Customer Churn Prediction Project

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN

BIG DATA ANALYTICS

Submitted by:

22BDA70157 – Diwakar Sherawat
22BDA70161 – Rony Biju
22BDA70132 – Abhay Manhas
22BDA70205 – Shruti Gupta
22BDA70198 – Sneha Narwal

Under the Supervision of:

Dr. Monica Luthra

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

August, 2025

Table of Contents

ABSTRACT:

This project focuses on leveraging advanced data science techniques to predict customer churn in the telecommunications sector. Customer churn refers to the phenomenon where customers discontinue their service, leading to revenue loss for companies. The project aims to build predictive models capable of identifying customers at high risk of churn, thereby enabling organizations to take proactive steps to retain them. The approach includes data acquisition, cleaning, exploratory data analysis (EDA), segmentation, and advanced churn modeling. Tools such as Python, Pandas, Matplotlib, and Seaborn are used for data analysis and visualization. The system provides actionable insights into key churn drivers like tenure, payment methods, and contract type. The expected outcome is a framework that enhances customer retention strategies and minimizes revenue leakage, ultimately contributing to long-term business sustainability.

1. INTRODUCTION

1.1 Problem Definition

Customer retention is a significant challenge for telecom companies, as losing existing customers is costlier than acquiring new ones. Traditional churn detection approaches rely on historical trends and lack predictive capabilities, making it difficult to intervene in time. There is a need for a robust, data-driven system that can predict churn with accuracy and provide insights for timely customer engagement.

1.2 Problem Overview

This project develops a predictive analytics system for customer churn detection using advanced data science techniques. It involves dataset preparation, exploratory analysis, customer segmentation, and predictive modeling to understand churn behavior. Insights are derived from demographics, customer lifecycle, and service usage patterns. The final output includes actionable retention strategies for at-risk customers.

1.3 Hardware Specification

- Processor: Intel i5 or above
- RAM: Minimum 8 GB
- Hard Disk: Minimum 250 GB
- Display: 13” or higher with 1366x768 resolution
- Internet Connectivity (for dataset access and package installations)

1.4 Software Specification

- Operating System: Windows 10 or above

- Programming Language: Python
- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- IDE: Jupyter Notebook / PyCharm
- Database: CSV / Relational Databases
- Version Control: Git & GitHub

2. LITERATURE SURVEY

2.1 Existing System

Existing churn prediction systems rely heavily on descriptive statistics and manual interventions. Telecom providers use CRM systems and dashboards that show churn rates but fail to provide predictive accuracy or individualized retention recommendations.

2.2 Proposed System

The proposed system uses advanced analytics and machine learning to predict churn risk. By analyzing behavioral data, demographics, and service attributes, it flags high-risk customers and recommends tailored retention strategies. Unlike existing systems, this model focuses on proactive prevention rather than reactive analysis.

2.3 Literature Review Summary **(Minimum 7 articles should refer)**

Year and Citation	Article/ Author	Tools/ Software	Technique	Source	Evaluation Parameter
2019	Customer Churn in Telecom - Patel et al	Python, Scikit-learn	Logistic Regression	IEEE	Accuracy
2020	Predictive Analytics for Churn - Kumar	R, Decision Trees	Classification	IJCA	Precision/Recall
2021	Telecom Churn Modeling - Li & Wang	Python, XGBoost	Ensemble Learning	Springer	AUC Score
2021	Churn Analysis using ML - Verma	Python, Random Forest	Feature Engineering	IJERT	Accuracy
2022	Big Data in Churn Prediction - Singh	Hadoop, Spark	Distributed ML	Elsevier	Scalability
2023	AI for Retention Strategies - Chen	Python, Deep Learning	Neural Networks	ACM	Automation
2024	Data-Driven Customer Retention - Sharma	Tableau, SQL	Visualization	IRJET	Interpretability

3. PROBLEM FORMULATION

Telecom companies face challenges in identifying and retaining customers at risk of churn. Current approaches fail to deliver accurate, actionable predictions. The formulated problem is to design a predictive analytics framework that:

- Detects high-risk customers.
- Identifies factors contributing to churn.
- Provides insights for proactive customer engagement strategies.

4. OBJECTIVES

- To collect and preprocess telecom customer data for churn analysis.
- To perform exploratory data analysis and identify churn trends.
- To segment customers based on demographics, tenure, and usage.
- To develop predictive models using machine learning techniques.
- To evaluate the models on accuracy, precision, and recall.
- To derive actionable retention strategies for at-risk customers.

5. METHODOLOGY

- **Requirement Analysis:** Identify key variables influencing churn (tenure, contract, payment type, demographics).

- **Data Acquisition & Cleaning:** Handle missing values, duplicates, and inconsistent records.

- **Exploratory Data Analysis:** Visualize churn trends using histograms, bar charts, and correlation matrices.

- **Segmentation:** Group customers based on tenure and monthly charges to detect vulnerable segments.

- **Model Development:** Build and train predictive models (Logistic Regression, Random Forest, XGBoost).

- **Evaluation:** Use metrics such as Accuracy, Precision, Recall, and AUC to evaluate performance.

- **Retention Strategies:** Propose actionable interventions like targeted offers, loyalty programs, or personalized plans.

6. EXPERIMENTAL SETUP

- **Platform:** Jupyter Notebook / Anaconda Distribution
- **Dataset:** Telecom Customer Churn Dataset (CSV format)
- **Tools:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Version Control:** GitHub for code management
- **Testing Tools:** Model evaluation using train-test split and cross-validation.

7. CONCLUSION

The proposed project offers a comprehensive predictive framework for customer churn analysis in telecom. By leveraging machine learning, it accurately identifies high-risk customers and provides insights into churn drivers. The system bridges the gap between raw data and actionable business strategies, ensuring improved retention and profitability. Future enhancements could include integration with real-time dashboards, deep learning models for higher accuracy, and automated alert systems for immediate interventions.

8. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

Chapter 1: Introduction
Problem statement, project overview, and specifications

Chapter 2: Literature Review
Existing vs proposed systems, related work

Chapter 3: Problem Formulation
Research gap and problem design

Chapter 4: Research Objectives
Defined aims and goals

Chapter 5: Methodologies
Techniques, models, and workflow

Chapter 6: Experimental Setup
Implementation environment and dataset

REFERENCES

[1] Patel, R., et al. (2019). “Customer Churn in Telecom using ML.” IEEE.
[2] Kumar, A. (2020). “Predictive Analytics for Churn.” IJCA.
[3] Li, Y., & Wang, Z. (2021). “Telecom Churn Modeling with XGBoost.” Springer.
[4] Verma, S. (2021). “Churn Analysis using Random Forest.” IJERT.
[5] Singh, P. (2022). “Big Data Applications in Churn Prediction.” Elsevier.
[6] Chen, D. (2023). “AI for Retention Strategies.” ACM.
[7] Sharma, K. (2024). “Data-Driven Customer Retention.” IRJET.