**PHASE-2**

**Projet title: predicting customer churn using machine learning to uncover hidden patterns**

**Github:** [**https://github.com/keerthi-0226/ebplDs- predicting-customer-churn-using-machine-learning-uncover-hidden-patterns.git**](%20https:/github.com/keerthi-0226/ebplDs-%20predicting-customer-churn-using-machine-learning-uncover-hidden-patterns.git)

**1.Problem Statement**

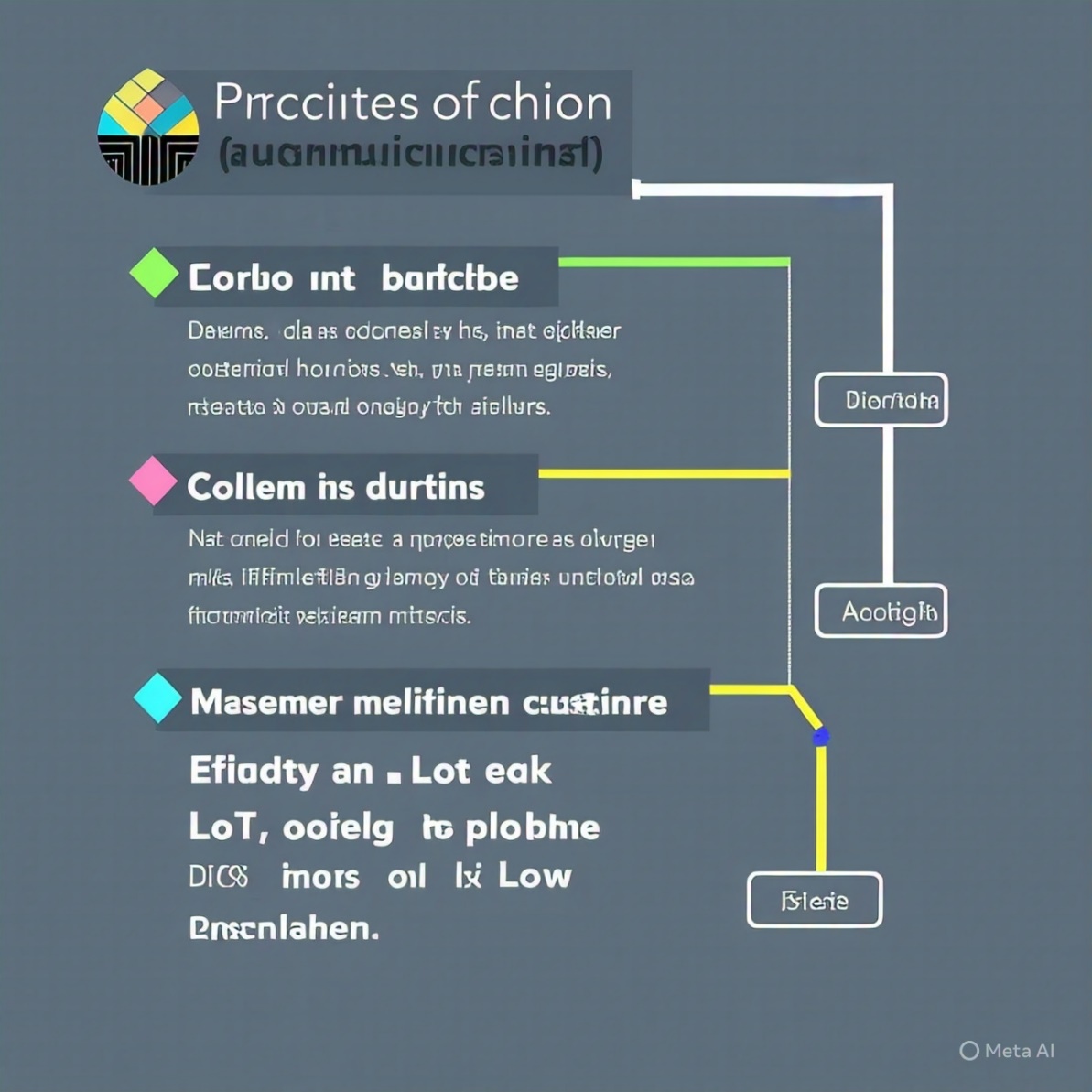
* The goal is to predict whether a customer is likely to stop using a company's service (i.e., churn).
* This binary classification task helps businesses proactively engage customers who are at risk of

Leaving.

**2.Project Objectives**

* Build a machine learning model to classify customers as churned or retained.
* Identify the key features contributing to churn.
* Uncover hidden patterns in behavioral and transactional data.
* Deploy a user-friendly interface for churn prediction (e.g., using Gradio or Streamlit).

**3.flowchart**

****

**4.Data Description**

* Time since last login
* Frequency of service use
* Customer support interactions
* Billing issues

**5.Data Preprocessing**

* Missing Values: Impute or drop nulls.
* Categorical Encoding:

One-hot for nominal features (e.g., Contract, PaymentMethod).

* Label encoding for binary (e.g., Gender, Partner).
* Feature Scaling: Use StandardScaler or MinMaxScaler for models like Logistic Regression or

SVM.

* Handling Imbalanced Data:

Check class distribution (e.g., churned vs. not churned).

Use SMOTE or class weighting if necessary.

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

* Churn Distribution: Pie or bar plot to check class balance.
* Correlation Matrix: Identify relationships among features.
* Feature vs. Churn Plots:
* Tenure vs. Churn
* MonthlyCharges vs. Churn
* Contract type vs. Churn

**7.Feature Engineering**

* Combine or transform features to create:

TotalChargesPerTenure = TotalCharges / Tenure

IsSeniorCitizen binary

HasMultipleServices count of subscribed services

**8.Model Building**

* Use classification algorithms:
* Baseline: Logistic Regression
* Advanced Models:

Random Forest

XGBoost / LightGBM

Support Vector Machine (SVM)

Cross-validation for robust performance estimation.

* Model Evaluation:

Confusion Matrix for class prediction breakdown.

Metrics:

Accuracy

Precision & Recall

F1 Score

ROC-AUC curve

**9. Visualization of Results & Model Insights**

* Once the model is trained (especially tree-based models like Random Forest or XGBoost),
* analyze which features are most important in predicting churn:

Interpret Feature Importance:

Use .feature\_importances\_ in tree models.

Visualize using a bar plot (e.g., seaborn.barplot).

* Common churn drivers:

Contract type (month-to-month contracts churn more)

High MonthlyCharges

Low Tenure

No tech support or online security

* Actionable Business Insights:

Offer incentives to long-tenured customers on monthly plans.

Promote bundled services (multiple services = lower churn).

Improve customer service where frequent tickets correlate with churn

**10.Team Members and Contributions**

* Clearly mention who worked on:
* **R.KEERTHANA:**
* Data cleaning
* **K.SAGUNTHALA:**
* EDA (Exploratory Data Analysis**)**
* **V.ASWINI:**
* Feature engineering
* **J.JANANI:**
* Model development
* Documentation and reporting