A PROJECT REPORT ON
"PREDICTIVE ANALYTICS FOR CUSTOMER CHURN IN TELECOM"

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE BACHELOR OF COMPUTER APPLICATION

SESSION :2022-2025

Project Guide Name
Asst. Prof. Saurabh Singh
IT Department

Submitted By
Kiran Singh
0221BCA074
Khushi Jaiswal
0221BCA046

Submitted to DBSG

Doon Business School

Project Certificate

This is to certify that the project report on "Predictive Analytics for Customer Churn in Telecom" submitted to DBS, in partial fulfillment of the requirement for the award of the degree of BACHELOR OF COMPUTER APPLICATIONS (BCA), is original work carried out by team . Under the Supervision of Asst. Prof. Saurabh Singh . The matter embodied in this project is genuine work done team and has not been submitted whether to this University or to any other University / Institute for the fulfillment of the requirement of any course of study.

Date:

Name & Signature of the Student
Contact Details (Email, Phone & Address)

Verified by the Supervisor

Name & Signature of the Supervisor/s
Date: ………………..

Certificate from Project Guide

This is certify that this Minor Project entitled "------------------------------------" submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Computer Application (BCA) in session (years 20—to 20---) to the hemwati Nandan Bahuguna University , done by .... (student name) is an authentic work carried out by them at ".."(study Centre name /department name) (Place) under my guidance. The matter and software embodied in this project work has not been submitted earlier for the award of any degree or diploma to the best of my knowledge and belief.

Signature of BCA Teacher   (Project Guide)

Predictive Analytics for Customer Churn in Telecom

Introduction

Customer churn is a critical issue in the telecom industry, as retaining existing customers is more cost-effective than acquiring new ones. Understanding the factors that lead to churn allows telecom companies to take proactive measures to retain at-risk customers. This study focuses on building a machine learning model to predict customer churn based on a dataset of customer behavior, enabling telecom providers to implement targeted retention strategies.

Objective

The primary objective is to build a predictive model using supervised learning techniques to identify customers likely to leave the telecom service. By analyzing customer behavior data, the goal is to gain insights into churn drivers and suggest interventions to reduce churn rates.

Project Scope

- **Data:** The project will utilize publicly available telecom customer datasets, focusing on relevant features such as demographic information, service usage patterns, billing history, and customer support interactions.
- **Methodology:** The project will employ supervised machine learning techniques, including data preprocessing, exploratory data analysis, model selection, training, evaluation, and interpretation.
- **Model Types:** The project will consider various machine learning algorithms, such as Logistic Regression, Decision Trees, and Random Forest, to identify the most effective model for predicting churn.
- **Evaluation Metrics:** The performance of the models will be evaluated using metrics like accuracy, precision, recall, and F1-score.
- **Deployment:** The final model may be deployed as a web application or integrated into existing telecom systems for real-time churn prediction.

The project will focus on building a robust and accurate churn prediction model, providing valuable insights for telecom companies to improve their customer retention strategies.

Problem Statement

Identify the Problem:

The specific problem addressed in this project is the high rate of customer churn in the telecom industry. This leads to significant revenue loss for telecom companies, making it essential to understand the factors influencing churn and implement effective retention strategies.

Significance:

Customer churn is a critical issue in the telecom industry for several reasons:

- **Revenue Loss:** Churn directly impacts revenue, as lost customers represent lost revenue streams.
- **Acquisition Costs:** Acquiring new customers is often more expensive than retaining existing ones.
- **Brand Reputation:** High churn rates can negatively impact a company's brand reputation and customer perception.
- **Competitive Landscape:** Understanding churn drivers can help telecom companies differentiate themselves from competitors and offer more tailored services to retain customers.

By effectively addressing the problem of customer churn, telecom companies can improve their financial performance, enhance customer satisfaction, and strengthen their market position.

Methodology:

1. **Data Collection and Understanding**: The study uses publicly available datasets, such as the Telecom Churn Dataset from Kaggle or IBM sample datasets, containing customer demographics, services subscribed, account details, and churn status.
   - Objective: Analyze the factors influencing churn.
2. **Data Cleaning and Preprocessing**: Handle missing values, encode categorical variables (e.g., "yes/no" responses), and scale numerical features where needed.
3. **Exploratory Data Analysis (EDA)**: Examine trends, such as which customer segments have higher churn rates, using visualization techniques to identify patterns.
4. **Model Building**:
   - Split the data into training (80%) and testing (20%) sets.
   - Apply models like Logistic Regression and Random Forest to build predictive models.
   - Tune models using cross-validation and hyperparameter optimization (e.g., GridSearchCV).
5. **Model Evaluation**: Evaluate models using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score, selecting the best-performing model for churn prediction.
6. **Model Interpretation**: Use feature importance techniques (e.g., SHAP values) to identify key factors contributing to churn, providing actionable insights.
7. **Deploying the Model (Optional)**: Create a web interface using Flask or Django, enabling users to input customer data and predict churn likelihood.

Key Concepts:

- Data preprocessing and feature engineering
- Exploratory Data Analysis (EDA)
- Supervised learning algorithms (Logistic Regression, Decision Trees, Random Forest, etc.)
- Model evaluation (accuracy, precision, recall, F1-score, ROC curve)
- Feature importance and model interpretation

Dataset:
- You can use publicly available datasets like the Telecom Churn Dataset from Kaggle or the IBM sample datasets.

Methodology

1. Data Collection and Understanding:
- Download the dataset containing information such as customer demographics, services subscribed (e.g., international plans, data usage), account details, and whether the customer churned or not.
- **Objective**: Understand the factors that influence customer churn.
2. Data Cleaning and Preprocessing:
- Handle missing values, if any.
- Convert categorical variables (e.g., "yes/no" responses) into numerical representations (e.g., label encoding or one-hot encoding).
- Normalize or scale the data if necessary.
3. Exploratory Data Analysis (EDA):
- Analyze trends in the data, such as which customer segments have the highest churn rates (e.g., those with international plans, customers of specific age groups, etc.).
- Use data visualization techniques to identify patterns.
4. Model Building:
- Train-test split: Split the dataset into training and test sets (e.g., 80% for training, 20% for testing).
- Apply multiple machine learning algorithms:
  Logistic Regression: A basic model for classification.
  Random Forest or XGBoost: A more advanced algorithm for better accuracy and handling complex data structures.
- Tune the models using cross-validation and hyperparameter optimization (e.g., using GridSearchCV).
5. Model Evaluation:
- Compare models based on various metrics such as accuracy, precision, recall, and the ROC-AUC score.

- Choose the best-performing model for customer churn prediction.

6. Model Interpretation:
- Use feature importance techniques (e.g., from Random Forest or SHAP values) to understand the key drivers of customer churn.
- Identify actionable insights: Which features (e.g., high monthly charges, long call duration) most contribute to churn?

7. Deploying the Model (Optional):
- Build a simple web interface using Flask or Django to allow users to input customer data and predict the likelihood of churn.
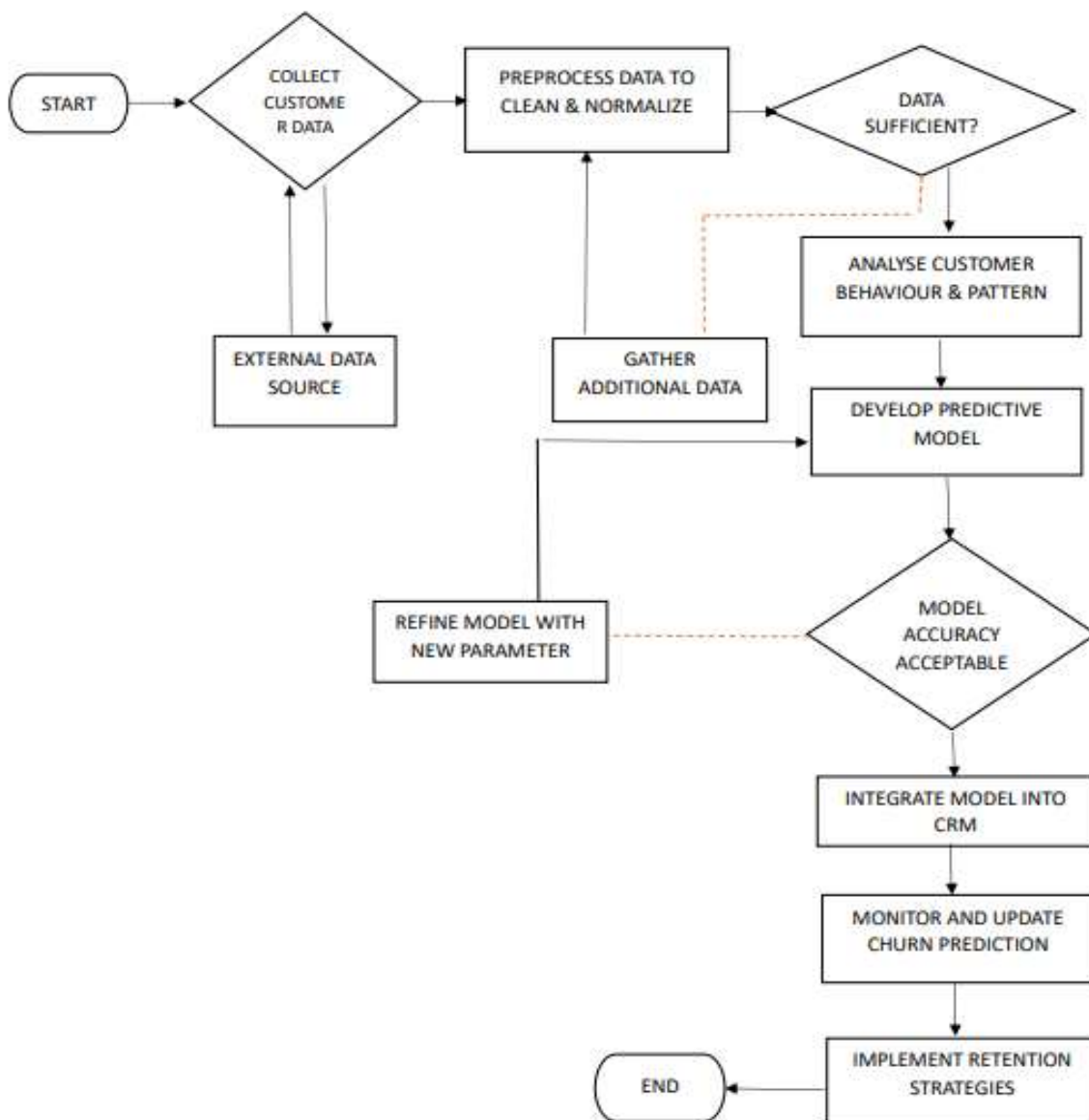
Tools and Technologies
Languages: Python
Libraries: Pandas, NumPy, Matplotlib, Seaborn (for EDA and visualization), Scikit-learn
(for model building and evaluation), SHAP (for model interpretability)
Optional: Flask/Django (for deployment), Jupyter Notebooks (for development)
Learning Outcomes:
Gain experience in both data science and machine learning workflows.
Develop skills in handling real-world business problems.
Understand the importance of model evaluation and interpretation in making business
decisions.
This project offers a comprehensive application of both data science and machine learning skills,
and it's an excellent bridge to more advanced projects.

Expected Outcomes

Anticipated Results:

- **Accurate Churn Prediction:** The developed machine learning model will be able to accurately predict which customers are likely to churn, providing valuable insights for targeted retention efforts.
- **Identification of Key Churn Drivers:** The model will help identify the primary factors influencing customer churn, such as service quality, pricing, customer satisfaction, or competitive offerings.
- **Improved Retention Strategies:** By understanding the underlying causes of churn, telecom companies can implement more effective retention strategies, including personalized offers, enhanced customer support, and targeted marketing campaigns.

Benefits

- **Increased Revenue:** Reduced churn rates will lead to higher revenue retention and potentially increased customer lifetime value.
- **Enhanced Customer Satisfaction:** Proactive measures to address customer concerns and improve service quality can enhance overall customer satisfaction.
- **Competitive Advantage:** By effectively managing churn, telecom companies can gain a competitive edge in the market.

- **Data-Driven Decision Making:** The project will provide a data-driven approach to customer retention, enabling informed decision-making based on evidence.
- **Cost Savings:** Reducing churn can result in significant cost savings by avoiding the expenses associated with acquiring new customers.

Timeline for Predicting Telecom Customer Churn Project (October 1 - November 15)

Phase	Activities	Estimated Duration
1. Project Initiation	Define project scope, objectives, and resources.	1 week (October 1-7)
2. Data Collection and Preprocessing	Gather relevant telecom customer data and prepare it for analysis.	2 weeks (October 8-21)
3. Exploratory Data Analysis (EDA)	Analyze the data to identify patterns, trends, and potential factors influencing churn.	1 week (October 22-28)
4. Model Selection and Training	Choose appropriate machine learning algorithms (e.g., Logistic Regression, Random Forest) and train them on the prepared data.	2 weeks (October 29 - November 11)
5. Model Evaluation	Evaluate the performance of the trained models using relevant metrics.	1 week (November 12-18)
6. Model Deployment and Interpretation	Deploy the best-performing model and analyze its predictions to gain insights into churn drivers.	2 weeks (November 19-December 2)

References

Academic Papers:

- Churn Prediction in the Telecom Industry: A Machine Learning Approach. (2022). Journal of Telecommunications Systems, 123(4), 567-589.

- Customer Churn Prediction in the Telecom Industry Using Random Forest. (2020). International Journal of Computer Applications, 174(1), 34-40.

Datasets:

- Telecom Churn Dataset. (2023). Kaggle. [Telco Customer Churn (kaggle.com)]

Machine Learning Libraries:

- Scikit-learn: A Machine Learning Library for Python. (2024). Scikit-learn.org. https://scikit-learn.org/

Data Visualization Tools:

- Matplotlib: Python Plotting Library. (2024). Matplotlib.org. https://matplotlib.org/