# **Problem Formulation: Internet Service Provider Customer Churn**

## **1. Business Problem Definition**

Customer churn occurs when a subscriber stops using an internet service provider (ISP) and either switches to a competitor or cancels their subscription. High churn rates lead to revenue losses, increased customer acquisition costs, and a lower customer lifetime value (CLV). Since acquiring new customers is significantly more expensive than retaining existing ones, ISPs must proactively identify at-risk customers and take strategic measures to retain them. The telecommunications industry is highly competitive, making customer retention essential for sustained growth. Common strategies for reducing churn include personalized offers, improved service plans, and loyalty rewards, but targeting the right customers is key to ensuring cost-effective retention efforts.

Addressing customer churn is critical because it directly impacts an ISP’s profitability and market position. Losing long-term subscribers results in revenue decline, while high acquisition costs make it unsustainable to replace lost customers continually. Additionally, churned customers may switch to competitors offering better plans, further intensifying market competition. Understanding the reasons behind customer churn helps ISPs enhance their service quality and offerings, leading to improved customer satisfaction and long-term business success.

**2. Key Business Objectives**

The primary objective of this project is to **develop an automated data pipeline and machine learning model to predict customer churn** and enable proactive intervention. The business objectives include:

1. **Identify At-Risk Customers**: Use predictive analytics to flag customers likely to churn, enabling early intervention
2. **Improve Customer Retention**: Provide insights to customer service teams for personalized retention strategies, such as tailored offers, discounts, and service adjustments
3. **Optimize Pricing and Service Plans**: Ensure customers receive competitive offers based on their usage patterns and preferences
4. **Enhance Business Decision-Making**: Use data-driven insights to allocate resources effectively and for strategic decision making.
5. **Automate and Scale the Process**: Develop an automatic, repeatable pipeline for continuous monitoring and predictions.

## **3. Data Sources and Attributes**

Dataset: The Internet Service Churn dataset from Kaggle will be used. Key Attributes are -

• Demographics: Gender, SeniorCitizen, Partner, Dependents.

• Account Information: Tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport.

• Service Usage: StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod.

• Target Variable (Churn): The binary churn indicator (Yes/No).

## **4. Expected Pipeline Outputs**

To ensure a robust and scalable solution, the pipeline will produce the following key outputs:

1. **Clean Datasets for Exploratory Data Analysis (EDA)**
   * Remove missing values, handle outliers, and standardize data.
   * Provide insights into churn patterns via visualizations.
2. **Transformed Features for Machine Learning**
   * Create new features (e.g., payment history, usage frequency, service issues, customer complaints etc.).
   * Scaled and normalized numerical features.
   * One-hot encoding or label encoding for categorical features.
3. **Deployable Churn Prediction Model**
   * A trained machine learning model (e.g., Logistic Regression, Random Forest, XGBoost) capable of predicting churn likelihood.
   * Model will be saved in a version-controlled repository for easy deployment.

## **5. Evaluation Metrics**

The effectiveness of the churn prediction model will be assessed using the following metrics:

* Accuracy
* Precision
* Recall
* F1-Score
* ROC-AUC Score