



Comparative Analysis of Machine Learning Algorithms for Customer Churn Prediction in the Telecommunications Industry

Zeny Widianingsih¹, Aldi Algino²

Sekolah Tinggi Manajemen Informatika dan Komputer IKMI, Indonesia

Email: zenywidianingsih939@gmail.com^{1*}, alginoaldi88@gmail.com²

ABSTRACT

This study jointly benchmarks Logistic Regression, Random Forest, and Support Vector Machine (SVM) for customer churn prediction and links the resulting feature importance profile to retention strategy. The IBM Telco Customer Churn dataset (7,043 records, 20 predictor features covering demographic, account, and service attributes) was preprocessed through missing value imputation, label encoding, and feature standardization, then split into 80% training and 20% testing subsets using stratified sampling. Model performance was evaluated using accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). Recall and AUC were prioritized because a missed churning costs more than an unnecessary retention offer, while AUC summarizes ranking quality across all intervention thresholds. Results show that Random Forest achieved the highest accuracy (80.62%) and AUC (0.842), followed closely by Logistic Regression, which recorded a slightly lower accuracy (79.99%) but the highest recall (0.548), while SVM obtained the lowest AUC (0.792) despite comparable accuracy. Random Forest feature importance analysis indicated that contract type, tenure, total charges, and monthly charges were the strongest predictive attributes, together accounting for more than half of the model's total importance; these are predictive associations rather than causal determinants of churn. These findings indicate that tree based ensembles offer a marginal advantage over linear and margin based classifiers on structured churn data, and that contract and tenure related variables deserve priority in retention strategy. The observed differences between models are descriptive, as performance was estimated from a single stratified train test split with lightly tuned hyperparameters and no significance testing.

Keyword: machine learning; customer churn prediction; random forest; telecommunications

INTRODUCTION

The telecommunications industry operates in a highly saturated and competitive market in which customer acquisition costs substantially exceed the cost of retaining existing subscribers (Amin et al., 2019). Customer churn, defined as the discontinuation of a customer's relationship with a service provider, directly threatens recurring revenue streams and long-term profitability, and has therefore been described as one of the most critical business intelligence problems in the telecommunication sector (Ahmed & Linen, 2017). Consequently, the ability to identify customers who are likely to churn before they actually leave has become a strategic priority for telecommunication operators, enabling proactive and cost-effective retention interventions rather than costly reactive win-back campaigns.

Machine learning has emerged as an effective approach for churn prediction because it can automatically learn complex, non-linear patterns from historical customer behavior, usage, billing, and subscription data (Ahmad et al., 2019). A wide range of algorithms have been applied to this problem, from simple linear classifiers such as Logistic Regression to ensemble tree-based methods such as Random Forest and margin-based methods such as the Support Vector Machine (Chang et al., 2024). Lalwani et al., (2022) reported that AdaBoost combined with Random Forest produced strong predictive performance on a benchmark telecom dataset, while Khoh et al., (2023) demonstrated that an optimized weighted ensemble of classical

machine learning models could further improve churn-detection accuracy on imbalanced telecom data.

More recently, Vishakha, (2026) benchmarked Logistic Regression, Random Forest, and XGBoost on the same IBM Telco Customer Churn dataset used in the present study, reporting that the simpler Logistic Regression model slightly outperformed the more complex ensemble methods in terms of ROC-AUC, illustrating that model complexity does not always translate into superior predictive performance on structured, tabular churn data. Wu et al., (2021) further proposed an integrated framework that combines churn prediction with customer segmentation, showing that prediction outputs become more actionable when they are linked to identifiable customer groups, an idea that motivates the retention oriented interpretation adopted in the present study.

Beyond classical machine learning, several studies have explored deep learning architectures for churn prediction. Karanovic et al., (2018); Wangperawong et al., (2016) applied convolutional and autoencoder-based deep learning models to churn data, while Islam & Kim, (2019) proposed a random-forest-based factor-identification framework achieving high accuracy on a large telecom dataset. Fujo et al., (2022); Sivasankar & Vijaya, (2019) reported that hybrid deep learning and fuzzy-clustering ensembles could further improve churn-detection performance, and Khattak et al., (2023); Saha et al., (2023) showed that composite and deep churn-prediction architectures could achieve higher accuracy than single classical models, albeit with reduced interpretability and higher computational cost. Pustokhina et al., (2023) further showed that combining metaheuristic optimization with extreme learning machines could improve churn classification, illustrating the continuing search for a favorable trade-off between accuracy, robustness, and practicality.

Despite this growing body of literature, comparative results across studies remain inconsistent, partly because of differences in preprocessing strategy, feature encoding, and evaluation protocol, and partly because model performance is highly dependent on the characteristics of the particular dataset under study (Ahmed & Linen, 2017). Moreover, Ascarza et al., (2016) demonstrated in a field experiment that proactive retention interventions can sometimes backfire if targeting is based on a poorly calibrated churn model, and Ascarza, (2018) further argued that targeting high-risk customers identified by a churn model is not always an effective retention strategy, underscoring the practical stakes of selecting an appropriate and well-validated prediction algorithm.

Poudel et al., (2024) additionally emphasized that interpretability is often as important as raw accuracy for churn models intended to support real business decisions. Although Vishakha, (2026) recently benchmarked Logistic Regression, Random Forest, and XGBoost on the same IBM Telco Customer Churn dataset, that study did not include a margin based classifier such as SVM, did not report confusion matrices that expose the false negative and false positive trade off, and did not translate the feature importance profile into segment level retention actions. There have therefore been limited studies that jointly (i) benchmark Logistic Regression, Random Forest, and SVM under an identical, reproducible preprocessing and evaluation pipeline on the widely used IBM Telco Customer Churn dataset, (ii) interpret model

error behavior in terms of retention cost, and (iii) connect the resulting feature importance profile back to concrete, actionable retention strategies.

Addressing this gap is urgent because telecommunication operators increasingly deploy churn models to trigger automated retention campaigns, so a poorly chosen or poorly understood model translates directly into wasted marketing budget or unnoticed customer loss (Lemmens & Gupta, 2020). Therefore, this research intends to provide a controlled comparative evaluation of these three algorithms under a single, consistent experimental pipeline, and to identify the customer attributes that most strongly predict churn behavior. The objectives of this research are (1) to compare the predictive performance of Logistic Regression, Random Forest, and SVM for customer churn prediction using accuracy, precision, recall, F1-score, and AUC, and (2) to identify the most influential features associated with customer churn to inform retention strategy.

The novelty of this study lies in three aspects. First, from a methodological standpoint, the three algorithms represent distinct learning paradigms (linear probabilistic, tree based ensemble, and kernel based margin classification) and are evaluated under a fully specified and identical pipeline, including the same imputation rule, the same encoding scheme, the same stratified split, and the same random seed, so that observed performance differences can be attributed to the learner rather than to pipeline variation, a confound that Grinsztajn et al., (2022) identified as a major source of inconsistent conclusions in tabular benchmarking.

Second, the study interprets model quality not only through aggregate metrics but through confusion matrices and a cost based reading of false negatives and false positives, which is rarely reported in comparative churn studies on this dataset. Third, the study explicitly distinguishes predictive importance from causal explanation when interpreting feature importance, in line with the explainability literature (Lundberg & Lee, 2017; Ribeiro et al., 2016), so that the retention recommendations derived from the model are framed as targeting priorities rather than as causal claims about why customers leave.

The expected benefits of this research are both theoretical and practical. Theoretically, the study contributes replication evidence on a widely used benchmark dataset, which strengthens the cumulative knowledge base on how classical learners behave on structured churn data and clarifies whether the Logistic Regression versus Random Forest ranking reported by Vishakha, (2026) is reproducible under a different pipeline. Practically, the findings provide telecommunication operators with limited analytics capacity with an evidence based basis for choosing an interpretable model whose error profile matches their retention budget, and with a short list of customer attributes that can be used to prioritize retention offers.

The broader implications are threefold: for retention strategy design, the results suggest that contract and tenure based segmentation should precede any price based intervention; for methodological benchmarking, the study demonstrates why single split comparisons should be complemented by resampling based validation and statistical testing before superiority claims are made; and for management, the results reinforce that model selection should be guided by the relative cost of missed churners and wasted offers rather than by accuracy alone.

METHOD

This study employed a quantitative, experimental research design based on supervised machine learning classification. The dataset used is the publicly available IBM Telco Customer Churn dataset, comprising 7,043 individual customer records and 20 predictor attributes covering demographic information (e.g., gender, senior citizen status, dependents), account information (e.g., tenure, contract type, payment method, monthly and total charges), and subscribed services (e.g., internet service, online security, tech support, streaming services). The target variable, Churn, is binary (Yes/No), with an overall churn rate of 26.54% in the dataset, indicating a moderately imbalanced classification problem.

This study involves no primary human subjects and no field data collection; it relies entirely on secondary, publicly available, and de-identified data, so no ethical clearance from an institutional review board was required. The research setting is therefore computational rather than geographical, and the unit of analysis is the individual customer record. The dataset was obtained by downloading the CSV file from the public IBM repository on GitHub (IBM, 2019), which is also mirrored on Kaggle, and the file was used without modification prior to the preprocessing steps described below. The data are released for open educational and research use, and no personally identifying information is contained in the records.

Data preprocessing consisted of the following steps: (1) the TotalCharges field, originally stored as text, was converted to numeric type and the small number of missing values arising from customers with zero tenure was imputed using the median value; (2) the non-informative customerID field was removed; (3) all categorical variables were transformed into numeric form using label encoding; and (4) numerical features were standardized using z-score normalization (StandardScaler) to ensure comparable feature scales, which is particularly important for distance- and margin-based algorithms such as SVM. The preprocessed dataset was then split into training (80%, $n = 5634$) and testing (20%, $n = 1409$) subsets using stratified random sampling to preserve the original churn ratio in both subsets.

To prevent data leakage, the median used for imputation and the mean and standard deviation used by the StandardScaler were estimated from the training subset only and then applied to the test subset, because statistics computed on the full dataset would allow information from the test records to influence the training process (Kapoor & Narayanan, 2023). Label encoding was selected because most categorical attributes in the dataset are binary (Yes/No) or have at most three or four levels, so the integer codes do not create high cardinality artifacts; however, it is acknowledged that for nominal attributes with more than two levels (e.g., payment method, internet service), label encoding imposes an arbitrary ordering that may slightly disadvantage the linear and kernel based models, and one hot or regularized target encoding could be explored as an alternative (Pargent et al., 2022). No resampling or class weighting was applied, because the study aims to establish a baseline comparison under the natural class distribution (churn rate 26.54%); the consequences of this choice for recall are examined in the Results and Discussion section, and class imbalance techniques such as SMOTE or cost sensitive weighting are recommended for follow up work (Fernández et al., 2018).

Three supervised classification algorithms were selected for comparison based on their prevalence and demonstrated effectiveness in prior churn-prediction literature: (1) Logistic Regression, a linear probabilistic classifier commonly used as an interpretable baseline; (2) Random Forest, an ensemble of decision trees trained via bagging, configured with 300 estimators and a maximum depth of 8 to reduce overfitting; and (3) Support Vector Machine (SVM) with a radial basis function (RBF) kernel, which maps features into a higher-dimensional space to find an optimal separating hyperplane.

The Random Forest configuration follows the evidence of Probst et al., (2019) that the number of trees mainly stabilizes variance beyond a few hundred estimators, whereas limiting tree depth is the most effective single control against overfitting on moderately sized tabular data; the remaining Random Forest settings, and all Logistic Regression and SVM settings, used scikit-learn default values (regularization strength $C = 1.0$ for both Logistic Regression and SVM, and $\gamma = \text{"scale"}$ for the RBF kernel). Hyperparameters were deliberately not tuned exhaustively so that the comparison reflects out of the box behavior under an identical pipeline rather than the outcome of unequal tuning effort. All models were implemented in Python 3 using the scikit-learn library with a fixed random seed (42) to ensure reproducibility, and default hyperparameters were used except where noted above. The research instrument in this study is therefore the software and computing environment itself, consisting of Python 3, the pandas library for data handling, and scikit-learn for model training and evaluation, executed on a standard personal computer without GPU acceleration.

Model performance was evaluated on the held-out test set using five standard classification metrics: accuracy (the proportion of correctly classified instances), precision (the proportion of predicted churners who actually churned), recall (the proportion of actual churners correctly identified), F1-score (the harmonic mean of precision and recall), and the Area Under the Receiver Operating Characteristic Curve (AUC), which summarizes the trade-off between true-positive and false-positive rates across all classification thresholds. Because churn prediction is typically used to trigger costly retention interventions, recall and AUC were treated as particularly important metrics, since failing to identify a true churner (a false negative) is generally more costly to a telecommunication operator than a false alarm. In addition, feature importance scores from the Random Forest model (based on mean decrease in Gini impurity) were extracted to identify the customer attributes most strongly associated with churn.

Following Lundberg & Lee, (2017); Ribeiro et al., (2016), these scores are interpreted as measures of predictive contribution within the fitted model and not as evidence of causal influence on churn. The validation strategy consisted of a single stratified hold out split; k-fold cross-validation was not applied, and no statistical significance test (e.g., McNemar's test or bootstrapped confidence intervals) was performed on the metric differences between models. Consequently, the reported metrics are point estimates from one partition, and small differences between models should be read as descriptive rather than as statistically confirmed (Raschka, 2018). A further limitation is that default or lightly tuned hyperparameters were used for all models rather than exhaustive grid search, which may cause some variance in the reported metrics.

RESULTS AND DISCUSSION

Table 1 summarizes the descriptive characteristics of the dataset after preprocessing, and Table 2 presents the performance of the three machine learning models on the test set.

Table 1. Dataset characteristics

Characteristic	Value
Total records	7043
Training set size	5634
Testing set size	1409
Overall churn rate	26.54%
Number of predictor features	19

Source: Data Processed

Table 2. Performance comparison of machine learning models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.7999	0.6447	0.5481	0.5925	0.8403
Random Forest	0.8062	0.6797	0.5107	0.5832	0.8423
Support Vector Machine	0.7999	0.6691	0.4866	0.5635	0.7916

Source: Data Processed

As shown in Table 2, Random Forest achieved the highest accuracy (0.8062) and AUC (0.8423), narrowly outperforming Logistic Regression (accuracy 0.7999, AUC 0.8403). Logistic Regression, however, obtained the highest recall (0.5481) among the three models, meaning it identified a larger proportion of actual churners than Random Forest (recall 0.5107) or SVM (recall 0.4866). SVM achieved comparable accuracy (0.7999) but the lowest AUC (0.7916), indicating weaker overall ranking ability between churners and non-churners despite feature standardization.

It should be stressed that these differences are small in absolute terms (0.63 percentage points in accuracy and 0.002 in AUC between Random Forest and Logistic Regression) and that no statistical significance test was performed on them; the ranking therefore describes the observed outcome on this particular test partition and should not be read as evidence that Random Forest is statistically superior. In relation to the first research objective, the results show that the three algorithms are closely matched on accuracy, that Random Forest and Logistic Regression are practically indistinguishable on AUC, and that the clearest separation appears on recall and on the AUC of SVM.

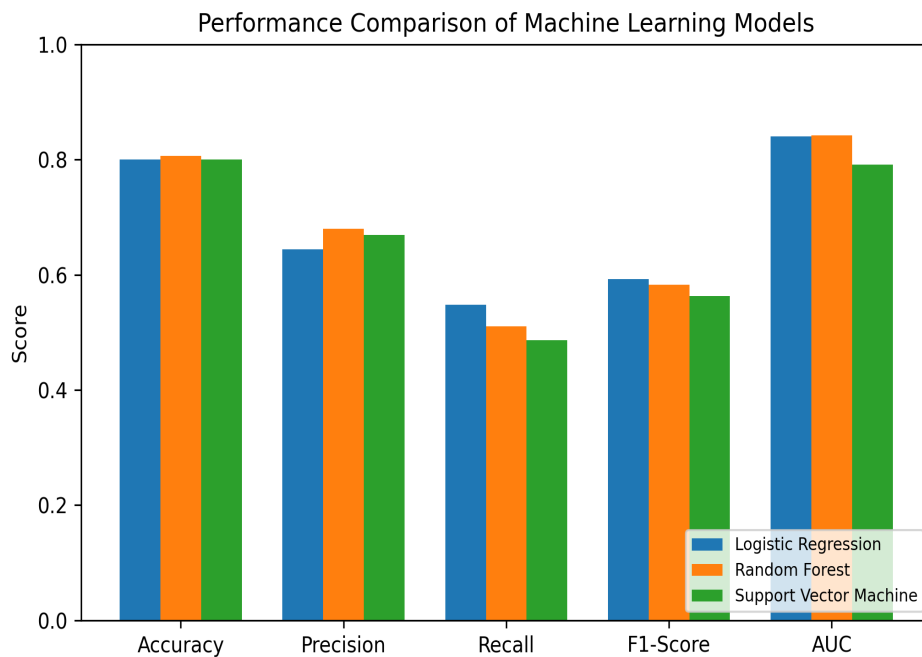


Figure 1. Performance comparison of Logistic Regression, Random Forest, and SVM across five evaluation metrics.

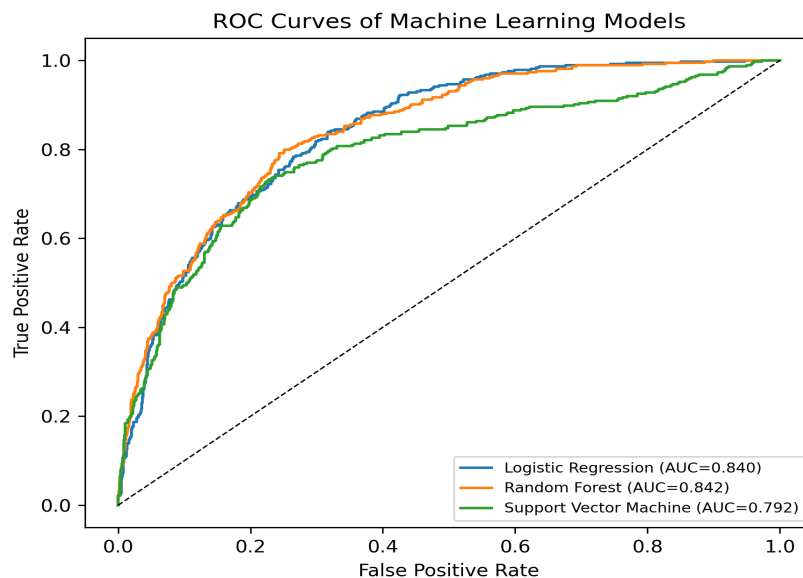


Figure 2. ROC curves of the three machine learning models on the test set.

These results are consistent with prior comparative studies on the IBM Telco Customer Churn dataset, which similarly report that Logistic Regression and Random Forest achieve closely matched accuracy and AUC, with the ranking between the two often depending on preprocessing choices and hyperparameter settings rather than a clear, dataset-independent superiority of one algorithm. The relatively modest recall values obtained by all three models (below 0.55) reflect the moderate class imbalance in the dataset (churn rate 26.5%) and suggest that, in operational deployment, probability thresholds should be tuned downward from the

default 0.5, or class-imbalance techniques such as class weighting or resampling should be applied, to increase the detection rate of true churners at an acceptable cost in additional false alarms.

Such threshold tuning presupposes that the predicted probabilities are reasonably calibrated; because Random Forest probabilities derived from vote fractions are known to be poorly calibrated relative to Logistic Regression, a calibration step such as Platt scaling or isotonic regression should precede any threshold adjustment in deployment (Guo et al., 2017). Figure 3 presents the ten most important features identified by the Random Forest model based on mean decrease in Gini impurity. Contract type, tenure, total charges, and monthly charges emerged as the four most influential predictors, together accounting for more than half of the model's total feature importance, followed by online security and technical support subscription status.

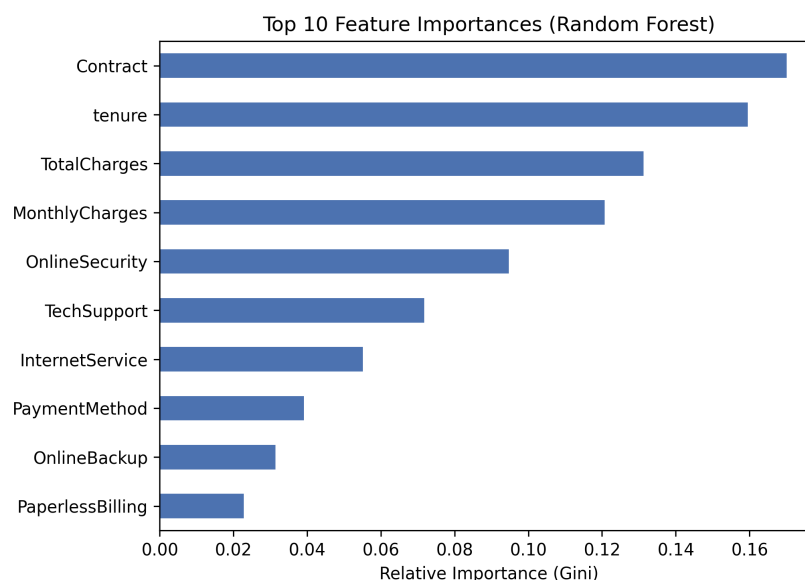


Figure 3. Top ten feature importances from the Random Forest model.

The dominance of contract type and tenure as predictors is consistent with, although it does not causally establish, the expectation that customers on flexible, month-to-month contracts and those with shorter tenure have both lower switching costs and less accumulated loyalty, making them more prone to churn than customers locked into one- or two-year contracts. The importance of monthly and total charges further suggests that price sensitivity is associated with churn decisions, while the relevance of online security and technical-support subscriptions indicates that customers who do not purchase value-added services may perceive lower overall service value and therefore be more likely to leave. Because Gini based importance measures the extent to which a feature reduces impurity in the fitted forest, it reflects predictive usefulness within this model and can be inflated for continuous or high cardinality variables such as total charges; it does not identify the mechanisms through which these attributes influence churn (Lundberg & Lee, 2017).

From a managerial perspective, these findings imply that telecommunication providers could prioritize retention offers, such as contract upgrade incentives, loyalty discounts, or bundled value-added services, toward month-to-month, short-tenure, and high-monthly-charge customers, who represent the segment at greatest predicted risk of churning; this addresses the second research objective by translating the importance ranking into targeting priorities, while the causal effectiveness of any specific offer would need to be verified through controlled experiments of the kind reported by (Ascarza et al., 2016). To better understand the practical error behavior of each model, Table 3 presents the confusion matrices obtained on the test set ($n = 1409$). A false negative in this context represents an actual churner that the model failed to flag, and therefore a lost opportunity for retention intervention, whereas a false positive represents a loyal customer incorrectly targeted with a retention offer, incurring an unnecessary marketing cost.

Table 3. Confusion matrices of the three machine learning models (test set, $n = 1409$)

Model	True Negative	False Positive	False Negative / True Positive
Logistic Regression	922	113	169 / 205
Random Forest	945	90	183 / 191
Support Vector Machine	945	90	192 / 182

Source: Data Processed

Random Forest produced the fewest false positives (90 misclassified loyal customers), whereas Logistic Regression produced the fewest false negatives among the three models (169 missed churners out of 374 actual churners in the test set), consistent with its higher recall reported in Table 2. This trade-off illustrates a common tension in churn-prediction deployment: Random Forest is more conservative and efficient in avoiding unnecessary retention costs, while Logistic Regression is more sensitive in capturing at-risk customers, which may be preferable when the cost of losing a customer greatly exceeds the cost of an unnecessary retention offer. This pattern is consistent with the precision-recall trade-off reported in recent comparative churn-prediction studies (Khoh et al., 2023; Lalwani et al., 2022).

To make the trade off in Table 3 operationally concrete, Table 4 converts the confusion matrices into an illustrative total misclassification cost under three cost ratios, where the cost of a false negative (a missed churner) is set at one, three, and five times the cost of a false positive (an unnecessary retention offer). The unit cost of a false positive is normalized to one, so the figures are relative rather than monetary. When both errors are weighted equally, Random Forest yields the lowest cost because of its lower false positive count; however, as soon as a missed churner is assumed to cost three times more than a wasted offer, which is a conservative assumption in subscription businesses where the lifetime value of a retained customer typically exceeds the cost of a discount, Logistic Regression becomes the least costly model, and its advantage widens as the ratio increases.

This cost sensitive reading shows that the practical superiority of a model depends on the operator's cost structure rather than on accuracy or AUC alone, which is consistent with the profit oriented view of churn management advocated by (Lemmens & Gupta, 2020). It also implies that, for any of the three models, lowering the decision threshold below 0.5 would shift errors from false negatives toward false positives and could reduce total cost further whenever the cost ratio is large.

Table 4. Illustrative total misclassification cost of the three models under different cost ratios (test set, $n = 1409$; cost of a false positive normalized to 1)

Model	Cost ratio FN : FP =	Cost ratio FN : FP =	Cost ratio FN : FP =
	1 : 1	3 : 1	5 : 1
Logistic Regression	282	620	958
Random Forest	273	639	1005
Support Vector Machine	282	666	1050

Source: Data Processed

Comparing the present findings with prior studies on similar or related telecommunication churn datasets provides useful context. Notably, Vishakha, (2026) benchmarked Logistic Regression, Random Forest, and XGBoost on the same IBM Telco Customer Churn dataset used in the present study and reported an almost identical Logistic Regression accuracy of 79.91% with an AUC of 0.8403, closely matching the Logistic Regression accuracy of 79.99% and AUC of 0.8403 obtained here, and likewise found that the simpler Logistic Regression model was competitive with, or even slightly ahead of, the ensemble methods on ROC-AUC.

This convergence across two independently conducted studies using the same dataset lends additional credibility to the present results and suggests that the relative ranking between Logistic Regression and Random Forest on this particular dataset is a robust, reproducible finding rather than an artifact of a specific train-test split. The accuracy and AUC levels obtained here are also broadly consistent with the range reported by Ahmad et al., (2019); Lalwani et al., (2022) for classical machine learning pipelines applied to the same dataset without extensive hyperparameter tuning or resampling. In contrast, studies that employ deep learning architectures, cost-sensitive resampling, or hybrid ensembles such as Islam & Kim, (2019); Khattak et al., (2023); Saha et al., (2023); Sivasankar & Vijaya, (2019) report noticeably higher accuracy figures, often above 90%, but typically at the cost of reduced model interpretability, higher computational requirements, and, in several cases, evaluation on different or preprocessed variants of the dataset that are not directly comparable.

This comparison suggests that the moderate performance gap between classical and deep learning approaches observed in the literature reflects not only algorithmic sophistication but also differences in feature engineering, class-imbalance handling, and validation protocol, reinforcing the argument of Poudel et al., (2024) that reported accuracy figures should always be interpreted alongside methodological transparency and interpretability requirements. The

present study nevertheless differs from Vishakha, (2026) in three respects that constitute its distinct contribution: it adds a kernel based SVM to the comparison and shows that, on this dataset, the RBF kernel does not improve ranking ability over the linear baseline; it reports confusion matrices and a cost based interpretation that reveal a recall advantage for Logistic Regression that aggregate AUC alone does not expose; and it explicitly frames feature importance as a targeting aid rather than as a causal explanation.

At the same time, the convergence should be interpreted cautiously, because both studies rely on a single hold out partition and neither reports confidence intervals; the appropriate next step is to replicate the present pipeline with repeated stratified k-fold cross-validation (e.g., 5 x 2 or 10 x 10 folds) and to test the Logistic Regression versus Random Forest difference with McNemar's test or a corrected resampled t test, which would determine whether the 0.002 AUC gap observed here is distinguishable from sampling noise (Raschka, 2018). From a managerial standpoint, the results reinforce the importance of aligning model choice with business objectives rather than optimizing accuracy alone. Ascarza et al., (2016) caution that poorly targeted retention campaigns triggered by a model with insufficient recall can waste marketing resources or even provoke reactance among customers who were not genuinely at risk.

Ascarza, (2018) similarly found that targeting only the customers flagged as highest-risk by a churn model is not always the most effective retention strategy, since some high-risk customers are difficult to retain regardless of intervention. The present findings suggest that Logistic Regression, despite its simplicity, may be preferable when the operational goal is to maximize the detection of at-risk customers (high recall), whereas Random Forest may be preferable when the goal is to minimize wasted retention spending on customers who would have stayed regardless (high precision).

Combining both perspectives, for instance by using Random Forest for coarse segmentation and Logistic Regression for fine-grained risk scoring within high-risk segments, could represent a practical compromise for telecommunication operators with constrained retention budgets. This recommendation is conditional on the cost structure analyzed in Table 4 and should not be generalized to operators whose retention offers are expensive relative to customer lifetime value, for whom the precision oriented behavior of Random Forest would be preferable.

It should be noted that the models in this study were evaluated using default or lightly tuned hyperparameters and a single stratified train-test split; therefore, the reported performance may not represent the maximum achievable accuracy for each algorithm. Specifically, the reported accuracy of 0.8062 for Random Forest and 0.7999 for Logistic Regression, and the AUC values of 0.8423 and 0.8403, are single partition point estimates whose sampling variability is unknown; a direct replication of this study using stratified 10-fold cross-validation on the same 7,043 records, with the same preprocessing and seed, would yield mean metrics with standard deviations and would establish whether the Random Forest advantage persists across folds.

Future work could additionally apply systematic hyperparameter optimization, class imbalance handling techniques such as SMOTE or class weighted loss (Fernández et al., 2018), and probability calibration prior to threshold selection (Guo et al., 2017), such as the optimized

weighted ensembling approach demonstrated by Khoh et al., (2023), in addition to exploring composite or deep learning architectures for tabular churn data, as demonstrated in related recent studies (Chang et al., 2024; Khattak et al., 2023; Saha et al., 2023), while bearing in mind the evidence of Grinsztajn et al., (2022) that tree based ensembles remain highly competitive with deep learning on medium sized tabular datasets of this kind.

CONCLUSION

This study compared Logistic Regression, Random Forest, and Support Vector Machine for customer churn prediction. Random Forest achieved the highest accuracy (80.62%) and AUC (0.842), while Logistic Regression achieved the highest recall (0.548), with both outperforming SVM in AUC. However, the differences between Random Forest and Logistic Regression were small and not statistically tested. Contract type, tenure, total charges, and monthly charges were the attributes most strongly associated with churn. These findings suggest that model selection should consider business objectives and the relative costs of missed churners and unnecessary retention offers rather than accuracy alone. Future research should apply cross-validation, statistical testing, class-imbalance handling, and probability calibration to improve model reliability and churn detection.

REFERENCES

- Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1), 28. <https://doi.org/10.1186/s40537-019-0191-6>
- Ahmed, A., & Linen, D. M. (2017). A review and analysis of churn prediction methods for customer retention in telecom industries. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICACCS.2017.8014605>
- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94, 290–301. <https://doi.org/10.1016/j.jbusres.2018.03.003>
- Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research*, 55(1), 80–98. <https://doi.org/10.1509/jmr.16.0163>
- Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research*, 53(1), 46–60. <https://doi.org/10.1509/jmr.13.0483>
- Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). Prediction of customer churn behavior in the telecommunication industry using machine learning models. *Algorithms*, 17(6), 231. <https://doi.org/10.3390/a17060231>
- Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61, 863–905. <https://doi.org/10.1613/jair.1.11192>
- Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1),

- 185–198. <https://doi.org/10.18576/isl/110120>
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems* (Vol. 35, pp. 507–520). <https://doi.org/10.48550/arXiv.2207.08815>
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning, PMLR* (Vol. 70, pp. 1321–1330). <https://doi.org/10.48550/arXiv.1706.04599>
- IBM. (2019). *Telco customer churn dataset*. IBM Developer. <https://github.com/IBM/telco-customer-churn-on-icp4d>
- Islam, M. R., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 7, 60134–60148. <https://doi.org/10.1109/ACCESS.2019.2914999>
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 100804. <https://doi.org/10.1016/j.patter.2023.100804>
- Karanovic, M., Popovac, M., Sladojevic, S., Arsenovic, M., & Stefanovic, D. (2018). Telecommunication services churn prediction: Deep learning approach. In *2018 26th Telecommunications Forum (TELFOR)* (pp. 1–4). IEEE. <https://doi.org/10.1109/TELFOR.2018.8612067>
- Khattak, A., Mehak, Z., Ahmad, H., Asghar, M. U., Asghar, M. Z., & Khan, A. (2023). Customer churn prediction using composite deep learning technique. *Scientific Reports*, 13, 17294. <https://doi.org/10.1038/s41598-023-44396-w>
- Khoh, W. H., Pang, Y. H., Ooi, S. Y., Wang, L. Y.-K., & Poh, Q. W. (2023). Predictive churn modeling for sustainable business in the telecommunication industry: Optimized weighted ensemble machine learning. *Sustainability*, 15(11), 8631. <https://doi.org/10.3390/su15118631>
- Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: A machine learning approach. *Computing*, 104(2), 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, 39(5), 956–973. <https://doi.org/10.1287/mksc.2020.1229>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). <https://doi.org/10.48550/arXiv.1705.07874>
- Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, 37(5), 2671–2692. <https://doi.org/10.1007/s00180-022-01207-6>
- Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, 17, 100567. <https://doi.org/10.1016/j.mlwa.2024.100567>
- Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies
-

- for random forest. *WIREs Data Mining and Knowledge Discovery*, 9(3), e1301. <https://doi.org/10.1002/widm.1301>
- Pustokhina, I. V., Pustokhin, D. A., Nguyen, P. T., Elhoseny, M., & Shankar, K. (2023). Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector. *Complex & Intelligent Systems*, 9(4), 3473–3485. <https://doi.org/10.1007/s40747-021-00353-6>
- Raschka, S. (2018). *Model evaluation, model selection, and algorithm selection in machine learning*. arXiv. <https://doi.org/10.48550/arXiv.1811.12808>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM. <https://doi.org/10.1145/2939672.2939778>
- Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., & El-kenawy, E.-S. M. (2023). Deep churn prediction method for telecommunication industry. *Sustainability*, 15(5), 4543. <https://doi.org/10.3390/su15054543>
- Sivasankar, E., & Vijaya, J. (2019). Hybrid PPFCM-ANN model: An efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. *Neural Computing and Applications*, 31(11), 7181–7200. <https://doi.org/10.1007/s00521-018-3548-4>
- Vishakha, F. N. U. (2026). *A comparative analysis of machine learning models for customer churn prediction in the telecommunications industry: A business analytics perspective*. SSRN. <https://doi.org/10.2139/ssrn.6819479>
- Wangperawong, A., Brun, C., Laudy, O., & Pavasuthipaisit, R. (2016). *Churn analysis using deep convolutional neural networks and autoencoders*. arXiv. <https://doi.org/10.48550/arXiv.1604.05377>
- 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. <https://doi.org/10.1109/ACCESS.2021.3084488>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).