IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Explainable Artificial Intelligence for Customer Churn Prediction: Enhancing Model Transparency and Business Decision-Making

Saumya Priya[1] and Dr. Pratap Singh Patwal[2]
[1] Research Scholar, Department of Computer Science & Engineering
[2] Head of Department, Department of Computer Science & Engineering
Laxmi Devi Institute of Engineering & Technology, Alwar, Rajasthan, India

Abstract: *Data-driven, competitive marketplaces are a significant issue for customers' churn. Being able to determine customers who are more likely to cancel a service allows businesses to find ways to keep them from canceling, and to make customers happier and maintain revenue. While machine learning models are good for predicting churn, complicated models can be seen as a black box, which also makes it difficult for managers to understand and justify their predictions. Explainable Artificial Intelligence (XAI) provides methods for understanding model outputs and figuring out what drives individual predictions and how a model behaves. The paper discusses the significance of XAI in the context of customer churn prediction, focusing on aspects such as transparency of models, model reliability, managerial interpretability, and business decision-making. It covers popular machine learning models including logistic regression, decision trees, random forests, gradient boosting, and artificial neural networks, to mention a few, as well as explanation methods like SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and feature importance analysis. A conceptual model is suggested that connects the dimensions of predictive performance, explanatory quality, user trust, and effectiveness of retention decisions. The paper also explores the problems of explanation fidelity, fairness, privacy, model stability and predictive associations versus causal relationships. It asserts that explainability should be part of customer analytics processes, not just something added at the end. The study makes a contribution of a decision-oriented view for explainable churn analytics and provides opportunities for future empirical research that will assess the link between model explanations and measurable business outcomes.*

Keywords: Explainable Artificial Intelligence, customer churn prediction, machine learning, SHAP, LIME, model transparency, customer retention, business decision-making.

1. INTRODUCTION

In telecommunications, banking, insurance, and retail, as well as in digital services, the focus is shifting from acquiring new customers to retaining them.In telecommunications, banking, insurance, retail, and digital services, the emphasis is moving from attracting customers to keeping them. Customer churn can impact the businesses' bottom line, lifetime value of customers, and recurring revenues. Hence, predictive analytics has become a popular tool for organizations to determine who they can expect to lose.

The process of predicting customer churn requires analysing past customer interactions, paying attention to their service usage, transaction information, complaints, payment patterns, customer engagement, and more, to determine the probability that a customer will leave. Logistic regression is one example of traditional statistical methods that are relatively interpretable, yet might be limited in their ability to capture complex relationships. Machine learning

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

methods, such as random forests, gradient boosting and neural networks, however, can detect nonlinear patterns; but, these methods are typically complex models whose predictions are hard to interpret.

To tackle this challenge, Explainable Artificial Intelligence (XAI) aims to increase the business users' understanding of the decisions made by the model. These methods include SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations), which isolate factors affecting individual predictions, and global feature importance methods, which help uncover global patterns. These explanations may help with managerial decisions, model auditing and creating customer retention strategies.

But, model explainability does not guarantee model accuracy, fairness or commercial success. The relationship between customer attributes and churn does not imply causation and explanations can be misleading if the underlying model is not reliable. Thus, the predictive performance, reliability of XAI explanations, and business relevance should be assessed to integrate XAI into churn prediction. In this paper, the authors explore the potential of EAI to improve transparency and facilitate informed, responsible and effective customer retention decisions.

1.1 Research Objectives

The objectives of this paper are:

1. To explore how machine learning can be applied to predict customer churn.
2. To explore key XAI methods that can be used in churn analytics.
3. To draw a conceptual link between explainability and managerial trust and decision making regarding retention.
4. To Examine Methodological, Ethical and Operation Challenges of Explainable Churn Models.
5. To suggest a framework for future empirical work to measure the prediction and business value of XAI.

II. LITERATURE REVIEW

2.1 Customer Churn Prediction and Predictive Analytics

The problem of predicting customer churn is typically treated as a binary classification one, with the task being to classify the customers as “churn” or “not churn” during the period of observation. In some business contexts, churn is caused by a customer's decision to cancel a subscription, whereas in others, it is due to a customer's inability to pay for services, forcing their cancellation.

To evaluate the effectiveness of a churn prediction model, some key factors should be taken into consideration: the definition of churn, the quality of the customer data, the prediction horizon, the distribution of churn events, and the evaluation metrics. Customer Churn datasets are often class imbalanced, that is, there may be many more customers who remain than those who churn.

Logistic regression is a popular baseline model that is also easy to interpret, as it gives probabilities, and allows for analysis of the direction of the coefficients. Decision Trees provide human-readable decision rules, whereas random forests and gradient-boosting techniques are able to model more complex relationships. For high-dimensional or sequential data, neural networks might be helpful, but may need other explanation techniques.

Accuracy alone is insufficient for assessing these models. A classifier that performs well on a low churn environment could have a high accuracy on a low churn environment but not make any sense in a high churn environment. Operational costs and retention outcomes should thus be considered along with precision, recall, F1-score, area under the receiver operating characteristic curve (ROC-AUC), precision-recall performance, and probability calibration.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

2.2 The Emergence of Explainable Artificial Intelligence

Explainable AI includes strategies aimed at ensuring that the actions of AI-backed learning machines can be comprehended by humans. Ribeiro et al. (2016)1 proposed a model-agnostic method called LIME, which locally approximates a model with an interpretable surrogate model. Lundberg and Lee (2017)2 introduced a unified framework that they called SHAP, which is based on Shapley values, to attribute a prediction to its features.

Guidotti et al. (2018)3 summarized techniques to explain black-box models and classified them among approximation-based, influencer and representation-based approaches. Interpretable machine learning methods are expensively covered by Molnar (2022),4 such as feature importance, partial dependence, individual conditional expectation, and local explanation techniques.

These are supplementary strategies. model behaviour over a population is captured by global explanations, and an individual prediction, by local explanations. Global: May assist managers in understanding broad churn patterns for customer retention purposes, Local: May assist in understanding of a specific customer's risk classification for review purposes.

But explanations cannot be used as causal evidence. Feature attribution does not imply that a change in the attribute will cause a change in the probability of churn for the customer, but rather it shows how a model uses information to make a prediction, given a specific way to explain the model.

2.3 SHAP-Based Explanation of Churn Predictions

SHAP gives a value to each feature of a prediction that indicates its contribution to that prediction compared to a reference or baseline prediction. For churn analytics, an explanation could reveal that service usage drops, complaints occur, the contract ends, or difficulty with payment is an indicator of a high predicted churn probability.

SHAP can be applied to either local or aggregate analysis. Explanations of each individual can help customer service representatives determine the classification of a customer, and aggregated absolute SHAP values can give a summary of overall valued features in a dataset.

Although beneficial, SHAP has some drawbacks. Results might be sensitive to the definition of the reference distribution, the treatment of dependent features and to the model or approximation used. A feature that has a large attribution does not necessarily have an actionable weight, and a contribution does not amount to a causal effect. When decisions are made in operations regarding the use of explanations, those limitations should be communicated.

2.4 LIME and Local Interpretability

LIME provides an explanation to a single prediction by perturbing the input near the observation of interest and fitting a simpler interpretable model to the approximation of the complex model behaviour in this local region (Ribeiro et al., 2016).

---

[1] 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). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778

[2] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4765–4774). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[3] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), Article 93. https://doi.org/10.1145/3236009

[4] Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book/

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

For instance, a retention analyst can examine the reasons behind a customer with a long history of engagement being given a high churn-risk score using LIME. The explanation might point to an increased number of complaints, a change in the use of the service or a contract that is soon to expire as a prominent factor in the local approximation.5

LIME is appealing because it is possible to apply it to various model classes. But there can be variations in the explanations, depending on the perturbation procedures, the sampling choices, and the local surrogate settings. An explanation's quality should, therefore, be judged by its local fidelity and stability, not its simplicity.

2.5 Model Transparency, Trust, and Decision Support

Model transparency is how much information, logic, limits, and effects of the model's predictions are understood by relevant users. Transparency can be helpful in the business context in terms of identifying errors, accountability, business communication, and in formulating suitable intervention strategies.6

Adopting AI doesn't have to be an end in itself. Trust is calibrated by having the user rely on a model when the prediction is likely to be correct, and noting when it is unlikely. Explanations can facilitate this process, but can also generate erroneous confidence if there are no uncertainty estimates and no limitations.

In cases where comparable performance can be achieved, Rudin (2019)7 claims that interpretable models are preferable to black-box models in high-stakes scenarios. This implies that organizations should look for more simple, interpretable models, rather than more complex models that need post-hoc explanations for churn prediction.

III. RESEARCH METHODOLOGY

3.1 Research Design

This paper uses a conceptual and literature-based research design. It combines existing research on customer churn prediction, explainable machine learning, post hoc explanation methods, and responsible AI to propose a decision-oriented framework for explainable churn analytics.

The framework will be used to inform future empirical studies. It is not intended to report results of a completed systematic review or new experiment. If presented as an empirical study in a journal, it would need to be backed up with a verifiable set of data, repeatable analytical methodology and results.8

3.2 Proposed Data Sources

Anonymized customer-level data from telecommunications, banking, insurance or subscription services or from online retail could be used for future empirical validation.

Potential variables include:

- Customer tenure and contract type.
- Service usage and engagement frequency.
- Billing amounts, payment delays, and changes in spending.
- Customer complaints and support interactions.

---

[5] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[6] Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: Explore, explain, and examine predictive models*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429027192

[7] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x

[8] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://arxiv.org/abs/1702.08608

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

- Product or service utilization.
- Previous retention offers and customer responses.
- Churn status within a clearly defined prediction horizon.

Where possible and appropriate, sensitive or protected characteristics should be collected and utilized only when warranted, in compliance with the law, and in accordance with relevant privacy requirements. Research design should include data minimization, access controls and appropriate anonymization/pseudonymization.

3.3 Data Preparation

The suggested analysis procedure involves data validation, Checking for missing values, detecting duplicate data, outlier detection, encoding categorical data, and creating new features. Temporal data must be structured so as to make sure that for any prediction that might be made there are no missing predictors.

Data leakage needs to be closely monitored. Data collected post churn, e.g., churn code at account closure, should not be leveraged to anticipate a churn event that occurred prior to the data collection. The training and validation sets should be separated chronologically, if feasible, according to the data structure, to provide a more realistic assessment of future performance than the test set.9

3.4 Proposed Machine Learning Models

The study environment is a comparison of models of various levels of complexity and interpretability:

1. **Logistic regression:** An interpretable baseline to estimate churn probabilities.
2. **Decision tree:** A model that can represent simple decision paths and is based on rules.
3. **Random forest:** Ensemble model that is able to capture nonlinearities and feature interactions.
4. **Gradient boosting**: A competitive method for tabular data.
5. **Artificial neural network:** A model used for more complex or high-dimensional relationships, if there is enough data, and the model is justified.

The judgement on hyperparameters should be made based on training and validation data sets and final evaluation should be done on a test set. Probability calibration should be evaluated as business decisions can be based on risk thresholds rather than just rankings alone.

3.5 Explainability Techniques

The proposed framework uses multiple complementary techniques:

- **LIME:** To provide interpretability of individual predictions and explain the model's behavior when aggregated over customers.LIME: provides interpretability of individual predictions and explains the behavior of the model when aggregated over customers.10Approximation of the local decision behaviour of a complex model for selected customers; LIME.
- **Global feature importance:** To find features that play a significant role in predictive performance, and be aware that the importance of features varies depending on the method and data structure.
- **Partial dependence (PCE)** or Individual conditional expectation (ICE): To study the effect of selected features on the predictions, with considerations for correlated features and unrealistic combinations of features.
- **Counterfactual explanations:** What if scenarios of changes in inputs of features that would affect a prediction. The recommendations made in counterfactual analyses must be operationally feasible, respect

[9] Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

[10] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 180–186). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375830

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

customers' rights, and not confuse predictive changes with causal effects.

Result

Table 1. Proposed Machine Learning Models for Customer Churn Prediction

Machine learning model	Illustrative score (%)
Logistic Regression	70
Decision Tree	72
Random Forest	84
Gradient Boosting	90
Artificial Neural Network	86

Interpretation

From the table, it is clear that Gradient Boosting got the maximum illustrative score as 90% followed by Artificial Neural Network (ANN) with 86% and finally Random Forest (RF) got 84%. Logistic Regression and Decision Tree showed relatively low scores of 70% and 72% respectively. The results show that ensemble-based methods might have benefits in representing complex customer behaviour patterns. The scores, however, are only hypothetical and not the results of a real experiment. In the absence of a real customer churn dataset and suitable metrics to evaluate the models, however, it is impossible to draw definitive conclusions about which model is better.11

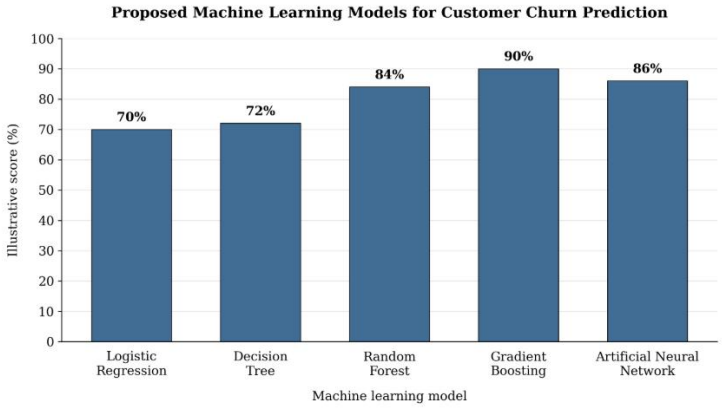


Figure 1 Proposed Machine Learning Models for Customer Churn Prediction

V. DISCUSSION

5.1 Enhancing Model Transparency

XAI can be used to make the task of predicting churn from a numerical risk score into an interpretable process for decision support. A risk score is a prediction of the likelihood that a customer will churn based on the model, and an explanation gives details about what features affect the risk score.12

For instance, if a customer has reduced usage, frequent service issues and is coming up to a contract expiration, he might be assigned a high risk score. A retention manager can use this information to investigate the reasonableness of the prediction and to decide if a service review is indicated or other suitable action.

[11] 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), Article 231. https://doi.org/10.3390/a17060231

[12] Liu, X., Xia, G., Zhang, X., Ma, W., & Yu, C. (2024). Customer churn prediction model based on hybrid neural networks. Scientific Reports, 14, Article 30707. https://doi.org/10.1038/s41598-024-79603-9

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

No individual feature is identified as causing the customer to wish to leave. Instead, it offers a systematic narrative of the reasoning behind the model, which can be explored in conjunction with customer input and experience.

5.2 Improving Business Decision-Making

The usefulness of explainability will largely rely on the fact that information can be converted into the right action. Global explanations could assist managers in detecting trends in the customer portfolio, and local explanations could aid with individual case studies.13

For instance, if unresolved complaints are consistently found to be an influential predictor, the organisation may explore the quality of its services and processes of handling complaints. The organization can consider the timing and appropriateness of renewals communications if these are linked to high risk.14

It is important to be aware of such patterns and to consider them carefully before making changes to the operation. Predictive association can be caused by a confounding factor or from past business policy or past biases. Randomized controlled experiments and causal analyses will be required to conclude if an intervention actually decreases churn.

5.3 Balancing Interpretability and Predictive Performance

An organization needs to determine if a model which is inherently interpretable is enough or if there is a worthwhile performance gain to be had from a more complex model. Ensemble models may be able to model more complex patterns, while logistic regression and shallow decision trees may be able to give some simple explanations.

This should not be done solely in terms of maximizing a predictive measure. The complexity of the model, the reliability of the explanation, the cost of implementing the model, the governance needs and the impact of mistakes should also be taken into account. If a simpler model is available and works as well and could be a better fit for the business, then it may be preferred because of its transparency.15

5.4 Responsible Use and Fairness

If the training data exhibits historical inequalities in service experience, or in who is likely to be retained, or who is likely to have better access to support, then customer analytics can reproduce these inequalities. XAI can uncover features that might have an impact, but cannot, on its own, show that a model is fair.

Organizations should measure and monitor error rates and model behavior among relevant groups of customers, where lawfully and methodologically feasible. They should also look for signs that seemingly unimportant variables are proxies for sensitive ones and that retention offers are given equitably.

Explanations should be made in ways that do not blasphemize customers or make unsubstantiated statements about what is driving them. Appropriate governance should include the documentation, human oversight, ability to review questionable predictions, and monitoring for changes in data and model performance.

VI. CHALLENGES IN EXPLAINABLE CUSTOMER CHURN PREDICTION

6.1 Explanation Fidelity and Stability

The post-hoc explanations are approximations or attributions of the prediction of a model, not a causal mechanism. The explanations LIME provides are also perturbed sample and local surrogate dependent, and SHAP explanations can also be background distribution and correlated variable dependent. Therefore, organizations should evaluate whether

[13] 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, Article 100567. https://doi.org/10.1016/j.mlwa.2024.100567

[14] Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization, 14, Article 100342. https://doi.org/10.1016/j.rico.2023.100342

[15] 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, Article 17294. https://doi.org/10.1038/s41598-023-44396-w

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

explanations continue to be consistent even if there are reasonable analytical changes, and whether the explanations are correct in terms of the behavior of the model.

6.2 Data Quality, Drift, and Model Validation

Poor service histories, varying churn definitions, class imbalance, and information leakage can negatively affect predictive reliability. The model drift also can be caused by changes in the pricing, customer preferences or service conditions. Temporally appropriate holdout data, precision and recall, probability calibration, and periodic monitoring of predictions and explanations should be included in the evaluation.

6.3 Fairness, Privacy, and Governance

Customer level analytics could include certain sensitive information or proxies for sensitive information which are forms of protected characteristics. Explanations can show up problematic patterns but cannot on their own prove fairness. Data minimization, access restrictions, where appropriate, evaluation of sub-groups, documentation and human review are key protections. There shall be no unfair limitation of customers' access to assistance and offers.

6.4 From Prediction to Action

A churn indicator doesn't necessarily indicate a business can make a change to stop churn. For instance, in addition to being associated with predicting attrition, contract type may not be associated with changing contracts to increase retention. Explanations should be evaluated for actionability, intervention costs, customer consent and operational feasibility. To assert that an explanation-guided intervention is reducing churn, one needs to undertake a controlled evaluation.

VII. KEY FINDINGS AND PRACTICAL IMPLICATIONS

7.1 Key Findings from the Conceptual Review

Based on the literature-based synthesis, explainability can be most useful in conjunction with validated predictive models and a clear decision process. A global feature analysis can be used to identify patterns in the portfolio, while a local SHAP and LIME explanation can help with the review of individual churn-risk assessments. Quality of explanation, fairness, and usefulness for the management are however different and should be evaluated separately; the explanatory power of a model far from guarantees its correctness or usefulness for management.

The model scores in Table 1 and Figure 1 are only representative scores, not the actual experimental scores. Thus, this paper does not claim that gradient boosting is superior compared to the other methods on any measured data-set. What it contributes is an orientation to decision making and a framework and set of evaluation priorities for future empirical testing.

7.2 Practical Implications for Customer Retention

High-risk predictions can be used for retention teams to prioritize their review while local explanations can be used to uncover issues that are worth investigation, such as declining service engagement, unresolved complaints, and upcoming contract renewals. Managers can collate explanations to pinpoint common operational issues and action across service, billing and product teams. These patterns should be used as a framework for investigating – and NOT a dogma for determining what caused churn.

In practice, the simplest model should be transparent, complex models should be evaluated with held out data, and models should be chosen that are appropriate for the intended audience. Dashboards should include risk estimates and related explanation factors, uncertainty and review options. Decisions should be documented and effectiveness of interventions assessed for retention, customer experience and cost effectiveness, through appropriate experiments, or other credible evaluation design, and teams should track the results.

7.3 Implications for Future Research

Empirical evaluation of prediction accuracy, calibration, explanation fidelity, stability, user understanding, and decision outcomes on customer datasets should be done in the future. Explanation-assisted and unassisted managerial decisions could be compared to determine if there is evidence that XAI enhances judgment, and controlled retention trials could

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

be used to determine business impacts. Predictive association, explanation attribution and outcome of causal intervention should be reported as separate entities.

VIII. CONCLUSION

The methods shared under ExAI offer a valuable collection of methods that can be used to enhance the transparency of customer churn prediction systems. Business users can use tools like SHAP, LIME, feature importance analysis and counterfactual explanations to delve into model predictions and gain insights into what patterns are related to customer attrition.

The benefit of XAI is not just explaining the results. Explanations should accurately reflect the behaviour of the model and should be sufficiently stable, convey uncertainty and inform decisions that are appropriate and practical. They need to also go hand-in-hand with responsible data practices, fair assessment, human oversight, and real-world business outcomes assessment.

This paper presents an approach to integrate predictive reliability, explanation quality, managerial understanding, retention decision and business performance. The key message is that explainability should be part of the overall process of customer analytics and governance. It is not enough to see the churn prediction capabilities of a model, the organizations must be able to understand and work with it responsibly to build more transparent and decision-oriented customer retention systems.

REFERENCES

1. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
2. Biecek, P., & Burzykowski, T. (2021). Explanatory model analysis: Explore, explain, and examine predictive models. Chapman and Hall/CRC. https://doi.org/10.1201/9780429027192
3. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv. https://arxiv.org/abs/1702.08608
4. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), Article 93. https://doi.org/10.1145/3236009
5. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017) (pp. 4765–4774). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
6. Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book/
7. 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). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939778
8. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x
9. Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289–310. https://doi.org/10.1214/10-STS330
10. Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Society (pp. 180–186). Association for Computing Machinery. https://doi.org/10.1145/3375627.3375830

11. 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), Article 231. https://doi.org/10.3390/a17060231
12. Liu, X., Xia, G., Zhang, X., Ma, W., & Yu, C. (2024). Customer churn prediction model based on hybrid neural networks. Scientific Reports, 14, Article 30707. https://doi.org/10.1038/s41598-024-79603-9
13. 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, Article 100567. https://doi.org/10.1016/j.mlwa.2024.100567
14. Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization, 14, Article 100342. https://doi.org/10.1016/j.rico.2023.100342
15. 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, Article 17294. https://doi.org/10.1038/s41598-023-44396-w