Customer Churn Prediction using Business Analytics & Machine Learning

An End-to-End Analytics Project

BUSINESS BACKGROUND

- Customer churn refers to customers discontinuing a service.
- High churn leads to:
  - Revenue loss
  - Increased acquisition cost
  - Lower customer lifetime value
- Businesses require a **proactive, data-driven approach** to identify at-risk customers before churn occurs.

PROBLEM STATEMENT

Organizations lack an effective analytical mechanism to identify customers who are likely to churn in advance. As a result, retention strategies are often reactive, untargeted, and inefficient.

This project addresses the need to:

- Predict customer churn
- Identify churn-driving factors
- Support targeted retention decisions

PROJECT OBJECTIVES

Predict customer churn

using historical customer data

Identify key factors

influencing churn

Segment customers

based on churn risk

Present insights

through an interactive dashboard for business decision-making

DATASET OVERVIEW

Dataset Used:

IBM Telco Customer Churn Dataset

Dataset Size:

- Total Customers: **7,043**

Key Data Categories:

- Customer tenure
- Contract type
- Internet and service usage
- Monthly charges
- Payment method
- Churn indicator (1 / 0)

ANALYTICAL APPROACH

Step-wise Methodology Followed:

1. Data loading and quality checks
2. Exploratory data analysis (EDA)
3. Feature engineering
4. Model building and evaluation
5. Model explainability
6. Risk segmentation and dashboard creation

Tools Used:

- Python (Google Colab)
- Machine Learning models
- Power BI

KEY EDA INSIGHTS

Key findings from exploratory data analysis:

- Churn is significantly higher among **month-to-month contract customers**
- Customers in the **0–12 months tenure group** show the highest churn
- Higher monthly charges are associated with increased churn
- Contract duration and tenure are strong churn indicators

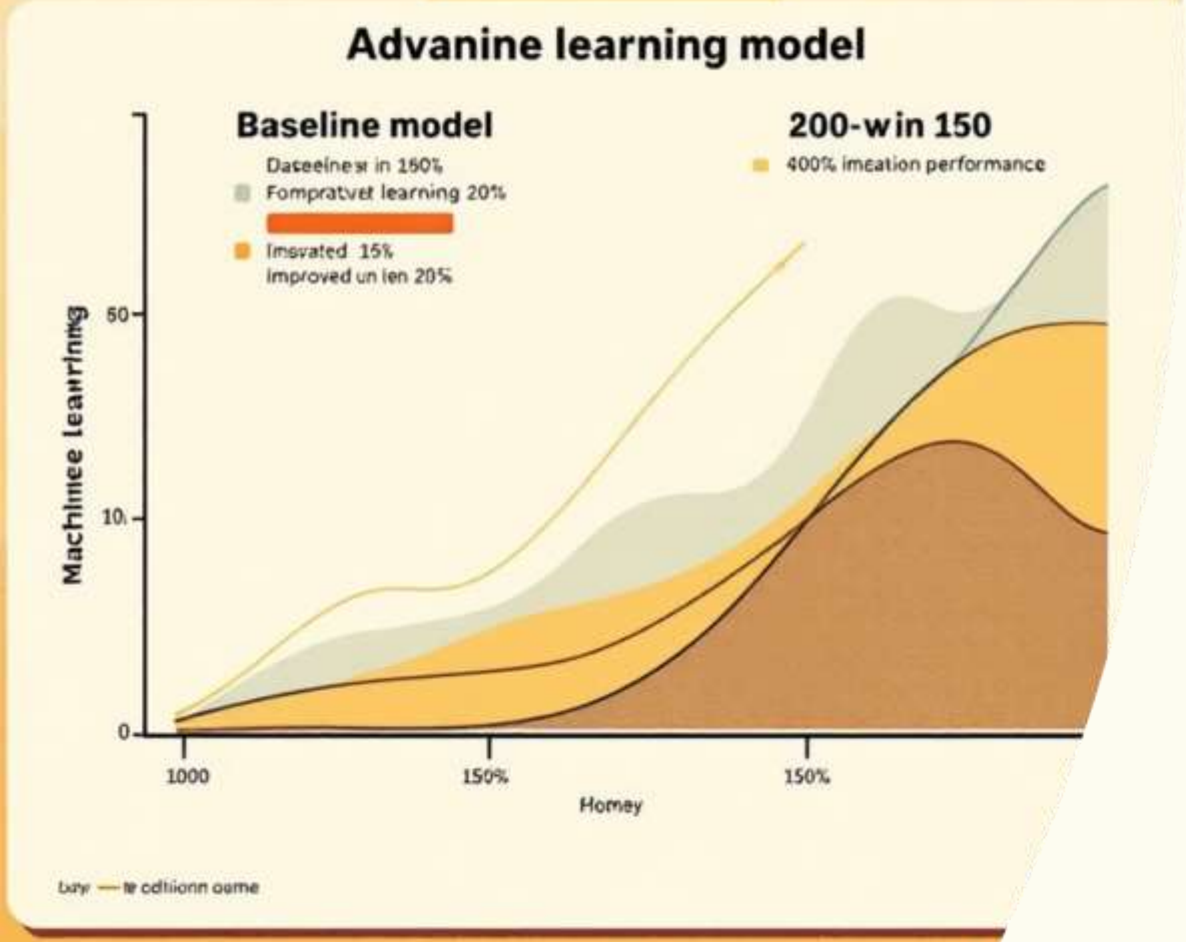
FEATURE ENGINEERING & MODELLING

Feature Engineering Performed:

- Encoding of categorical variables
- Creation of tenure-based groupings
- Preparation of model-ready dataset

Models Implemented:

- Logistic Regression (final selected model)
- Advanced Machine Learning model (evaluated for comparison)



MODEL EVALUATION

- Models were evaluated using classification metrics
- Baseline model demonstrated better predictive capability than Advanced model
- Churn probabilities were generated for each customer
- Model outputs were suitable for business interpretation

MODEL EXPLAINABILITY & RISK SEGMENTATION

Explainability Approach:

- Feature importance analysis
- Interpretation of churn-driving variables

Risk Segmentation Created:

- Low Risk
- Medium Risk
- High Risk

This segmentation enables focused retention strategies.

POWER BI DASHBOARD OUTPUT

Our interactive Power BI dashboard provides a comprehensive overview of customer churn, enabling business users to quickly grasp key trends and identify at-risk customers.

7,043

Total Customers

Monitored in the dataset

1,869

Churned Customers

Who discontinued service

27%

Overall Churn Rate

Indicating the scale of the problem

The dashboard further visualizes critical insights, including:

- Customer distribution by **churn risk segment**
- Churn rates analyzed by **contract type** and **tenure group**
- A detailed **customer-level churn probability table** for targeted interventions

BUSINESS INSIGHTS & CONCLUSION

Key Business Insights:

- Early-tenure customers are more likely to churn
- Month-to-month contracts exhibit the highest churn
- High-risk segment represents a smaller but critical group

Conclusion:

- The project successfully predicts churn and identifies key drivers
- Risk segmentation supports targeted retention actions
- The dashboard provides actionable insights for business teams

Future Scope:

- Real-time churn prediction
- Integration with CRM systems
- Testing retention strategies using experiments