# Project Report

# TELECOM CUSTOMER CHURN PREDICTIVE MODEL

SUBMITTED IN THE PARTIAL FULFILLMENT REQUIREMENT FOR THE AWARD OF DEGREE OF

# Bachelor of Technology

(Computer

Science

and Engineering)

SUBMITTED BY

**NISHCHAY VASHISHTHA (230737)**

**DIVIDH VIJAY (230730)**

**DAKSH NEHRA (230723)**

UNDER THE SUPERVISION OF

**DR. SHILPA MAHAJAN**

School of Engineering and Technology

Logo, company name

Description automatically generated

# BML MUNJAL UNIVERSITY Gurugram, Haryana - 122413

# Dec 2023

**CANDIDATE’S DECLARATION**

We hereby certify that We have worked on project entitled, ”**Telecom Customer Churn Predictive Model**”, in partial fulfillment of requirements for the award of Degree of **Bachelor of Technology** in **Computer Science Engineering** department at **BML Munjal University**, is an authentic record of our own work carried out during a period from August, 2024 to December, 2024 under the supervision of DR. SHILPA MAHAJAN.

## Nishchay Vashishtha (230737)

## Dividh Vijay (230730)

## Daksh Nehra (230723)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

## DR. SHILPA MAHAJAN

Assistant Professor

BML Munjal University

Customer churn is a critical challenge for telecom companies, leading to significant financial losses, operational disruptions, and missed growth opportunities. As competition intensifies and service offerings become increasingly similar, retaining existing customers has become a strategic priority. Addressing this issue requires accurate and reliable predictive models that can identify customers at risk of leaving, enabling proactive retention strategies tailored to their needs.

Decision tree-based models, including Random Forest, have emerged as powerful tools in the realm of churn prediction. These models are favoured for their balance of accuracy, interpretability, and ease of implementation. Decision trees provide a straightforward, visual representation of the decision-making process, allowing stakeholders to understand the factors contributing to customer churn. Random Forest, an ensemble of decision trees, enhances predictive performance by reducing overfitting and capturing complex relationships within the data.

The application of decision tree-based models highlights the transformative role of predictive analytics in converting raw customer data—such as usage patterns, demographics, and feedback—into actionable business insights. These models enable telecom companies to proactively identify customers at risk of churning and implement timely interventions, such as personalized offers, loyalty programs, or enhanced customer service.

Interpretability is a key advantage of these models, empowering decision-makers to understand the drivers of churn and fostering trust in the predictions. By identifying key churn indicators, such as service quality, pricing issues, or usage trends, telecom companies can design targeted retention strategies that optimize resource allocation and improve customer satisfaction.

In an era of fierce competition and evolving customer expectations, churn prediction serves as a cornerstone for data-driven customer relationship management. Decision trees and Random Forest models provide telecom companies with the tools to strengthen their market position, reduce costs associated with customer acquisition, and deliver superior customer experiences. These predictive capabilities ensure sustained growth and profitability while maintaining a competitive edge in a dynamic industry landscape.

We are highly grateful to **DR. SHILPA MAHAJAN, ASSISTANT PROFESSOR** BML Munjal University, Gurugram, for providing supervision to carry out the seminar/case study from August-December 2024.

**DR. SHILPA MAHAJAN** has provided great help in carrying out my work and is acknowl- edged with reverential thanks. Without the wise counsel and able guidance, it would have been impossible to complete the training in this manner.

We would like to express thanks profusely to thank **DR. SHILPA MAHAJAN**, for stimulating us time to time. We would also like to thank entire team of BML Munjal University. We would also thank our friends who devoted their valuable time and helped us in all possible ways towards successful completion.

## NISHCHAY VASHISHTHA (230737)

## DIVIDH VIJAY (230730)

## DAKSH NEHRA(230723)

**Figu re No. Figure Description Page No.**

## Table No. Table Description Page No.

**Abbreviation Full Form**

**SVM** Support Vector Machine

**abbreviation** Full Form

**abbreviation** Full Form

## Contents Page No.

[***Candidate’s Declaration***](#_bookmark0)**i**

[Abstract](#_bookmark1) ii

[Acknowledgement](#_bookmark2) iii

[List of Figures](#_bookmark3) iv

[List of Tables](#_bookmark4) v

[List of Abbreviations](#_bookmark5) vi

1. [Introduction to Organisation](#_bookmark6) 1
2. [Introduction to Project](#_bookmark7) 2

[2.1 Overview](#_bookmark8) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[2.2 Existing System](#_bookmark9) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[2.3 User Requirement Analysis](#_bookmark10) . . . . . . . . . . . . . . . . . . . . . . . . . . 2

[2.4 Feasibility Study](#_bookmark11) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

1. [Literature Review](#_bookmark12) 3

[3.1 Comparison](#_bookmark13) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

* 1. [Objectives of Project (Must be clearly, precisely defined and Implementa-](#_bookmark14)

[tion must be done.)](#_bookmark14) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1. [Exploratory Data Analysis](#_bookmark15) 5

[4.1 Dataset](#_bookmark16) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

* 1. [Exploratory Data Analysis and Visualisations](#_bookmark17) . . . . . . . . . . . . . . . 5

[4.3 Related Sections](#_bookmark18) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

1. [Methodology](#_bookmark19) 6
   1. [Introduction to Languages (Front End and Back End)](#_bookmark20) . . . . . . . . . . . 6
   2. [Any other Supporting Languages/ packages](#_bookmark21) . . . . . . . . . . . . . . . . 6

[5.3 User characteristics](#_bookmark22) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.4 Constraints](#_bookmark23) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.5 Use Case Model/Flow Chart/DFDS](#_bookmark24) . . . . . . . . . . . . . . . . . . . . . 6

[5.6 Database design](#_bookmark25) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.7 Table Structure](#_bookmark26) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.8 ER Diagrams](#_bookmark27) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.9 Assumptions and Dependencies](#_bookmark28) . . . . . . . . . . . . . . . . . . . . . . . 6

[5.10 ML algorithm discussion](#_bookmark29) . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

[5.11 Implementation of Algorithm with Screen Shots/ Figures ( Each Figure](#_bookmark30)

[must be numbered and Description of Figure must be provided)](#_bookmark30) . . . . . 6

1. [Results](#_bookmark31) 7
2. [Conclusion and Future Scope](#_bookmark32) 8

[7.1 Conclusion](#_bookmark33) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

[7.2 Future Scope](#_bookmark34) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

# Introduction to Organisation

In today’s highly competitive telecommunications industry, customer retention is a key factor in sustaining profitability and growth. Telecom companies operate in a dynamic environment where customer expectations, technological advancements, and market competition are constantly evolving. To address these challenges, organizations within the telecom sector are increasingly turning to data-driven solutions to predict and prevent customer churn—one of the most critical issues faced by the industry.

This project, **"Telecom Customer Churn Predictive Model,"** is conducted as part of a university-level initiative aimed at understanding how organizations leverage advanced analytics and machine learning to identify potential churners and implement proactive retention strategies.

# Introduction to Project

## Overview

In today’s competitive telecommunications industry, customer retention is essential for sustaining profitability and market share. Customer churn—the loss of existing subscribers—can have a significant negative impact on revenue, as acquiring new customers is often more expensive than retaining current ones. Predictive analytics, driven by large volumes of customer data, has emerged as a crucial tool for identifying customers likely to churn and enabling telecom companies to take proactive retention measures. This literature survey reviews recent advancements in churn prediction, providing an overview of the key methodologies, performance metrics, and challenges identified in the field. The focus is particularly on the effectiveness of ensemble machine learning techniques and the growing role of explainable AI in making predictions more interpretable for business decision-makers.

## Existing System

## User Requirement Analysis

## Feasibility Study

# Literature Review

## Add the summaries of research work you referred for your work. Either discuss the research in chronological order or by topic. Do cite all the research articles at the end of report in Bibliography section. For example the below paragraphs discuss the research in a flow of the topic

It has been very crucial task for researchers to identify the current and historical trends. Further, extraction from this information to focus on innovative methods and concepts also requires critical thinking. Keeping in view the volume of unstructured data, it is a tedious task to extract the desired information. But with technological advances, various methods have been developed for extracting the relevant information from voluminous data. The data from various bibliographic sources viz. research articles, book chapters, patents, and technical reports, can be summarized by identification of topics [[1](#_bookmark35), [2](#_bookmark36)].

Topic modeling is an automated process which is based on algorithmic-based analysis [[3](#_bookmark37), [4](#_bookmark38)]. From a corpus, this method is used to identify the patterns. Further, semantic meaning is also added to the corpus’s vocabulary. Topic modeling can be applied by two methods viz. topic analysis and clustering but topic analysis is considered better choice for research trends’ identification [[5](#_bookmark39)]. The main difference between the two methods is the assignment of a document to the topic or cluster. A document can be assigned to multiple topics in topic analysis, whereas it joins exactly one cluster in clustering.

## Comparison

This section should have table where you need to compare the results from different prominent research of area of interest.

## Objectives of Project (Must be clearly, precisely defined and Implemen- tation must be done.)

This section should elaborate the objectives based on the gaps from the previous re- searchers.

# Exploratory Data Analysis

## Dataset

Discussion on the dataset, scraping methods, pipelines etc.

## Exploratory Data Analysis and Visualisations

Details analysis of the data and you could include feature engineering into this section

## Related Sections

# Methodology

This section discusses the methodology to develop the ML model. It may include the following but not limited to:

## Introduction to Languages (Front End and Back End)

## Any other Supporting Languages/ packages

## User characteristics

## Constraints

## Use Case Model/Flow Chart/DFDS

## Database design

## Table Structure

## ER Diagrams

## Assumptions and Dependencies

## ML algorithm discussion

* 1. **Implementation of Algorithm with Screen Shots/ Figures ( Each Figure must be numbered and Description of Figure must be provided)**

# Results

**Conclusion and Future Scope**

## Conclusion

## Future Scope

1. Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1). <https://doi.org/10.1186/s40537-019-0191-6>
2. Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. *Algorithms*, *17*(6).

<https://doi.org/10.3390/a17060231>

1. Kashyap, S. R. (2023). *Customer Churn Prediction Using Ensemble Techniques on Telco Dataset* (Vol. 10). JETIR.

[www.jetir.orgd376](http://www.jetir.orgd376)

1. Sook Ling, L., Mustafa, N., & Abdul Razak, S. F. (2021). Customer churn prediction for telecommunication industry: A Malaysian Case Study. *F1000Research*, *10*. <https://doi.org/10.12688/f1000research.73597.1>
2. Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, *7*, 60134–60149. https://doi.org/10.1109/ACCESS.2019.2914999
3. Umayaparvathi, V., & Iyakutti, K. (2016). A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. In *International Research Journal of Engineering and Technology*. http://www.fuqua.duke.edu/centers/ccrm/index.html