

Predicting Customer Churn in Telecom Using Logistic Regression and Decision TreesAuthor: **Lea Araby¹***Affiliation:*

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

Abstract:

Customer churn is costly for telecom operators, and managers need models they can understand and act on. This study compares two interpretable classifiers, logistic regression and a decision tree, for predicting churn, and examines how the classification threshold affects recall and precision. Using the public IBM Telco Customer Churn dataset (7,032 customers after cleaning, of whom 26.6% churned), the models were trained on 80% of the data and evaluated on a held-out 20%, with bootstrap confidence intervals and an exact McNemar test used to assess differences between them. Churn was most strongly associated with month-to-month contracts (42.7% churn, versus 2.9% for two-year contracts), fiber optic internet service (41.9%), electronic-check payment (45.3%) and short tenure (median 10 versus 38 months). Logistic regression achieved slightly higher accuracy (79.7% versus 77.9%) and AUC (0.835 versus 0.822) than the decision tree, but the difference was not statistically significant (McNemar $p = .061$). Lowering the threshold from 0.5 to 0.3 raised the logistic regression's recall from 56.4% to 75.9%, while precision fell from 63.4% to 50.7%. The findings suggest that a simple, interpretable model is sufficient for this problem, that the threshold should reflect the cost of retention offers, and that contract length and early-tenure customers are natural targets for retention efforts. The data come from a fictional single-company sample, and the results show associations, not causes.

Keywords:

Customer churn; logistic regression; decision tree; telecommunications; classification threshold; customer retention

Introduction:

Telecommunications is a highly competitive industry where most potential customers already hold a subscription, so growth depends on keeping existing customers as much as on winning new ones. Classic work in services marketing showed that even small improvements in retention can translate into large gains in profit (Reichheld & Sasser, 1990). Telecom firms hold detailed records of each customer's contract, services, billing and tenure, which makes customer churn (the loss of customers who stop using a service) a natural setting for predictive analytics. Churn prediction models aim to identify the customers with the highest propensity to leave, so that retention campaigns can be directed at them and made more efficient (Verbeke et al., 2012).

Many algorithms can be applied to churn prediction, but a prediction is only useful if managers can understand and act on it. Logistic regression and decision trees remain attractive because their outputs (odds ratios and if-then rules) can be explained to non-technical staff (De Caigny et al., 2018). Two practical issues deserve more care in applied work. First, differences in accuracy between models can be reported as point estimates without asking whether they exceed sampling variation. Second, because churners are a minority, overall accuracy can look good even when many churners are missed, so the decision threshold matters as much as the model. This project addresses both issues using a public telecom dataset.

The study asks three research questions. RQ1: Which customer characteristics are most strongly associated with churn? RQ2: How do logistic regression and a decision tree compare in predicting churn on unseen customers, and is any difference statistically meaningful? RQ3: How does the classification threshold change the balance between precision and recall? Section 2 reviews the literature, Section 3 describes the data and methods, Section 4 reports the results, and Section 5 discusses implications and limitations.

Literature Review:

Customer churn has long been treated as a question of profit. Reichheld and Sasser (1990) reported that profits rise as a customer's relationship with a firm lengthens, so small gains in retention can yield large gains in profit, and Ascarza et al. (2018) review the key issues in retention management since then. Retention is a particular concern in mature mobile markets (Ahn et al., 2006). Studying the Korean mobile market, Ahn et al. found that call-quality factors influenced churn, that members of card programmes and heavy users were more likely to churn, and that partial defection (a shift to non-use or suspension) helps explain eventual defection. Call quality and usage are not available in the Telco dataset used here, which contains contract, service, billing and demographic variables only, so this study asks how far such account-level characteristics can explain and predict churn.

A large part of the literature asks which algorithm predicts churn best. In a tournament in which academics and practitioners built models on the same data, Neslin et al. (2006) concluded that the

choice of method matters, since differences in accuracy could change the profitability of a campaign by hundreds of thousands of dollars. Vafeiadis et al. (2015) compared several machine-learning techniques on a telecom dataset, and Lemmens and Croux (2006) applied bagging and boosting of classification trees to churn. Alongside accuracy, comprehensibility is valued: De Caigny et al. (2018) describe decision trees and logistic regression as two very popular churn algorithms with strong performance and good comprehensibility, although trees struggle with linear relations and logistic regression with interactions. Rudin (2019) argues that for high-stakes decisions inherently interpretable models should be preferred to explanations of black boxes. Churn management is not high-stakes in her sense, but the principle that decision-makers should see why a customer is flagged supports comparing these two models.

Beyond basic account-level predictors, churn management must be understood within the broader context of a firm's data-driven culture and macroeconomic environment. The successful deployment of predictive analytics, knowledge management, and big data capabilities is increasingly recognized as a primary driver of product development and overall firm performance (Azoury et al., 2024; Subrahmanyam et al., 2024). Furthermore, consumer economic decisions—including the likelihood of maintaining or canceling a telecom subscription—are significantly shaped by macroeconomic interrelatedness and generational behaviors, which can be effectively modeled using logistic regression and ensemble machine learning techniques (Bejjani et al., 2024; Mawad & Freiha, 2024; Nasser et al., 2025). During periods of economic vulnerability, integrating these predictive frameworks serves as a critical smart decision support system, allowing operators to forecast financial distress and navigate industry disruptions proactively (Elia et al., 2021; Subrahmanyam & Beainy, 2027).

However, realizing the financial benefits of these predictive models requires navigating the ethical and organizational complexities of AI adoption. The integration of artificial intelligence into business culture frequently introduces challenges regarding algorithmic transparency, internal controls, and corporate governance (Athari et al., 2023; Diab & El Hajj, 2024; Saber & Sassine, 2022). Research demonstrates that the drivers of AI adoption rely heavily on the perceived trust and expertise of the professionals utilizing the tools (Jabbour Al Maalouf et al., 2026). Therefore, rather than relying on opaque solutions, managers must leverage their intrinsic business experience by utilizing highly interpretable models like decision trees and logistic regression (Chaanine et al., 2023). Maintaining this transparency ensures that targeted retention campaigns remain authentic, properly governed, and avoid triggering consumer confusion or expectation disconfirmation (Daou et al., 2025).

How a model is judged matters as much as which one is used. Verbeke et al. (2012) note that churn models are usually evaluated with statistical measures although their purpose is cost reduction, which can lead to suboptimal model selection, and they benchmarked classifiers on eleven real telecom datasets with a profit-based measure. Class imbalance is a further problem characteristic of churn, and Burez and Van den Poel (2009) address ways of handling it. Saito and Rehmsmeier

(2015) show that ROC plots can be deceptive on imbalanced data and that precision–recall information is more informative, which is why precision and recall are reported here alongside AUC. When two classifiers are compared on one test set, Dietterich (1998) found McNemar's test to have low Type I error and to be the only acceptable test for algorithms that can be run only once.

This study does not propose a new algorithm. Its contribution is to combine, in one reproducible analysis of a free dataset, a comparison of two baseline models with confidence intervals and a paired significance test, an explicit look at how the threshold changes recall and precision, and an interpretation in retention-management terms.

Methodology:

The study is a quantitative, predictive (supervised classification) design. The data are the IBM Telco Customer Churn sample dataset distributed through Kaggle (BlastChar, n.d.), which describes a fictional telecom company with 7,043 customers and 21 variables. The data are public and anonymised, so no personal data were handled; because the company is fictional, the findings illustrate the method rather than describe a real firm. AI tools were used as a complementary aid for brainstorming, coding and drafting, and all reported figures come from the analysis of this dataset.

The outcome is Churn (1 = left within the last month, 0 = stayed). The 19 predictors are demographics (gender, senior citizen, partner, dependents); account features (tenure, contract type, paperless billing, payment method, monthly charges, total charges); and subscribed services (phone service, multiple lines, internet service, online security, online backup, device protection, technical support, streaming TV, streaming movies). The customer identifier was excluded. TotalCharges was converted from text to numeric; eleven blank values, all belonging to customers with zero tenure, were removed (0.16%), leaving 7,032 customers. The categories "No internet service" and "No phone service" were recoded as "No" to avoid collinearity with InternetService and PhoneService. Categorical variables were dummy-coded (first category as reference), and numeric variables were standardised for the logistic regression.

Churners and non-churners were compared with Mann–Whitney U tests for numeric variables and chi-square tests (with Cramér's V) for categorical variables, at $\alpha = .05$. For prediction, the data were split by stratified random sampling into a training set (80%, $n = 5,625$) and a test set (20%, $n = 1,407$), with the test set used only for final evaluation. Logistic regression was fitted with a negligible penalty ($C = 10^6$), so coefficients are effectively maximum-likelihood estimates, and exponentiated coefficients are read as odds ratios. A CART decision tree was tuned by five-fold stratified cross-validation on the training set, choosing maximum depth (2, 3, 4, 5, 6, 8, 10) and minimum leaf size (10, 25, 50, 100) to maximise AUC.

Performance was measured on the test set at a 0.5 probability threshold, with churn as the positive class, using accuracy, precision, recall, F1 and AUC, and compared with a majority-class baseline. Uncertainty was assessed with 95% percentile bootstrap intervals (2,000 resamples) for AUC and the AUC difference, and an exact McNemar test on the two models' predictions

(Dietterich, 1998). As a sensitivity analysis, a lower threshold of 0.3, chosen in advance and not tuned on the test set, and balanced class weights were also evaluated. Analyses used Python with pandas, scikit-learn (Pedregosa et al., 2011), SciPy and Matplotlib, with all random seeds fixed at 42 for reproducibility.

Results:

Of the 7,032 customers analysed, 1,869 (26.6%) had churned. The training and test sets had the same churn rate (26.6%), so a model that always predicted "no churn" would reach 73.4% accuracy on the test set; this is the baseline for the accuracy figures below.

4.1 Which customers churn? (RQ1)

Table 1. *Churn by customer characteristics (N = 7,032)*

Variable	Result (churn rate for categories)	p	Cramér's V
Tenure (months), median	Stayed 38 vs. churned 10	< .001	—
Monthly charges (USD), median	Stayed 64.45 vs. churned 79.65	< .001	—
Total charges (USD), median	Stayed 1,683.60 vs. churned 703.55	< .001	—
Contract	Month-to-month 42.7%; one year 11.3%; two year 2.9%	< .001	0.41
Internet service	DSL 19.0%; fiber optic 41.9%; none 7.4%	< .001	0.32
Payment method	Electronic check 45.3%; other methods 15.3%–19.2%	< .001	0.30
Paperless billing	Yes 33.6%; no 16.4%	< .001	0.19
Technical support	No 31.2%; yes 15.2%	< .001	0.16
Dependents	No 31.3%; yes 15.5%	< .001	0.16
Senior citizen	Yes 41.7%; no 23.7%	< .001	0.15
Gender	Female 27.0%; male 26.2%	.490	0.01

Note. Numeric variables: Mann–Whitney U test; categorical variables: chi-square test. Only the ten strongest and gender are shown; 14 of 16 categorical variables were significant (all $p < .001$); PhoneService ($p = .350$) was not.

Churners had much shorter tenure and higher monthly charges than customers who stayed (Table 1, Figure 1). The strongest categorical associations were with contract type ($V = .41$), internet service ($V = .32$) and payment method ($V = .30$): month-to-month customers, fiber optic

customers and electronic-check payers churned at 42.7%, 41.9% and 45.3% respectively (Figure 2). Gender showed no association with churn.

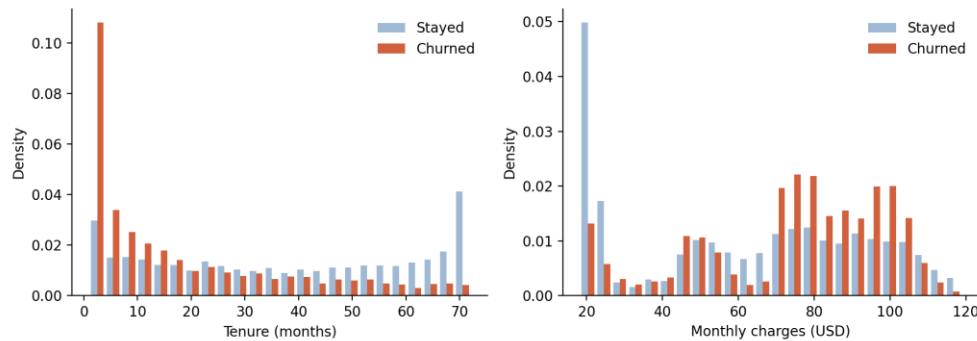


Figure 1. *Distribution of tenure and monthly charges by churn status.*

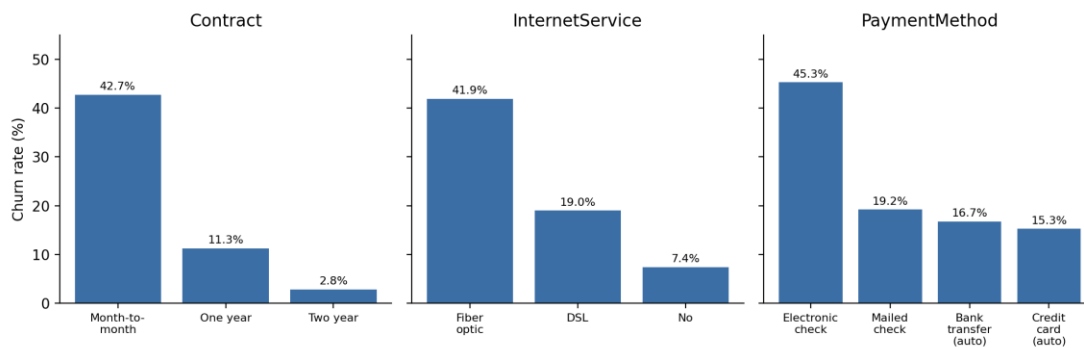


Figure 2. *Churn rate by contract type, internet service and payment method.*

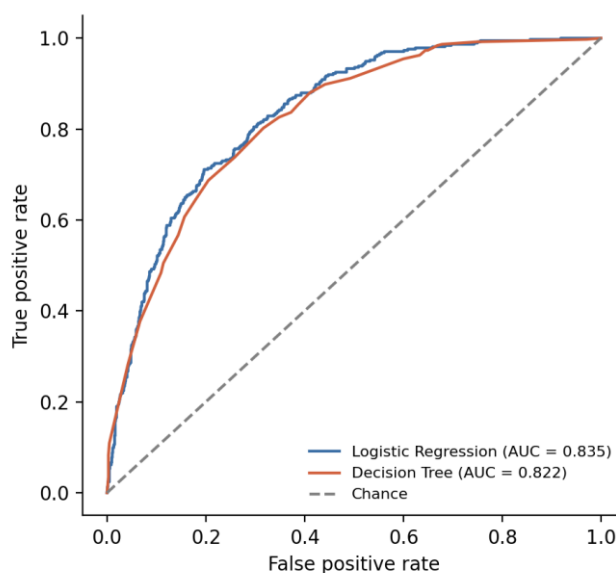
4.2 Model comparison (RQ2)

The tree selected by cross-validation had a maximum depth of 5 and at least 50 customers per leaf. On the test set, logistic regression scored slightly higher on every measure except recall (Table 2, Figure 3): accuracy 79.7% vs. 77.9%, AUC 0.835 vs. 0.822. Both exceeded the 73.4% baseline. Cross-validated AUC on the training set showed the same order (0.846 vs. 0.832). The bootstrap 95% intervals for AUC were [0.812, 0.857] for logistic regression and [0.799, 0.844] for the tree. The AUC difference of 0.013 had a 95% interval of [-0.001, 0.026], and the McNemar test (logistic regression alone correct on 102 customers, the tree alone on 76) gave $p = .061$. The difference between the models was therefore not statistically significant at the .05 level.

Table 2. Test-set performance ($n = 1,407$; churn is the positive class)

Model and threshold	Accuracy	Precision	Recall	F1	AUC
Logistic regression, 0.5	79.7%	63.4%	56.4%	0.597	0.835
Decision tree, 0.5	77.9%	58.7%	56.7%	0.577	0.822
Logistic regression, 0.3	74.0%	50.7%	75.9%	0.608	0.835
Decision tree, 0.3	71.5%	47.8%	80.2%	0.599	0.822

Note. Baseline accuracy (always predict no churn) = 73.4%. AUC does not depend on the threshold.

**Figure 3.** ROC curves of the two models on the test set.

In the logistic regression, the largest odds ratios were for fiber optic service, streaming TV, streaming movies and phone service, and the smallest for having no internet service, monthly charges, tenure and a two-year contract (Table 3). The negative conditional coefficient for monthly charges, although churners pay more on average, reflects overlap with the services customers subscribe to. In the decision tree, tenure and fiber optic service together accounted for about 79% of total importance.

Table 3. *Largest predictors in each model*

Predictor	Logistic regression odds ratio	Decision tree importance
InternetService = No	0.07	0.04
InternetService = Fiber optic	14.49	0.34
MonthlyCharges	0.10	0.03
tenure	0.24	0.45
Contract = Two year	0.24	0.02
StreamingTV = Yes	2.70	< .01
StreamingMovies = Yes	2.63	< .01
PhoneService = Yes	2.43	0.01

Note. Eight largest coefficients ranked by absolute value. Numeric predictors are standardised (odds ratio per one SD). Reference categories: month-to-month contract, DSL internet.

4.3 Effect of the threshold (RQ3)

Lowering the threshold from 0.5 to 0.3 raised recall from 56.4% to 75.9% for logistic regression and from 56.7% to 80.2% for the tree, while precision fell to 50.7% and 47.8% and accuracy to 74.0% and 71.5% (Table 2). Balanced class weights gave similar results (recall 79.7% and 78.9%; precision 48.9% and 47.7%). Under these alternative rules the two models performed almost identically.

Discussion and Conclusion:

To answer RQ1, churn was with a strong association with contract type, internet service and payment method, and churners had far shorter tenure (median 10 vs. 38 months). In answer to RQ2, logistic regression was slightly better than the decision tree (accuracy 79.7% vs. 77.9%; AUC 0.835 vs. 0.822), and both beat the 73.4% baseline, but the difference was not statistically significant (McNemar $p = .061$). Reporting intervals and a paired test prevents over-reading such small gaps (Dietterich, 1998). In answer to RQ3, the threshold mattered more than the model: at 0.3, logistic regression's recall rose from 56.4% to 75.9% while precision fell from 63.4% to 50.7%. A single accuracy figure is therefore an incomplete summary for an imbalanced outcome (Saito & Rehmsmeier, 2015).

For a retention team, these associations suggest hypotheses to test rather than proven causes. Month-to-month customers were 55.1% of the base but 88.6% of churners, so incentives for longer contracts are an obvious lever. Churn is concentrated early: 55.5% of churners left within 12 months, and customers with 12 months' tenure or less churned at 47.7% vs. 17.1% for others, pointing to onboarding. Customers combining a month-to-month contract, fiber optic service and electronic-check payment (1,307 customers, 18.6% of the sample) churned at 60.4% and are a natural priority group. At a threshold of 0.5 the logistic regression flags 333 test customers and

captures 211 of the 374 who left; at 0.3 it flags 560 and captures 284. The better setting depends on the cost of a retention offer relative to the revenue lost when a customer leaves.

Limitations should be kept in mind. The data are a fictional IBM sample from one company at one point in time, so results may not generalise; there is no information on reasons for leaving or the cost of retention offers, so financial value could not be estimated; only two model families and a limited tuning grid were used; correlated predictors (for example, monthly charges and services) complicate the reading of individual coefficients; the results rest on one train-test split; and the analysis shows associations, not causes. Future work could compare these models with bagging, boosting and random forests (Lemmens & Croux, 2006), use real operator data over several periods for time-to-churn analysis, evaluate models in monetary terms with retention costs and customer lifetime value (Verbeke et al., 2012), and test the suggested interventions in field experiments.

References

- 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.
<https://doi.org/10.1016/j.telpol.2006.09.006>
- Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrifft, R. (2018). In pursuit of enhanced customer retention management: Review, key issues, and future directions. *Customer Needs and Solutions*, 5(1–2), 65–81. <https://doi.org/10.1007/s40547-017-0080-0>
- Athari, S. A., Saliba, C., Khalife, D., & Salameh-Ayanian, M. (2023). The role of country governance in achieving the banking sector's sustainability in vulnerable environments: New insight from emerging economies. *Sustainability*, 15(13), 10538.
- Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The Influence of a Data-Driven Culture on Product Development and Organizational Success through the Use of Business Analytics. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 15(2), 123-134.
- Bejjani, G., Toros, E., Elia, J., & Jabbour Al Maalouf, N. (2024). Technology analytic adeptness for examining the interrelatedness between the main macroeconomic variables in Lebanon employing a vector autoregression (VAR). *Journal of Decision Systems*, 1-30.
- BlastChar. (n.d.). Telco customer churn [Data set]. Kaggle.
<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

- Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. <https://doi.org/10.1016/j.eswa.2008.05.027>
- Chaanine, N., Mhanna, R. Y., Verdie, J. F., & Beainy, R. (2023). Leveraging business experience and knowledge to boost entrepreneurial achievement. *International Journal of Work Innovation*, 4(2), 104-115.
- Daou, L., Sayegh, E., Atallah, E., Jabbour Al Maalouf, N., & Sarkis, N. (2025). Greenwashing as a Barrier to Sustainable Marketing: Expectation Disconfirmation, Confusion, and Brand–Consumer Relationships. *Sustainability*, 17(22), 9979. <https://doi.org/10.3390/su17229979>
- De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269(2), 760–772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- Diab, B., & El Hajj, M. (2024). Ethics in the age of artificial intelligence: unveiling challenges and opportunities in business culture. *Cogent Business & Management*, 11(1). <https://doi.org/10.1080/23311975.2024.2408440>
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7), 1895–1923. <https://doi.org/10.1162/089976698300017197>
- Elia, J., Toros, E., Sawaya, C., & Balouza, M. (2021). Using Altman Z”-score to predict financial distress: Evidence from Lebanese alpha banks. *Management Studies and Economic Systems*, 6(1/2), 47-57.
- Jabbour Al Maalouf, N., Ficani, A., & Elia, J. (2026). Drivers of AI adoption in auditing: a quantitative exploration of perceived benefits and trust across auditor profiles. *Cogent Business & Management*, 13(1). <https://doi.org/10.1080/23311975.2026.2709916>
- Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2), 276–286.
- Mawad, J. L. J., & Freiha, S. S. (2024). The Role of Influencers in Shaping the Economic Decisions of Consumers Using the Logistic Regression Approach—Does the Generation Factor Matter?. *Sustainability*, 16(21), 9546. <https://doi.org/10.3390/su16219546>
- Nasser, A., Yammine, J., Harouny, J., & Khalife, D. (2025). Predicting Economic Quads Through Asset Returns Using Ensemble Machine Learning. *IEEE Access*.
- 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. *Journal of Marketing Research*, 43(2), 204–211. <https://doi.org/10.1509/jmkr.43.2.204>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D.,

- Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Reichheld, F. F., & Sasser, W. E., Jr. (1990). Zero defections: Quality comes to services. *Harvard Business Review*, 68(5), 105–111.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Saber, A. M., & Sassine, M. E. (2022). Reflections on internal control and internal audit roles for ensuring financial transparency toward Lebanese NGOs financial donors. *EuroMed Journal of Management*, 4(3), 186-198.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), Article e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Subrahmanyam, S., & Beainy, R. H. (2027). Smart Decision Support for Uncertainty and Disruptions in Industry 4.0 Operations. In *Smart Operational Research Solutions in the Era of Industry 4.0 and Sustainability* (pp. 167-198). IGI Global Scientific Publishing.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024, April). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9.
- 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. *European Journal of Operational Research*, 218(1), 211–229. <https://doi.org/10.1016/j.ejor.2011.09.031>