Predicting Customer Churn for Business Growth

(Proyect-TelecomX-1-ETL-Latam-Cynthia_-Challenge Telecom X análisis de evasión de clientes - Parte 2)

Agenda

Project Overview

01

Analytical Summary

Overview of churn prediction pipeline

02

Project Outline

Steps taken in the analysis

03

Model Comparison

Performance evaluation of ML models

04

Key Factors

Drivers of customer churn

05

Patterns Identified

High-risk and protected segments

06

Strategies to Reduce Churn

Actionable retention plans

07

Conclusion

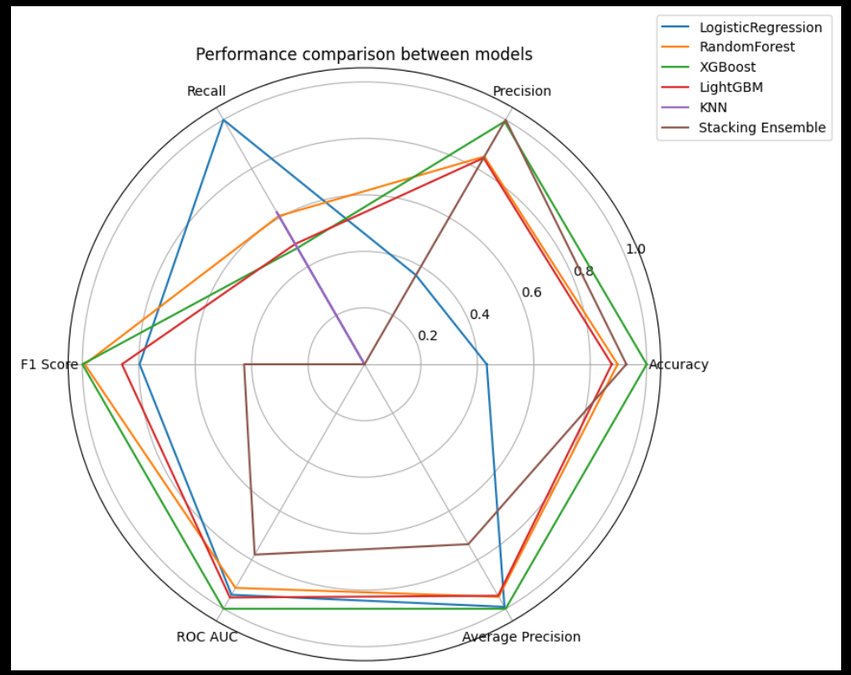
Summary of findings and next steps

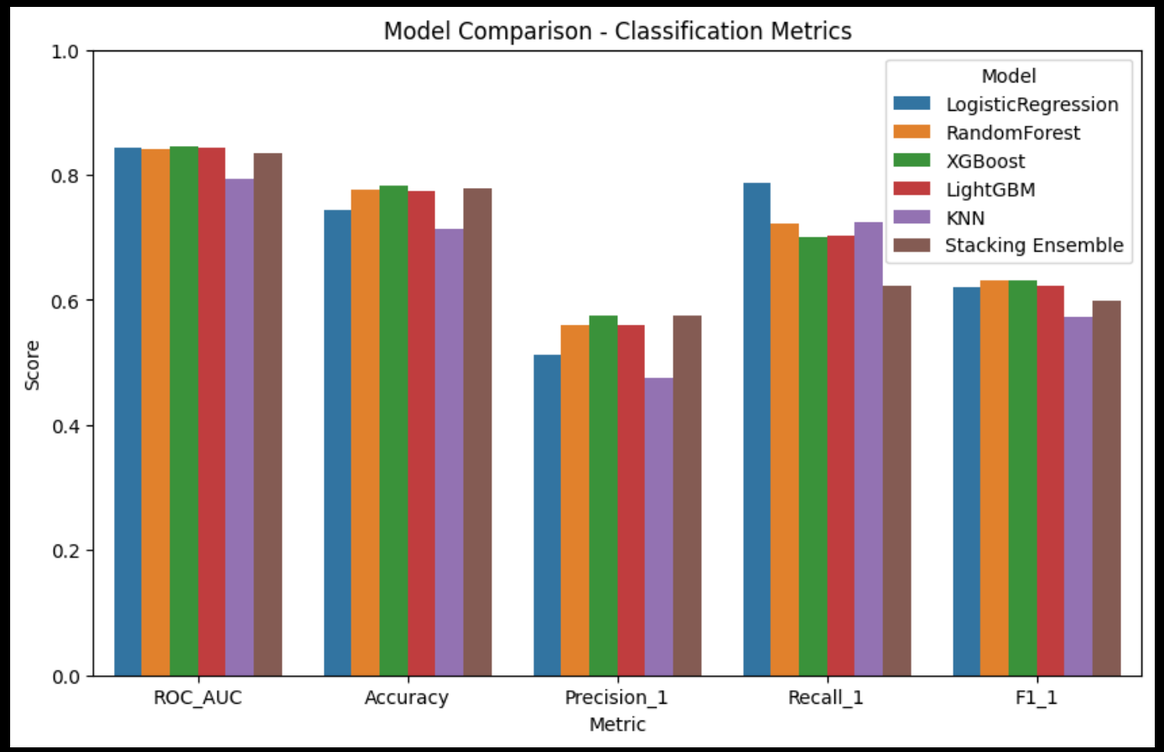
Section 1

Analytical Summary

A robust pipeline was built to anticipate customer churn using machine learning techniques and explanatory analytics.

The XGBoost model performed best, standing out for its robustness, balance, and ability to correctly identify the positive class (churn).





Section 2

Project Outline: Our Methodical Approach

Our analysis followed a comprehensive, multi-stage process to ensure accuracy and interpretability:

1. **Exploratory Analysis & Data Preparation**
   Initial data understanding and cleaning.
2. **Correlation Analysis & Balancing**
   Identifying relationships and addressing data imbalance.
3. **Multi-Model Training & Optimization**
   KNN, LG RF, RL, XGB, and Ensemble models with hyperparameter tuning (GridSearchCV).
4. **Model Comparison & Interpretation**
   Evaluating performance and understanding feature contributions (SHAP & Feature Importance).
5. **Final Report & Strategy**
   Consolidating insights into actionable recommendations.

Section 3

Model Comparison: Performance Metrics

After rigorous evaluation, the XGBoost model demonstrated superior overall balance and predictive power.

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Observations
XGBoost	0.783	0.575	0.701	0.631	**0.845**	Good overall balance, excellent ROC AUC, and high recall
LightGBM	0.774	0.560	0.703	0.623	0.843	Very similar to XGBoost, good recall and precision
RandomForest	0.776	0.560	0.722	0.631	0.841	High recall, good overall balance
LogisticRegression	0.744	0.511	0.786	0.620	0.843	Better recall (sensitivity), lower precision
Stacking Ensemble	0.779	0.578	0.623	0.600	0.834	Better precision, slightly lower recall and F1
KNN	0.714	0.475	0.725	0.574	0.794	Lower accuracy and precision, but acceptable recall

Section 3

Confusion Matrix Insights

Analysis of the Confusion Matrix confirms XGBoost's effectiveness in balancing churn identification and false alarms.

Model	TN	FP	FN	TP	Observations
XGBoost	**841**	**194**	112	262	Better TN (low FP), FN similar to LightGBM
LightGBM	828	207	111	263	Good Balance, somewhat more FN than KNN, but less FP
RandomForest	823	212	104	270	Balanced, FN slightly better than XGBoost
LogisticRegression	754	281	**80**	294	Better FN (fewer false negatives), more FP (risk of higher costs)
Ensemble (Stacking)	865	**170**	141	233	Better TN (less FP), more FN (risk of losing churning customers)
KNN	735	300	103	**271**	Good Recall (high TP), many False Positives (high FP)

> 🎯 Selected Model: XGBoost. Robust, well-balanced, and ideal for minimizing customer loss (high recall) while avoiding false alarms (reasonable accuracy). Adapts well to various business cases.

Section 4

Key Churn Factors

We utilized SHAP and correlation analysis to identify variables influencing churn probability.

Factors Increasing Churn:

- **Contract Month to Month:** Low commitment, easy to leave.
- **Internet Fiber Optic:** Customers with more alternatives.
- **Payment Electronic Check:** Volatile profile, lower contractual commitment.
- **Paperless Billing:** Digital customer, often attentive to new offers.

Factors Decreasing Churn:

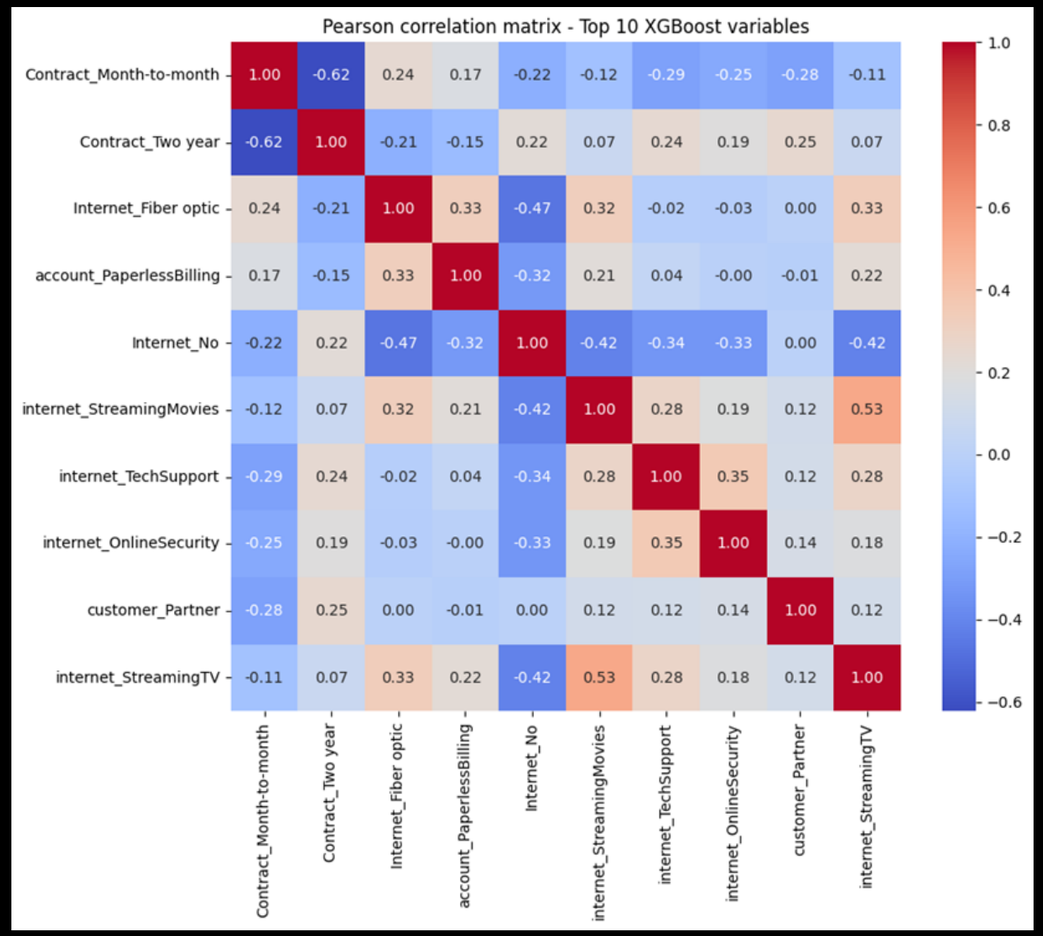
- **Customer Tenure:** Longer tenure, greater loyalty.
- **Two-year Contract:** Long contracts limit exits.
- **Internet TechSupport:** Technical support adds value, reduces churn.
- **Internet OnlineSecurity:** Additional protections build trust.
- **No Internet:** Less incentive to leave.

Mixed Effects:

Internet StreamingMovies & StreamingTV: Effect depends on other variables (e.g., contract type, additional services).

Section 5

Identified Churn Patterns



High-Risk Segment

- Monthly contracts, low tenure.
- Fiber optics, e-check payments, e-invoicing.
- **Profile:** Digital-savvy, short-term commitment, potentially flighty.

Protected Segment

- Two-year contracts, long tenure.
- Utilize additional services (tech support, online security).
- **Profile:** Loyal, value-added service-oriented, stable.

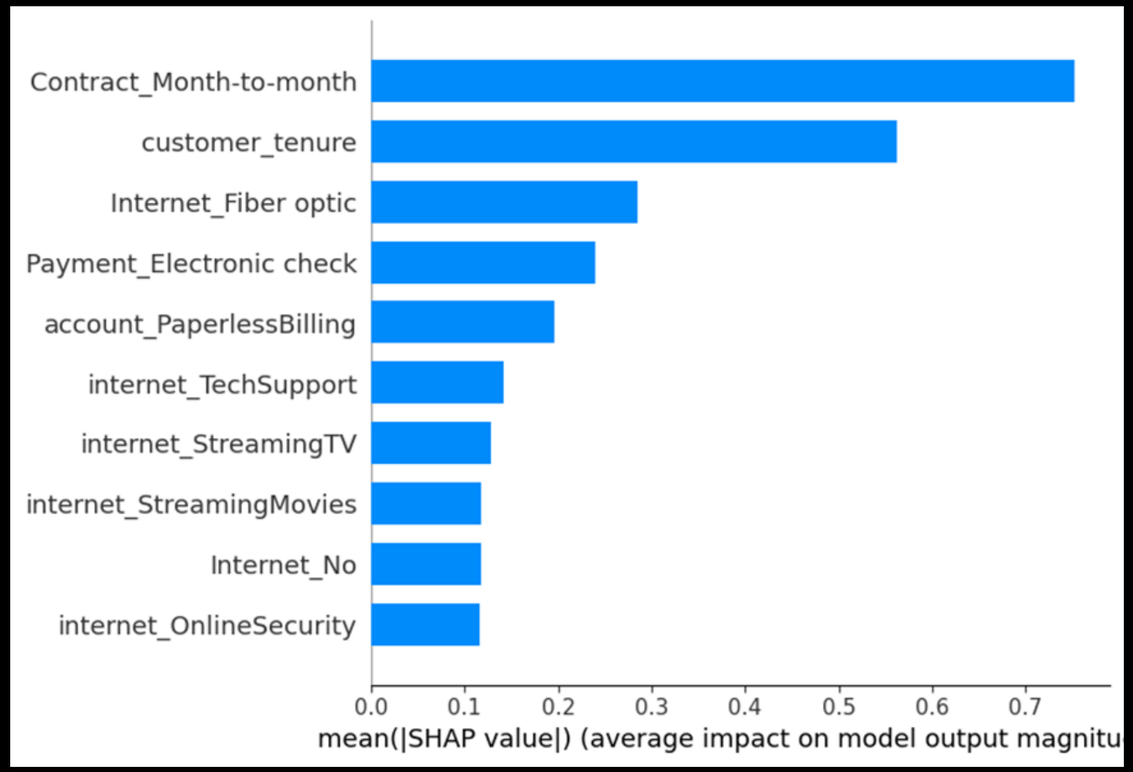
Key Interactions: Long contract + additional services maximize retention. New customer + monthly contract + digital services = highest risk.

Section 6

Actionable Churn Reduction Strategies

High-Retention Customers

- Enhance technical support & security.
- Offer upgrades and bundled packages.
- Proactive loyalty programs.



High-Risk Customers

- Incentivize longer contracts with benefits.
- Improve fiber service experience & promotions.
- Promote automatic payment methods.
- Personalize digital communications.

General Strategies

- Improve onboarding for initial engagement.
- Use predictive models for early detection.
- Optimize digital self-service and support.
- Personalize streaming offers based on profile and bundle with other services.

Conclusion

Driving Loyalty, Reducing Churn

- ## Robust Pipeline

  Advanced preprocessing, variable selection, and class balancing significantly improved predictive capacity.

- ## Key Factors Validated

  Contract type, payment method, customer tenure, and additional services are primary churn influences.

- ## Interpretable Insights

  Combined SHAP with non-linear models for clear, explainable churn drivers.

- ## Actionable Strategies

  Derived strategies are clear, actionable, and ready for immediate implementation by marketing and retention teams.

This combined approach provides a solid foundation for designing customized interventions to increase loyalty and reduce customer churn.