

Predicting Telecom Customer Churn and Translating Risk into Retention Action: Evidence from the Orange Telecom Dataset

Jiayi Wan

School of Mathematics, The University of Edinburgh, Edinburgh, UK

Email: jwanstudio05@gmail.com

Manuscript received August 2, 2026; accepted October 9, 2026; published October 27, 2026.

Abstract—Telecom churn models can identify customers who are likely to leave, but their predictions do not always indicate which customers should receive retention efforts first. This study investigates whether churn predictions can be used to distinguish customers with different retention priorities. Using the Orange Telecom Customer Churn Dataset, logistic regression and random forest models with balanced class weights were compared on the test set based on accuracy, precision, recall, F1-Score and ROC-AUC. Among the examined variables, higher customer service calls, daytime usage and international-plan subscription were positively associated with churn, whereas voice-mail-plan subscription was negatively associated. Random forest achieved higher precision, F1-Score and ROC-AUC than logistic regression while maintaining similar recall, mainly due to fewer false-positive predictions. The predicted probabilities separated customers into four risk tiers, with observed churn increasing from 2.4% and 3.2% in the Low- and Medium-risk tiers to 37.8% and 98.5% in the High- and Critical-risk tiers. High-risk customers had more customer service calls on average, whereas Critical-risk customers had a higher proportion of international-plan subscribers. These findings suggest that combining churn probabilities with customer characteristics helps translate predicted churn risk into more targeted retention actions.

Keywords—customer churn, telecom industry, machine learning, customer retention, random forest

I. INTRODUCTION

Customer churn refers to customers discontinuing their relationship with a company's products or services. In the telecommunications industry, such churn threatens revenue stability and long-term profitability because relatively low switching barriers enable customers to change providers in response to differences in pricing, service quality, and changing usage needs. Retaining existing customers is generally more cost-effective than acquiring new ones, making churn prediction a key component of customer relationship management.

This study aims to predict telecom customer churn and examine how prediction results can support enterprise decision-making. Specifically, it examines customer characteristics associated with telecom churn and compares the predictive performance of different machine learning models for identifying customers at risk of churn. Finally, it explores how predicted churn risk can be translated into targeted customer retention strategies.

Existing research often treats churn prediction as the end point of analysis, rather than examining how prediction outputs can support targeted customer retention strategies.

This study addresses this gap by combining predictive modelling with probability-based customer risk profiling to support differentiated customer retention decisions.

The remainder of this paper is organized as follows. It first reviews the literature on telecom customer churn, predictive modelling, and customer retention to establish the research context and identify the research gap. The methodology describes the dataset, data preprocessing, feature engineering, model development, and evaluation procedures. The empirical findings are presented based on the three research questions and discussed in relation to previous studies, with implications for churn management. The paper concludes by summarizing the main findings, limitations, and directions for future research.

II. LITERATURE REVIEW

Customer churn has long been treated as both a predictive analytics problem and a customer relationship management challenge in the telecommunications industry.

Customer usage behavior is an important source of information for churn prediction (Wei & Chiu, 2002), particularly in terms of usage intensity, service experience, and subscription arrangements. High levels of usage are commonly interpreted as a favorable sign of strong customer attachment. However, Ahn *et al.* (2006) argue that under usage-based pricing, longer call durations may lead to higher bills or excess charges. These additional costs may increase churn risk when customers perceive insufficient value. Customer service interactions represent another important factor influencing customer loyalty. Ahn *et al.* (2006) note that repeated complaints increase the risk of customer churn, while Keramati and Ardabili (2011) further suggest that even a single complaint may substantially raise the likelihood of departure. Gerpott *et al.* (2001) argue that higher overall satisfaction strengthens customer loyalty, which in turn reduces the probability of churn, while Bolton (1998) adds that early service experiences are therefore especially important. These findings suggest that identifying dissatisfaction before customers leave is important for improving customer retention. Customer service calls reflect not only contact frequency but also how effectively operators resolve customer concerns. Hemsathapat and Dagli (1999) identified high international call usage, international-plan subscription, and frequent customer service calls as characteristics associated with higher churn risk. Wong (2011) argues that the decisive issue may not be whether a customer subscribes to a particular plan, but

whether that plan aligns with the customer's actual usage patterns and payment needs.

Based on these churn-related factors, previous studies have applied various statistical and machine learning methods for prediction. Logistic regression is commonly used as the benchmark approach for churn prediction because of its interpretability and probabilistic outputs (Neslin *et al.*, 2006), although it is less capable of modelling nonlinear relationships. Tree-based ensemble methods such as random forest and XGBoost have been therefore increasingly adopted to overcome these limitations. Many studies agree that ensemble learning demonstrates superior predictive performance compared with traditional statistical models, particularly for complex customer behavior (Burez & Van den Poel, 2009; Ahmad *et al.*, 2019).

Class imbalance presents an additional methodological problem as churners normally account for only a small proportion of customers, relying solely on classification accuracy may produce misleading conclusions. Verbeke *et al.* (2012) recommend evaluating models using precision, recall, F1-Score and ROC-AUC to obtain a more comprehensive assessment of predictive performance.

However, prediction accuracy alone may not be sufficient for retention decisions. Verbeke *et al.* (2012) show that selecting churn models solely through statistical performance measures may produce suboptimal retention outcomes because intervention costs, response rates and customer value are not incorporated into conventional model evaluation. This shifted attention from predicting churn towards determining which customers should receive retention interventions. Wu *et al.* (2021) extend this logic by integrating churn prediction with customer segmentation, allowing model outputs to support more differentiated churn management.

Although previous studies have demonstrated strong predictive performance, relatively limited attention has been paid to how these predictions can inform practical customer retention strategies. Existing research often stops at developing accurate prediction models rather than translating churn probabilities into differentiated retention strategies. In addition, churners are commonly treated as a homogeneous group despite the possibility that different service- and price-related pathways may lead to churn.

III. METHODOLOGY

A. Dataset and Sample

This study used the cleaned version of the publicly available Orange Telecom Customer Churn Dataset obtained from Kaggle. The dataset comprised predefined training and test sets containing 2666 and 667 customer records, respectively. Each record contained 19 predictor variables describing customer characteristics, service subscriptions, call usage, and customer service interactions. Churn was defined as a binary outcome indicating whether a customer had left the service provider. The training sample contained 14.6% churners, indicating an imbalanced class distribution. Model performance was evaluated on the predefined test set.

B. Data Preprocessing

All preprocessing and variable-selection decisions were based exclusively on the training data. International plan and Voice mail plan were recoded as binary indicators. Churn was converted to integer form.

Predictor redundancy was assessed before exploratory analysis. Within each call category, minutes and charge variables were nearly perfectly collinear ($r \approx 0.9999$). Charges were calculated as a fixed linear transformation of minutes. The four charge variables were therefore removed. Number of voicemail messages also showed substantial structural dependence on Voice mail plan, as all customers without the plan recorded zero messages. Therefore, the count variable was excluded in favor of the more interpretable subscription indicator.

Variables with limited analytical value were then removed. Area code showed no association with Churn ($p = 0.906$). State was also excluded despite its statistically significant association ($p = 0.0047$), because its 51 categories would substantially increase model dimensionality while offering limited value for individual-level retention interventions.

C. Exploratory Analysis and Feature Engineering

Exploratory analyses examined churn rates, Pearson correlations, and feature distributions across churn groups. Total day, evening, and night calls, together with account length, showed near-zero correlations with Churn and substantial distributional overlap between churners and non-churners. These variables were therefore excluded from subsequent modelling. In contrast, customer service calls exhibited a pronounced non-linear relationship with churn. Churn rates remained stable at approximately 10%–14% among customers making zero to three service calls but increased sharply thereafter, reaching 52.9% among customers making four or more calls, compared with 11.3% among those making fewer than four calls. Based on this threshold, a binary feature, `high_service_calls`, was created and coded as 1 for four or more calls and 0 otherwise, allowing logistic regression to capture the observed discontinuous increase in churn risk.

D. Model Development

Two supervised classification models, logistic regression and random forest, were developed using the same eight predictors: International plan, Voice mail plan, Total day minutes, Total evening minutes, Total night minutes, Total international minutes, Total international calls, and `high_service_calls`. For logistic regression, predictor values were standardized using parameters estimated exclusively from the training set and subsequently applied unchanged to the test set. Feature scaling was not applied to random forest because tree-based models are insensitive to differences in predictor scales. Given the imbalanced outcome distribution, both models were fitted using balanced class weights, assigning greater importance to observations from the minority churn class.

E. Model Evaluation and Risk Profiling

Model performance was evaluated on the held-out test set using accuracy, precision, recall, F1-Score, and ROC-AUC. To further interpret model behavior, logistic regression

coefficients and random forest feature-importance scores were compared to examine the consistency of key predictors across models. Based on the performance comparison, the better-performing model was used to generate customer-level churn probabilities, which were classified into four risk tiers: Low (<0.20), Medium ($0.20\text{--}0.40$), High ($0.40\text{--}0.70$), and Critical (≥ 0.70). As these thresholds were predefined rather than optimized through cost-sensitive analysis, the resulting risk tiers were evaluated by comparing observed churn rates across groups. Mean predictor values were then calculated within each tier to compare customer characteristics across risk groups.

IV. RESULTS

A. Customer Characteristics Associated with Churn

Customer service contact, daytime usage and subscription status were the characteristics most strongly associated with customer churn. Customer service calls showed a clear threshold pattern, with churn increasing from 11.3% among customers making fewer than four calls to 52.9% among those making four or more calls. Customers subscribing to an international plan had a substantially higher churn rate of 43.7%, compared with 11.3% for non-subscribers. Customers with a voice mail plan experienced a lower churn rate of 8.9%, compared with 16.7% for customers without the plan. Among continuous usage variables, total daytime minutes showed the strongest association with churn.

In the logistic regression model, `high_service_calls`, total daytime minutes and international-plan subscription were identified as positive predictors of churn, whereas voice-mail-plan subscription was negatively associated with churn. Similarly, the random forest model identified total daytime minutes and `high_service_calls` as the two most important predictive features.

B. Predictive Performance of Logistic Regression and Random Forest

Table 1 compares the out-of-sample performance of logistic regression and random forest on the held-out test set.

Table 1. Test-set performance of logistic regression and random forest

Metrics	Logistic Regression	Random forest	Difference
Accuracy	0.847	0.948	+0.101
Precision	0.479	0.812	+0.333
Recall	0.832	0.821	-0.011
F1-Score	0.608	0.817	+0.209
ROC-AUC	0.855	0.912	+0.057

Random forest outperformed logistic regression on four of the five evaluation metrics while maintaining a comparable level of recall. The largest improvement was observed in precision, which increased from 0.479 to 0.812 (+0.333). F1-Score and ROC-AUC also improved, indicating stronger overall predictive performance. Random forest was selected as final model to generate customer-level churn probabilities, and logistic regression was used as an interpretable benchmark.

Recall was the only metric for which logistic regression performed marginally better, identifying 83.2% of actual churners compared with 82.1% for random forest, a

difference of only 1.1 percentage points. Confusion matrix analysis showed that, although random forest correctly identified one fewer churning, it reduced the number of false-positive predictions from 86 to 18. The results therefore indicate that random forest substantially reduced false-positive classifications while maintaining a comparable recall.

C. Churn Risk Segmentation and Customer Profiles

Table 2 presents the distribution of customers and observed churn rates across the four predefined risk tiers generated from the random forest predicted probabilities.

Table 2. Customer distribution and observed churn rate by predicted risk tier

Risk Tier	Predicted Probability	Number of Customers	Observed Churn Rate (%)
Low	<0.20	495	2.4
Medium	$0.20\text{--}0.40$	62	3.2
High	$0.40\text{--}0.70$	45	37.8
Critical	>0.70	65	98.5

Observed churn rates were 2.4% and 3.2% in the Low and Medium tiers, respectively, before increasing to 37.8% in the High tier and 98.5% in the Critical tier. Observed churn rates increased monotonically from the Low to the Critical tier.

Table 3 compares the characteristics of customers in the High and Critical risk tiers.

Table 3. Comparison of high- and critical-risk customer profiles

Characteristic	Overall Average	High Risk	Critical Risk
Customer service calls	1.56	2.82	2.37
International-plan subscription (%)	7.9	11.0	25.0
Total daytime minutes	180.9	221.4	223.7

High- and Critical-risk customers showed similar daytime usage but differed in customer service calls and international-plan subscription. Customers in the High tier made more service calls than those in the Critical tier, averaging 2.82 calls compared with 2.37, while both exceeded the overall average of 1.56. In contrast, international-plan subscription was much more concentrated in the Critical tier, where 25.0% of customers subscribed, compared with 11.0% in the High tier and 7.9% overall. Mean daytime usage was similar in the two upper tiers, at 221.4 min for High-risk customers and 223.7 min for Critical-risk customers, compared with an overall mean of 180.9 min.

The two upper risk tiers represented distinct customer profiles. The High tier was characterized by more frequent customer service calls together with greater daytime usage. The Critical tier showed similarly high daytime usage but a substantially higher proportion of international-plan subscribers.

V. DISCUSSION

A. Interpreting the Characteristics Associated with Churn

Customer service experience was an important indicator of churn risk. The sharp increase in churn among customers making four or more service calls suggests that repeated

contact may act as an early warning signal of relationship deterioration. Repeated calls may also reflect unresolved problems that require repeated customer contact.

Heavier service use and premium service adoption did not necessarily correspond to stronger customer retention. The positive association between daytime usage and churn challenges the assumption that heavier usage necessarily indicates stronger customer attachment. Customers with high daytime usage may incur higher usage-related costs and therefore be more sensitive to price or perceived value. A similar pattern was observed among international-plan subscribers, whose higher churn rate suggests that subscribing to additional services does not necessarily guarantee retention.

By contrast, the negative association between voice-mail-plan subscription and churn may reflect broader differences in customer engagement rather than the effect of the service itself. Customers subscribing to the voice-mail plan may have a more established relationship with the provider or make greater use of the available services.

B. Interpreting the Model Comparison

Customer churn in this study was associated with combinations of customer behaviors and service characteristics rather than a single dominant predictor. This pattern helps explain why random forest achieved stronger performance, as its ensemble structure is better suited to capturing nonlinear interactions among predictors.

The main advantage of random forest was not a substantial increase in the number of churners identified, but a large reduction in false-positive predictions. Logistic regression correctly identified one additional churner but misclassified considerably more non-churners as churners. Therefore, random forest achieved a better balance between identifying actual churners and avoiding unnecessary customer interventions.

Reducing false-positive predictions has direct implications for retention resource allocation. In telecommunications settings where churners represent a minority of customers, unnecessary interventions may consume resources without improving retention outcomes. By reducing false positives while maintaining comparable recall, random forest allows retention efforts to focus more selectively on customers with higher estimated churn risk.

Although random forest achieved stronger predictive performance, logistic regression remained valuable for interpreting the factors associated with churn through its coefficient estimates. The two models therefore served complementary purposes within this study. Logistic regression supported explanation of the key drivers of churn, whereas random forest provided the more reliable basis for customer risk identification and subsequent retention decisions.

C. Translating Churn Probabilities into Retention Decisions

The marked increase in observed churn across the four risk tiers indicates that the random forest probabilities provided a reliable ranking of customer risk. This ordering is more useful for retention management than a binary churn prediction because it allows intervention intensity to reflect both the estimated probability of churn and the likely

cost of intervention. Low- and Medium-risk customers may require only routine monitoring or low-cost communication, whereas High- and Critical-risk customers require more intensive retention efforts. Risk segmentation therefore supports the allocation of limited retention resources rather than treating all predicted churners as equally urgent.

The distinction between the High and Critical tiers also shows that similar levels of churn risk may arise from different customer profiles. High-risk customers made more customer service calls, while Critical-risk customers were more strongly characterized by international-plan subscription and high daytime usage. These profiles suggest at least two potential pathways to churn. The first is a service-friction pathway, in which repeated customer service contact may reflect unresolved complaints or dissatisfaction with the resolution process. The second is a value-mismatch pathway, in which heavy usage and international-plan subscription may raise concerns about price, plan suitability, and perceived value.

These pathways require different retention responses. For customers in the service-friction pathway, generic discounts are unlikely to address the underlying source of dissatisfaction. A more appropriate response would be to review recent service contacts, resolve outstanding issues and prioritize customers with repeated complaints for proactive follow-up. The purpose of this intervention would be to restore confidence in the service relationship.

For customers in the value-mismatch pathway, retention efforts should focus on whether the existing tariff matches actual usage. High daytime usage may indicate higher usage-related charges, while international-plan customers may have needs that are not fully addressed by their current plans. Service providers should track customers' recent usage patterns and consider whether alternative packages better meet their needs. Adjusting plans or offering temporary price discounts can further enhance perceived value for customers. These interventions would target perceived value.

Risk tiers can be used to match retention efforts with customer risk levels. High-risk customers may require lower-cost interventions, such as targeted messages, plan reviews or scheduled follow-up, whereas Critical-risk customers may justify direct contact from a retention specialist or more substantial personalized offers. This approach also helps avoid allocating costly incentives when service recovery or plan adjustment may be sufficient.

D. Implications, Limitations and Future Research

The study shows that churn prediction can support retention decisions rather than simply identify customers who are likely to leave. This is achieved by combining predictor interpretation, model comparison and risk segmentation to translate predicted churn risk into differentiated retention decisions.

Several limitations should be considered when interpreting these findings. First, the analysis was based on a simplified dataset that lacked information on customer revenue, contract conditions, billing, competitors' offers, complaint content and longitudinal behavioral changes. Second, the predefined risk thresholds were not optimized according to customer lifetime value, intervention cost or

expected treatment response. Third, only logistic regression and random forest were evaluated, leaving the performance of more advanced ensemble methods unexplored. Finally, the proposed retention strategies were derived from associations between customer profiles and churn outcomes rather than direct evidence of intervention effectiveness.

Future research should use data from multiple telecom providers and incorporate richer behavioral and commercial information. It should also evaluate whether different retention strategies are effective for different customer profiles.

VI. CONCLUSION

This study examined how machine learning can support both telecom customer churn prediction and the translation of predicted risk into retention decisions. Repeated customer service contact, higher daytime usage and international-plan subscription emerged as the strongest indicators of increased churn risk, whereas voice-mail-plan subscription was associated with lower risk. Random forest achieved stronger overall predictive performance than logistic regression, particularly by substantially reducing false-positive classifications while maintaining a comparable level of recall. Predicted probabilities also separated customers into different risk groups, with observed churn rising sharply from the Low and Medium tiers to the High and Critical tiers.

Importantly, customers with elevated risk did not form a homogeneous group. High-risk customers were more strongly characterized by frequent customer service contact, whereas Critical-risk customers showed a greater concentration of international-plan subscription, with both groups exhibiting high daytime usage. These differences suggest that churn probabilities alone are insufficient for retention planning. Combining predicted risk with customer profiles allows retention priorities and interventions to be tailored more effectively.

Within the constraints of a single public dataset, limited commercial information, predefined risk thresholds, and comparison of only two predictive models, the study nevertheless demonstrates that predicted churn probabilities become more useful for retention planning when interpreted alongside customer profiles. Future research should use richer behavioral and commercial data to evaluate whether profile-specific retention strategies improve customer

outcomes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Ahmad, A. K., Jafar, A., & Aljoumaa, K. 2019. Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data*, 6: 28.
- Ahn, J. H., Han, S. P., & Lee, Y. S. 2006. Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, 30(10–11): 552–568.
- Bolton, R. N. 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, 17(1): 45–65. doi: 10.1287/mksc.17.1.45
- Burez, J., & Van den Poel, D. 2009. Handling class imbalance in customer churn prediction. *Expert Syst. Appl.*, 36: 4626–4636.
- Gerpott, T. J., Rams, W., & Schindler, A. 2001. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, 25(4): 249–269. doi: 10.1016/S0308-5961(00)00097-5
- Hemsathapat, K., & Dagli, C. H. 1999. Neural network model for detecting rare patterns in data: A telecommunication example. *Proceedings of the 1999 Artificial Neural Networks in Engineering Conference, ANNIE'99*.
- Keramati, A., & Ardabili, S. M. S. 2011. Churn analysis for an Iranian mobile operator. *Telecommunications Policy*, 35(4): 344–356.
- Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. 2006. Defection detection: measuring and understanding the predictive accuracy of customer churn models. *J. Mark. Res.*, 43: 204–211.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. 2012. New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *Eur. J. Oper. Res.*, 218: 211–229.
- Wei, C. P., & Chiu, I. T. 2002. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Syst. Appl.*, 23: 103–112.
- Wong, K. K. K. 2011. Getting what you paid for: Fighting wireless customer churn with rate plan optimization. *Journal of Database Marketing & Customer Strategy Management*, 18(2): 73–82.
- Wu, S., Yau, W. C., Ong, T. S., & Chong, S. C. 2021. Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access*, 9: 62118–62136.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).