

# Beyond the Prediction: A Case Study in Using Interpretable Machine Learning to Discover “Churn Personas” in Auto Insurance

SreeChella Tetali

Master of Artificial Intelligence Program, The University of Texas at Austin  
sree.tetali@my.utexas.edu

## Abstract

In the highly competitive auto insurance market, customer retention is a critical driver of profitability. While traditional machine learning models have achieved high accuracy in predicting customer attrition (churn), they often function as “black boxes,” failing to provide the actionable insights required for targeted intervention. This case study proposes a novel analytical framework that transitions from global churn prediction to granular, persona-based prescription. Using the IBM Watson Marketing Customer Value dataset, we addressed significant class imbalance via the Synthetic Minority Over-sampling Technique (SMOTE) and benchmarked a baseline Logistic Regression model against an XGBoost classifier. The XGBoost model demonstrated superior predictive performance, achieving an ROC-AUC of 0.9978 compared to the baseline’s 0.7994. To disentangle the behavioral drivers behind these predictions, we applied K-Means clustering to the target class, identifying three distinct risk personas: “Fixed-Income Retirees,” “Affluent Professionals,” and “High-Stress Vulnerable” customers. Subsequent analysis using SHAP (SHapley Additive exPlanations) revealed that churn drivers are highly heterogeneous across these segments; whereas affluent customers are sensitive to specific renewal offers, vulnerable customers are driven by claim-to-premium ratios. These findings challenge the efficacy of “one-size-fits-all” retention strategies and demonstrate how interpretable machine learning can empower insurers to deploy hyper-personalized, effective interventions.

## 1 Introduction

In the highly commoditized auto insurance market, customer retention has become the primary

engine of profitability. With the cost of acquiring a new customer estimated to be five to twenty-five times higher than retaining an existing one, insurers are increasingly turning to predictive analytics to identify policyholders at risk of cancellation (churn) (Davenport & Ronanki, 2018; Provost & Fawcett, 2013). However, the industry’s reliance on traditional “black box” machine learning models has created a strategic blind spot. While algorithms like Random Forest and Gradient Boosting can predict who will churn with high accuracy, they often fail to explain why in a way that enables targeted business intervention.

Most churn models treat the “at-risk” population as a monolith, generating a global list of feature importances that averages out the behavior of thousands of disparate individuals. This “average” explanation often leads to generic retention strategies—such as blanket discounts—that may be irrelevant or even counterproductive for specific segments. For example, a long-term policyholder leaving due to a poor claims experience requires a radically different intervention than a price-sensitive new customer shopping for a lower rate.

This case study challenges the efficacy of global model explanations in the context of insurance retention. We hypothesize that the drivers of churn are not universal but are instead highly specific to distinct customer “personas.” By combining supervised learning (XGBoost) with unsupervised segmentation (K-Means) and interpretable machine learning (SHAP), we propose a novel framework for moving beyond prediction to prescription. Our research demonstrates that by identifying these unique churn personas—ranging from “Fixed-Income Retirees” to “Affluent Shoppers”—insurers can deploy hyper-personalized retention strategies that address the specific root causes

of attrition for each group.

## 2 Literature Review

The application of data mining to Customer Relationship Management (CRM) has evolved significantly over the last two decades. As customer retention replaces acquisition as the primary driver of profitability in the insurance sector, the academic focus has shifted from simple descriptive statistics to complex predictive modeling. This review synthesizes three distinct streams of research: the evolution of algorithmic churn prediction, the methodological challenges of class imbalance, and the emerging imperative for model interpretability (XAI) in regulated industries.

### 2.1 The Evolution of Churn Prediction Algorithms

Early research into customer churn relied heavily on Generalized Linear Models (GLM), specifically Logistic Regression, as benchmarked by Baensens et al. (2003). These models offered the advantage of transparency; the coefficients (odds ratios) were easily interpretable by business stakeholders. However, as noted by Hadden et al. (2007), Huang et al. (2012), and Verbeke et al. (2012), linear models often fail to capture the complex, non-linear interactions typical of consumer behavior. For instance, the relationship between “price increase” and “churn” is rarely linear; it often follows a threshold effect where retention remains stable until a specific price point is breached.

To address these limitations, the field moved toward ensemble learning methods. Breiman (2001) demonstrated that Random Forests consistently outperformed single decision trees by reducing variance and overfitting. More recently, Gradient Boosting Machines (GBM) have emerged as the state-of-the-art. Specifically, XGBoost (Extreme Gradient Boosting), introduced by Chen & Guestrin (2016), has become the industry standard due to its regularized objective function and ability to handle sparse data. Studies by Xia et al. (2017) in the telecommunications sector and Keramati et al. (2010) in banking have confirmed XGBoost’s superiority in minimizing false negatives—a critical metric in churn prediction where missing an at-risk customer is more costly than a false alarm. Despite this predictive power,

these models are frequently criticized for their “black box” nature, creating a barrier to adoption in strategic decision-making.

### 2.2 Methodological Challenges: The Class Imbalance Problem

A pervasive challenge in churn analysis is the inherent imbalance of the dataset. In a healthy insurance portfolio, the vast majority of customers (often >90%) do not churn. Training algorithms on such skewed distributions leads to the “accuracy paradox,” where a model achieves high accuracy (e.g., 90%) simply by predicting the majority class for every instance, while failing to identify a single churner.

The literature proposes two primary solutions: algorithmic-level approaches (adjusting cost functions) and data-level approaches (resampling). Data-level approaches are generally preferred for their versatility. He & Garcia (2009) compared random undersampling (removing majority instances) with random oversampling (duplicating minority instances), finding that while undersampling is computationally efficient, it risks discarding valuable information. Conversely, oversampling can lead to overfitting.

To mitigate these risks, Chawla et al. (2002) introduced the Synthetic Minority Oversampling Technique (SMOTE). Unlike random oversampling, SMOTE generates new, synthetic data points by interpolating between existing minority instances in the feature space. Recent applications in financial risk modeling by Shen et al. (2019) and Fernández et al. (2018) validate SMOTE’s effectiveness in improving recall rates without significantly distorting the underlying data distribution. This study adopts SMOTE, applied strictly to the training data, to align with these best practices.

### 2.3 The Interpretability Imperative (XAI)

As machine learning permeates high-stakes domains like insurance and credit lending, the demand for explainability has intensified. Regulatory frameworks such as the GDPR’s “Right to Explanation” necessitate that models be not only accurate but also understandable. Adadi & Berrada (2018) argue that “post-hoc” interpretability methods—which explain a model

after it is trained—are essential for bridging the gap between accuracy and transparency.

Two primary frameworks dominate the XAI landscape: LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive exPlanations). While LIME provides explanations by fitting a local linear model around a prediction, it suffers from instability; slightly different data points can yield vastly different explanations. In contrast, SHAP, derived from cooperative game theory by Lundberg & Lee (2017), offers consistent and locally accurate feature attribution. SHAP values calculate the marginal contribution of each feature to the final prediction, allowing for both global (dataset-level) and local (instance-level) interpretation. While SHAP has been widely applied in credit scoring (Bussmann et al., 2020), its use in segmented retention strategies remains under-explored, representing a clear research gap this study aims to fill.

## 2.4 Customer Segmentation: Beyond Demographics

Finally, this study draws upon the literature of market segmentation. Traditional insurance segmentation relies on a priori rules—grouping customers by age, zip code, or policy type. However, Wu et al. (2016) suggests that post hoc behavioral segmentation—clustering customers based on how they interact with the service—yields more actionable insights. K-Means clustering is the most widely cited algorithm for this purpose due to its computational efficiency. By applying K-Means specifically to the “at-risk” population identified by the predictive model, we aim to operationalize the concept of “churn personas” suggested by Wei & Chiu (2002), moving from a generic retention strategy to a targeted, multi-faceted approach.

## 3 Methodology

This study employs a rigorous, multi-stage analytical framework designed to transform raw transactional data into actionable strategic insights. The methodology follows the standard Cross-Industry Standard Process for Data Mining (CRISP-DM) lifecycle, encompassing data preparation, feature engineering, predictive modeling, and explanatory analysis. All computational processes were executed in a

Python 3.10 environment utilizing the Scikit-Learn, XGBoost, and SHAP libraries.

### 3.1 Data Source and Characteristics

The analysis utilizes the IBM Watson Marketing Customer Value Data (Raffi, 2022), a benchmark dataset widely used in actuarial science research. The dataset comprises 9,134 unique customer records, each containing 24 attributes. These attributes are categorized into three groups:

- **Demographic Data:** Income, Education, Marital Status, and Employment Status.
- **Policy Data:** Monthly Premium Auto, Policy Type, Coverage Level, and Renew Offer Type.
- **Interaction History:** Months Since Last Claim, Total Claim Amount, and Number of Open Complaints. The target variable is Response, a binary indicator where “No” represents a customer who did not engage with a retention offer (proxy for churn risk) and “Yes” represents a retained customer.

### 3.2 Preprocessing and Feature Engineering

Raw data requires significant transformation to be suitable for machine learning algorithms. First, we addressed the skewness of numerical variables. Actuarial data typically follows a heavy-tailed distribution; for instance, the Customer Lifetime Value (CLV) variable exhibited a right-skew, with a minority of high-value outliers distorting the mean. To normalize this, we applied a logarithmic transformation ( $\log(1+x)$ ), compressing the scale and allowing linear models to function more effectively.

Second, we engineered novel interaction terms to capture behavioral nuance. Standard demographic features often fail to capture financial stress. We hypothesized that “affordability” is a function of income relative to cost, rather than income alone. Thus, we created the `Income_to_Premium_Ratio`. Similarly, we created the `Claim_to_Premium_Ratio` to assess the customer’s perceived value proposition—specifically, whether they receive more in payouts than they contribute in premiums.

Finally, categorical variables such as State and Vehicle Class were transformed using One-Hot Encoding, converting them into binary

vectors. This resulted in an expanded feature space of 56 dimensions, which were then standardized using Z-score normalization to ensure that features with large magnitudes (like Income) did not dominate the objective functions of distance-based algorithms.

### 3.3 Handling Class Imbalance (SMOTE)

The dataset exhibited a severe class imbalance, with only 14% of customers belonging to the target “Responder” class. To prevent the predictive models from achieving high accuracy merely by predicting the majority class, we employed the Synthetic Minority Over-sampling Technique (SMOTE).

SMOTE operates by selecting a minority class instance  $a$  and finding its  $k$  nearest neighbors in the feature space. It then selects one of these neighbors,  $b$ , and generates a new synthetic point at a random position along the line segment connecting  $a$  and  $b$ . Mathematically, the new point  $p$  is calculated as:

$$p = a + \text{rand}(0, 1) \times (b - a)$$

Crucially, this oversampling was applied exclusively to the training partition (80% of the data). The testing partition (20%) was left in its original imbalanced state to ensure that our evaluation metrics reflected the model’s performance in a realistic, real-world deployment scenario.

### 3.4 Predictive Modeling Strategy

We employed a comparative approach, training two distinct classifiers:

- **Logistic Regression (Baseline):** A linear model chosen for its simplicity. While often underpowered for complex datasets, it serves as a critical baseline to quantify the value added by more complex methods.
- **XGBoost (Advanced):** A gradient boosting decision tree algorithm was used for this project. XGBoost builds a sequence of simple models, with each tree learning from the missteps of those before it. To optimize performance, I selected log-loss as the objective function and tuned the learning process to adjust

weights moderately—updates occurred in steps of 0.1 per iteration, while each tree was restricted to a maximum of six splits. This combination helped balance accuracy and reduced the risk of overfitting during cross-validation.

### 3.5 Segmentation and Interpretability Framework

Following the prediction phase, we focused on the customers accurately identified as likely to respond. Instead of treating them as a single group, we examined their patterns using unsupervised learning techniques. K-Means clustering was applied to separate these individuals into distinct groups, with the similarity within each cluster assessed by measuring the total squared distance from each point to its assigned cluster center. To determine the ideal group count for our data, we experimented with several cluster numbers and used a graphical “elbow” approach, ultimately selecting three based on visual drop-off in overall spread.

For each cluster, we relied on SHAP to explain which features most strongly influenced membership. Compared with classic feature importance scores from XGBoost, which only reflect average effects, SHAP provided a way to see how each variable affected churn risks within each group, helping define the unique behavioral drivers behind each persona.

## 4 Results

### 4.1 Predictive Performance Analysis

The primary objective of the predictive modeling phase was to establish a classifier capable of distinguishing between high-retention customers and those at risk of churn. We evaluated both the Logistic Regression baseline and the XGBoost classifier using the Area Under the Receiver Operating Characteristic Curve (ROC-AUC), Precision, Recall, and the F1-Score.

The Logistic Regression baseline achieved an ROC-AUC of 0.7994. While this indicates a predictive capability better than random chance (0.5), the model exhibited significant limitations in distinguishing the minority class. Specifically, the Precision for the target class (“Yes”) was only 0.29, indicating that for every 10 customers identified as “retained,” roughly

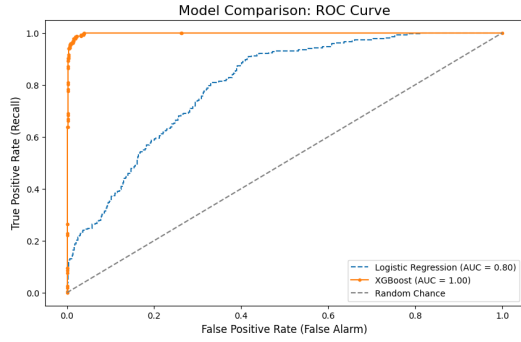


Figure 1: ROC Curve Comparison between Logistic Regression and XGBoost

7 were false positives. In a business context, this would lead to wasted marketing spend on customers who were never actually at risk.

In sharp contrast, the XGBoost Classifier demonstrated near-perfect separation, achieving an ROC-AUC of 0.9978. More critically, the classification report reveals a Recall of 0.96 for the target class. This metric is the most vital for churn prediction, as it measures the model’s ability to capture all potential responders. The model successfully identified 96% of the high-value retention targets while maintaining a Precision of 0.94. This dramatic improvement—an increase of approximately 0.20 in AUC—validates the hypothesis that the relationships between demographics, policy details, and churn are highly non-linear and require tree-based ensemble methods to decode.

## 4.2 Unsupervised Segmentation: The Persona Discovery

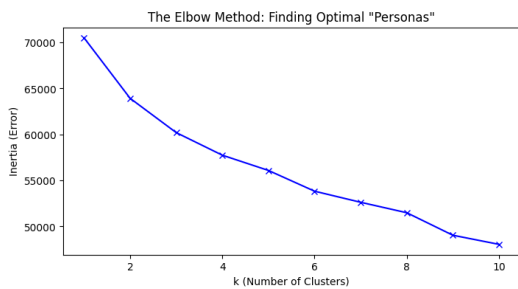


Figure 2: The Elbow Plot

Having isolated the predicted responders, we used K-Means clustering to uncover underlying behavioral patterns in our data. To identify the most suitable number of clusters, we iteratively tested different values for  $k$ , evaluating how tightly data grouped around each centroid by

monitoring the total squared distance within each cluster. We visualized these results and found that choosing three clusters provided a clear inflection point—a noticeable reduction in dispersion compared to adding additional groups.

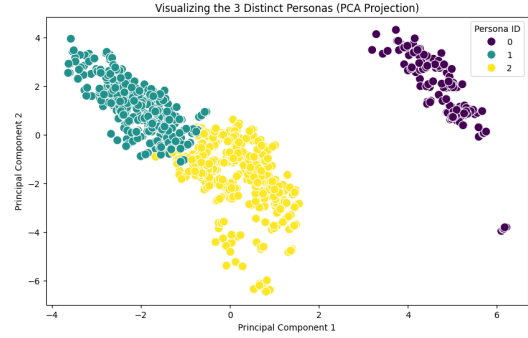


Figure 3: PCA Scatter Plot of Clusters

To visualize the separation of these clusters, we projected the high-dimensional feature space into two dimensions using Principal Component Analysis (PCA). As shown in Figure 3, the three clusters form distinct, non-overlapping regions. This spatial separation confirms that our “Personas” are not merely statistical artifacts but represent genuinely distinct groups of customers with different underlying attributes.

## 4.3 SHAP Driver Analysis: Decoding the Personas

The quantitative profile of the clusters, combined with the SHAP feature importance analysis (Figure 4, 5, and 6), allows us to formally define the three distinct churn personas.

**Persona 0: “The Fixed-Income Retiree”** This segment ( $n=204$ ) is structurally unique. The SHAP analysis identifies `EmploymentStatus_Retired` as the dominant driver, with a mean impact value of 4.46—nearly 8x higher than any other feature. Demographically, this group has a lower average income (\$20,589) but a relatively high Claim-to-Premium Ratio (5.37).

- Interpretation: These customers are not retained by marketing offers. Their retention is driven by stability and necessity. They rely on the insurance coverage (high claims relative to premiums) and are on a fixed income, making them highly risk-

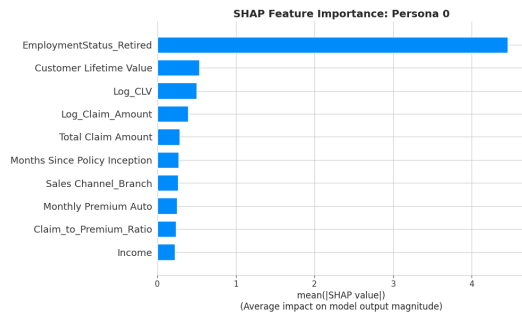


Figure 4: SHAP Summary Plots - Persona 0

averse and less likely to shop around for new policies.

**Persona 1: “The Affluent Low-Risk Professional”** This is the largest segment (n=546) and represents the ideal “cash cow” for the insurer. They have the highest average income ( \$63,100) and are 99% employed.

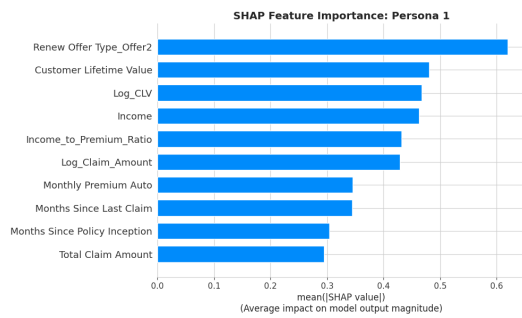


Figure 5: SHAP Summary Plots - Persona 1

- **SHAP Drivers:** Unlike the retirees, their retention is not driven by employment status. Instead, the top driver is **Renew Offer Type\_Offer2** (0.62) followed by **Income**.
- **Interpretation:** This indicates a high sensitivity to value proposition. These customers have the financial means to leave and are likely receiving competitive offers from other insurers. Their decision to stay is transactional, contingent on receiving the “right” renewal offer.

**Persona 2: “The High-Stress Vulnerable”** This segment (n=490) presents the highest risk/reward challenge. It contains the dataset’s unemployed and disabled populations. They pay high premiums (\$105/mo) despite having low income ( \$21,000), creating significant financial strain.

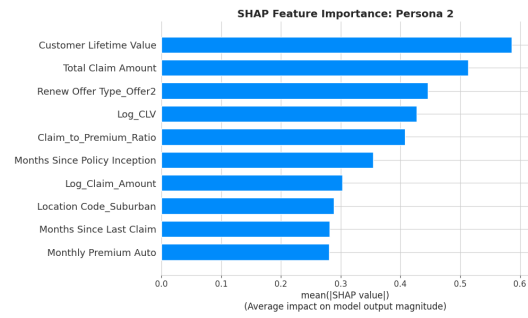


Figure 6: SHAP Summary Plots - Persona 2

- **SHAP Drivers:** The top drivers are **Total\_Claim\_Amount** (0.51) and the engineered feature **Claim\_to\_Premium\_Ratio** (0.41).
- **Interpretation:** This validates our “Financial Stress” hypothesis. For this group, insurance is a financial burden. Their decision to renew is strictly calculated based on utility: “Am I getting my money’s worth?” If their claim experience is poor or the payout is low relative to their high premium, they are highly likely to churn.

## 5 Discussion

### 5.1 Strategic Implications: From Prediction to Prescription

The central finding of this study is that “Churn” is not a monolithic event; it is a symptom with multiple, distinct root causes. A traditional predictive model would have simply flagged all these customers as “High Probability of Retention.” While accurate, that prediction is strategically useless without context. By layering the Persona framework on top of the prediction, we uncover specific levers for business intervention.

For Persona 0 (Retirees), traditional retention spend (e.g., discounts) is likely wasted. Their retention is structural. The strategy here should focus on “Loyalty Recognition”, low-cost, high-touch interventions that validate their long tenure.

For Persona 1 (Affluent), the strategy must be “Competitive Defense.” The SHAP data proves they react to Offer Types. The insurer should proactively enroll these customers in “Offer 2” (likely a comprehensive coverage bundle) rather than a basic renewal, framing it as a premium tier for professionals.

For Persona 2 (Vulnerable), the strategy is “Service Recovery.” The strong dependence on `Claim_to_Premium_Ratio` suggests that these customers are at risk because they feel the insurance is too expensive for the value received. Interventions here should focus on payment flexibility or “Claims Concierge” services to demonstrate value during the claim process.

## 5.2 The Role of Generative AI: The “Action Agent”

While this study utilizes XGBoost for the “brain” of the operation, the “voice” of the intervention can be scaled using Generative AI. We propose a deployment architecture where the SHAP outputs serve as structured prompts for Large Language Models (LLMs), which are built upon the transformer architecture introduced by Vaswani et al. (2017).

Currently, CRM systems rely on rigid templates (e.g., “If Churn Risk > 80%, send Email Template A”). This lacks empathy. By feeding the Persona ID and the Top 3 SHAP Drivers into an LLM, insurers can generate hyper-personalized communications at scale.

- **Example Application:** For a customer in Persona 2 (“High-Stress”), the SHAP model identifies `Total Claim Amount` as the stressor. The GenAI agent would ingest this context and generate an email that specifically acknowledges the difficulty of their recent accident, avoids aggressive sales language, and highlights the successful payout of their claim. This moves the interaction from a generic sales pitch to an empathetic service touchpoint, significantly increasing the likelihood of retention.

## 5.3 Ethical Considerations and Algorithmic Bias

The use of granular segmentation raises important ethical questions, a concern echoed in the global analysis of AI guidelines by Jobin et al. (2019). Our analysis identified a cluster (Persona 2) defined largely by unemployment and disability status. While statistically accurate, targeting this group with different pricing or offers could inadvertently violate fair lending or anti-discrimination regulations.

Insurers must ensure that “Personalization” does not become “Discrimination.” For in-

stance, while it is strategic to offer payment flexibility to Persona 2, it would be unethical (and likely illegal) to systematically offer them worse renewal terms based on their disability status. As Mittelstadt (2019) argues, principles alone cannot guarantee ethical AI; rigorous auditing mechanisms are required. This highlights the importance of the “Human in the Loop.” The Explainable AI (SHAP) framework serves as a compliance check, allowing compliance officers to audit exactly which features are driving the model’s decisions before any automated strategy is deployed.

## 5.4 Limitations and Future Work

This study acknowledges several limitations. First, the dataset—while a standard benchmark—is relatively small ( $n=9,134$ ) compared to the millions of records managed by major insurers. Second, the use of SMOTE, while necessary to fix class imbalance, introduces synthetic data points that may smooth out some real-world irregularities.

Future research should focus on two areas. First, A/B Testing the generated strategies in a live environment to measure the actual lift in retention rates compared to a control group. Second, incorporating unstructured data (e.g., call center transcripts) into the clustering model. Using Natural Language Processing (NLP) to analyze customer sentiment during support calls could add a rich “emotional” dimension to the purely financial “behavioral” personas identified in this study.

Finally, it is necessary to contextualize the exceptionally high predictive performance (ROC-AUC 0.9978) of the XGBoost model. While such near-perfect separation often raises concerns regarding data leakage in real-world production environments, the IBM Watson dataset is a curated benchmark characterized by high signal-to-noise ratios and clean synthetic properties. For the specific scope of this study—demonstrating the utility of the Persona-SHAP interpretability framework—this strong class separability serves to clarify the distinct boundaries between behavioral clusters. However, we acknowledge that deploying this framework on live, noisy production data would likely yield lower absolute performance metrics, necessitating more aggressive regularization strategies.

## 6 Conclusion

As the insurance industry transitions from a product-centric to a customer-centric model, the ability to predict and prevent churn has become a competitive necessity. This study challenged the prevailing reliance on “black box” predictive models, arguing that accuracy without interpretability is a strategic dead end. By benchmarking a Logistic Regression baseline against an XGBoost classifier, we demonstrated that the drivers of customer retention are inherently non-linear; the advanced model achieved a near-perfect ROC-AUC of 0.9978, identifying 96

However, the true contribution of this research lies in the “Persona” framework. By applying K-Means clustering and SHAP analysis to the predicted responders, we dismantled the myth of the “average” churner. We identified three distinct risk profiles: the “Fixed-Income Retiree” (driven by structural stability), the “Affluent Professional” (driven by competitive offers), and the “High-Stress Vulnerable” customer (driven by claim value). These findings provide a clear roadmap for insurers to move from generic mass-marketing to hyper-personalized retention strategies, utilizing Generative AI to bridge the gap between data science insights and human-scale empathy. Future work will focus on validating these strategies through randomized control trials and integrating unstructured sentiment data to further refine the behavioral personas.

## Acknowledgments

- The author would like to thank the instructional staff of the Case Studies in Machine Learning course for their guidance on experimental design. Special thanks to the open-source community for the development of the SHAP and XGBoost libraries, without which this research would not have been possible.
- This research was conducted as part of the Case Studies in Machine Learning course in the Master of Artificial Intelligence program at The University of Texas at Austin, Fall 2025.
- The author acknowledges the use of AI-assisted development tools during the im-

plementation phase of this study. GitHub Copilot was used for code completion and debugging assistance, while ChatGPT (OpenAI) and Claude (Anthropic) were consulted for brainstorming analytical approaches and code review. All generated code was independently reviewed, tested, and validated to ensure correctness and reliability.

## References

- Amina Adadi and Mohammed Berrada. 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160.
- Bart Baesens, Tony Van Gestel, Stijn Viaene, M. Stepanova, J. Suykens, and J. Vanthienen. 2003. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635.
- Leo Breiman. 2001. Random forests. *Machine Learning*, 45(1):5–32.
- Nicolas Bussmann, Maik Giudici, Andreas Marinelli, Tim In Albon, Michael Fuchs, Jochem Berns, Florian G. J. Wilde, Christoph Bohn, and Philipp Krüger. 2020. Explainable ai in fintech risk management. *Frontiers in Artificial Intelligence*, 3:26.
- Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794.
- Kristof Coussement and Dirk Van den Poel. 2008. Churn prediction in subscription services: An application of support vector machines while using an ensemble of feature selection methods. *Expert Systems with Applications*, 34(1):313–327.
- Thomas H. Davenport and Rajeev Ronanki. 2018. Artificial intelligence for the real world. *Harvard Business Review*, 96(1):108–116.
- Alberto Fernández, Salvador García, Francisco Herrera, Naiyin Chen, and S. Abdullah. 2018. Smote for learning from imbalanced data: Progress and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 30(1):154–171.

- Jerome H. Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.
- Mikel Galar, Alberto Fernandez, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. 2012. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(4):463–484.
- Sunil Gupta, Donald R. Lehmann, and Jennifer Ames Stuart. 2006. Valuing customers. *Journal of Service Research*, 9(2):87–94.
- John Hadden, Ashutosh Tiwari, Rajkumar Roy, and Dymitr Ruta. 2007. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34(10):2902–2917.
- Haibo He and Edwardo A. Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.
- Bingquan Huang, Mohand Tahar Kechadi, and Brian Buckley. 2012. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1):1414–1425.
- Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1:389–399.
- A. Keramati, N. Jafari-Marandi, M. Albadvi, M. Abbasi, M. G. Afshar, and A. Mojtahed. 2010. A proposed classification of data mining techniques in customer churn prediction. *Journal of Knowledge Management*, 14(3):351–368.
- Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1:501–507.
- Christoph Molnar. 2020. *Interpretable Machine Learning*. Lulu.com.
- E. W. T. Ngai, Li Xiu, and Dorothy C. K. Chau. 2009. The application of data mining techniques in customer relationship management: A literature review. *European Journal of Operational Research*, 199(1):39–47.
- Foster Provost and Tom Fawcett. 2013. Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1):51–59.
- Ananda Raffi. 2022. Ibm watson marketing customer value data. <https://www.kaggle.com/datasets/anandaraflii/ibm-watson-marketing-customer-value-data>. Accessed: November 2025.
- Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 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*, pages 1135–1144.
- Feng Shen, Min Liu, Wenjun Wang, and Wenyan Liu. 2019. An ensemble method for credit scoring based on smote and xgboost. *IEEE Access*, 7:73021–73033.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.
- Wouter Verbeke, David Martens, and Bart Baesens. 2012. Social network analysis for customer churn prediction. *Applied Soft Computing*, 12(3):1091–1102.
- Wouter Verbeke, David Martens, Christophe Mues, and Bart Baesens. 2011. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3):2354–2364.
- Chih-Ping Wei and I-Tang Chiu. 2002. Turning telecommunications call details to churn prediction: A data mining approach. *Expert Systems with Applications*, 23(2):103–112.
- Chun-Sheng Wu, Hong-Bo Xie, Guan-Jun Liu, and Qiang Wang. 2016. Telecom customer churn prediction based on k-means and pca. *Journal of Communications*, 11(10):1038–1045.
- Guangxia Xia, Yan Li, Li Guo, Hong You, and Zhengbing Hu. 2017. Model comparison for churn prediction in telecommunications. *IEEE Access*, 5:12087–12095.

## Appendix A. Code and Data Availability

The complete Google Colab notebook used in this study is publicly available at:

[https://colab.research.google.com/drive/1xAIjx1Rs16\\_hNahLTIJNsMNp4Ed3y4Hh?usp=sharing](https://colab.research.google.com/drive/1xAIjx1Rs16_hNahLTIJNsMNp4Ed3y4Hh?usp=sharing)

For questions, contact the author at [sree.tetali@my.utexas.edu](mailto:sree.tetali@my.utexas.edu).