Customer Churn Prediction for a Telecom Company

Presented by:

1. Elsayed Ashraf Bakry (Data Preparation & EDA & Model Training & Validation & Presentation)

2. EL-Hussein Salah (Checking Errors & Deployment via Web Interface & & Promo Video)

(Team 13), GTC Internship

The Critical Challenge: Understanding Customer Churn

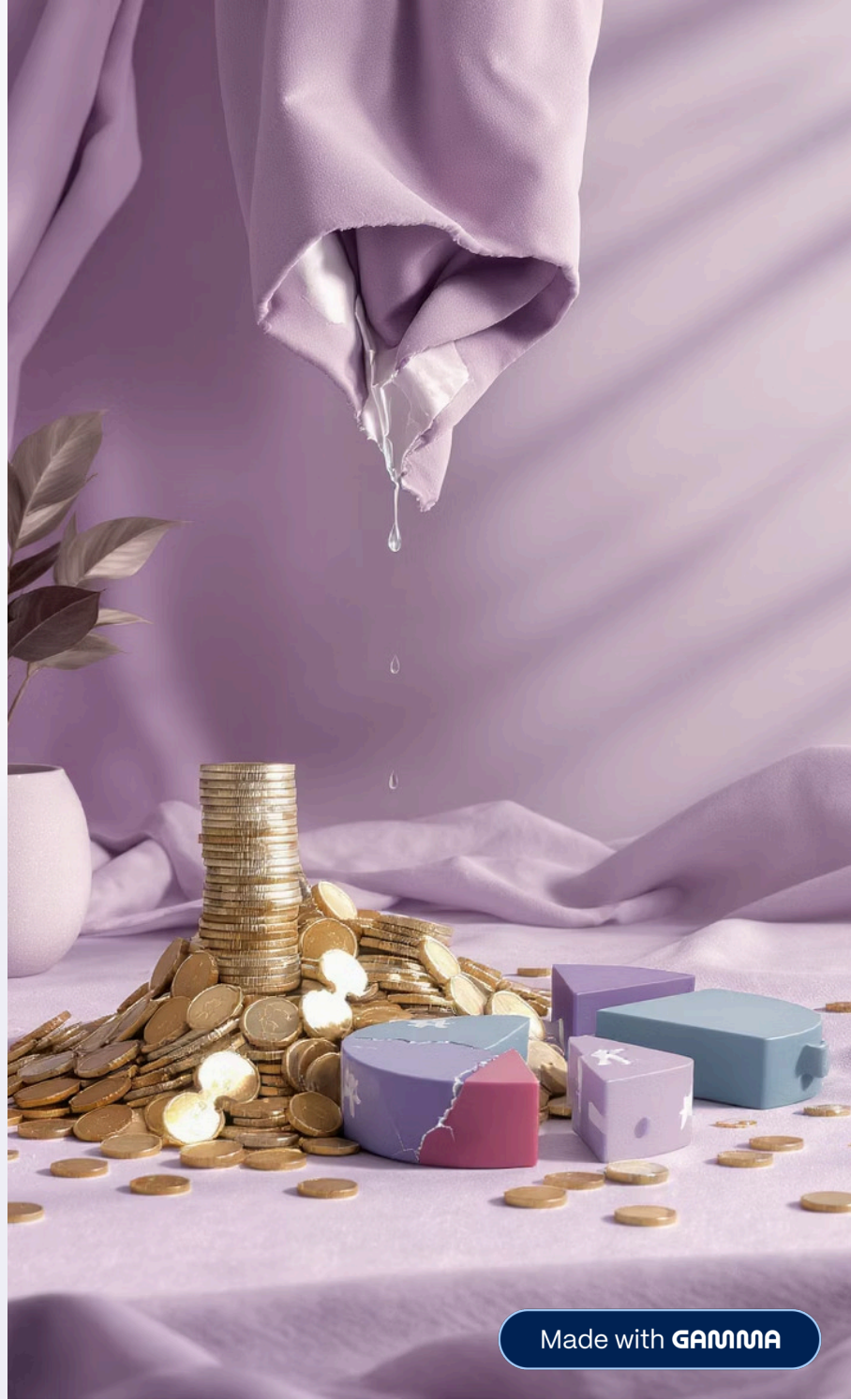
Customer churn is a significant threat to profitability and growth in the highly competitive telecom sector.

Why it Matters

- High acquisition costs for new customers
- Lost revenue from departing subscribers
- Negative impact on market share and brand perception

Proactive Solutions

- Identify at-risk customers before they leave
- Implement targeted retention strategies
- Optimize customer experience and service offerings



Business Impact: Costs vs. Benefits

1 The Cost of Churn

- Up to 5x more expensive to acquire a new customer than retain an existing one.
- Lost lifetime value of a customer.
- Increased marketing and sales expenses.

2 Benefits of Retention

- Increased customer loyalty and advocacy.
- Stable, predictable revenue streams.
- Enhanced brand reputation and market position.

Effective churn prediction translates directly into significant cost savings and revenue growth.

Dataset Overview: Telco Customer Churn

Leveraging a comprehensive Kaggle dataset to build robust prediction models. 📊

Dataset Source

Publicly available "Telecom Customer Churn" from Kaggle.

Size & Scope

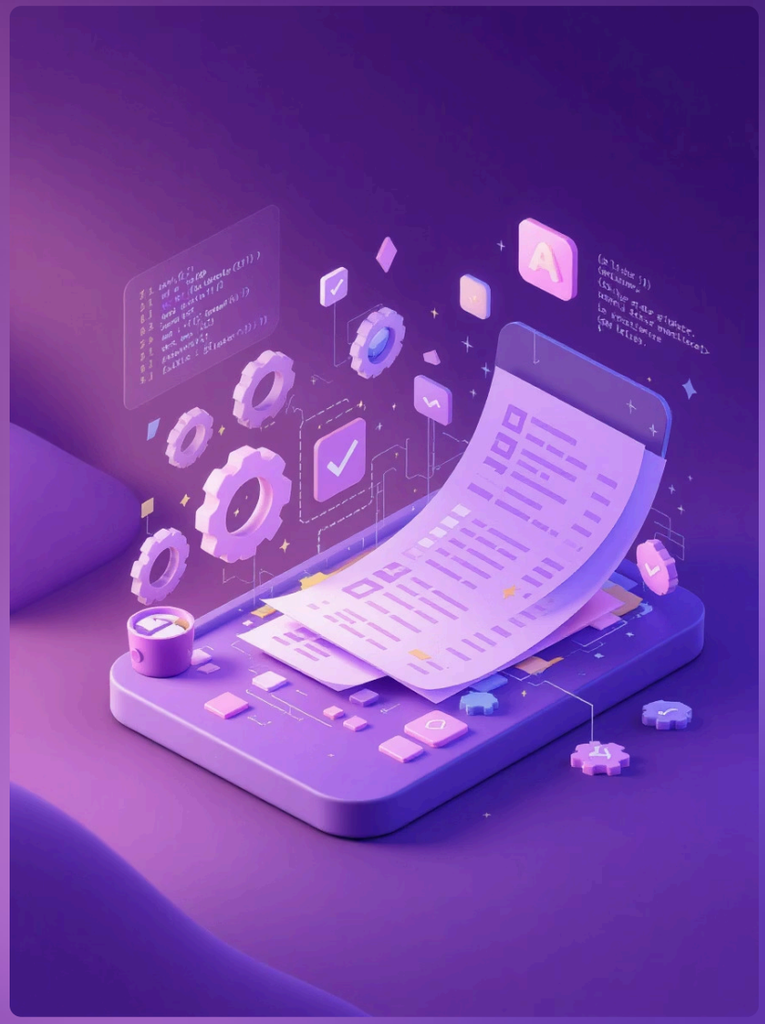
7,043 customer records with 21 features.

Key Features

Demographics, services, contract details, monthly charges, total charges.

Target Variable

"Churn" (Yes/No) indicating customer departure.



Data Preparation: Laying the Foundation

Rigorous data cleaning and preprocessing ensure model accuracy and reliability.

01

Missing Values

Addressed 'TotalCharges' missing values by imputation or removal.

02

Irrelevant Columns

Dropped 'customerID' as it holds no predictive power.

03

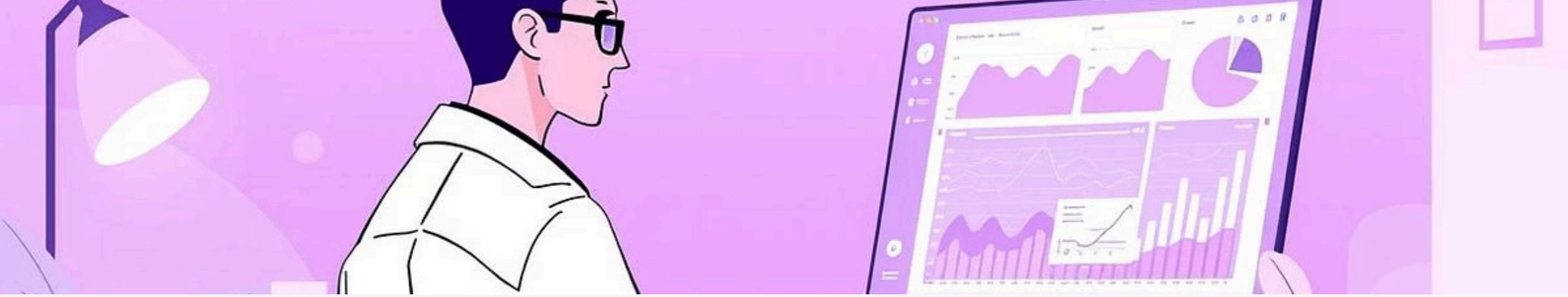
Data Type Conversion

Ensured all features are in appropriate numerical or categorical formats.

04

Categorical Encoding

Converted categorical variables into numerical representations (e.g., One-Hot Encoding).



Exploratory Data Analysis: Uncovering Insights

Initial analysis revealed critical patterns and imbalances in the dataset. 📊

Churn Imbalance

Significant class imbalance: 73.5% 'No' churn vs. 26.5% 'Yes' churn.

Tenure Insights

Customers with shorter tenures show higher churn rates.

Contract Type

Month-to-month contracts have a disproportionately high churn rate.

Charges Analysis

Higher monthly charges correlate with increased churn, especially for non-long-term contracts.

Feature Engineering: Enhancing Predictive Power

Creating new, more informative features from existing data. 🤖

Tenure Grouping

Categorized tenure into bins (e.g., 0-12 months, 13-24 months) to capture non-linear effects.

ServiceCount

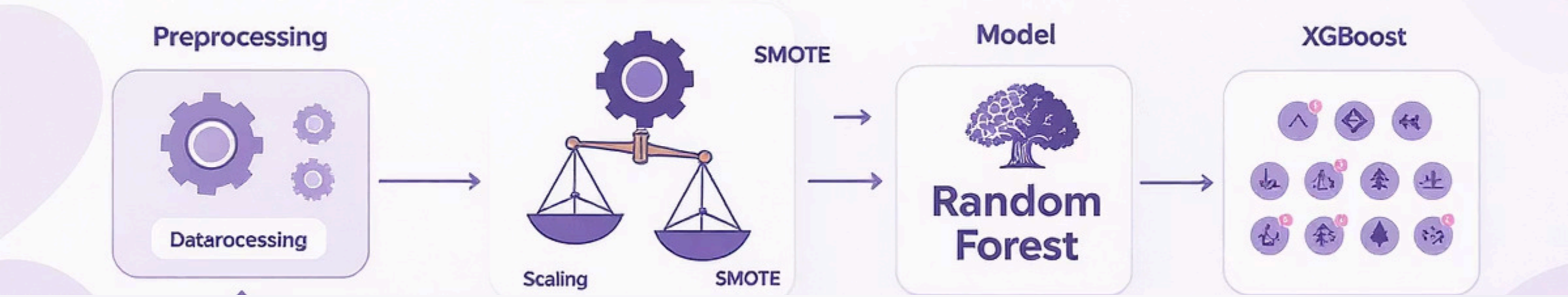
Aggregated the total number of services each customer subscribes to.

Charge Ratios

Calculated ratios like 'MonthlyCharges / TotalCharges' to represent spending patterns.

Average Monthly Spend

Derived average monthly spend over tenure for a clearer financial footprint.



Modeling Approach: Building Robust Predictors

A systematic approach to model selection and training for optimal performance. 🤖

Preprocessing Pipeline

Standardized and scaled numerical features to prevent bias.

SMOTE for Imbalance

Applied Synthetic Minority Over-sampling Technique (SMOTE) to balance the churn classes.

Model Selection

Evaluated Random Forest and XGBoost for their strong predictive capabilities.

Hyperparameter Tuning

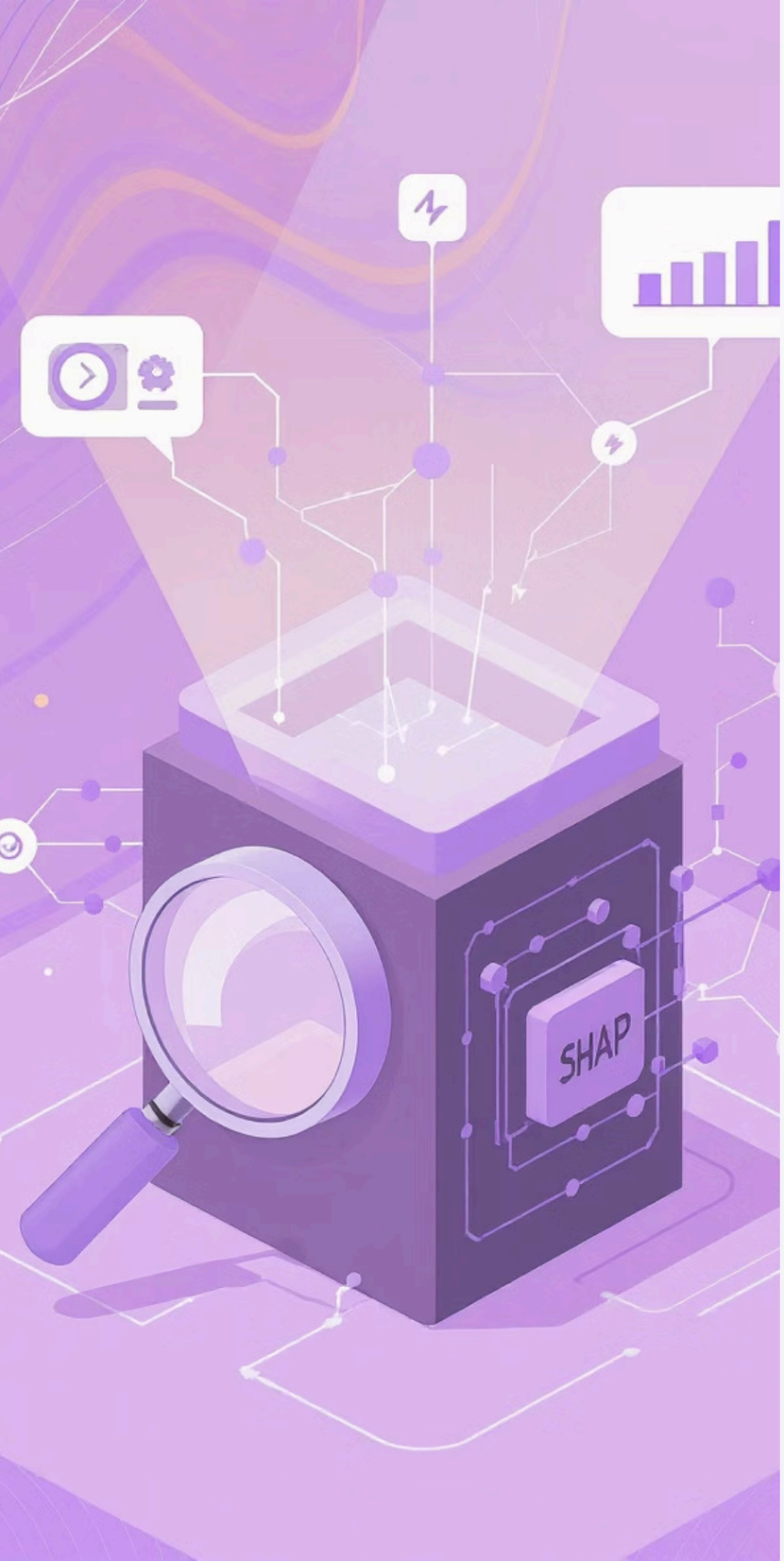
Optimized model parameters using techniques like GridSearchCV for best performance.

Model Evaluation: Quantifying Performance

Cross-validation and key metrics ensure reliable and generalizable models. 📊

Accuracy	0.78	0.81
Precision	0.62	0.68
Recall	0.55	0.61
F1-score	0.58	0.64

XGBoost demonstrated superior performance across key metrics, making it the preferred model for deployment.



Model Interpretability: Understanding the "Why"

Beyond predictions, understanding feature impact is crucial for actionable insights. 🤖

Feature Importance

Identified top drivers of churn: Contract type, Tenure, Monthly Charges, and Fiber Optic service.

SHAP Visualizations

Utilized SHAP (SHapley Additive exPlanations) for both global and local interpretations.

Global Explanations

Aggregated feature impacts across the entire dataset to show overall trends.

Local Explanations

Provided insights into why a specific customer is predicted to churn, enabling personalized interventions.

Interactive Demo: Customer Input Interface

Our user-friendly Streamlit application provides a seamless way for business users to input customer data and instantly predict churn risk. 🚀

Intuitive Form Design

An easy-to-use form ensures business stakeholders can effortlessly enter customer parameters without technical expertise.

Dropdown Menus for Categorical Data

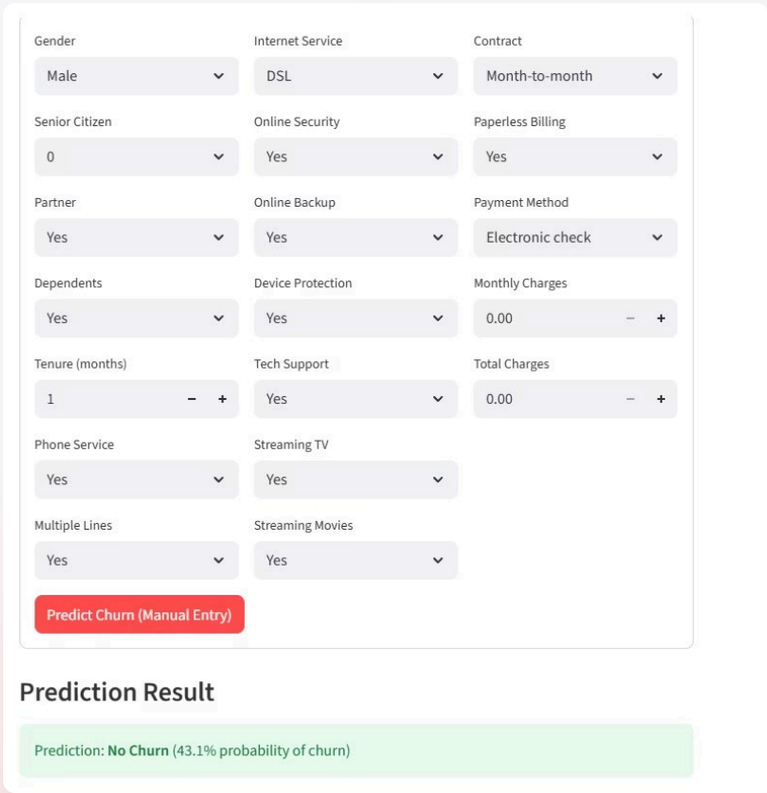
Effortlessly select values for categorical features like contract type, internet service, and payment method from clear dropdown lists.

Direct Numerical Inputs

Input specific values for key metrics such as Monthly Charges, Total Charges, and Tenure using dedicated numerical fields.

Simple 'Predict' Button

Initiate churn risk assessment with a single click, providing immediate results for individual customer scenarios.



Results & Business Impact

Our predictive model delivers strong performance, translating directly into tangible business value.

Model Results: Robust & Reliable

High Accuracy

Achieved high predictive accuracy with Random Forest & XGBoost, ensuring reliable churn identification.

Balanced Performance

Balanced class performance using SMOTE, effectively identifying both churners and non-churners.

Actionable Insights

Reliable interpretability with SHAP to understand the "why" behind each prediction.

Business Impact: Value & Growth

Proactive Retention

Implement targeted retention strategies to proactively engage at-risk customers.

Reduced Costs

Lower customer acquisition costs by retaining existing customers more effectively.

Stable Growth

Foster stable revenue growth and enhance brand reputation through improved customer loyalty.

Conclusion & Next Steps

This project successfully developed and deployed a robust churn prediction solution, providing significant value for strategic customer retention.

Project Conclusion

End-to-End Pipeline

Built a comprehensive churn prediction pipeline, from raw data to actionable insights.

Accurate & Interpretable Models

Delivered highly accurate, interpretable, and business-ready machine learning models.

Actionable Insights

Provided clear, actionable insights to guide targeted customer retention strategies.

Future Direction & Next Steps

dd

Deploy Streamlit App

Roll out the interactive Streamlit application to production for broader business user access.

∫

Integrate with CRM

Seamlessly integrate churn predictions into existing CRM systems for proactive interventions.

dd

Explore Deep Learning

Investigate advanced deep learning models to potentially enhance prediction accuracy further.

Monitor & Retrain

Establish a continuous monitoring and retraining process with new data to maintain model performance.