

# Project Report: Customer Churn Prediction Using Machine Learning

## 1. Project Title

### Customer Churn Prediction in a Telecom Company Using Machine Learning

## 2. Project Objective

To predict whether a telecom customer is likely to **churn (leave the service)** using historical customer data. This helps companies take proactive retention actions.

## 3. Dataset Overview

- **Source:** [Telco-Customer-Churn.csv](#)
- **Records:** 7,043 customers
- **Columns:** 21
- **Target Variable:** Churn (Yes or No)

### Sample Features:

Feature	Description
gender	Male or Female
SeniorCitizen	0 or 1
Partner, Dependents	Yes/No
tenure	No. of months the customer stayed
InternetService, PhoneService	Type of services
MonthlyCharges, TotalCharges	Billing info
Contract, PaymentMethod	Type of contract & payment
Churn	Target: Did the customer churn?

## 4. Data Preprocessing

### Cleaning

- Dropped irrelevant column: customerID
- Replaced blank strings in TotalCharges with 0.0 and converted to float
- Checked and confirmed no missing values remain

### Encoding

- Applied **Label Encoding** to all categorical features using LabelEncoder

- Saved encoders to encoders.pkl for later prediction use

## Imbalance Handling

- Class distribution: Churn = No (73.5%), Yes (26.5%)
- Used **SMOTE (Synthetic Minority Oversampling Technique)** to balance the classes

## 5. Exploratory Data Analysis (EDA)

### Numerical Features


- Plotted **histograms** and **boxplots** of tenure, MonthlyCharges, TotalCharges
- Generated correlation heatmap

### Categorical Features

- Count plots showed patterns such as:
  - Month-to-month contracts → Higher churn
  - Fiber optic users → More likely to churn
  - Electronic check payment → Higher churn

## 6. Model Training & Evaluation

### Algorithms Used:

Model	Accuracy (CV)
Decision Tree	78%
XGBoost	83%
<b>Random Forest</b>	<b>84%</b>  Selected

### Final Evaluation (Test Set)

Metric	Value
Accuracy	77.85%
Precision (Churn)	58%
Recall (Churn)	59%
F1 Score (Churn)	58%
Confusion Matrix	

[[878 158]

[154 219]]

## 7. Model Deployment

 **Files Saved**

- customer\_churn\_model.pkl → Trained Random Forest model
- encoders.pkl → Saved label encoders

## Prediction System

Built a prediction system where the user can enter the following values:

- Gender, Tenure, Internet service, Billing info, etc.

# Sample input structure

```
{
  'gender': 'Female',
  'SeniorCitizen': 0,
  'Partner': 'Yes',
  ...
}
```

## Output:

- Prediction: Churn or No Churn
- Probability: e.g. [0.78, 0.22]

## 8. Interactive System (User Input)

The model was extended to allow **dynamic user input** through the console:

```
input("Enter tenure: ")
```

Then:

- Transformed via saved encoders
- Predicted using the trained model
- Displayed result with probability

## 9. Conclusion

- Successfully built a **customer churn prediction system** with ~78% accuracy
- Model identifies churn-risk customers using contract, usage, and billing data
- Can be deployed in **CRM or business intelligence dashboards** for proactive retention

## 10. Future Enhancements

- Use **GridSearchCV** or **Optuna** for hyperparameter tuning
- Add **SHAP** or **LIME** for explainable AI
- Deploy as a web app using **Flask** or **Streamlit**
- Incorporate **real-time data pipeline** for live predictions

## 11. Project Files

File	Description
Telco-Customer-Churn.csv	Raw dataset
Churn_Prediction_Telco.ipynb	Jupyter/Colab notebook
customer_churn_model.pkl	Trained Random Forest model
encoders.pkl	Label encoders for categorical features