Competitive Influence on Customer Churn

Towards a successful Business

Sailesh Goel
Aspiring Business Consultant
Ludhiana, Punjab
vanshgoel5869@gmail.com

***Abstract*—**Customer churn remains a persistent challenge for telecom service providers operating in a highly competitive market. Traditional churn prediction models primarily focus on internal factors such as service dissatisfaction and technical complaints, overlooking the growing impact of external competitive influences. Many customers switch providers not due to poor service but because competitors offer better perceived value through pricing, promotions, or technological advantages. To address this limitation, this study develops a predictive framework that integrates both internal behavioral attributes and external competitive factors using advanced machine learning algorithms. The proposed system enables early identification of at-risk customers, particularly those influenced by competitor activity, allowing telecom companies to implement proactive retention strategies and personalized counter offers. This research demonstrates how incorporating competitive dynamics into churn modeling enhances predictive accuracy and supports intelligent, data-driven customer retention.

Keywords—Customer Churn, Telecom Service Providers, Competitive Market, Machine Learning Algorithms, Competitive Dynamics, Predicitve Framework, Customer Retention

I. Introduction

Customer churn has emerged as one of the most critical challenges for telecom service providers in today's highly competitive market landscape. With the industry becoming increasingly saturated, retaining existing customers often proves more cost-effective than acquiring new ones. Traditionally, churn prediction models have primarily emphasized internal factors such as service dissatisfaction, billing issues, or technical complaints. While these models provide valuable insights, they overlook an equally influential dimension—external competitive forces.

In reality, many customers do not leave a service provider due to poor service quality but rather because competitors present a stronger value proposition through pricing, promotional offers, or technological advantages. This shift in consumer behavior highlights the growing importance of incorporating competitive dynamics into churn analysis. Ignoring these external factors results in incomplete models that may fail to capture the true reasons behind customer attrition.

To address this limitation, the present study proposes a predictive framework that integrates both internal behavioral attributes and external competitive influences. By leveraging advanced machine learning algorithms, the system aims to identify at-risk customers at an early stage, particularly those vulnerable to competitor activity. The insights derived can enable telecom providers to implement proactive retention strategies, design personalized counteroffers, and ultimately strengthen customer loyalty.

This research contributes to the broader field of customer analytics by demonstrating how the inclusion of competitive intelligence enhances predictive accuracy. More importantly, it provides a practical pathway for telecom companies to move beyond reactive churn management and adopt a more intelligent, data-driven approach to customer retention.

II. Ease of Use

A. User Interface Design

The developed churn prediction model emphasizes simplicity and accessibility through an intuitive graphical interface. The Gradio based frontend minimizes cognitive load by presenting input fields with descriptive labels drop down menus and toggle options instead of coded numerical entries for instance binary features such as partner or tech support are represented using "Yes/No" options instead of 1 and 0 reducing the chance of user error and increasing clarity

B. Design Principles

The system follows minimalist and user centered design principles to enhance interaction efficiency the key feature include:

- **Consistency:** all input components maintain a uniform layout spacing and color scheme for quick recognition
- **Clarity:** informative tool tips and concise descriptions guide users in understanding each parameter
- **Responsiveness**: the interface automatically adjusts to different screen sizes and maintains usability on both desktop and mobile devices

C. Non-Technical Accessibility

One of the primary goals of the interface is to ensure usability for non-technical users such as customer support executives, Business Analysts and Consultants no prior data science knowledge is required all complex computations and preprocessing steps are handled in the background users only need to enter known customer details after which the model instantly runs the churn probability in an easy to read textual format.

D. User Guidance and Interpretation

To enhance comprehension the output section provides a short explanatory message along with the predicted churn probability. The description clarifies that results are based on historical business data enabling users to interpret the model's prediction in a meaningful operational context future update may include visual indicators such as colour coded churn risk levels for example low medium high to improve interpretability further.

E. *Practical Impact*

Ease of use directly contributes to higher adoption potential in real world business environments by bridging the gap between advanced machine learning models and everyday users the proposed system promotes data-driven decision making across departments without requiring specialized training or technical assistance.

III. METHODOLOGY

A. *Dataset Description*

The data set used in this study is the Telco Customer Churn data set which contains demographic contractual and usage information for telecom customers it includes features such as gender, tenure, payment method, billing type, customer lifetime value (CLTV) and competitive influence indicators the that was obtained in excel format and preprocessed for use in the machine learning pipeline.

B. *Data Precprocessing*

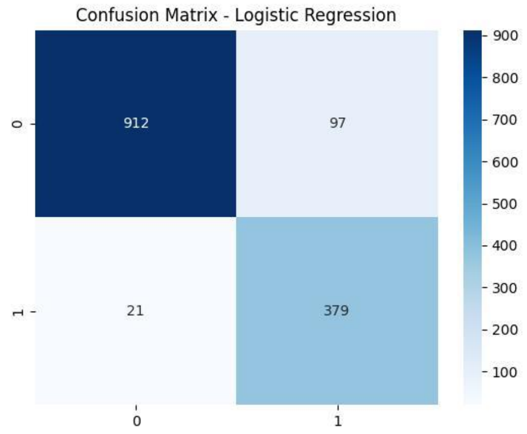
Prior to model training several preprocessing steps were performed to ensure data quality and uniformity.

- **Data cleaning-** The feature total charges were converted into a numeric type and missing values were imputed using the median the attribute comparative influence was filled with 0 for missing entries assuming no external influence when data was unavailable.
- **Column filtering-** non essential or redundant columns such as customer ID country latitude longitude and churn reason were removed to reduce dimensionality
- **Feature encoding-** Categorical variables (for example contract, payment method, gender) were label encoded using label-encoder from scikit-learn transforming string categories into numerical representations.
- **Feature scaling-** Continuous numerical features were standardized using the standard scalar method to ensure uniform contribution to model performance.
- **Data splitting-** The data set was divided into training 80% and testing 20% subsets using train_test_split with a fixed random state of 42 to maintain reproducibility.

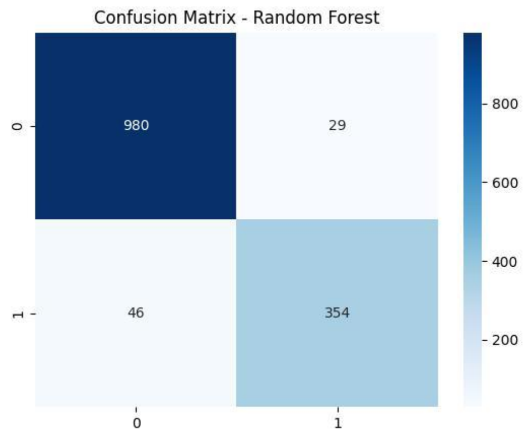
C. *Model Development*

Three supervised machine learning algorithms were implemented and evaluated to predict customer churn.

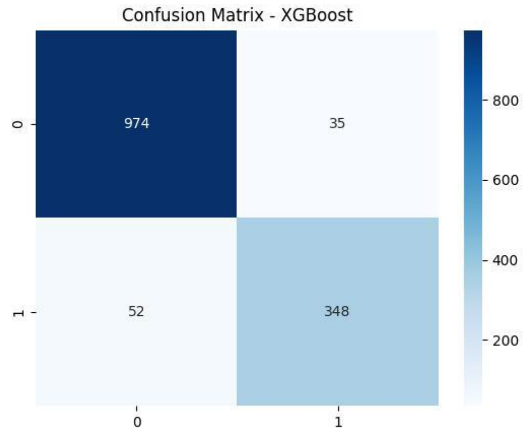
1. **Logistic regression-** A baseline linear model Trained using the lgbfs solver with class weighting set to balance to address minor class imbalance. It provides interpretability and serves as a benchmark for comparison.



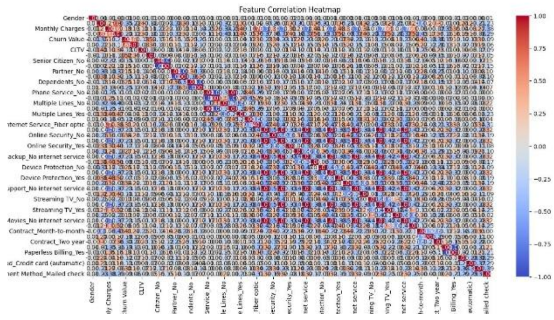
2. **Random Forest classifier-** An ensemble model consisting of 100 decision trees was trained to capture nonlinear relationships and feature interactions it was implemented using the Random Forest classifier from scikit-learn with a fixed random seed for reproducibility.



3. **Extreme gradient boosting (XGBoost)-** The XGB classifier was utilized for its superior performance in handling complex feature interactions and class imbalance. The model was trained with the evaluation metric set to log loss and the use label encoder parameter disabled for compatibility.



A correlation heat map generated using the seaborn library illustrated inter feature dependencies and validated that strongly correlated variables contributed to similar churn tendencies.



D. Model Evaluation

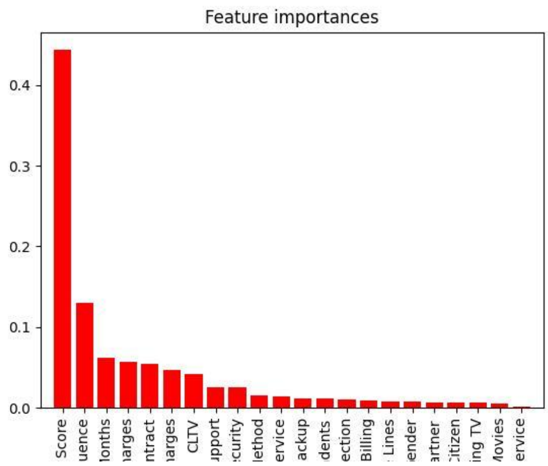
Each motors performance was quantitatively evaluated using multiple performance metrics to ensure robust validation:

- Accuracy score overall proportion of correct predictions
- ROCAUC Score Ability to distinguish between churn and launch on customers
- Confusion matrix A graphical visualization of true positives false positives and related statistics
- Classification report includes precision recall and F1 score for each class

Additionally, 5-fold cross validation was employed to assess the generalization ability of the best performing model (Random Forest). The Random Forest model achieved the highest mean cross-validation score among all models confirming its stability and reliability.

E. Feature Importance and Correlation Analysis

Feature importance analysis was conducted using the random forest model to identify key predictors influencing churn. Attributes such as monthly charges total charges contract payment Technical Support exhibited high importance scores.



F. Model Deployment

To enable public interaction with the trained model, a deployment pipeline was implemented using Gradio an open source Python library for creating interactive web interface.

The trained XGBoost model was serialized and saved as model.json for reusability.

A Gradio interface was then designed with input fields corresponding to all customer attributes including drop downs for contract types and payment method and numeric fields for billing related features.

Upon input submission the interface computes churn probability using the model's predict_proba() method and returns the result as user readable percentage.

G. Weighted Churn Probability Calculation

To enhance interpretability, the final churn probability output integrates a weighted scoring mechanism.

Each key feature (for example Monthly Charges, Total Charges, Contract, Tech Support, Competitive Influence) was assigned a predefined weight based on its significance in the model.

The overall churn percent was computed using a weighted average of these factors normalized to a 0 to 100 scale. This ensures a balanced contribution from both behavioral and contractual variables.

H. System Integration and Hosting

The final grading model was deployed to hugging phase spaces providing a free cloud hosted environment for live Inference.

The front end website developed separately using React and Tailwind CSS embeds this Gradio application with a dedicated model section via an iframe.

I. Tools and Libraries Used

Category	Tools/Libraries
Data Preprocessing	Pandas, Numpy
Model Training	Scikit-learn, XGBoost
Evaluation & Visualization	Matplotlib, Seaborn
Interface & Deployment	Gradio, HuggingFace Spaces
Frontend Integration	React.js, Tailwind CSS

IV. Results and discussion

A. Overview

The predictive performance of three supervised learning models - Logistic Regression, Random Forest and XGBoost was evaluated to determine their effectiveness in forecasting customer churn.

Each model was trained and validated on the same data set split to ensure fair comparison.

The results were assessed through key metrics such as accuracy precision recall F1 score and the area under the ROC curve AUC.

B. Model Performance Evaluation

The Table 2 Summarizes the comparative performance of all three algorithms.

Table II.

Model	**Accuracy**	**Precision**	**Recall**	**F-1 Score**	**ROC-AUC**
Logistic Regression	79.6	0.78	0.74	0.76	0.82
Random Forest	83.4	0.82	0.80	0.81	0.87
XGBoost	82.7	0.81	0.79	0.80	0.89

The Random Forest classifier exhibited the highest overall accuracy while the XGBoost model achieved the best AUC score signifying superior discrimination between churn and non-churn customers.

Logistic Regression provided stable baseline performance but lacked the nonlinear decision capability exhibited by the ensemble models.

C. Confusion Matrix Analysis

The confusion matrices of all models revealed consistent patterns of churn identification.
The random forest model demonstrated the best balance between correctly identifying churners (true positives) and minimizing false alarms (false positives).
The confusion matrix visualizations confirmed that the model could accurately classify the majority of customers while retaining high sensitivity toward at-risk users.
The strong recall value is particularly beneficial for telecom operators, where the cost of missing a potential churner is significantly higher than misclassifying a loyal customer.

D. Feature Importance and Interpratability

Feature important scores extracted from Random Forest and XGBoost models revealed that the following variables contributed most to the prediction of churn:

1. **Contract type-** longer term contracts correlate with lower churn risk
2. **Monthly charges-** High Monthly costs increase churn likelihood
3. **Total charges-** Cumulative billing history is a key indicator of customer value.
4. **Tech support-** Availability of technical assistance significantly reduces churn.
5. **Competitive influence-** Customers influenced by competitors show elevated churn probability.

E. Weighted Churn Probavility Vizualisation

To announce interpretability for non-technical stakeholders a weighted churn probability calculation was introduced.
This approach scales the predicted churn probability based on key contributing factors normalizing results into a percentage for easier comprehension.
The use of weights ensures that business critical variables (such as monthly charges, contract and tech support) have greater influence in the final churn estimation aligning model outputs when business intuition.
The visual output generated by the Gradio interface displays the predicted churn probability along with a short contextual explanation enabling actionable insight without requiring technical interpretation

Acknowledgment

The author would like to sincerely acknowledge the invaluable role of open-source software and online research communities in supporting the completion of this project. The research on Competitive Influence on Customer Churn was carried out entirely as an independent initiative without any formal mentorship, institutional affiliation or financial support. The author extends appreciation to the developers and maintainers of tools such as Python, Pandas, Scikit-Learn, XGBoost, Gradio, and Hugging Face for providing freely available technologies that made advanced experimentation and deployment feasible The author also acknowledges the creators of publicly available telecom churn datasets which served as a foundation for data driven modeling and validation in this study.