

PROBLEM DEFINITION:

Customer churn prediction is a crucial challenge for businesses across various industries. It involves developing a machine learning model that can predict whether a customer is likely to churn, i.e., cancel their subscription or leave a service. This prediction is based on various customer attributes and behavior. Retaining existing customers is often more cost-effective than acquiring new ones, making it essential for companies to identify and proactively address churn risk. The ultimate goal of this model is to enhance customer retention and improve business profitability.

DATASET CONTENT:

Dataset Information:

The dataset contains information about customers and their interactions with a service provider. It is designed to predict customer behavior in terms of churn (whether a customer will leave the service provider).

Columns in the Dataset:

Churn: This column indicates whether a customer left within the last month. It serves as the target variable for prediction (1 for churn, 0 for no churn).

Gender: Whether the customer is a male or a female

SeniorCitizen: Whether the customer is a senior citizen or not (1, 0)

Partner: Whether the customer has a partner or not (Yes, No)

Dependents: Whether the customer has dependents or not (Yes, No)

Tenure: Number of months the customer has stayed with the company

PhoneService: Whether the customer has a phone service or not (Yes, No)

MultipleLines: Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService: Customer's internet service provider (DSL, Fiber optic, No)

OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup: Whether the customer has online backup or not (Yes, No, No internet service)

DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service)

TechSupport: Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract: The contract term of the customer (Month-to-month, One year, Two year)

CheckPaperlessBilling: Whether the customer has paperless billing or not (Yes, No)

PaymentMethod: The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

MonthlyCharges: The amount charged to the customer monthly

TotalCharges: The total amount charged to the customer

Churn: Whether the customer churned or not (Yes or No)

DESIGN THINKING:

1. Data Collection:

The dataset used in this project comes from Kaggle . There are 21 training variables (features) that can be used for predicting customer churn. These variables include customer-related attributes and information about services and billing. The dataset link is [Telco-customer-churn](#).

2.Data Preprocessing:

Handle missing values and outliers.

Encode categorical variables using techniques like one-hot encoding.

Normalize or scale numerical features if needed.

3.Exploratory Data Analysis (EDA):

Visualize the data to understand the distribution of features.

Identify patterns and relationships between customer attributes and churn.

Analyze correlations and trends.

4.Feature Engineering:

Create new features or derive meaningful insights from existing ones.

Feature selection to choose the most relevant attributes for prediction.

5.Machine Learning or Statistical Modeling:

Split the dataset into training and testing sets.

Select appropriate machine learning algorithms for binary classification (e.g., logistic regression, decision trees, random forests, XGBoost).

Train multiple models and evaluate their performance using metrics such as accuracy, precision, recall, F1-score, and ROC AUC.

6.Model Evaluation:

Choose the best-performing model based on evaluation metrics.

Perform cross-validation to ensure model generalization.

Interpret model results to understand which customer attributes are influential in predicting churn.

7.Model Tuning:

Fine-tune hyperparameters to improve model performance.

Address issues like overfitting or underfitting.

8.Model Deployment:

Deploy the trained model using a web framework like Flask for real-time predictions.

Develop an API for integration with business systems.

9.Continuous Monitoring:

Continuously monitor model performance and retrain it as needed with new data.

Update the model to adapt to changing customer behavior.

10.Interpretation and Insights:

Provide actionable insights to the business based on model predictions.

Identify strategies for customer retention and churn prevention.