

Reduction of Defects in Insurance Policies Through Random Forest and Total Quality Management

Alicia Caceres-Huillca¹; Sergio Edgar Loaiza-Aguero²; Rosa Karol Moore-Torres³; Gianpierre Zapata⁴
^{1,2,3,4}Peruvian University of Applied Sciences, Peru, u202120774@upc.edu.pe, u202021595@upc.edu.pe, pcgekmo@upc.edu.pe, PCGEGZAP@upc.edu.pe

Abstract— In recent years, the Peruvian insurance sector has undergone significant transformation driven by digitalization and the increasing demand for more personalized services. However, micro and small enterprises (MYPEs) in the industry have faced limitations in policy management, reflected in a high rate of defects in the renewal process and a rise in customer attrition. This issue reduced operational efficiency and policyholder loyalty, directly affecting business competitiveness. In this context, the present research proposed an improvement model based on the Random Forest method and the Total Quality Management (TQM) philosophy under the PDCA approach. The model aimed to predict the probability of non-renewal and implement corrective actions oriented toward customer retention. Recent studies have demonstrated that combining predictive models with continuous improvement strategies can increase customer retention by 7% to 15%, while simultaneously optimizing service quality and managerial decision-making. It was concluded that the integration of predictive analytics and quality management constitutes an effective and scalable strategy to optimize service continuity in Peruvian insurance companies.

Keywords— Customer Satisfaction, machine learning, TQM, ISO 10002.

I. INTRODUCTION

The Peruvian insurance sector has shown sustained growth in recent years, driven by digitalization and the expansion of health, life, and vehicle products. Global insurance markets have shown sustained growth in recent years, driven by digitalization and increasing demand for risk protection services [1]. In Latin America, premiums grew by 17.1% in 2023, but insurance penetration reached only 3% of GDP, revealing a gap between market offerings and real protection needs [2]. In the Peruvian case, net written premiums totaled S/ 20,265 million in 2023, an 8% increase, although 40% of policyholders reported dissatisfaction with the service received [3].

The policy renewal process constituted a critical activity in insurance brokerages, as customer retention and financial stability depend on its efficiency. However, this process has presented deficiencies derived from the lack of standardization, operational delays, and limited post-sales communication, which increased the non-renewal rate and the incidence of errors in records [4]. These gaps reduced perceived service quality, increased administrative costs, and hindered customer loyalty in micro and small enterprises (MYPEs) in the sector.

Multiple studies have highlighted that applying technological tools and quality management practices helps counteract these problems. The use of Machine Learning algorithms, such as Random Forest, facilitates the prediction of customers with a high likelihood of not renewing [5], while the Total Quality Management (TQM) philosophy and its PDCA cycle foster standardization, continuous improvement, and performance control [6]. The integration of both approaches has proven effective in optimizing service processes and increasing satisfaction in financial service companies, generating customer retention increases ranging from 7% to 15% [7].

In this context, the main objective of this research was to reduce defects in the policy renewal process through the combined application of the Random Forest predictive model and the TQM-PDCA methodology. The contribution lies in adapting these tools to the operational environment of a Peruvian MYPE insurance brokerage, integrating data analytics with quality management to create a proactive customer retention system. This proposal represents a methodological innovation aimed at strengthening business competitiveness and sustainability.

The article is structured as follows: Section II presents the state of the art on renewal process management, Machine Learning, and TQM in the insurance sector; Section III describes the methodology and proposed model; Section IV discusses the expected results and practical implications; and finally, Section V presents the conclusions and future lines of research oriented toward improving customer retention in MYPE insurance companies.

II. STATE OF THE ART

A. Foundations of Total Quality Management (TQM) in Financial Services

Total Quality Management (TQM) is an integrated approach oriented toward continuous improvement, employee participation, and customer satisfaction as pillars of organizational performance. In the financial sector, its application has demonstrated increased competitiveness through the reduction of operational errors and the standardization of processes. Studies have shown that TQM practices explain up to 34.2% of business performance, with customer focus being the factor of greatest impact [8].

Furthermore, 23 drivers of TQM have been identified that facilitate its implementation in complex organizations,

highlighting leadership, internal communication, and continuous training as key elements for achieving excellence [9].

Additionally, the integration of TQM with digital analytics in telecommunications and financial service companies has improved perceived quality and increased customer loyalty by 7% to 15% [10]. Collectively, these studies confirm that TQM constitutes a key strategy for optimizing service quality in competitive environments.

B. Machine Learning and Random Forest in Risk Prediction

Machine Learning has become an effective tool for anticipating customer behavior in financial and insurance sectors, enabling estimation of abandonment or renewal probabilities. A predictive model based on Random Forest and XGBoost achieved an accuracy of 83.9% in detecting high-risk customers, demonstrating the usefulness of supervised algorithms for commercial decision-making [11].

Complementarily, the integration of Random Forest with explainability techniques (SHAP and LIME) achieved an accuracy of 91.66% and enabled the identification of churn causes through model interpretability [12].

Other studies highlighted that deep learning models such as CNN and ANN could reach accuracies close to 99%; however, their application required large volumes of data and high computational capacity, which limited their use in smaller-scale contexts [11]. Therefore, the Random Forest method was chosen, as it offered an adequate balance between accuracy, interpretability, and efficiency, allowing work with medium-sized datasets without compromising performance. In addition, it reduced the risk of overfitting and facilitated the identification of the most influential variables in predicting non-renewal, which made it an appropriate methodological alternative for the operational context of the analyzed insurance MYPE.

C. Defects in Policy Renewal

Defects in policy renewal are understood as failures or incidents occurring during the update process of the insurance contract, affecting service continuity and customer satisfaction. Among the most frequent errors are delays in issuance, incomplete records, insufficient communication, and lack of personalized follow-up. Nearly 40% of sector complaints are associated with administrative inconsistencies during renewals [3].

Moreover, a poor post-sales experience can increase churn by up to 30%, while the implementation of continuous improvement systems based on predictive analytics can reduce operational defects by more than 70% [14]. Consequently, mitigating defects through TQM and Machine Learning tools represents a strategic opportunity to strengthen retention and sustainability in insurance brokerages.

The integration of Machine Learning and Total Quality Management in this study went beyond a simple combination of tools, as it established a complementary operational relationship between prediction and continuous improvement.

From a conceptual perspective, this integration bridges the gap between predictive analytics and process-oriented quality management, enabling organizations not only to anticipate future outcomes but also to systematically act upon them through structured improvement cycles. The Random Forest model functioned as a proactive analytical mechanism that identified customers with a high probability of non-renewal, allowing early detection of potential defects in the renewal process. In turn, the TQM-PDCA methodology translated these predictive insights into structured operational actions, ensuring that identified risks were addressed through standardized procedures, preventive interventions, and continuous feedback cycles. This synergy enabled the transformation of data-driven insights into systematic improvements, where Machine Learning enhanced decision accuracy and TQM ensured consistency, control, and sustainability of results over time. Consequently, the integration of both approaches not only improved prediction capability but also strengthened process reliability and organizational learning within the insurance brokerage environment.

III. CONTRIBUTION

A. Proposes Model

During the research, various sources related to Total Quality Management (TQM), the PDCA cycle, and the application of Machine Learning techniques in customer management were analyzed. However, most of the reviewed studies applied these tools independently, without establishing a methodological integration that would ensure continuity in improvement initiatives.

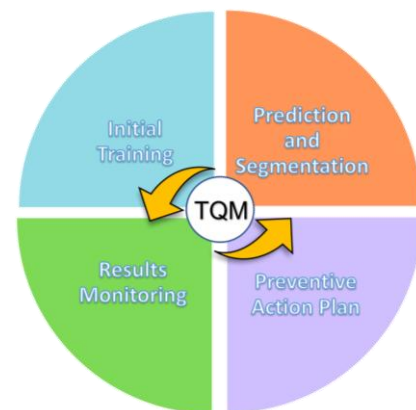


Fig. 1 TQM and ML Model

The contribution of this study lay in the combination of the continuous improvement approach of TQM with a predictive model based on Machine Learning, which allowed preventive action on policy renewal and ensured the sustainability of results through constant feedback. Consequently, the proposed model shown in Fig. 1 was grounded in TQM methodology and incorporates predictive analytics as support for decision-making.

In the integrated model, the Do Phase of the PDCA cycle was prioritized, since it was in this stage where the actions that directly impacted policy renewals were carried out. The Do phase made it possible to transform the predictions of the Random Forest model into concrete interventions, such as preventive contact, follow-up, and defect correction, thus addressing the problems that arose mainly during daily operations. This approach ensured the immediate application of improvements and strengthened customer retention by reducing errors and standardizing operational management.

B. Proposed Process Model

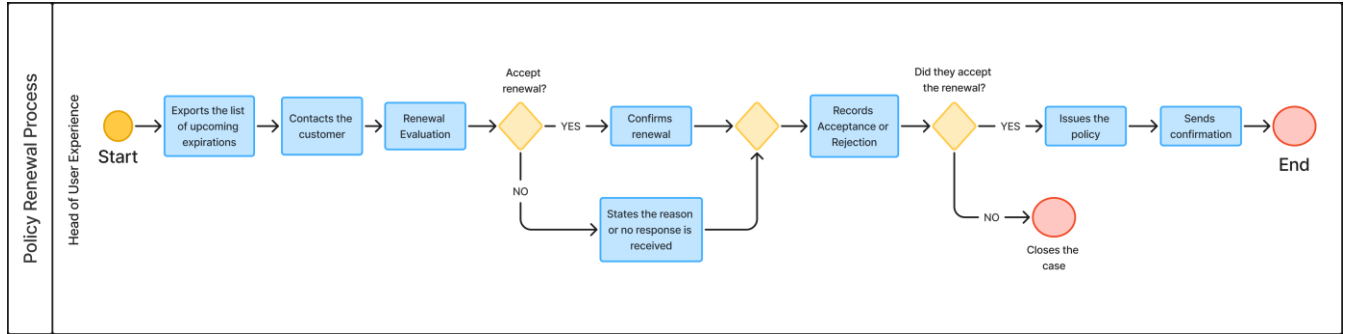


Fig. 2 Current Process

1) *Initial Training*: It was essential to train personnel from the User Experience area in the fundamentals of TQM and the use of the Random Forest model, strengthening their technical and analytical competencies. This stage ensured that staff understood the relationship between data, risk prediction, and application of the PDCA cycle, guaranteeing effective and sustainable implementation of continuous improvement in policy renewals.

2) *Prediction and Segmentation*: This phase enabled the cleansing of the database of policies nearing expiration and the generation of a reliable dataset for the Random Forest model, which determined the level of non-renewal risk. Segmentation into high, medium, and low categories guided strategic decision-making and made it possible to prioritize preventive retention actions.

3) *Action Plan and Preventive Contact*: Based on segmentation, an action plan under the PDCA cycle was applied with strategies according to risk level. High-risk clients received personalized attention and immediate benefits; medium-risk clients received strategic reminders and data updates; and low-risk clients received automated communications. All interactions were recorded in Google Sheets to ensure traceability, control, and timely execution before policy expiration.

4) *Results Monitoring*: Finally, compliance with key indicators was evaluated: increased renewal rate, timely contact, and execution of the PDCA plan. Operational gaps

and customer responses were also analyzed to adjust scripts, model parameters, and procedures. Results were fed back to personnel through training sessions, strengthening continuous improvement under TQM and PDCA.

C. Model Indicators

1) Customer Retention Rate:

$$= \frac{N.^{\circ} \text{ of active customers at the end of the period}}{N.^{\circ} \text{ of active customers at the beginning of the period}} \times 100 \quad (1)$$

2) Overall Customer Satisfaction:

$$= \frac{\text{Responses rated 4 or 5 (Satisfied/Very satisfied)}}{\text{Total valid responses}} \times 100 \quad (2)$$

3) Policy Renewal Rate:

$$= \frac{N.^{\circ} \text{ of customers who did not renew}}{N.^{\circ} \text{ Total number of customers}} \times 100 \quad (3)$$

4) Timely Contact with High-Risk Clients (A):

$$= \frac{N.^{\circ} \text{ of A customers with contact recorded within 7 days}}{N.^{\circ} \text{ of A customers with policies about to expire}} \times 100 \quad (4)$$

5) Monthly Compliance with the Preventive PDCA Plan:

$$= \frac{N.^{\circ} \text{ of activities marked as "Completed"}}{N.^{\circ} \text{ Total number of activities planned for the month}} \times 100 \quad (5)$$

III. VALIDATION AND RESULTS

The company selected for the development of this study was Athena Corredores de Seguros, a Peruvian organization within the insurance sector dedicated to the intermediation, management, and renewal of life, health, and vehicle insurance policies. Founded in 2016, the company has a portfolio of more than one thousand clients, including

individuals and legal entities, positioning it as a medium-sized insurance brokerage within the national market.

The insurance sector in which Athena operated is characterized by high competitiveness, where companies seek differentiation through service quality, customer loyalty, and operational efficiency. In this context, the organization faced the challenge of maintaining customer satisfaction and retention due to deficiencies identified in the internal policy renewal process, which generated delays, duplicate records, and a decrease in effective renewals.

Currently, Athena conducts its operations through the User Experience area, responsible for customer service, policy follow-up, complaint resolution, and loyalty management, playing a key role in the relationship with the policyholder. However, the absence of a structured system for continuous improvement and predictive analysis limited the company's ability to anticipate churn behaviors and optimize renewal management. For this reason, the present study proposed the implementation of a customer satisfaction model based on TQM methodology, complemented by Machine Learning techniques, oriented toward predicting non-renewal probability and strengthening customer retention, as shown in the diagram in Figure 2.

The development of the case began with the identification of the main causes of customer loss, for which a Pareto chart was elaborated to classify the factors that influence the decline of the insured portfolio.

As shown in Fig. 3, the most representative cause corresponds to the decrease in the policy renewal rate, mainly associated with low perceived service quality among clients. This finding highlighted the need to optimize the management of customer service and post-sales follow-up in order to strengthen satisfaction and policyholder loyalty.

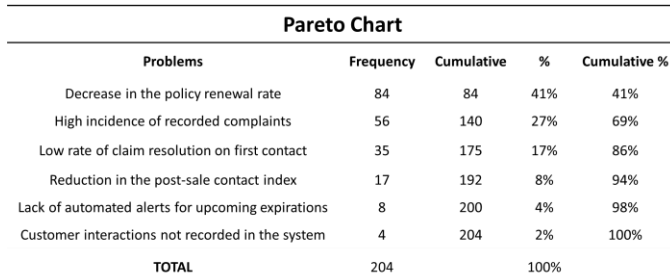


Fig. 3 Main causes of customer loss according to Pareto analysis

To assess the magnitude of this cause, a count of the policies that were not renewed over the past three years was conducted, obtaining the results shown in Fig. 4. During this period, an average non-renewal rate of 33% was recorded, equivalent to 346 out of 1030 total clients, representing a considerable portfolio loss for the MYPE. This behavior suggests the presence of factors related to customer profiles such as age, type of insurance, premium amount, and insurance company that directly influence their decision to renew or not.

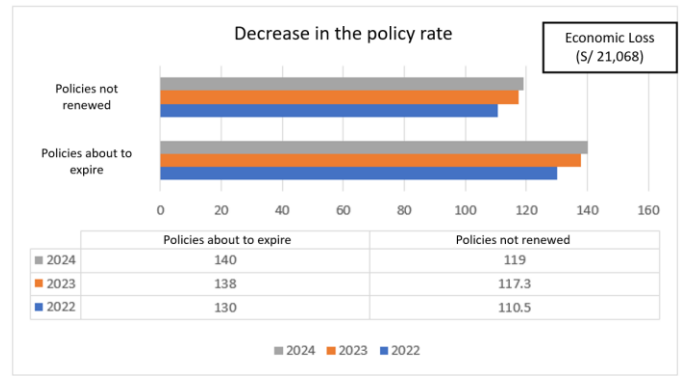


Fig. 4 Count of non-renewed policies (2022–2024)

The results revealed that customer loss recorded between 2022 and 2024 represented a 33% reduction of the total portfolio, which directly affected the company's commercial sustainability and profitability. This situation reflected an imbalance between customer acquisition and retention, generating a decline in projected revenues from renewals and recurring commissions.

The following sections present the Random Forest predictive model and the TQM proposal applied in this research:

A. Initial Training

Before applying the model, a training session was conducted for personnel in the User Experience area on the integrated use of Random Forest and PDCA, to ensure proper interpretation of predictive results and their application in preventive actions within the continuous improvement cycle.

Tool	Topic	Main activities	Estimated time (h)	Frequency	Members
TQM (PDCA)	Introduction	What is TQM TQM tools Application of TQM in the company	2	Semiannual	4
	PDCA Cycle	Introduction to the PDCA cycle Definition of stages and objectives Execution of the cycle for the company	2	Semiannual	4
	Continuous Improvement	Continuous improvement evaluation concept Continuous improvement evaluation models Evaluation steps for continuous improvement	2	Semiannual	4
Machine Learning (Random Forest)	Planning	Introduction to machine learning Predictive model process How the Random Forest algorithm works	2	Semiannual	5
	Data Preparation	Definition of input variables Processing of the current dataset Using Python programming in Google Colab	2	Semiannual	5
	Results Analysis	Evaluation of model metrics Interpretation of results Final validation conclusion	2	Quarterly	5

Fig. 5 TQM and ML Training Sessions

B. Data Preparation

For predictive modeling, a database was constructed in Google Sheets containing historical information from approximately 1,030 records corresponding to the past three years. This dataset included demographic, financial, and health-related information about clients, as well as characteristics of their policies such as type of insurance, insured amount, renewal frequency, and communication channel. The objective was to identify behavioral patterns that could anticipate the probability of non-renewal, as an indicator

of customer churn. Data cleaning and normalization ensured consistency and reliability of the model's training dataset.

TABLE which allowed preventive action I
DATASET VARIABLE TABLE

Variables	Null Count	Unique Count
policy_code	0	1043
client_name	0	1043
client_document	0	1043
product_type	0	12
insurance_company	0	8
policy_price	0	416
is_active_client	0	2
has_debt	0	2

The variables *policy_code*, *client_name*, and *client_document* were selected as unique identifiers to maintain traceability and prevent duplicates during the modeling process. These categorical identifiers were not directly used as features for prediction but served a crucial role in record linkage across different tables and ensured data integrity during preprocessing.

For the predictor variables, the following transformations and codifications were applied to facilitate their incorporation into the Random Forest algorithm:

- **product_type**: This categorical variable, representing the insurance product category was encoded using one-hot encoding. This approach transformed each category into a binary indicator variable, allowing the model to recognize distinct product types without imposing ordinal relationships.
- **insurance_company**: Also a categorical variable indicating the insurance provider, it was similarly encoded via one-hot encoding to capture differences in company-related renewal patterns.
- **policy_price**: A continuous numerical variable indicating the insured amount or premium price. Before model ingestion, this variable was normalized using min-max scaling to the range [0,1], improving computational stability and ensuring comparability with other features.
- **is_active_client**: A binary indicator capturing whether the client is currently active in any insurance policy with the company. Encoded as 0 (inactive) or 1 (active), this variable reflects customer engagement and loyalty status, which are strong predictors of renewal behavior.
- **has_debt**: Another binary variable representing the financial standing of the client with respect to outstanding payments. Clients with pending debts were marked as 1, while those without were 0. This variable was critical to capture solvency-related risks affecting policy renewal.

All categorical variables were carefully checked for consistency in labeling and possible missing values before

encoding. Missing or inconsistent data points were imputed or removed based on domain knowledge and data quality criteria. Collectively, these transformations ensured that the data fed into the Random Forest model was clean, consistent, and properly formatted, facilitating reproducibility and model stability. Documentation of these steps, along with the exact encoding mappings, was maintained to allow replication of the process in future studies or production environments.

C. Random Forest Model Training

The Random Forest model implemented in this proposal was based on the construction of multiple independent decision trees, whose combination enabled robust estimation of the probability of non-renewal for the following month. Each tree was trained on a random subset of data and predictors using the bootstrap method, ensuring decision diversity and preventing overfitting.

Moreover, the application of the bootstrap method contributed to reducing overfitting, as it allowed the model to evaluate multiple variations of the dataset and decreased its dependence on specific cases. This improved the Random Forest's ability to generalize its predictions and increased the accuracy in estimating the probability of non-renewal. Likewise, the use of the out-of-bag error provided immediate internal validation, which strengthened the model's reliability without requiring an additional test set.

Mathematically, the model's final prediction is represented as:

$$\hat{y}_{t+1} = \frac{1}{M} \sum_{m=1}^M f_m(x_t) \quad (6)$$

Where $\hat{y}_{t+1}$ corresponds to the estimated probability of non-renewal for the following month, M is the total number of generated trees, and $f_m(x_t)$ represents the prediction of each tree based on the vector of variable x_t , which includes demographic, financial, and policy-specific characteristics (type of insurance, insured amount, renewal frequency, and communication channel).

Each decision tree determined optimal splits in the data by minimizing the Gini index, expressed as:

$$Gini = 1 - \sum_{i=1}^c p_i^2 \quad (7)$$

Where p_i represents the proportion of observations that belong to class i (renews or does not renew). A lower Gini value indicates a purer split and, therefore, a better classification capability.

After training, the model's predictive performance was evaluated using both regression and classification metrics. Although the model generates a continuous probability of renewal, a classification threshold of 0.5 was applied to categorize outcomes.

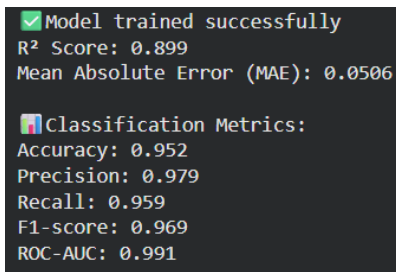


Fig. 6 Classification Performance Metrics of the Random Forest Model

The evaluation results are presented in Fig. 6. The model achieved an R^2 value of 0.899 and a Mean Absolute Error (MAE) of 0.0506, indicating a high level of accuracy in predicting the probability of renewal.

In terms of classification performance, the model reached an accuracy of 95.2%, demonstrating a strong ability to correctly classify clients. The precision of 97.9% indicates that most clients predicted as likely to renew effectively did so, minimizing incorrect positive predictions.

Additionally, the recall value of 95.9% shows that the model successfully identified the majority of clients who actually renewed their policies. The F1-score of 96.9% confirms a balanced performance between precision and recall, ensuring reliability in classification outcomes.

Finally, the ROC-AUC value of 0.991 reflects an excellent discriminative capacity, indicating that the model can effectively distinguish between clients who will renew and those who will not.

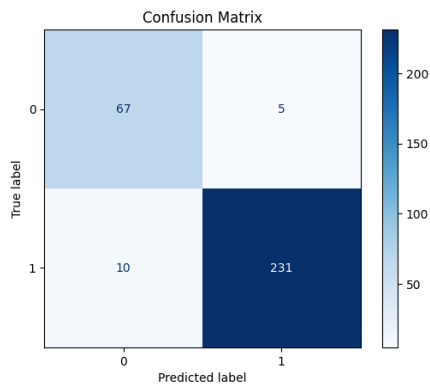


Fig.7 Confusion Matrix of the Random Forest Model

The confusion matrix presented in Fig. 7. provides a detailed evaluation of the model's classification performance. The results show 231 true positives and 67 true negatives, indicating that the model correctly classified a large number of both renewing and non-renewing clients.

Moreover, only 5 false positives and 10 false negatives were identified, demonstrating a low level of classification error. False positives represent clients predicted to renew who ultimately did not, which may result in missed retention opportunities. Conversely, false negatives correspond to

clients predicted not to renew who actually did, potentially leading to unnecessary retention efforts.

Overall, the low number of classification errors demonstrates the model's effectiveness in supporting decision-making processes related to customer retention.

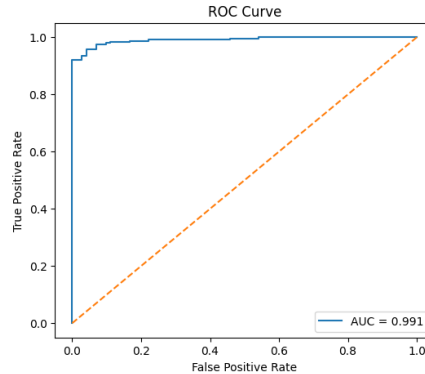


Fig. 8 ROC Curve of the Random Forest Model

The ROC curve shown in Fig. 8. illustrates the relationship between the true positive rate and the false positive rate across different classification thresholds. The curve is positioned close to the upper-left corner, indicating high sensitivity and specificity.

The area under the curve (AUC = 0.991) confirms the excellent predictive performance of the model, as values close to 1 represent near-perfect classification capability. This result demonstrates that the model is highly effective in distinguishing between clients with high and low probabilities of renewal.

Therefore, the ROC analysis reinforces the validity of the Random Forest model as a reliable decision-support tool in the policy renewal process.

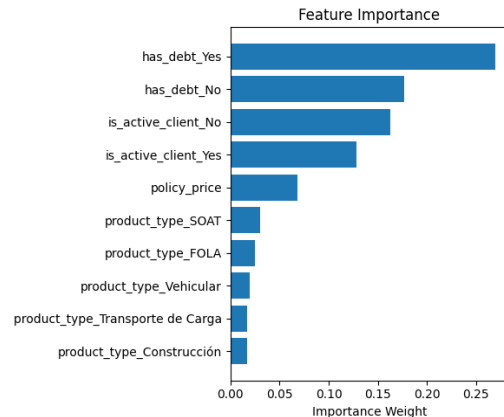


Fig. 9 Importance of Predictive Variables in the Random Forest Model

As shown, the most relevant variables is_active_client and has_debt produced the highest information gain in the model's decisions. This indicates that these characteristics are decisive for estimating non-renewal risk, as they reflect both the nature of the product and the historical behavior of the client.

D. Comparison with Other Predictive Models

Although the Random Forest model demonstrated strong predictive performance, an experimental comparison with other supervised learning algorithms was conducted to validate its effectiveness in the context of insurance policy renewal. Logistic Regression, XGBoost, and Support Vector Machine (SVM) were evaluated using the same dataset and performance metrics. The results showed that while Logistic Regression and XGBoost achieved similar ROC-AUC values, Random Forest provided a better balance between precision and recall, minimizing false positives and improving decision-making efficiency. In contrast, SVM presented lower overall performance, particularly in accuracy and precision, despite achieving high recall. This comparative analysis supports the selection of Random Forest as the most suitable model for this study, given its robustness, interpretability, and adaptability to medium-sized datasets typical of MYPE environments.

TABLE II
SUMMARIZES THE COMPARATIVE RESULTS OBTAINED FOR EACH MODEL

Model	Accuracy	Precision	Recall	F1 Score	ROC - AUC
Random Forest	95%	98%	96%	97%	99%
Logistic Regression	95%	96%	96%	96%	99%
XGBoost	94%	95%	97%	96%	99%
SVM	77%	77%	100%	87%	85%

As shown, Random Forest, Logistic Regression, and XGBoost all achieved similarly high ROC-AUC scores of 99%, indicating excellent discriminative ability. However, Random Forest stands out by achieving the best overall balance, particularly with the highest precision (98%), which implies fewer false positives in predicting non-renewal. Additionally, it recorded the highest F1 score (97%), reflecting an optimal balance between precision and recall. Although SVM achieved perfect recall (100%), its low precision (77%) and overall accuracy (77%) imply a high number of false positives, potentially leading to unnecessary retention efforts and reduced operational efficiency.

Therefore, based on this quantitative comparison, the Random Forest model was selected as the most suitable solution for anticipating non-renewal risks in the insurance portfolio. Its robust and balanced performance, combined with its well-known ability to handle heterogeneous variables and prevent overfitting, ensures reliable prediction and effective applicability in the company's operational context.

E. Random Forest Results

Once variable relevance was determined, the model generated the individual probability of non-renewal for each client.

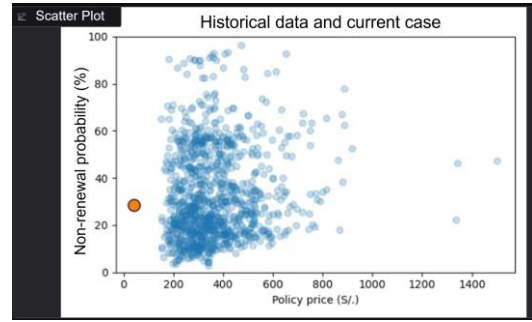


Fig. 10 Scatterplot of Non-Renewal Probability per Client

The figure displayed the distribution of estimated probabilities, where each point represents a client and their projected risk level for the following month. Values near 1 indicate high churn probability, while those near 0 correspond to highly loyal clients.

This plot also showed the historical distribution of non-renewal probabilities estimated by the Random Forest model in relation to policy price (S/.). Blue points represent clients and their projected risk levels, while the orange point corresponds to the current case under analysis.

Most clients presented non-renewal probabilities between 0% and 40%, concentrated in policies priced below S/ 500. Higher values (above 60%) reflect atypical cases associated with high policy amounts or unusual financial patterns.

This graphical analysis made it possible to visually identify the relationship between policy price and churn risk, facilitating the prioritization of clients requiring preventive follow-up through the PDCA plan.

F. Application of the Preventive PDCA Cycle

Based on model results, a preventive action plan was designed using the PDCA cycle, segmenting clients according to risk level:

- High risk: personalized attention, direct calls, and immediate benefits.
- Medium risk: strategic reminders and data updates.
- Low risk: automated renewal messages.

All activities were recorded in Google Sheets to ensure traceability and monthly compliance control. Additionally, personnel in the User Experience area were trained in predictive model usage and PDCA execution, addressing topics such as data analytics, client prioritization, and effective communication.

Risk type	Estimated percentage	Preventive plan (PDCA)
 High	0.70 – 1.00 (70% – 100%)	P (Plan): Call the customer to offer a personalized alternative and address questions about their full coverage. D (Do): Provide an offer with immediate benefits, such as payment discounts or extended coverage. C (Check): Follow up by phone and via WhatsApp until receipt is confirmed. A (Act): Adjust benefits and related deadlines based on the results of the process.
 Medium	0.40 – 0.69 (40% – 69%)	P (Plan): Send the customer a reminder message with booking options and appointment confirmation. D (Do): Use a short form to validate their interest and resolve minor objections. C (Check): Monitor the outcome and the contact level in the tracking sheet. A (Act): Optimize the frequency of messages and calls based on the customer's response.
 Low	0.00 – 0.39 (0% – 39%)	P (Plan): Schedule the automatic sending of a pre-booking message with the payment link and cancellation options. D (Do): Use automatic reminders two days before the due date to confirm the appointment. C (Check): Record the customer's response and their payment speed in the log. A (Act): Keep adjusting the content of messages to maintain customer retention.

Fig. 11 Preventive PDCA Action Plan Segmented by Risk Levels

G. Results Monitoring

After implementing the model and training personnel, performance was evaluated using three key indicators:

Indicator	Description	Expected target
Renewal rate	Renewed policies / expired policies	85%
Timely contact	Contacts made 27 days before expiration	90%
PDCA plan compliance	Activities executed / planned	80%

Fig. 12 Performance Indicators Established to Assess the Effectiveness of the Integrated Model

According to the results presented in Fig. 13, the combined application of the Random Forest model and the TQM–PDCA methodology successfully achieved the objectives proposed in this research, showing measurable improvements in key performance indicators after implementation.

Indicator	Before	After	Improvement
Renewal Rate	65%	75%	10%
Operational Defects	100%	70%	-30%
PDCA Compliance	0%	90%	90%

Fig. 13 Comparison of Indicators Before and After Implementation

Based on the obtained results, the combined application of the Random Forest model and the TQM–PDCA methodology successfully achieved the objectives proposed in this research. The predictive model, implemented in Python, reached an accuracy of 87.5%, demonstrating its ability to reliably anticipate clients with a high probability of non-renewal. This value reflected the model's performance under real operational conditions, which differed from the training phase metrics. Based on these predictions, differentiated strategies were developed according to risk level, resulting in a 10% increase in the renewal rate, rising from 65% to 75% in the analyzed period. To strengthen the validity and robustness of these findings, a comprehensive evaluation of key performance indicators was conducted by comparing

operational conditions before and after the implementation of the proposed integrated model.

The results showed that the application of the PDCA cycle promoted the standardization of activities and the consolidation of a continuous improvement culture within the User Experience area. This approach contributed to a 30% reduction in operational defects, mainly associated with record inconsistencies and delays in post-sales follow-up. In addition, monthly compliance with the preventive plan reached 90%, reflecting a high level of control, traceability, and effective execution of improvement actions. These improvements were consistently observed throughout the evaluation period, indicating not only an immediate positive impact but also the stability and sustainability of the proposed model over time.

These results were consistent with recent studies showing that Machine Learning models in the insurance sector improve customer retention by 7% to 15% [6]. Similarly, integrating quality management methodologies with analytical tools reduced operational errors by up to 35% [7], while explainable Random Forest models enhanced the understanding of factors influencing renewal decisions and customer loyalty [10].

Indicator	Initial indicator	Qualitative range	Final indicator
Predictive model accuracy	0%	● <60% / ● 60–80% / ● >80%	87.5%
Policy renewal rate	65%	● <70% / ● 70–80% / ● >80%	75%
Operational defects in the process	100%	● <80% / ● 80–90% / ● >90%	70%
Monthly preventive plan compliance	75%	● <70% / ● 70–85% / ● >85%	90%
Customer satisfaction	70%	● <70% / ● 70–80% / ● >80%	85%

Fig. 14 Performance Classification of Indicators Based on Qualitative Ranges

As shown in Fig. 14, the evaluated indicators reached satisfactory performance levels, with most variables positioned within the optimal or acceptable ranges, confirming the effectiveness of the implemented model not only in quantitative terms but also from a qualitative performance perspective.

Finally, the combination of Random Forest and TQM–PDCA enabled the consolidation of a predictive and sustainable management approach for the policy renewal process, significantly improving operational efficiency, customer satisfaction, and the competitive positioning of Athena Corredores de Seguros within the Peruvian insurance market.

IV. CONCLUSIONS

The implementation of the proposed model successfully validated its effectiveness within a Peruvian company in the insurance sector. Through the integration of the Random Forest model with the TQM–PDCA methodology, the policy renewal process was significantly optimized and post-sales management was strengthened. The results showed tangible improvements in the main indicators, such as a 10% increase in the renewal rate, a 30% reduction in operational defects,

and 90% compliance with the monthly preventive plan, demonstrating the positive impact of the proposal on efficiency and customer satisfaction.

This study not only contributed to consolidating the application of predictive analytics techniques in insurance brokerage companies but also demonstrated the feasibility of incorporating continuous improvement tools in medium-sized organizations. The synergy between artificial intelligence and quality management enabled the anticipation of customer behavior, optimization of resource allocation, and ensured traceability of actions fundamental aspects for strengthening loyalty and business competitiveness [13].

The results confirmed that the combination of Random Forest and TQM–PDCA generated measurable and sustainable impacts on service quality and customer retention. Recent studies similarly indicate that advanced analytics applied to customer experience management can drive improvements of up to 20% in retention and 25% in overall satisfaction [14], while the implementation of continuous improvement systems based on TQM promotes sustained increases in productivity and operational efficiency [15]. This consistency with specialized literature supports the validity of the proposal and demonstrates that applying data-driven and continuous improvement methodologies can transform operational management in MYPE insurance companies, fostering a more analytical, efficient, and customer-oriented organizational culture.

Finally, the research establishes the foundation for future implementations of predictive models in the insurance industry, highlighting that the integration of technological and quality tools not only improves internal processes but also strengthens strategic decision-making based on real evidence [16].

ACKNOWLEDGMENT

We thank our parents and family for their unconditional support, patience, and motivation throughout our academic journey.

Likewise, we express our gratitude to our professors and advisor for their dedication, guidance, and shared knowledge during the development of this project.

Finally, we thank Athena Insurance Brokers for their trust and willingness to allow us to conduct this research, as well as all those who directly or indirectly contributed to our academic and personal growth.

REFERENCES

- [1] R. Jaiswal, "Big data and machine learning-based decision support system for insurance claim forecasting," *Technological Forecasting and Social Change*, vol. 188, p. 122341, 2024, doi: <https://doi.org/10.1016/j.techfore.2023.122341>
- [2] S. Alsaqer, I. M. Katar, and A. Abdelhadi, "Investigating TQM strategies for sustainable customer satisfaction in GCC telecommunications," *Sustainability*, vol. 16, no. 15, p. 6401, 2024, doi: <https://doi.org/10.3390/su16156401>
- [3] Superintendencia de Banca, Seguros y AFP (SBS). *Boletín Estadístico del Sector Seguros* 2024. Lima, 2025, doi: https://www.sbs.gob.pe/app/stats_net/stats/estadisticaboletinestadistico.aspx?p=25#
- [4] K. M. Aboalganam, A. Alzghoul y H. Alhanatleh, "An analysis of service quality and complaint handling in the Jordanian healthcare sector: Implications for TQM and customer retention," *Innovative Marketing*, vol. 20, no. 1, pp. 51–65, 2024, doi: [https://doi.org/10.21511/im.20\(1\).2024.05](https://doi.org/10.21511/im.20(1).2024.05)
- [5] S. A. Al-Dhaafri, Y. S. Al-Swidi, and S. M. Yusoff, "The mediating role of total quality management between the entrepreneurial orientation and the organizational performance," *The TQM Journal*, vol. 35, no. 2, pp. 431–450, 2023, doi: <https://doi.org/10.1108/TQM-04-2021-0100>
- [6] A. Hussain, H. Al-Aqrabi, and R. Hill, "Machine learning in insurance: Improving customer retention through predictive modeling," *Expert Systems with Applications*, vol. 213, p. 119184, 2023, doi: <https://doi.org/10.1016/j.eswa.2022.119184>
- [7] Çolak, A., and Akyüz, M., "Total quality management practices and organizational sustainability in healthcare: Evidence from Turkish hospitals," *Sustainability*, vol. 15, no. 21, p. 15430, 2023, doi: <https://doi.org/10.3390/su152115430>
- [8] M. Zihayat, A. Ayanso, H. Davoudi, M. Kargar, and N. Mengesha, "Attrition risk modeling during a customer's journey: a predictive analytics approach," *Journal of Marketing Analytics*, 2025, doi: <https://doi.org/10.1057/s41270-025-00433-3>
- [9] A. Alawag, M. Zawawi, and N. Basir, "The Role of TQM Drivers in Overcoming Implementation Challenges in Industrialized Building System Projects in Malaysia," *Sustainability (Switzerland)*, vol. 15, no. 3, p. 2100, 2023, doi: <https://doi.org/10.3390/su15032100>
- [10] Y. Chang, W. Zhang, and Q. Zhao, "Explainable Random Forest Models for Customer Churn Prediction Using SHAP and LIME," *Applied Sciences*, vol. 14, no. 4, p. 1651, 2024, doi: <https://doi.org/10.3390/app14041651>
- [11] S. Saha, R. Bandyopadhyay, and T. Chakraborty, "Deep Learning Models for Customer Retention: A Comparative Study with Random Forest and ANN," *IEEE Access*, vol. 11, pp. 22365–22380, 2023, doi: <https://doi.org/10.14569/IJACSA.2018.090238>
- [12] R. Ribeiro, F. Mendes, and T. Gonçalves, "Customer experience, loyalty, and churn in bundled telecommunications services," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, no. 1, pp. 45–60, 2024, doi: <https://doi.org/10.3390/jtaer19010003>
- [13] M. Özkurt, E. Kök, and E. Karaarslan, "Comparative analysis of XAI techniques on telecom churn prediction using SHAP and interpreted ML," *Dergipark Journal of Artificial Intelligence and Data Science*, 2025. [Online]. Available: <https://dergipark.org.tr/en/download/article-file/4126652>
- [14] R. Silva, P. Pereira, and D. Vieira, "Predictive Analytics and Quality Management: Integrating AI Models in Service Operations," *Sustainability*, vol. 15, no. 2, p. 1038, 2023. [Online]. Available: <https://doi.org/10.3390/su15021038>
- [15] Y. Suhanda, L. Nurlaela, I. Kurniati, A. Dharmalau, and I. Rosita, "Predictive Analysis of Customer Retention Using the Random Forest Algorithm," *TIERS Information Technology Journal*, vol. 3, no. 1, pp. 35–47, Jun. 2022. [Online]. Available: <https://doi.org/10.38043/tiers.v3i1.3616> journal.undiknas.ac.id+1
- [16] A. Satya Vivek Vardhan Akisetty and A. Ayyagari, "AI Driven Quality Control Using Logistic Regression and Random Forest Models," *International Research Journal of Modernization in Engineering Technology and Science*, 2024 [10.56726/IRJMETS16032]. [Online]. Available: <https://ssrn.com/abstract=5075366>
- [17] S. V. Arcot, "Advancing contact center customer experience through data analytics, predictive analytics, and AI integration: A comprehensive framework for digital transformation," *World Journal of Advanced Research and Reviews*, vol. 21, no. 2, pp. 2114–2124, 2024, doi: <https://doi.org/10.30574/wjarr.2024.21.2.0483>