BY: KAUSHIK KARMAKAR

CHURN PREDICTION SYSTEM

Problem statement:

To develop a Churn Prediction System that identifies which customers are likely to stop using a service. This kind of model is crucial in industries like telecom, SaaS and banking where retaining customers is more profitable than acquiring new ones.

Here, we will use a real customer data to build, train and evaluate a machine learning model and then present the findings with an analysis dashboard that a business decision maker could act on.

Objective:

1) **Predict which customers are likely to churn.**
2) **Quantify churn risk using probability scores.**
3) **Enable data driven retention decisions.**

Key Outcomes:

1) **Built and evaluated multiple machine learning models to predict churn.**
2) **Selected a high-performing model based on ROC-AUC and recall.**
3) **Segmented customers by churn risk and quantified revenue at risk.**
4) **Developed an interactive Streamlit dashboard for decision support.**

DATASET OVERVIEW

Source: Kaggle's Telco Customer Churn dataset

Features: customerID, gender, *SeniorCitizen, Partner,* Dependents, *tenure,* PhoneService, *MultipleLines,* InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, *TechSupport, StreamingTV,* StreamingMovies, *Contract,* PaperlessBilling, *PaymentMethod,* MonthlyCharges, TotalCharges

Target variable: Churn

Shape of the dataset: There are 7043 rows and 21 columns.

Link: Telco Customer Churn

DATA PREPARATION

1) Converted numerical fields previously having datatype object. (TotalCharges)
2) Converted categorical field which was previously assigned as int64 for encoding (SeniorCitizen).
3) Removed features which are not used for prediction in model (customerID).
4) Preformed exploratory data analysis (EDA) and identified churn pattern.
5) Encoded all categorical fields with the help of a Label Encoder so that data could be fitted to the classification models.

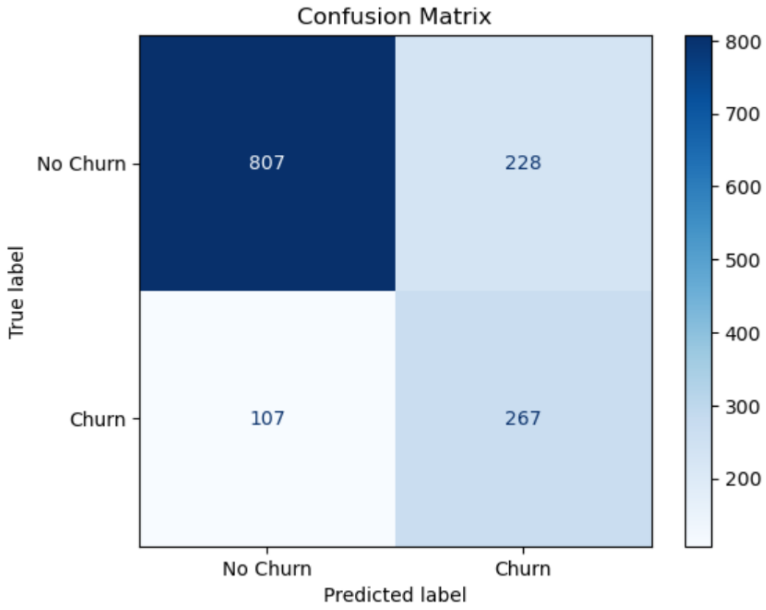
MODELLING APPROACH

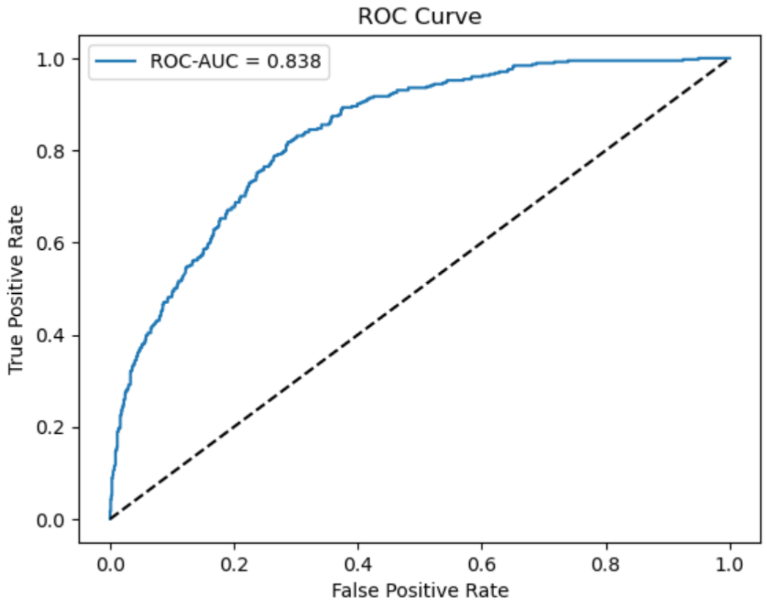
1) Stratified and preformed train test split to handle class distribution in the train and test splits.
2) Class imbalance is handled using smote inside cross validation pipelines
3) Model training compared among classification models: XGBoost Classifier, Random Forest Classifier, Support Vector Classifier, Logistic Regression and Decision Tree.
4) Models are been passed through RandomizedSearchCV and best model is selected as per their cross validation ROC-AUC

MODEL PERFORMANCE

The final model achieved has:

1) Strong ROC-AUC Score
2) Improved recall for churned customer.

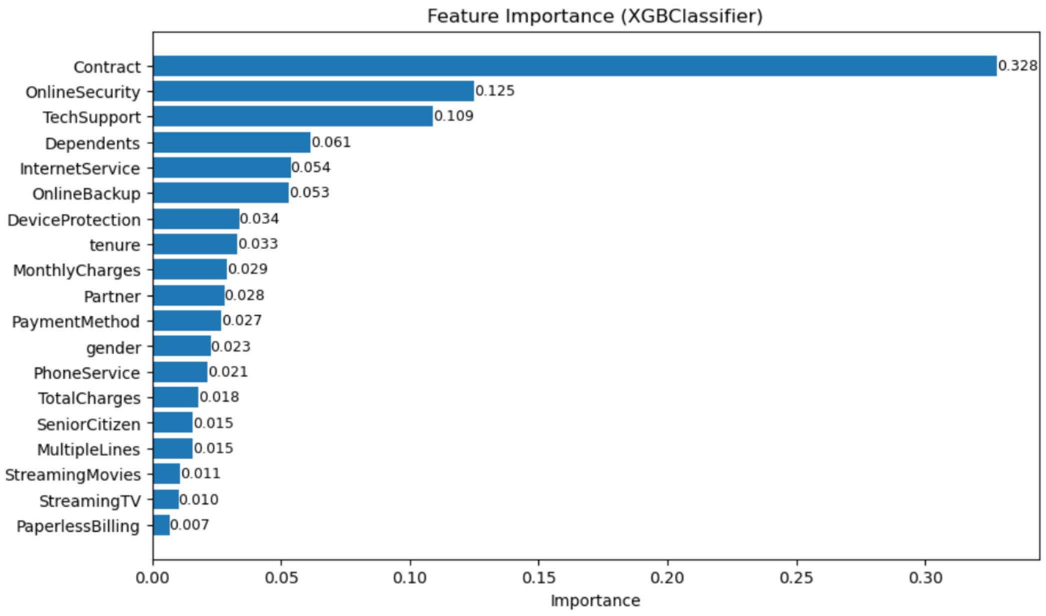




KEY CHURN DRIVES

Feature importance analysis revealed that churn is strongly influenced by:

1) Contract
2) Online Security Availability
3) Tech Support Availability
4) Dependents
5) Internet Service Used

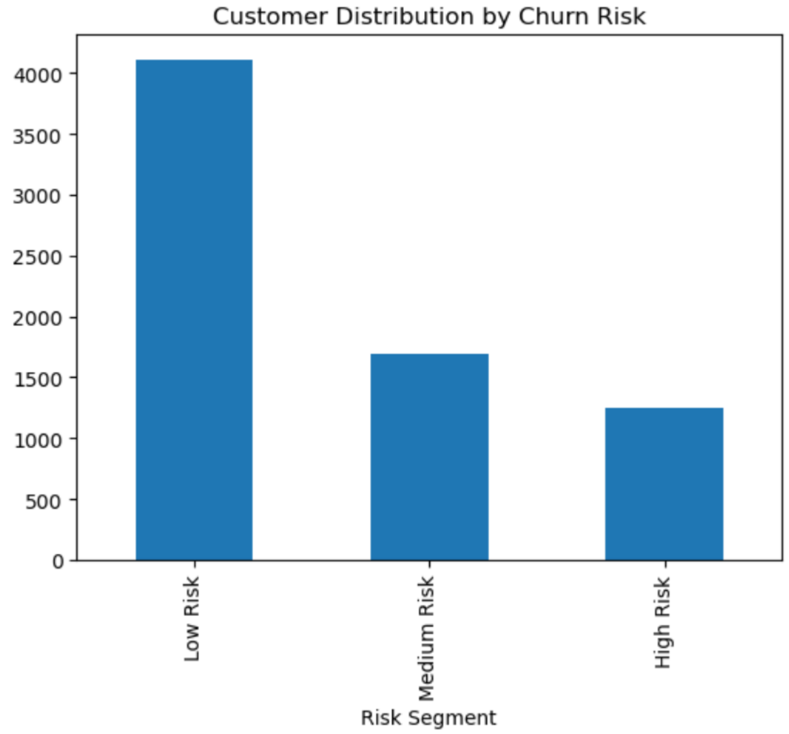


BUSINESS IMPACT AND RETENTION STRATEGY

Churn Risk Segment:

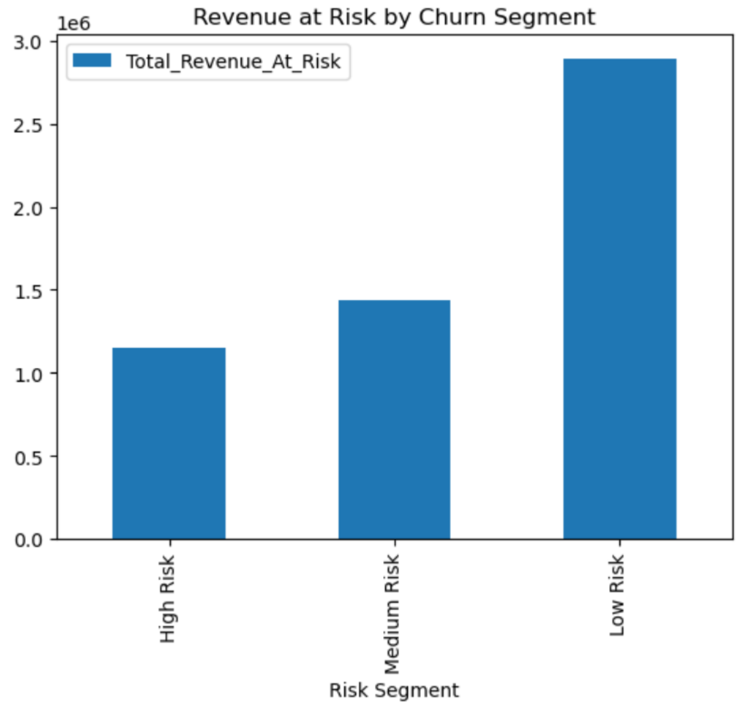
Customers were ranked by predicted churn probability and segmented into:

1) High Risk: > 0.7
2) Medium Risk: 0.4 to 0.7
3) Low Risk: < 0.4



REVENUE AT RISK

1) Monthly subscriptions when magnified for an annual year were used to find revenue at risk.
2) A larger number of high-risk customers represents disproportionality in large share of revenue at risk.



RETENTION SIMULATION

A stimulation has been performed which shows amount that can be recovered. Here we took reference of targeted top 10% high risk customers with 30% retention success.

A targeted retention campaign could potentially recover approximately ₹ 197067.78 in monthly revenue, based on churn risk ranking and conservative assumptions.

Recommendations

1) Prioritize retention efforts on high-risk customers.
2) Focus on contract upgrades and loyalty incentives.
3) Use churn probabilities to optimize marketing spend.

Assumptions and Limitations:

1) Revenue estimated using MonthlyCharges may not be accurate.
2) Retention success rate is assumed 30% which is not necessary in practical case.
3) Model performance depends on historical data.

Conclusion

This project demonstrates an end-to-end churn prediction system trained on publicly available Kaggle's Telco Customer Churn dataset that goes beyond model accuracy to deliver actionable business insights. By combining machine learning, expandability, and business impact analysis, the system support informed retention decision making.

THANK YOU