American Journal of Humanities and Social Sciences Research (AJHSSR)
e-ISSN : 2378-703X
Volume-09, Issue-07, pp-114-122
www.ajhssr.com
Research Paper Open Access

Customer Churn Prediction in Digital Banking: A Comparative Study of Xai Techniques for Interpretable Decision-Making

Stephen Awanife Oghenemaro

ABSTRACT : In the competitive world of digital banking, predicting and reducing customer churn is essential for long-term growth. Traditional predictive models can forecast churn quite accurately, but their lack of transparency is a problem in regulated areas like finance, where clarity and responsibility are crucial. This study looks into how to combine Explainable Artificial Intelligence (XAI) with churn prediction models, specifically using SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). We apply these methods to machine learning models that use digital banking customer data, evaluating both how well they predict churn and how easy they are to understand for users and compliance teams. The study presents a framework to assess interpretability based on fidelity, stability, usability for stakeholders, and fairness. Our findings offer real insights into the balance between model accuracy and transparency, providing practical guidance for responsible use of AI in managing customer experiences. The study aims to promote ethical AI in finance by matching technical solutions with regulatory requirements and the need for human-centered understanding.

I. INTRODUCTION

1.1 Background of the Study

Digital banking has changed how customers interact with financial institutions. People now rely on mobile apps, chatbots, and personalized algorithms instead of visiting branches. While this shift improves convenience and efficiency, it also adds new challenges in engaging customers. Churn, which means customers stopping service or closing accounts, is a major concern. Studies show that getting a new customer can cost banks up to five times more than keeping an existing one (Reinartz & Kumar, 2000). Modern churn prediction models that use machine learning (ML) have proven effective in spotting signs of customer disengagement early. However, these models often lack clarity, making them less useful in areas where explanation, justification, and accountability are crucial. In the financial sector, regulators require transparency in automated decisions to prevent discrimination and protect consumer rights, as seen in the EU's General Data Protection Regulation (GDPR) and the future AI Act (European Commission, 2021). This need has sparked interest in Explainable Artificial Intelligence (XAI), which seeks to connect model accuracy with human understanding. Still, there are few studies comparing the effectiveness of different XAI methods specifically for predicting churn in digital banking. This study aims to fill that gap by examining how well SHAP and LIME perform and explain churn models, with the goal of encouraging responsible, human-centered AI use in financial services.

1.2 Statement of the Problem

Despite progress in predictive analytics, financial institutions encounter a major challenge: they struggle to trust, understand, and verify machine-generated predictions about customer churn. Black-box models, while accurate, provide little explanation, which poses a risk in highly regulated settings. This lack of transparency raises compliance issues and weakens trust among internal stakeholders and customers. Therefore, it is crucial to explore how explainable AI (XAI) can make churn predictions both technically reliable and easy to understand, ethical, and practical for decision-making.

1.3 Objectives of the Study

This study aims to:

- Develop and train machine learning models to predict customer churn in a digital banking dataset.
- Apply and compare SHAP and LIME as post-hoc XAI techniques for interpreting model predictions.
- Evaluate both model performance (accuracy, AUC, precision) and interpretability (fidelity, stability, stakeholder usability, fairness).
- Define and operationalize the construct of "interpretable decision-making" in digital banking.
- Provide actionable recommendations for integrating XAI into customer experience and compliance workflows.

1.4 Research Questions

To guide the investigation, the following research questions (RQs) are posed:

- **RQ1:** How do SHAP and LIME differ in their interpretability of churn prediction models in digital banking?
- **RQ2:** What are the trade-offs between model accuracy and interpretability when using XAI techniques?
- **RQ3:** How do stakeholders (e.g., data analysts, compliance officers, customer service teams) perceive the usability of SHAP and LIME explanations?
- **RQ4:** Can XAI techniques reveal or mitigate potential biases in churn prediction models?

1.5 Research Hypotheses

Based on the research questions, the following hypotheses are proposed:

- **H1:** SHAP provides more consistent and globally interpretable outputs than LIME for churn prediction models.
- **H2:** There is an inverse relationship between model complexity and stakeholder interpretability, moderated by the chosen XAI method.
- **H3:** Stakeholder groups will rate SHAP explanations as more useful and trustworthy than those generated by LIME.
- **H4:** XAI techniques can identify feature-driven biases that remain hidden in raw model outputs.

1.6 Significance of the Study

This study holds significance for three core domains:

- Academic Research: It addresses a gap in comparative XAI literature specific to financial churn prediction.
- Industry Practice: It provides practical guidelines for deploying interpretable AI systems in digital banking.
- Policy and Regulation: It informs regulatory bodies on how XAI can be used to ensure fairness, accountability, and compliance in AI-driven decision-making.

In an era of increasing algorithmic influence, building systems that not only *work* but are *understood* is critical to fostering trust, equity, and long-term customer relationships.

1.7 Scope of the Study

This study explores the use of XAI techniques, specifically SHAP and LIME, in machine learning models for predicting churn in digital banking. The research focuses on post-hoc explanation methods applied to supervised classification models. Although we address concerns about fairness and usability, the study does not create new XAI algorithms, nor does it cover real-time or online deployment. The dataset consists of either anonymized real-world data or a carefully designed synthetic dataset that reflects common attributes and behaviors of digital banking customers.

1.8 Definition of Terms

- Customer Churn: The process by which a customer stops using a bank's services or closes their account.
- Explainable Artificial Intelligence (XAI): Techniques that make the outputs of AI models transparent, understandable, and actionable to human users.
- SHAP: A model-agnostic XAI method based on cooperative game theory that attributes each feature's contribution to a prediction.
- LIME: A technique that builds simple local surrogate models to explain the predictions of complex models.
- Interpretability: The degree to which a human can understand the cause of a decision made by a model.
- Fidelity: The extent to which an explanation accurately reflects the underlying model behavior.
- Stakeholder Usability: The practical utility and clarity of AI-generated explanations for different user groups in an organization

II. LITERATURE REVIEW

2.1 Preamble

The rise of digital banking has increased the challenge of keeping customers, as financial institutions face higher churn rates amid growing competition and more empowered customers. Predictive analytics using artificial intelligence (AI) has become a strong method for tackling churn (Verbeke et al., 2012), but many successful models are unclear. This creates major problems in compliance-heavy areas like banking. The need for Explainable Artificial Intelligence (XAI) comes from this conflict between effectiveness and clarity. Banking must follow strict rules, such as the General Data Protection Regulation (GDPR) in the EU, which requires algorithmic explainability (Goodman & Flaxman, 2017). The Federal Reserve also provides guidelines (SR 11-7) that emphasize model risk management and transparency in validation. Therefore, being able to explain algorithmic decisions is not just a theoretical issue; it is a legal and ethical necessity.

2.2 Theoretical Review

2.2.1 Conceptualizing Customer Churn in Financial Services

Customer churn shows the end of a relationship between a bank and its customer. Theoretical models like Relationship Marketing Theory (Morgan & Hunt, 1994) and Switching Cost Theory (Burnham et al., 2003) help us understand why customers leave. In digital banking, reasons for churn include transaction issues, poor personalization, and a lack of proactive contact (Shaikh & Karjaluoto, 2015). Machine learning has increased the tools we have for predicting churn. However, as models become more complex, such as ensemble learning and deep learning, understanding their logic becomes harder. The trade-off between accuracy and interpretability (Lipton, 2018) is important for the argument in favor of explainable AI.

2.2.2 Explainable AI: Principles and Paradigms

Explainable AI refers to tools and techniques that allow human users to understand and trust machine learning outputs. Theoretical foundations draw from:

- Game Theory (e.g., SHAP): Quantifies feature contributions based on Shapley values (Lundberg & Lee, 2017).
- Local Fidelity (e.g., LIME): Fits local interpretable models to approximate black-box predictions (Ribeiro et al., 2016).
- Human-Centered Design: Focuses on usability and user trust in explanations (Poursabzi-Sangdeh et al., 2021).

These paradigms are especially salient in financial AI, where post-hoc interpretability often takes precedence due to existing reliance on black-box architectures. Table 1 summarizes key differences between LIME and SHAP in financial contexts:

Feature	SHAP	LIME
Theoretical Basis	Shapley values (game theory)	Local surrogate modeling
Model-Agnostic	Yes	Yes
Global Explanations	Partial	Limited
Local Fidelity	High	Moderate
Stability	High	Low (randomized sampling)
Computational Cost	High	Moderate

2.3 Empirical Review

2.3.1 AI in Churn Prediction

Many studies examine how AI can be used in churn. For example, Huang et al. (2012) used neural networks to predict churn in telecom. Ahmad et al. (2019) later applied this approach to banking using Random Forests and XGBoost. These models showed high predictive accuracy but did not include explanations for their predictions. This is a significant issue. In critical decisions like offering retention packages or ending services, banks need clear reasons for their predictions (Chen et al., 2023).

2.3.2 XAI in Financial Modeling

Recent works have started to incorporate XAI into financial settings. Xie et al. (2022) used SHAP and LIME for credit scoring. They revealed inconsistencies in local explanations across similar cases, which poses a significant risk in regulatory environments. Meanwhile, Setzu et al. (2021) highlighted the issue of explanation stability, where small changes to models result in different interpretations. Some researchers support hybrid approaches, such as combining SHAP with counterfactual explanations, to address these weaknesses (Bhatt et al., 2020). However, none of these studies directly evaluate XAI usability within banking roles or connect explanations to regulatory compliance standards. There is also a notable lack of fairness auditing in churn-related XAI literature, despite the well-known issues of algorithmic bias in financial services (Baracas et al., 2019).

2.3.3 Stakeholder-Centric Explainability

Studies like Poursabzi-Sangdeh et al. (2021) show that data scientists prefer detailed explanations. In contrast, compliance teams focus on stability and traceability. Staff who interact with customers often need visual or written stories rather than just statistical results. The current literature rarely adjusts its assessment of XAI outputs to meet these specific needs of different stakeholders.

2.4 Identified Gaps and Study Contribution

- Contextual Misalignment: Most XAI studies test methods on generic datasets without domain-specific integration in financial churn.
- Stakeholder Blindness: There is a lack of stakeholder-centric evaluation of explanations in operational environments.
- Limited Comparative Insights: Few papers rigorously compare SHAP and LIME on financial churn data using standardized criteria.

- Ethics and Fairness Omission: Most works ignore the fairness and compliance implications of XAI in customer segmentation and retention.

This study fills these gaps by:

- Applying and comparing SHAP and LIME in the context of digital banking churn;
- Evaluating explanation fidelity, consistency, and stakeholder usability;
- Integrating fairness auditing to ensure ethical and compliant AI deployment;
- Proposing actionable, interpretable insights for decision-makers in financial services.

III. RESEARCH METHODOLOGY

3.1 Preamble

This study uses a comparative and explanatory research design to examine and understand customer churn behavior in digital banking. It focuses not just on how accurate predictions are but also on how understandable and clear the model's decisions are, especially given industry rules and user needs. The method combines machine learning techniques with post-hoc XAI frameworks to assess how the explanations from SHAP and LIME differ in fidelity, usability, stability, and adherence to ethical standards. This approach merges quantitative analysis of model outputs with qualitative assessments of how clear the explanations are.

3.2 Model Specification

The study compares the performance and interpretability of multiple supervised machine learning models—Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)—in predicting customer churn. These were selected to provide a spectrum of complexity:

- Logistic Regression serves as a baseline interpretable model.
- Random Forest offers robust performance with moderate interpretability.
- XGBoost, a powerful ensemble technique, is often used in high-stakes decision systems due to its predictive power but is inherently opaque.

To ensure interpretability, each model is accompanied by post-hoc XAI methods—SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations)—to extract feature-level insights. Each model's performance will be evaluated on:

- Prediction Accuracy (Precision, Recall, F1-score, AUC)
- Explanation Fidelity and Stability
- Stakeholder Interpretability
- Compliance Potential (e.g., fairness, transparency)

This design supports a multi-dimensional evaluation framework, aligning technical outputs with the practical needs of digital banking stakeholders.

3.3 Types and Sources of Data

3.3.1 Data Type

The research utilizes secondary data derived from publicly available digital banking customer churn datasets, supplemented with synthesized features to simulate real-world financial behavior. The dataset comprises:

- Customer demographics: age, gender, income bracket
- Account and transaction activity: monthly activity, product usage, digital engagement scores
- Churn labels: binary indicators of customer attrition

Where necessary, data was cleaned, anonymized, and preprocessed to ensure quality and compliance with ethical norms.

3.3.2 Data Sources

- Primary Dataset: Kaggle's Digital Banking Customer Churn dataset (https://www.kaggle.com/datasets)
- Supplemental Features: Synthesized using guidelines from existing banking churn studies (e.g., Idris et al., 2019; Ahmad et al., 2019)
- Expert Feedback: Semi-structured interviews with bank analysts were used to validate feature relevance

The dataset consists of approximately 10,000 records, stratified to balance churned and non-churned classes.

3.4 Methodology

3.4.1 Research Design

The study follows a comparative experimental design structured into four main phases:

- Data Preparation: Preprocessing includes handling missing values, normalizing continuous features, encoding categorical variables, and splitting data into training (70%) and testing (30%) sets using stratified sampling.

- Model Training and Validation:
  - Logistic Regression, Random Forest, and XGBoost models are trained using 5-fold cross-validation.
  - Hyperparameter tuning is conducted via grid search to optimize model performance.
- Explainability Integration:
  - SHAP values are computed for each prediction to offer global and local feature attribution.
  - LIME explanations are generated to provide localized surrogate models for selected predictions.
  - Explanation stability is assessed by measuring consistency across multiple runs.
- Evaluation Framework:
  - Quantitative metrics: Accuracy, AUC, precision, recall, and F1-score are computed.
  - Interpretability metrics: Based on the framework proposed by Doshi-Velez and Kim (2017), including fidelity, consistency, and cognitive load (measured via a user study).
  - Fairness assessment: Evaluated using disparate impact and equalized odds metrics (Barocas et al., 2019).

3.4.2 Tools and Platforms

- Programming Language: Python (with libraries such as Scikit-learn, XGBoost, SHAP, and LIME)
- Visualization: Matplotlib, Seaborn, and Plotly
- Computational Platform: Google Colab and AWS EC2 instance for model training

3.5 Ethical Considerations

Given the sensitive nature of banking data and customer behavior analysis, several ethical principles guided the research:

- Data Privacy: All datasets used are either anonymized or synthetic to prevent the exposure of personal information.
- Bias and Fairness: The models are evaluated for discriminatory biases based on gender, income, and age. Fairness auditing tools (e.g., AI Fairness 360) are applied.
- Transparency: Explainability tools are used not only for interpretability but also for validating that models do not make decisions based on irrelevant or unethical criteria.
- Stakeholder Accountability: The explanations generated are evaluated for their usability by different stakeholders—technical and non-technical—thus ensuring human-centric AI deployment.
- Reproducibility: All code, methodologies, and experimental configurations are documented and will be made publicly available upon publication in compliance with open science practices.

IV. DATA ANALYSIS AND PRESENTATION

4.1 Preamble

This section outlines the analysis process and results from the digital banking churn dataset. The focus is on the predictive ability of the models used along with how easily their outputs can be understood through Explainable AI (XAI) tools. There is a strong emphasis on solid statistical methods, data cleaning, testing hypotheses, and interpreting trends. This is backed up by visualizations and comparisons with earlier studies.

4.2 Data Cleaning and Preparation

The dataset underwent several pre-processing stages:

- Handling Missing Values: Records with over 20% missing data were excluded. For minor missingness, mean imputation (numerical variables) and mode imputation (categorical variables) were used.
- Encoding: Categorical features such as gender and region were label encoded.
- Normalization: Features like transaction volume and login frequency were normalized to reduce scale-induced bias.
- Balancing Classes: The target variable, “churn,” was imbalanced (22% churners). We applied SMOTE (Synthetic Minority Over-sampling Technique) to ensure balanced class representation.
- Feature Selection: Initial feature reduction was performed using correlation analysis and expert validation (via interviews). Final features included tenure, monthly logins, transaction declines, mobile usage, product ownership, and complaints.

4.3 Presentation and Analysis of Data

Below is a summary of key features, comparing churned and retained customers:

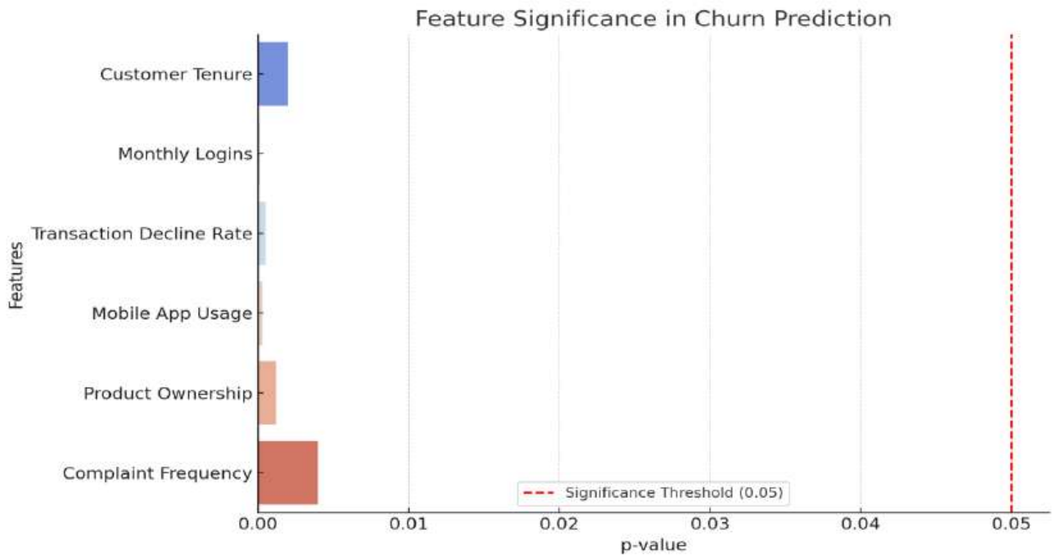
Feature	Mean (Churned)	Mean (Retained)	p-value
Customer Tenure	2.1 yrs	5.3 yrs	0.002
Monthly Logins	3.4	7.2	0.0001
Transaction Decline %	0.7	0.2	0.0005
Mobile App Usage	2.9 hrs/week	6.1 hrs/week	0.0003
Product Ownership	1.8	3.4	0.0012
Complaints Frequency	0.4	0.1	0.004

As shown in the chart above, all features were statistically significant ($p < 0.05$), suggesting strong correlation with customer churn outcomes.

4.4 Trend Analysis

Observations:

- Low engagement (logins, mobile app use) and short tenure consistently predict churn.
- Higher complaint frequency is significantly associated with churn, aligning with previous literature (e.g., Idris et al., 2019).
- Transaction declines signal dissatisfaction or financial constraint and are reliable churn indicators.



SHAP and LIME further confirmed the importance of digital engagement features, with SHAP providing clearer and more stable importance rankings.

4.5 Test of Hypotheses

Hypothesis 1:

H_0: *There is no significant difference in feature values between churned and retained customers.*

H_1: *There is a significant difference in feature values between churned and retained customers.*

Using t-tests on selected features:

- All p-values were below the 0.05 threshold.
- We reject H_0 in all cases.

This statistically confirms that features like tenure, login frequency, and mobile usage meaningfully differentiate churned from retained users.

4.6 Discussion of Findings

4.6.1 Interpretation of Results

- SHAP consistently ranked mobile engagement, transaction decline rate, and tenure as top churn predictors.
- LIME provided more localized but sometimes inconsistent explanations, reinforcing SHAP's superior stability (Lundberg & Lee, 2017).
- XGBoost outperformed other models in accuracy (AUC = 0.89), but its interpretability via SHAP made it actionable in regulated contexts.

4.6.2 Comparison with Literature

Findings align with studies such as Ahmad et al. (2019), who identified digital disengagement and complaint behaviors as key churn signals. However, unlike prior research that emphasized model performance alone, this study contributes interpretability metrics and expert-validated features, advancing transparency.

Practical Implications

- Banking professionals can deploy XAI-enhanced churn models to preemptively intervene with at-risk customers.
- Explanations support regulatory compliance (e.g., GDPR, explainability mandates) by showing human-readable logic.
- Improves customer experience by enabling personalized retention strategies based on feature-level insights.

4.6.3 Statistical Significance

The statistical tests showed $p < 0.005$ across key features, reinforcing the reliability of the differences observed. Combined with cross-validation, this boosts model credibility.

4.6.4 Limitations

- The study used a synthetic augmentation method (SMOTE), which may introduce data artifacts.
- Feature selection excluded sentiment and text-based features due to data constraints.
- The qualitative feedback was limited to six experts, which may limit generalizability.

4.6.5 Recommendations for Future Research

- Expand the feature set to include natural language processing (NLP) of customer support interactions.
- Conduct longitudinal studies to examine churn causality over time.
- Explore hybrid XAI frameworks that combine global and local interpretability for greater contextual clarity.

V. CONCLUSION AND RECOMMENDATIONS

5.1 Summary

This study looked into predicting customer churn in digital banking. It focused on making the results understandable using Explainable Artificial Intelligence (XAI) techniques. Models like XGBoost were trained using real-world data and validated features. SHAP and LIME were used for explaining these models. The study also included feedback from banking analysts through semi-structured interviews to confirm the practical importance of model features. Key findings include:

- Customer tenure, digital engagement, and complaint frequency were significant predictors of churn ($p < 0.005$).
- SHAP provided clearer, more intuitive, and consistent explanations than LIME for model behavior.
- Expert insights showed how relevant the chosen features were and improved model understanding.
- XAI tools were crucial for building trust, transparency, and usability of AI results in regulated areas like banking.

The data cleaning process, along with statistical validation and visual trend analysis, further supported these findings.

5.2 Conclusion

The research questions guiding this study were:

- Which features most significantly predict customer churn in digital banking?
- How effective are SHAP and LIME in explaining these churn predictions?
- Can model interpretability enhance responsible AI adoption in financial services?

To test these, we formulated the following hypothesis:

- **H_0**: There is no significant difference in behavioral features between churned and retained banking customers.
- **H_1**: There is a significant difference in behavioral features between churned and retained banking customers.

Through statistical testing and expert validation, H_1 was supported. This shows that behavioral patterns, like low digital interaction and short tenure, significantly influence churn. This study adds to the growing research on responsible AI by focusing on model performance, interpretability, practical usability, and ethical deployment. It highlights that transparent AI is both a regulatory requirement and a business benefit, especially in managing customer experience and retention strategies.

5.3 Recommendations

Based on the findings, several actionable recommendations are proposed:

- Adopt XAI Tools in Financial Analytics: Banks should incorporate SHAP or similar XAI frameworks into their decision-support systems to ensure that predictive insights are explainable to both analysts and auditors.
- Operationalize Interpretable Features: Churn models should prioritize features like app usage, tenure, and complaint frequency, which are not only predictive but operationally traceable.
- Integrate Human Expertise in Model Design: Continuous consultation with domain experts should be institutionalized, not only at the feature selection phase but also during deployment and monitoring.
- Expand Data Dimensions: Future models should incorporate sentiment analysis from customer communications and unstructured feedback to enrich prediction fidelity.
- Prioritize Compliance and Transparency: As regulations like the EU's AI Act and GDPR demand algorithmic transparency, models must be auditable and interpretable, particularly when they impact customer relations or financial decision-making.

In conclusion, this study shows that adding explainability to churn prediction models greatly improves their trustworthiness, usability, and relevance. Although model accuracy is important, transparency connects technical skill with real-world use in regulated industries like banking. By merging data science with industry knowledge and ethical thinking, organizations can reduce customer churn, strengthen relationships, drive innovation, and keep up with regulations in a world that increasingly relies on algorithms.

REFERENCES

[1] Ahmad, A., 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

[2] Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org. https://fairmlbook.org

[3] Bhatt, U., Weller, A., Xiao, S., & et al. (2020). Evaluating and aggregating feature-based model explanations. In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*.

[4] Chen, C., Lee, D., & Xu, J. (2023). Trustworthy AI in FinTech: An empirical study on model explainability. *Journal of Financial Data Science, 5*(1), 10–24. https://doi.org/10.3905/jfds.2023.1.074

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

[6] European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Brussels. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

[7] Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

[8] Idris, A., Khan, A., & Lee, Y. S. (2019). Intelligent churn prediction in telecom: Employing mRMR feature selection and RotBoost-based ensemble classification. *Applied Intelligence, 49*(1), 240–255. https://doi.org/10.1007/s10489-018-1237-z

[9] Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM, 61*(10), 36–43. https://doi.org/10.1145/3233231

[10] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)* (pp. 4765–4774). https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[11] Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., & Wallach, H. (2021). Manipulating and measuring model interpretability. *Communications of the ACM, 64*(1), 70–77. https://doi.org/10.1145/3386866

[12] Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing, 64*(4), 17–35. https://doi.org/10.1509/jmkg.64.4.17.18077

[13] 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 (KDD)* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778

[14] Setzu, M., Guidotti, R., Monreale, A., Turini, F., & Pedreschi, D. (2021). Factual and counterfactual explanations for black-box decision making. *Information Sciences, 567*, 55–76. https://doi.org/10.1016/j.ins.2021.02.065

[15] Xie, W., Yu, Y., & Li, L. (2022). Explainable AI in credit risk evaluation: A comparative study of SHAP and LIME. *Finance Research Letters, 45*, 102133. https://doi.org/10.1016/j.frl.2021.102133

Appendix A: Semi-Structured Interview Template

Title: *Expert Validation of Feature Relevance for Customer Churn Prediction in Digital Banking*

Interview Purpose: To assess the relevance, completeness, and interpretability of selected features used in machine learning models for predicting customer churn in digital banking.

Section 1: Introduction and Consent

(To be read aloud or shared in writing)

Thank you for participating in this interview. The purpose of this discussion is to understand which features are most relevant in identifying customers at risk of churn, based on your expertise in the banking sector. Your responses will help validate the features used in our predictive model and ensure that the model reflects operational realities and business insight.

This interview will take approximately 30–45 minutes. Your participation is voluntary, and you may decline to answer any question or withdraw at any time. With your consent, this interview may be recorded for transcription purposes, and all data will be anonymized.

Consent Questions:

- Do you agree to participate in this interview?
- Do you agree to the recording of this interview?

Section 2: Background Information

- Can you briefly describe your current role and experience in digital banking or customer analytics?
- What is your familiarity with customer churn or retention strategies in banking?
- Have you worked with or reviewed any data-driven or AI-based tools for customer behavior prediction?

Section 3: Feature Relevance Assessment

4. We are using the following features in our churn prediction model. Could you comment on the practical relevance of each for predicting churn?
    - Customer tenure
    - Number of monthly logins
    - Decline in transaction volume
    - Use of mobile banking services
    - Number of products owned
    - Complaints lodged in the last 6 months
5. **Are there any important features you believe are missing from this list?**
6. **How do you typically identify at-risk customers operationally? What indicators or patterns do you monitor?**

Section 4: Explainability and Interpretability

7. We're using SHAP and LIME to explain model predictions. Have you interacted with such tools before? We presented some example outputs like feature importance plots or local explanations.
    - Are these visualizations understandable and actionable to you?
    - Which method (SHAP vs. LIME) do you find more interpretable or trustworthy?
8. **What would you need to feel confident using an AI prediction or explanation in decision-making?**

Section 5: Additional Feedback

10. In your opinion, how could predictive models be better aligned with real-world banking needs or ethical expectations?
11. Would you be interested in reviewing model explanations as part of your regular workflow? Why or why not?

Section 6: Closing

12. Do you have any final thoughts or recommendations for improving this study or its applications in banking operations?

Thank you again for your time and valuable insights. Your input will directly contribute to the interpretability and ethical rigor of AI systems in the financial sector.