**Phase-1**

Predicting customer churn using machine learning to uncover hidden patterns

**NAME: R Dhavan**

**REGISTER NUMBER: 422623104007**

**INSTITUTION: University college of engineering panruti DEPARTMENT: Computer science and engineering DATE OF SUBMITION**: 29/04/2025

🧩 **Problem Statement**

Customer churn, the phenomenon where customers stop doing business with a company, significantly impacts a company's revenue and growth. Identifying customers likely to churn allows companies to take proactive measures to retain them. However, churn patterns are often complex and not easily detectable with traditional analysis methods. Therefore, we aim to leverage machine learning to uncover hidden patterns and accurately predict customer churn.

# 🎯 Objectives of the Project

* Develop a machine learning model that predicts the probability of customer churn.
* Identify key factors influencing customer churn through data-driven insights.
* Provide actionable recommendations to reduce churn rates.
* Create a deployable solution for real-time or batch prediction of customer churn.

# 🔭 Scope of the Project

* Focus on building a predictive model for customer churn using historical customer data.
* Target sectors can include telecom, banking, subscription services, or e- commerce.
* Deliver insights into customer behavior and churn patterns.
* Limited to structured data (tabular format) for the current phase; future extensions may include unstructured data (e.g., customer reviews).
* Deployment may be on a local server, cloud environment, or integrated into existing CRM systems.

🗂️ **Data Sources**

* **Primary Source**: Publicly available datasets like the Telco Customer Churn dataset (Kaggle, IBM).
* **Secondary Sources**: Synthetic data generation, internal CRM databases (if working in a corporate setting).
* **Attributes**: Customer demographics, account information, service usage patterns, contract details, payment history, etc.

# 🛠️ High-Level Methodology

## 📥 Data Collection

* Gather historical customer data with churn labels (churned/not churned).
* Integrate External and Derived Data Sources

## 🧹 Data Cleaning

* Handle missing values, duplicates, and outliers.
* Correct data inconsistencies.

## D Exploratory Data Analysis (EDA)

* Visualize key variables and their relationships to churn.
* Analyze distributions and correlations.

## 🧬 Feature Engineering

* Create new features such as tenure groups, payment method categories, or interaction scores.
* Encode categorical variables, normalize numerical features.

## 🧠 Model Building

* Train multiple machine learning models (e.g., Logistic Regression, Decision Trees, Random Forest, XGBoost, Neural Networks).
* Hyperparameter tuning using GridSearchCV or RandomizedSearchCV.

## ✅ Model Evaluation

* Metrics: Accuracy, Precision, Recall, F1 Score, ROC-AUC.
* Cross-validation to ensure model robustness.

## 📈 Visualization & Interpretation

* Feature importance plots.
* SHAP values for explainability.
* Churn prediction dashboard (optional).

🚀 **Deployment**

* Package the model into a Flask/Django API.
* Deploy on AWS/GCP/Azure or integrate into company systems.

# 🧰 Tools and Technologies

\_□\_□ ¿ **Programming Language**

* Python

\_□¯ **Notebook/IDE**

* + Jupyter Notebook
  + VS Code
  + PyCharm

\_µ–'–‘ \_lµH **Libraries**:

* + - Data Manipulation: pandas, NumPy
    - Visualization: matplotlib, seaborn, plotly
    - Machine Learning: scikit-learn, xgboost, lightgbm
    - Explainability: SHAP, LIME
    - Deployment: Flask, FastAPI

O☼◦**Optional Tools for Deployment**:

* + - Docker for containerization
    - AWS EC2, Google Cloud App Engine, or Heroku for cloud deployment
    - Git/GitHub for version control
    - MLflow or DVC for model tracking

**Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| Team members | Role | Responsibilities |
| M Hariprasath (Project lead) | Data collection and Data cleaning | Perform data cleaning and transformation |
| M Santhosh | Feature Engineering and Deployment | Define scope, timeline and Ensure data compliance and ethics. |
| R Dhawan | Model Evaluation | Conduct EDA and feature engineering and interpret and explain model results |
| V Navaneetham | Visualization & Interpretation | Ensure dashboard is user friendly and responsive and Handle basic back end/frontend integration |
| S Kanagavally | Exploratory Data Analysis (EDA) and Model Building | Prepare models for production and Ensure scalability and reliability |