**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**II SEMESTER 2020-21**

## DSE CL ZG628T DISSERTATION

**Dissertation Outline**

1. Student details

Name - Samarth Malhotra

Student ID - 2018ap04535

WILP email id - 2018ap04535@wilp.bits-pilani.ac.in

Mobile - 9167329289

1. Date - 3/4/2021
2. Dissertation Title: Customer Churn Prediction in Telecom using Machine Learning
3. Supervisor details

Name - Mayank Jain

Role in organization - Principal Software Developer

Qualification - B.E Computer Science

Experience - 13 years 5 months

Email id - mayank.nirmal@gmail.com

Mobile - 8805413880

1. Problem statement (what is the problem being addressed)

Customer churn is the likelihood of a customer to leave a brand, stop using its services and switching over to other providers. It is a major challenge in businesses with subscription-based model and has direct impact on the revenue of the company, especially in the telecom field. The cost of churn includes both the loss of revenue and the marketing costs involved in replacing those customers with new ones, therefore, predicting and preventing customer churn has a potential revenue source therefore the telecom companies must make an effort to retain their customers.

In the face of stiff competition in the market, the customers have very wide choice and often they switch over from one product to another, there is always a search for better options.

There can be several factors responsible for customer churn, including:

* The availability of quality services
* Low-cost alternatives
* Better features and content
* Customer experience
* Availability of self-service options
* Easy access to the maintenance staff
* Network coverage

The above-mentioned list is not exhaustive, the inventory could vary depending upon service provider and would require domain knowledge.

Customer churn prediction modelling aims to understand the customer’s behavior and attributes (gender, age, dependents, financial status), also the likelihood to switching of the brand, possible reasons and the remedial measures to retain the customer.

With a better understanding and an insight into potential customers leaving the brand in the volatile market condition, the brand can take a suitable action after the analysis which will lead to most retention impact on the customer.

1. Business process flow, if any

None

1. Objective of the project (Expected outcome)

The main objective of the present project is to design a churn prediction model that could help telecom operators to foresee the customer behavior and accurately predict the customers who are likely to churn.

In order to know the customer behavior, the relevant historical data will be used and as the current research in the field confirms machine learning could be efficiently applied to predict the customer churn and take the retention measures.

Principal objectives:

1. Create visualizations to showcase how each feature is affecting the target class
2. Create multiple machine learning models to predict the target variable and evaluate them with multiple metrics (AUC Score, precision, recall, f1 score)
3. Identify the features which are important for the chosen model
4. Uniqueness of the project

In many organizations the customer churn is reactive in the sense that when customer calls to end the subscription only then offers are rolled out to retain the customers.

In this project we are aiming to make this process proactive by actively predicting unhappy customers in advance and making necessary adjustments to retain them.

1. Benefit to the organization

Oracle provides end to end cloud solutions to telecommunication providers, it spans everything from capturing the network calling data to billing and processing payments to generating audit reports.

The project will directly benefit Oracle in providing up to date information about customer churn to its client i.e., telecom operators and operators in turn will ensure that the customer churn could be prevented in time by opting for retention strategies which will have direct impact on their revenue.

1. Scope of work

* Acquisition of data
* Exploration of data
* Cleaning of data
* Implementing the code for churn prediction using python
* Visualizations
* Report Creation

1. Resources needed for the project, including people, hardware, software, etc.

* Telecom Customer Data
* Visualization libraries
* Windows Machine
* Python libraries

1. Potential challenges & risks in doing the project

As the features are unknown to us that are useful to predict customer churn, we have to consider all the data which is huge in volume therefore it’s difficult to perform feature engineering using tradition approaches. Usually, the data for the analysis is consolidated from multiple sources which is of different variety and causes a formidable challenge.

Many a time the dataset is unbalanced which means there can be possibility of having one class as 10% or less in compared to the other class.

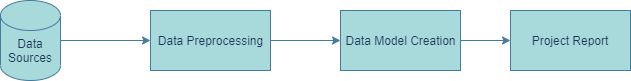
The telecom customer datasets are usually full of features and therefore feature engineering process gets complicated.

1. Background of previous work done in the chosen area

In the previous work many machine learning approaches have been applied to predict churn in telecom, some of these approaches have used advanced methodology of data mining to predict churn on prepaid customers also some have used neural networks for large telecom companies.

There has been proposed approaches on large social networking dataset analyzed using big data platforms for the customer churn.

1. Solution architecture



1. Detailed Plan of Work (as follows)

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Task** | **Expected date of completion** | **Names of Deliverables** |
| 1 | Acquisition of telecom data | 6/5/2021 | csv |
| 2 | Preprocessing of data | 8/5/2021 | Python code |
| 3 | Exploration of data | 10/5/2021 | Python code and visualizations |
| 4 | Feature Engineering | 14/5/2021 | Python code |
| 5 | Model Creation / Evaluation | 31/05/2021 | Python code |
| 6 | Identifying Important Features | 10/6/2021 | Python code and visualizations |
| 7 | Project Report | 10/7/2021 |  |