the Zip codes and cities that have the highest churn rate.
   1. **Data Mining Success Criteria**

Our success criteria will be measured by the following:

1. Our data should be clean enough to enable great machine learning models results
2. Our data should be enough to give the most insights.
   1. **Producing a Project Plan**

The project plan will use the Kanban framework. We shall use the Jira platform to display the boards.

**2.0 Data Understanding**

**2.1 Data Understanding Overview**

**2.2 Data Description**

We have 7043 rows and 33 columns in our dataset. Most of our datatypes are categorical hence we shall need to do encoding on some of the features to use in the Machine Learning Model.

Below are featured descriptions of the features:

* **CustomerID**: A unique ID that identifies each customer.
* **Count**: A value used in reporting/dashboarding, to sum up, the number of customers in a filtered set.
* **Country**: The country of the customer’s primary residence.
* **State**: The state of the customer’s primary residence.
* **City**: The city of the customer’s primary residence.
* **Zip Code**: The zip code of the customer’s primary residence.
* **Lat Long**: The combined latitude and longitude of the customer’s primary residence.
* **Latitude**: The latitude of the customer’s primary residence.
* **Longitude**: The longitude of the customer’s primary residence.
* **Gender**: The customer’s gender: Male, Female
* **Senior Citizen**: Indicates if the customer is 65 or older: Yes, No
* **Partner**: Indicate if the customer has a partner: Yes, No
* **Dependents**: Indicates if the customer lives with any dependents: Yes, No. Dependents could be children, parents, grandparents, etc.
* **Tenure Months**: Indicates the total amount of months that the customer has been with the company by the end of the quarter specified above.
* **Phone Service**: Indicates if the customer subscribes to home phone service with the company: Yes, No
* **Multiple Lines**: Indicates if the customer subscribes to multiple telephone lines with the company: Yes, No
* **Internet Service**: Indicates if the customer subscribes to Internet service with the company: No, DSL, Fiber Optic, Cable.
* **Online Security**: Indicates if the customer subscribes to an additional online security service provided by the company: Yes, No
* **Online Backup**: Indicates if the customer subscribes to an additional online backup service provided by the company: Yes, No
* **Device Protection**: Indicates if the customer subscribes to an additional device protection plan for their Internet equipment provided by the company: Yes, No
* **Tech Support**: Indicates if the customer subscribes to an additional technical support plan from the company with reduced wait times: Yes, No
* **Streaming TV**: Indicates if the customer uses their Internet service to stream television programming from a third party provider: Yes, No. The company does not charge an additional fee for this service.
* **Streaming Movies**: Indicates if the customer uses their Internet service to stream movies from a third party provider: Yes, No. The company does not charge an additional fee for this service.
* **Contract**: Indicates the customer’s current contract type: Month-to-Month, One Year, Two Years.
* **Paperless Billing**: Indicates if the customer has chosen paperless billing: Yes, No
* **Payment Method**: Indicates how the customer pays their bill: Bank Withdrawal, Credit Card, Mailed Check
* **Monthly Charge**: Indicates the customer’s current total monthly charge for all their services from the company.
* **Total Charges**: Indicates the customer’s total charges, calculated to the end of the quarter specified above.
* **Churn Label**: Yes = the customer left the company this quarter. No = the customer remained with the company. Directly related to Churn Value.
* **Churn Value**: 1 = the customer left the company this quarter. 0 = the customer remained with the company. Directly related to Churn Label.
* **Churn Score**: A value from 0-100 that is calculated using the predictive tool IBM SPSS Modeler. The model incorporates multiple factors known to cause churn. The higher the score, the more likely the customer will churn.
* **CLTV**: Customer Lifetime Value. A predicted CLTV is calculated using corporate formulas and existing data. The higher the value, the more valuable the customer. High-value customers should be monitored for churn.
* **Churn Reason**: A customer’s specific reason for leaving the company. Directly related to Churn Category.

**3.0 Methodology**

We shall use the Logistic Regression model as our base model then thereafter, we shall use the Adaboost and XgBoost Models