Project Synopsis

for

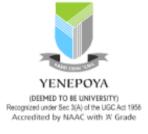# customer churn prediction

Hiza Rafi
22BDACC101
Bachelors of computer application
Bigdata analytics, cloud computing, cyber security with IBM
Yenepoya Institute of Arts Science Commerce and Management
Mangalore
05-03-2025

Under the guidance of

Ms. Bhoomika
Faculty, Computer Science Department
Yenepoya Institute of Arts Science Commerce and Management
Mangalore

Submitted to

YENEPOYA
(DEEMED TO BE UNIVERSITY)
Recognized under Sec 3(A) of the UGC Act 1956
Accredited by NAAC with 'A' Grade

YENEPOYA INSTITUTE OF ARTS, SCIENCE, COMMERCE AND MANAGEMENT
BALMATTA, MANGALORE
YENEPOYA (DEEMED TO BE UNIVERSITY)

I.   TITLE OF THE PROJECT

Customer Churn Prediction

II.   STATEMENT OF THE PROBLEM

Customer churn is a significant challenge for businesses, particularly in subscription-based services. The problem involves identifying customers who are likely to discontinue using a company's service. Predicting churn early allows businesses to take proactive measures to retain customers and reduce revenue loss. This project aims to develop a machine learning model that can analyze customer data and predict whether a customer is likely to churn or not.

III.   WHY THIS PARTICULAR TOPIC CHOSEN?

Customer retention is a crucial business concern, and this project provides a real-world application of machine learning. The topic is chosen because:

- It is **easy to implement** with beginner-friendly ML models.

- It has **high practical value** in industries such as telecom, banking, and e-commerce.

- It helps in understanding **data analytics, visualization, and model training**.

- It is a **high-scoring project** with clear objectives and scope.

IV.   OBJECTIVE AND SCOPE

Objective:
- To build a machine learning model that predicts customer churn.
- To identify the key factors affecting churn.
- To provide insights using data visualization techniques.
- To develop a simple web application for predictions.

Scope:
- The project will focus on customer data collected from telecom industries.
- The model will be trained on structured data containing customer details, usage patterns, and payment history.

- The project will not include external factors like customer emotions or social media sentiment analysis.

V. METHODOLOGY

1. **Data Collection:** Use publicly available datasets such as the Telco Customer Churn dataset from Kaggle.
2. **Data Preprocessing:** Handle missing values, convert categorical variables, and normalize numerical data.
3. **Exploratory Data Analysis (EDA):** Use visualizations to understand trends and relationships in the data.
4. **Model Development:** Train models such as **Logistic Regression, Decision Tree, and Random Forest** to classify churn.
5. **Model Evaluation:** Use accuracy, precision, recall, and F1-score to evaluate performance.
6. **Web Application:** Use Flask to create a simple user interface where customer details can be input to get churn predictions.

VI. PROCESS DESCRIPTION

- **Module 1**: **Data Collection & Cleaning** – Load dataset, clean missing values, and preprocess categorical variables.
- **Module 2: Data Analysis & Visualization** – Generate insights using correlation heatmaps, bar charts, and histograms.
- **Module 3: Model Training** – Train multiple machine learning models and choose the best one.
- **Module 4: Model Testing & Evaluation** – Measure accuracy and optimize performance.
- **Module 5: Web Application** – Deploy the model using Flask for user interaction.

VII. RESOURCES AND LIMITATIONS

Resources Required:
- **Hardware:** Laptop with at least 8GB RAM and a good processor.
- **Software:** Python, Jupyter Notebook, VS Code, Flask (optional).
- **Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn.

Limitations:

- The dataset used may not generalize to all industries.
- The project focuses only on structured tabular data and does not consider external market trends.
- The accuracy of predictions depends on the quality of available data.

VIII.   TESTING TECHNOLOGIES USED

- **Unit Testing:** Checking data preprocessing and feature engineering outputs.
- **Integration Testing:** Ensuring that different components (data preprocessing, model training, and prediction) work together correctly.
- **Model Evaluation:** Using confusion matrix, accuracy, precision, recall, and F1-score to verify performance.
- **User Testing (If Web App is Developed):** Checking functionality for different customer inputs and ensuring a smooth user experience.

IX.   CONCLUSION

This project provides a **practical and easy-to-implement** machine learning solution for customer churn prediction. The model will help businesses take **proactive steps** to reduce customer loss, making it a valuable project with **real-world applications**. Additionally, the project introduces students to essential concepts in **data science, analytics, and machine learning**, making it an **ideal beginner-level project** with good scoring potential.