Telco Customer Churn Prediction & Analysis

Team: CLS ONL3_AIS4_G1 – Team 2

Team Members:

1- Shahd Ashraf Ramadan

2- Aya Anwar Mahmoud

3- Aya Hesham Saleh

4- Ayat Mohammad Mekky

5- Dorothy Gerges Dawod

6- Mohammad Naser Ibrahim

Supervisor: Eng. Khaled El-Liethy

Agenda Overview

1. Introduction

Customer churn is a major challenge for telecom companies, as losing customers directly **reduces revenue** and increases **acquisition costs**. Predicting churn enables proactive **retention strategies**, helping maintain **loyalty** and **maximize revenue**. This project implements a full **machine learning pipeline**, from **data cleaning** to **deployment**, combining **technical rigor with business understanding** to identify **patterns behind churn** and provide **actionable insights**.

2. Objectives

- Analyze **customer behavior** to identify **key churn drivers**.
- **Preprocess**, **engineer**, and **select** the most **relevant features** for modeling.
- Train and compare six **classification models: Logistic Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost**.
- Evaluate models using **accuracy, precision, recall, F1-score, and AUC**.
- Derive **actionable business insights** to **guide retention strategies** effectively.

3. Stakeholder Analysis

Stakeholder	Role	Needs/Expectations
Telecom Management	Decision Makers	Accurate churn predictions
Customer Service Team	Implement Retention Actions	Clear insights for high-risk customers
Data Science Team	Model Development	Clean data and reliable model
IT / DevOps	Deployment	Easy integration and scalable system
Customers	Indirectly affected	Improved service and offers

4. Dataset Overview

Dataset Name: Telco Customer Churn Dataset

Source: Kaggle – Telco Customer Churn

Dataset Design:

Column	Type	Description
customerID	String	Unique identifier
gender	String	Customer gender
seniorCitizen	Int	Binary flag
partner	Int	Binary flag
dependents	Int	Binary flag
tenure	Int	Months as customer
contract	String	Contract type
paymentMethod	String	Payment method
monthlyCharges	Float	Monthly subscription
totalCharges	Float	Total charges
churn	Int	Target variable

Churn rate: ~26.5% (moderately imbalanced)

5. Data Preprocessing & Cleaning

0

Duplicates

11

Missing Values Removed

TotalCharges

(from categorical to numeric)

Data Type Conversion

Outliers:

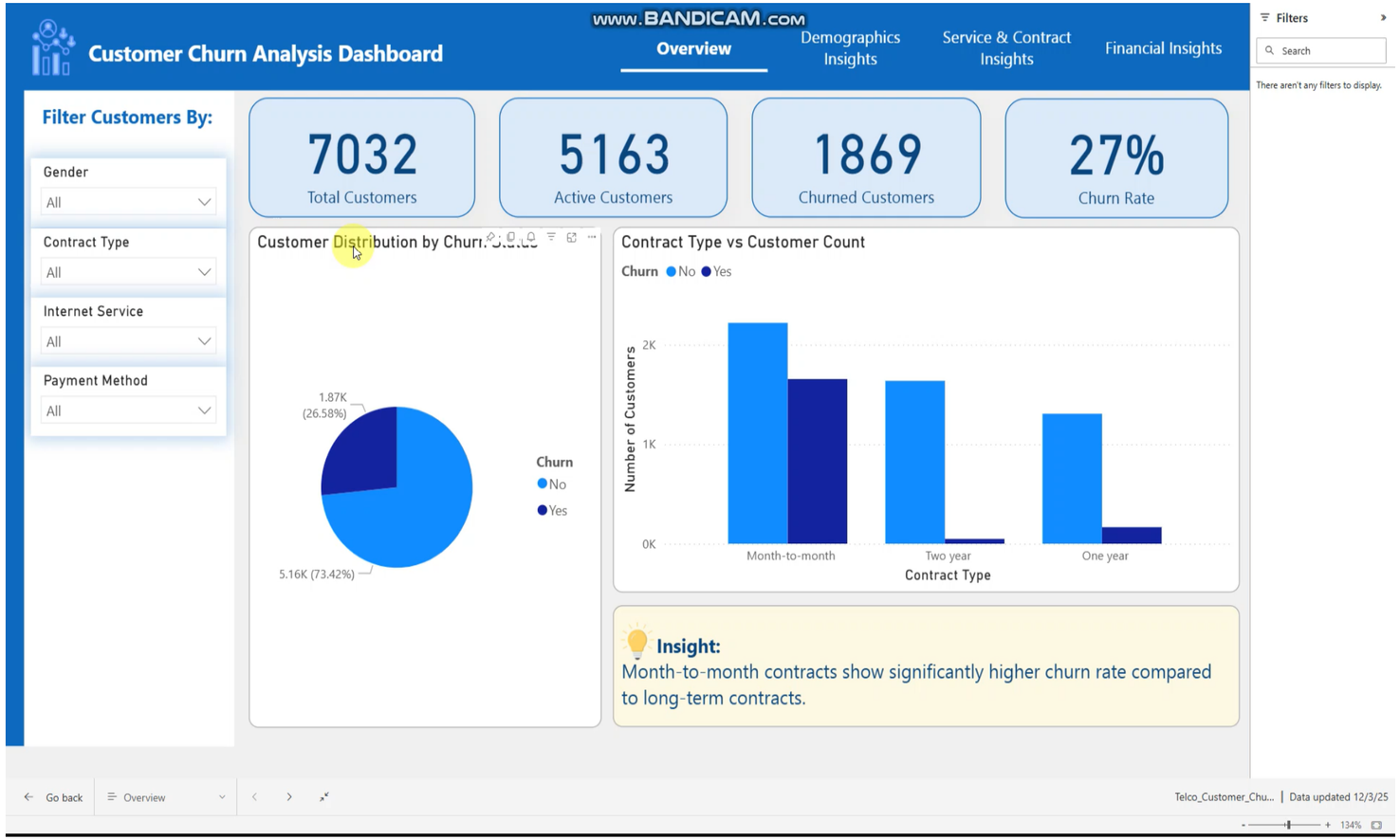
- Numeric features inspected
- No extreme values removed (represent real customer behavior)

Encoding: Applied only for models that cannot handle categorical features

Dataset: Saved for reproducibility: cleaned_Telco-Customer-Churn.csv

6. Exploratory Data Analysis (EDA)

6.1: Dashboard Demo:



Note: We performed the exploratory analysis using both Python (Plotly) and Power BI.
For this presentation, Power BI visuals are used as they provide clearer and more intuitive business insights.

6. Exploratory Data Analysis (EDA)

6.2 Insights:

- **Month-to-month customers have the highest churn rate**, making contract type one of the strongest churn drivers.
- **Customers paying through Electronic Check churn significantly more** than those using other payment methods.
- **Short-tenure customers (0–12 months) contribute the largest portion of churn**, indicating early dissatisfaction.
- **Fiber Optic users show higher churn rates than DSL customers**, suggesting potential service or pricing issues.
- **Senior citizens and customers with single-line contracts exhibit higher churn levels**, highlighting vulnerable customer segments.

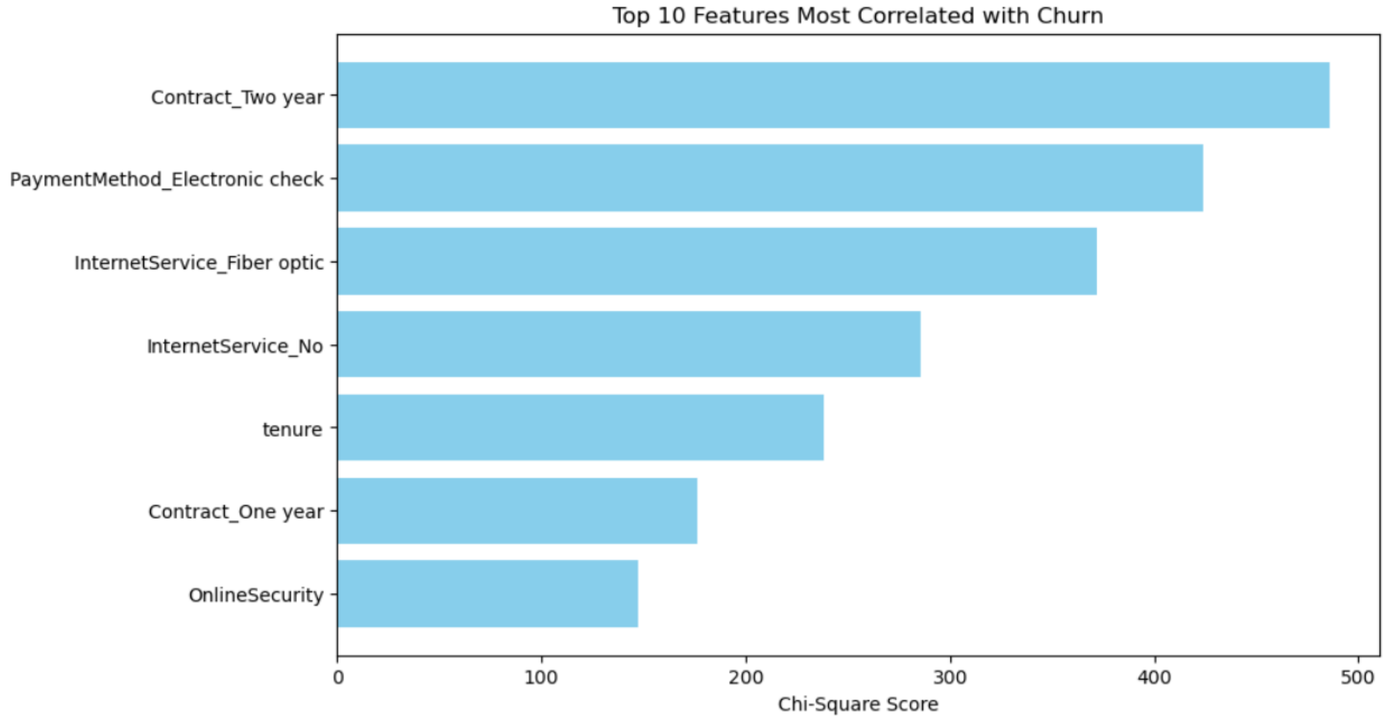
7. Statistical Tests & Feature Selection

Chi-Square Test (Categorical Features)
Significant association with churn ($p < 0.05$):

- **Contract**
- **PaymentMethod**
- **InternetService**

Mann–Whitney U Test (Numeric Features)
Significant differences between churn vs non-churn groups:

- **Tenure**
- **MonthlyCharges**



8. Feature Engineering

1. Binary Encoding

- Converted binary categorical variables to numeric values:
  Yes → 1, No → 0
- Improves model interpretability and performance.

2. Tenure Grouping

- Created tenure bands to capture customer lifecycle stages:
  0–12 months, 13–24 months, 25–48 months, 49–72 months
- Helps identify patterns related to early and long-term churn.

3. AvgChargesPerMonth

- New feature calculated as:
  AvgChargesPerMonth = TotalCharges / Tenure
- Represents the true average monthly spending for each customer.

4. ServiceCount

- Counted the number of active services per customer.
- Useful for analyzing relationships between service bundles and churn.

5. Interaction Feature

- Created an interaction term:
  Tenure × MonthlyCharges
- Highlights customers with low tenure but high monthly charges, who are typically more likely to churn.

6. One-Hot Encoding

- Applied one-hot encoding to multi-category variables:
  Contract, PaymentMethod, InternetService
- Avoids introducing unintended ordinal relationships.

9. Modeling & Evaluation

Model	Accuracy	AUC	Precision (Class 0)	Recall (Class0)	F1 (Class0)	Precision (Class 1)	Recall (Class1)	F1 (Class1)
Logistic Regression	0.7235	-----	0.90	0.71	0.79	0.49	0.77	0.60
Random Forest	0.7285	-----	0.86	0.76	0.80	0.49	0.65	0.56
Gradient Boosting	0.7235	0.825	0.90	0.71	0.79	0.49	0.77	0.60
XGBoost	0.7776	-----	0.83	0.88	0.85	0.60	0.50	0.55
LightGBM	0.7832	0.822	0.84	0.87	0.85	0.60	0.55	0.57
CatBoost	0.7500	0.931	0.85	0.81	0.83	0.53	0.60	0.56

CatBoost chosen for final deployment due to:

- Excellent handling of categorical features
- High AUC & consistent cross-validation performance
- Strong recall for churners

10. Business Impact & Recommendations

Early churn prediction enables targeted retention campaigns, allowing the company to proactively address high-risk customers.

Using CatBoost along with recommended actions can reduce churn by ~6%, improving revenue stability and customer lifetime value.

Key churn drivers identified:

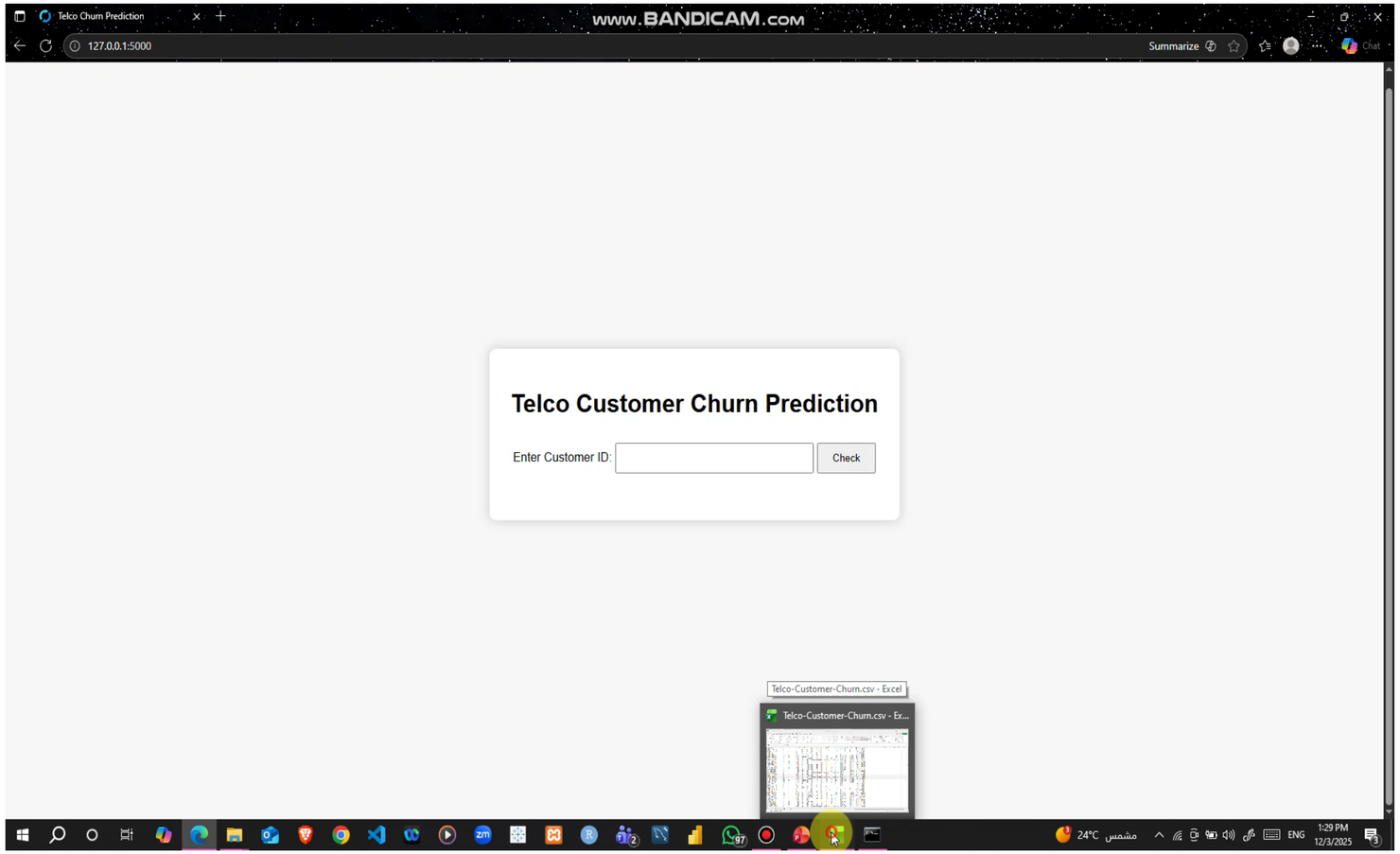
- Contract Type
- Payment Method
- Customer Tenure
- Service Usage Patterns

Recommended Actions:

- Offer personalized retention incentives for month-to-month or high-risk customers.
- Promote automatic payment methods to reduce churn.
- Focus support and engagement on short-tenure and high-spending customers.
- Enhance services for Fiber Optic users and senior citizens, addressing specific needs.

11. Web Application

11.1 Demo:



- **Enter Customer ID → show churn prediction & probability**
- **Invalid Customer IDs are detected and appropriate error messages are returned without crashing the system**

11. Web Application

11.1 Implementation / Tools:

Backend

Frontend

ML Model

12. Conclusion

In this project, we built a complete machine learning pipeline to predict customer churn for a telecom company using Python, Power BI, and CatBoost. We identified the main factors driving churn, engineered meaningful features, and created a web application ready for deployment. By predicting churn early and applying targeted retention strategies, the company can reduce churn by around 6%, stabilize revenue, and increase customer lifetime value. This solution combines technical accuracy with practical business insights, making it both actionable and reliable.

Thank You