**PHASE-3**

**Project Title:**

**Predicting customer churn using machine learning**

**Student Name:** R.KEERTHANA

**Register Number:** 511523205024

**Institution:** P. T. Lee Chengalvaraya Naicker College of Engineering and Technology.

**Department:** B. Tech Information Technology.

**Year:** Second year.

**Date of Submission**: 14.05.2025.

**Githup link:** <https://github.com/keerthi-0226/EbplDs_predicting-customer-chum-using-machine-learning-uncover-hidden-patterns.git>

**Dataset link:** [**https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset**](https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset)

**1.Problem Statement:**

* In highly competitive industries like telecommunications, customer retension is a major challenge.companies lose significant revenue when customers discontinue their services (churn). Traditional methods of identifying at-risk customers are often reactive and inefficient.
* The goal of this project is to develop a machine learning-based solution that can accurately predict whether a customer is likely to churn based on their historical and behavioral data. By leveraging customer demographics, account information, service usage, and billing details, we aim to build a predictive model that enables the company to take proactive retention measures.
* This project will:
* Analyze the customer dataset to identify churn patterns.
* Preprocess and prepare the data for machine learning.
* Train, evaluate, and optimize a classification model.
* Provide actionable insights based on model interpretation.
* **Key Objective**: Build a model that predicts customer churn (Yes/No) and identify key factors influencing the churn decision.

1. **Abstract:**

* Customer churn is a critical concern for businesses in competitive markets, especially in the telecommunications sector. Retaining existing customers is often more cost-effective than acquiring new ones. Therefore, the ability to predict churn can provide significant strategic advantages. This project focuses on building a machine learning model to predict customer churn using the Telco Customer Churn dataset.

1. **System Requirements:**
2. **Hardware Requirements**

|  |  |  |
| --- | --- | --- |
| **Component** | **Minimum Requirement** | **Recommended Requirement** |
| **Processor(CPU)** | **Dual-core 2.0 GHz** | **Quad-core 2.5+GHz** |
| **RAM** | **4 GB** | **8 GB or more** |
| **Storage** | **2 GB of free space** | **SSD with 10+ GB of free space** |
| **GPU(optional)** | **Not required (unless using deep learning)** | **NVIDIA GPU for acceleration** |

**2. Software Requirements**

|  |  |
| --- | --- |
| Component | Version/Notes |
| Operation System | Windows 10/11.macOS. or any Linux distribution |
| Python | 3.7 or higher |
| Jupyter Notebook | (Optional) For interactive development |
| IDE | VS code / Pycharm / Jupyter / Google Colab |

**4. Objectives:**

* The primary objective of this project is to develop a machine learning model that accurately predicts whether a customer is likely to churn (i.e., stop using a service) based on historical customer data. By analyzing customer demographics, service usage, account information, and billing behavior, the model aims to:
* Identify patterns and key indicators of customer churn.
* Classify customers into churn and non-churn categories.
* Help businesses proactively retain at-risk customers.
* Provide actionable insights for improving customer satisfaction and loyalty.

1. **Flowchart of the Project Workflow:**

* (1). Data Collection-Gather customer-related data (e.g., demographics, usage history, billing information)(2).Data Preprocessing-Clean missing and inconsistent data,Encode categorical variables,Scale numerical values for better model performance(3). Exploratory Data Analysis (EDA)-Visualize distributions and trends,Understand relationships between variables,Identify outliers or anomalies(4).Feature Selection-Choose important predictors using correlation, feature importance, or statistical tests(5).Split Dataset-Divide the data into training and testing subsets (e.g., 80/20 split)(6).Model Selection & Training-Choose appropriate algorithms (e.g., Random Forest, Logistic Regression),Train the model using training data(7).Model Evaluation-Evaluate model using accuracy, confusion matrix, precision, recall, and F1-score(8).Churn Prediction on New Data-Use the trained model to predict churn likelihood for new/unseen customer data(9). Model Deployment-Save the trained model,Integrate it into business systems or web applications for real-time use(10).End-The process concludes; model is ready for production use.

**Flow Chart:**

|  |
| --- |
| **Start** |

|  |
| --- |
| **Data Collection** |

|  |
| --- |
| **Data Preprocessor** |

|  |
| --- |
| **Exploratory Data Analytics** |

|  |
| --- |
| **Feature Selection** |

|  |
| --- |
| **Train/Test Slipt** |

|  |
| --- |
| **Model Selection / Training** |

|  |
| --- |
| **Model Evaluation** |

|  |
| --- |
| **Make Prediction** |

|  |
| --- |
| **Save & Deploy Model** |

|  |
| --- |
| End |

**6. Dataset Description**

* **Demographic Features:**
* **Customer ID: Unique identifier for each customer**
* **Age: Customer's age**
* **Gender: Customer's gender**
* **Location: Customer's geographic location**
* **Behavioral Features:**
* **Usage patterns: Frequency, duration, or volume of product/service usage**
* **Purchase history: Frequency, recency, and monetary value of purchases**
* **Customer support interactions: Number and nature of support requests**
* **Transactional Features:**
* **Account status: Active, inactive, or cancelled**
* **Subscription plan: Type of plan or service subscribed to**
* **Payment history: Timeliness and method of payments**
* **Churn Indicator:**
* **Churn status: Binary indicator (0/1) or categorical label (e.g., "active," "churned") indicating whether the customer has churned**
* **Additional Features:**
* **Customer feedback: Ratings, reviews, or sentiment analysis of customer feedback**
* **Marketing interactions: Number and type of marketing communications**
* **The dataset would typically be structured as a table with each row representing a customer and each column representing a feature. The target variable would be the churn status, and the goal would be to predict this variable based on the other features.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Customer ID** | **Age** | **Gender** | **Usage** | **Purchases** | **Churn Status** |
| **1** | **35** | **Male** | **100** | **5** | **0** |
| **2** | **42** | **Female** | **50** | **2** | **1** |
| **3** | **28** | **Male** | **200** | **10** | **0** |

* **Sample Dataset:**

**7.Data processing:**

* **Data Cleaning:**  Handle missing value
* **Data Transormation:** Encode categorical variables
* **Feature Engineering:** Create new features from existing ones
* **Data Split:** Split data into training andbtesing sets

**8. Exploratory Data Analysis (EDA)**

* **Univariate Analysis:**
* Demograhic variables
* Behavioral variables
* Transactional variables
* **Bivariate Analysis:**
* Correlation between variables
* Relationalship between variables and churn status

**9. Feature Engineering**

* **Feature Creation:**
* Creation new features from existing ones
* Extract information from existing features
* **Feature Selection:**
* Identify the most relevant features that contribute to churn
* Remove redundant or irrelevant features

**10. Model Building**

* **Models Tried**:

Logistic Regression

Random Forest

Decision tree

Gradient Boosting

Support Vector Machines

* **Why These Models**:

**Logistic Regression**: A popular algorithm for binary classification problems like customer churn prediction.

**Random Forest**: An ensemble method that combines multiple decision trees to improve prediction accuracy.

**Decision Trees:** Can handle complex interactions between features and are easy to interpet.

**11. Model Evaluation**

* Evaluating the models performance using the evaluation merics.
* **MetricValue:**

Accuracy=0.85

Precision=0.80

Recall=0.90

F1-Score=0.85

ROC AUC Score=0.92

**12. Deployment**

* **Deployment Method**: RESTful API,Batch Processing,Real-time Streaming,Model-as-a-Service.
* **Public Link**: <https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset>
* **Sample Prediction:**

|  |  |
| --- | --- |
| Customer ID | Predicted Probability of Churn |
| 12345 | 0.75 |
| 67890 | 0.30 |
| 34567 | 0.90 |

**13. Source Code:**

**Goole colab phase 3 link:**

**https://colab.research.google.com/**

**14. Future Scope:**

* **Advanced Machine Learning Techniques:** Research is ongoing into using more complex machine learning models like deep learning and ensemble methods to improve prediction accuracy. Techniques like Swish Recurrent Neural Networks (S-RNN) and BiLS-CNN are being explored for their potential in predicting customer churn.
* **Explainable AI (XAI):** As machine learning models become more complex, there's a growing need to understand how they're making predictions. XAI can help businesses understand why customers are churning and make data-driven decisions.
* **Real-time Prediction:** With the increasing availability of real-time data, machine learning models can be used to predict customer churn in real-time, enabling businesses to take proactive measures to retain customers.
* **Industry-specific Solutions:** Different industries have unique characteristics that affect customer churn. Developing industry-specific machine learning models can help businesses better understand and address customer churn.
* **Integration with Business Systems:** Integrating machine learning models with existing business systems can help automate the prediction and retention process, making it more efficient and effective.
* **Telecom Industry**: Predicting customer churn in the telecom industry can help businesses identify high-risk customers and develop targeted retention strategies.
* **E-commerce:** Machine learning models can help e-commerce businesses predict customer churn and develop personalized marketing campaigns to retain customers.
* **Insurance Industry:** Predicting customer churn in the insurance industry can help businesses identify customers who are likely to switch to competitors and develop strategies to retain them.

**15.Team Members and Contributions**

* Clearly mention who worked on:
* **J.JANANI:**
* Data collection,cleaning.
* **V.ASWINI:**
* Feature engineering,visualitation.
* **R.KEERTHANA:**
* Optional deployment.
* **K.SAGUNTHALA:**
* Final report,presentation.