1. Problem Statement

predicting customer churn using machine learning to uncover hidden patterns

2. Project Objectives

1. Predict customer churn using machine learning models.

2. Identify key factors influencing customer churn.

3. Uncover hidden behavioral patterns in customer data.

4. Evaluate model accuracy and performance.

*5. Provide insights to support churn reduction strategies*

**3. Flowchart of the Project Workflow**

*2. Data Preprocessing*

*(Cleaning, handling missing values, encoding)*

*↓*

*3. Exploratory Data Analysis (EDA)*

*(Discover patterns, visualize churn trends)*

*↓*

*4. Feature Engineering*

*(Create meaningful input features)*

*↓*

*5. Model Selection & Training*

*(Apply ML algorithms like Logistic Regression, Random Forest, etc.)*

*↓*

*6. Model Evaluation*

*(Use metrics like accuracy, precision, recall, F1-score)*

*↓*

*7. Behavioral Pattern Analysis*

*(Interpret model to find churn drivers)*

*↓*

*8. Deployment & Insights*

*(Deploy model, suggest churn-reduction strategies)*

4. Data Description

*The dataset used for this project consists of customer information from a subscription-based service.*

*It includes demographic, behavioral, and service usage features that help in identifying patterns*

*related to customer churn. Each row represents a unique customer.*

*1. CustomerID – Unique identifier for each customer*

*2. Gender – Gender of the customer (Male/Female)*

*3. Age – Age of the customer*

*2. Data Preprocessing*

*(Cleaning, handling missing values, encoding)*

*↓*

*3. Exploratory Data Analysis (EDA)*

*(Discover patterns, visualize churn trends)*

*↓*

*4. Feature Engineering*

*(Create meaningful input features)*

*↓*

*5. Model Selection & Training*

*(Apply ML algorithms like Logistic Regression, Random Forest, etc.)*

*↓*

*6. Model Evaluation*

*(Use metrics like accuracy, precision, recall, F1-score)*

*↓*

*7. Behavioral Pattern Analysis*

*(Interpret model to find churn drivers)*

*↓*

*8. Deployment & Insights*

*(Deploy model, suggest churn-reduction strategies)*

4. Data Description

*The dataset used for this project consists of customer information from a subscription-based service.*

*It includes demographic, behavioral, and service usage features that help in identifying patterns*

*related to customer churn. Each row represents a unique customer.*

*1. CustomerID – Unique identifier for each customer*

*2. Gender – Gender of the customer (Male/Female)*

*3. Age – Age of the customer*

5. Data Preprocessing

*Data Processing Steps:*

*1. Data Cleaning:*

*Remove duplicate records*

*Handle missing or null values (imputation or removal)*

*Correct inconsistent data entries*

*2. Data Transformation:*

*Convert categorical variables to numerical using encoding techniques (e.g., Label*

*Encoding, One-Hot Encoding)*

*Normalize or scale numerical features using methods like Min-Max Scaling or*

*Standardization*

*1. Understanding Data Distribution:*

*Analyze the distribution of numerical features such as age, tenure, monthly charges, and total*

*charges.*

*Use histograms and boxplots to detect skewness and outliers.*

*2. Churn Rate Overview:*

*Calculate and visualize the overall churn rate.*

*Use pie charts or bar plots to show the proportion of churned vs. non-churned customers.*

*3. Categorical Variable Analysis:*

*Explore the impact of categorical features (e.g., gender, contract type, payment method) on*

*churn.*

*Use count plots and stacked bar charts for comparison.*

6. Exploratory Data Analysis (EDA)

*1. Understanding Data Distribution:*

*Analyze the distribution of numerical features such as age, tenure, monthly charges, and total*

*charges.*

*Use histograms and boxplots to detect skewness and outliers.*

*2. Churn Rate Overview:*

*Calculate and visualize the overall churn rate.*

*Use pie charts or bar plots to show the proportion of churned vs. non-churned customers.*

*3. Categorical Variable Analysis:*

*Explore the impact of categorical features (e.g., gender, contract type, payment method) on*

*churn.*

*Use count plots and stacked bar charts for comparison*

7. Feature Engineering

*1. Encoding Categorical Variables:*

*Apply Label Encoding or One-Hot Encoding to convert categorical features (e.g., Gender,*

*Contract Type, Payment Method) into numerical format for model compatibility.*

*2. Creating Derived Features:*

*Calculate Average Monthly Spend = TotalCharges / Tenure*

*Create Engagement Score based on usage indicators (e.g., logins, service subscriptions)*

*3. Binning and Grouping:*

*Group continuous features like Tenure or Age into bins (e.g., New, Loyal, At-risk) to simplify*

*patterns*

*Convert MonthlyCharges into spending categories (Low, Medium, High)*

8. Model Building

*1. Train-Test Split:*

*Split the preprocessed dataset into training and testing sets (e.g., 80% training, 20%*

*testing) to evaluate model performance on unseen data.*

*2. Baseline Model Creation:*

*Start with a simple model like Logistic Regression to establish a baseline for comparison.*

*3. Advanced Model Training:*

*Train multiple machine learning algorithms to improve prediction accuracy:*

9. Visualization of Results & ModelInsights

Visualization of Results & Model Insights:

1. Churn Distribution:

Bar chart showing percentage of churned vs. retained customers.

2. Feature Importance:

Horizontal bar plot of top features influencing churn (e.g., tenure, contract type, monthly charges).

3. Confusion Matrix:

Visual display of true positives, false positives, true negatives, and false negatives.

4. ROC Curve:

Line graph showing the trade-off between true positive rate and false positive rate for model

evaluation.

1. Programming Language:

Python

2. Data Manipulation & Analysis:

Pandas, NumPy

3. Data Visualization:

Matplotlib, Seaborn, Plotly

4. Machine Learning Libraries:

Scikit-learn

XGBoost / LightGBM (for advanced models)