Customer Churn Prediction with Retention Strategy Simulation using Machine Learning

Problem Statement

Customer churn leads to significant revenue loss in subscription-based businesses.
Traditional churn models focus only on prediction accuracy, failing to explain why customers leave and what actions should be taken.

Objectives

- Predict customer churn accurately

- Identify key drivers behind churn

- Segment customers by risk level

- Simulate retention strategies

- Quantify business impact (revenue saved)

Dataset Description

- Source: Telco Customer Churn Dataset

- Records: ~7,000 customers

- Features:

  o Demographics

  o Services subscribed

  o Billing & payment details

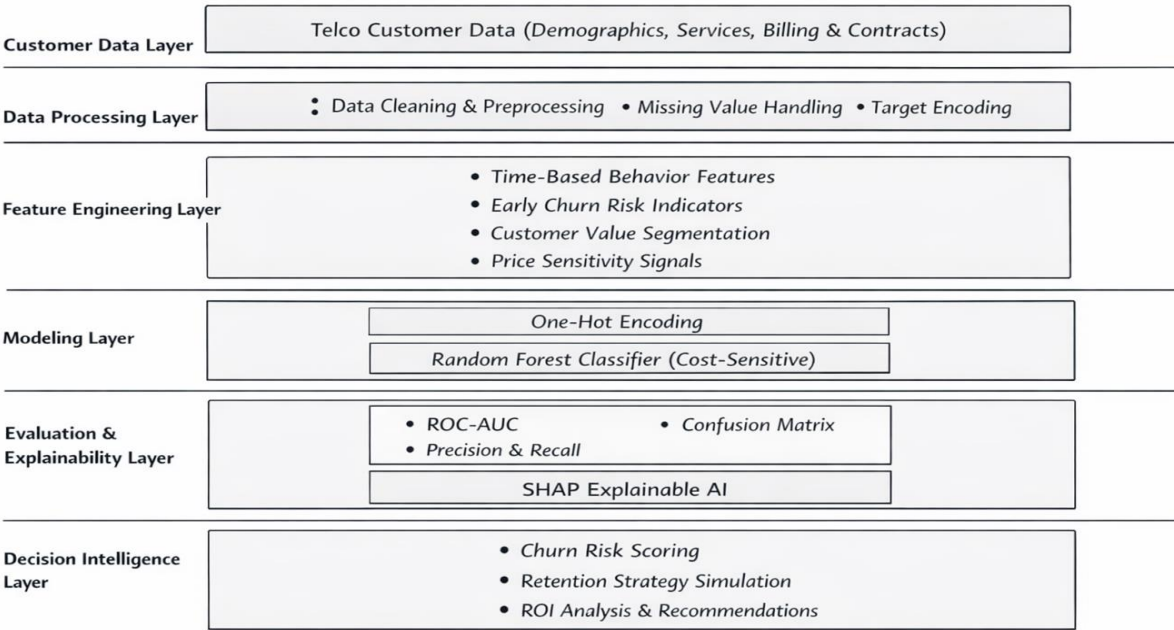 o Contract information

- Target: Churn (Yes / No)

Methodology

◆ **Data Preprocessing**

- Missing value treatment

- Data type correction

- Target encoding

- Removal of non-predictive identifiers

- ◆ **Feature Engineering (Key Contribution)**
  - Time-aware features without raw time series
  - Early churn risk identification
  - Customer value modeling
  - Behavioral proxies aligned with industry practices

- ◆ **Modeling**
  - Random Forest (cost-sensitive)
  - Class imbalance handled using class weights
  - Hyperparameters tuned for recall & ROC-AUC

- ◆ **Evaluation Metrics**
  - ROC-AUC
  - Precision, Recall, F1-score
  - Confusion Matrix
  - Business-oriented error interpretation

- ◆ **Explainable AI**
  - SHAP used for:
    - Global feature importance
    - Individual churn explanations
  - Enables trust & actionability

System Architecture

Customer Data Layer	Telco Customer Data (Demographics, Services, Billing & Contracts)
Data Processing Layer	⦂ Data Cleaning & Preprocessing • Missing Value Handling • Target Encoding
Feature Engineering Layer	• Time-Based Behavior Features • Early Churn Risk Indicators • Customer Value Segmentation • Price Sensitivity Signals
Modeling Layer	One-Hot Encoding Random Forest Classifier (Cost-Sensitive)
Evaluation & Explainability Layer	• ROC-AUC • Confusion Matrix • Precision & Recall SHAP Explainable AI
Decision Intelligence Layer	• Churn Risk Scoring • Retention Strategy Simulation • ROI Analysis & Recommendations

Retention Strategy Simulation

- Customers segmented into:
    - High risk
    - Medium risk
    - Low risk
- Retention actions simulated:
    - Discounts
    - Contract upgrades
    - Support outreach
- ROI calculated using:
    - Retention success rate
    - Offer cost
    - Customer lifetime value

Results

📊 Model Performance

- ROC-AUC ≈ 0.85
- High recall on churn class
- Stable generalization

💼 Business Impact

- Significant churn reduction potential
- Positive net revenue gain
- Optimized retention spending

Key Insights

- Month-to-month contracts drive churn
- Early tenure customers are most vulnerable
- High charges increase churn probability
- Targeted retention is more cost-effective than mass offers

Project Novelty

✓ Explainable churn modeling

✓ Time-aware behavior simulation

✓ Cost-sensitive learning

✓ Retention ROI simulation

✓ Modular ML pipeline

This project goes beyond prediction and acts as a decision-support system.

Tools & Technologies

- Python

- Pandas, NumPy

- Scikit-learn

- SHAP

- Matplotlib, Seaborn

- Anaconda

- VS Code & Jupyter

Conclusion

This project demonstrates how machine learning can be transformed into a business-ready product by combining prediction, explainability, and decision intelligence.
It reflects real-world ML system design and is suitable for enterprise deployment with minimal extension.