Predictive Modeling for Retaining Members of a Religious Community Using Machine Learning Techniques: A Web Platform

Jaime Blanco-López*, Saray Galeano-Ospino, Giancarlo Cabrera-Rey and Hamlet Josiah Pirazan-Pomare

Corporación Universitaria Adventista, Medellín, Colombia

Abstract

Member attrition in religious communities poses a challenge to their sustainability and growth, making it necessary to have tools that can help anticipate attrition and support decision-making. Methodology: This study employs a quantitative, non-experimental, cross-sectional design, using historical data on the membership of a religious community in Colombia. A predictive model was built using the Random Forest algorithm, which was integrated into a web platform and validated through a pilot usability test with community leaders. Results: The model achieves an overall accuracy of 91% and an AUC-ROC of 0.96, demonstrating high predictive power. The variable importance analysis identifies the year of baptism, the church, and the association as the main factors associated with retention. Discussion: The results confirm the effectiveness of using machine learning techniques to predict retention in religious contexts, extending their application beyond traditional business environments. Conclusions: The developed model and platform constitute a reliable and functional tool to support preventive strategies aimed at improving member retention in religious communities.

Keywords

Predictive model, Member retention, Religious community, Machine learning, Random forest, Attrition, Data analytics, Web platform

1. Introduction

Retaining members, employees, or customers has long been a priority for organizations [1]. Religious communities also face significant challenges in maintaining member engagement, as attrition is influenced by interpersonal experiences, changing beliefs, and the search for new forms of transcendence [2]. This situation poses challenges for church leaders and families, who must monitor participation and strengthen commitment to ensure community growth [3].

Because church members generate behavioral and organizational records, their data can be analyzed to anticipate attrition [4]. Predictive analytics, widely adopted in sectors such as telecommunications, retail, finance, and services, uses data mining and machine learning (ML) to identify individuals at risk of leaving, enabling proactive retention strategies that are more cost-effective than acquiring new members [5, 6]. Applying these approaches to nonprofit organizations, such as religious communities, represents a valuable strategy for improving membership retention.

This study proposes a ML based solution to predict member attrition in a religious community, enabling leaders to make timely preventive decisions. A quantitative, non-experimental, cross-sectional design was employed using a pre-existing, private dataset that was not authorized for public release by the participating institution due to confidentiality policies. The methodology included data preprocessing, development of a predictive model based on the Random Forest algorithm, its integration into a

IWSAI 2026: 1st International Workshop on Systems, Applications, Artificial Intelligence and Innovation, September 16–18, 2026, Piura, Peru

*Corresponding author.
jaime.blanco@unac.edu.co (J. Blanco-López); sagaleano@unac.edu.co (S. Galeano-Ospino); giancarlo.cabrera@unac.edu.co (G. Cabrera-Rey); hamletj.pirazanp@unac.edu.co (H. J. Pirazan-Pomare)
0000-0001-9149-7910 (J. Blanco-López); 0000-0003-1755-781X (S. Galeano-Ospino); 0000-0002-0339-6894 (G. Cabrera-Rey); 0009-0001-2201-0424 (H. J. Pirazan-Pomare)

three-tier web platform (Next.js/React, Flask/Python, PostgreSQL), and a pilot usability evaluation with church leaders in Colombia using the System Usability Scale (SUS).

The Random Forest model demonstrated strong predictive performance, achieving 91% accuracy and an Area Under the ROC Curve (AUC) of 0.96. It also obtained an F1-score of 0.80 for identifying attrition cases (class 0), supporting its usefulness for preventive decision-making. Variable importance analysis identified the year of baptism as the most influential predictor, followed by the church and the association to which the member belongs.

1.1. Background

This section presents previous work on the analysis of attrition in religious communities. It reviews research and proposed solutions that have addressed this phenomenon from various perspectives, primarily through predictive models, which provide significant value to the selected communities.

Gonzalez-Nucamendi et al. [7] conducted a study in which they applied data analysis tools to identify undergraduate students at high risk of dropping out and to understand the associated factors. The study utilized a database containing information on more than 14,000 students, of whom only 8.5% had dropped out. After careful cleaning and selection of variables, clustering, classification, and ML techniques were applied. Among the eight models tested, the Random Forest algorithm performed best, achieving 50% accuracy in predicting dropout and exceeding 70% in overall retention and accuracy metrics, thanks to a small fine-tuning of the probability threshold.

Afzal et al. [8] conducted a study focused on predicting customer churn in the telecommunications sector, using various ML models on a public dataset of customers in Iran. They compared five of the most used algorithms: logistic regression, k-nearest neighbors (k-NN), Naïve Bayes, decision tree, and Random Forest. By analyzing true and false positive and negative metrics, they evaluated the performance of each model. The results showed that Random Forest was the most accurate, achieving a prediction rate of 86.94%, outperforming the rest, especially the decision tree, which achieved 80.04%. This established Random Forest as the most effective model in this case.

Arefin et al. [9] analyzed customer churn prediction in the banking sector using ML and deep learning models. They evaluated seven algorithms, including logistic regression, Random Forest, XGBoost, LightGBM, and deep neural networks. To improve performance, a technique called SMOTE was applied, which helps balance classes in imbalanced datasets. Without SMOTE, Random Forest was the most accurate model, with an 85.41% accuracy rate and an AUC of 92.65%. After applying SMOTE, several models exceeded 84% accuracy, with LightGBM standing out as the most effective.

Unlike educational and commercial settings, where the literature has established the use of ML to predict student dropout and customer churn, predicting disengagement in religious communities represents an unexplored research gap. This context is of particular interest, given that the disengagement of believers is driven by complex socio-relational and institutional dynamics: the perception of judgment and condemnation by adult members, the low appeal of liturgical services, the loss of representation due to the generational gap, the termination of roles in key ministries or groups, and critical transitions to higher education [10]. The integration of these qualitative and institutional factors into ML frameworks opens the door to identifying patterns of risk for disengagement and developing early preventive interventions within church organizations.

The remainder of this paper is organized as follows: Section 2 describes the methodology, Section 3 presents the results, Section 4 discusses the findings, and Section 5 concludes the study and outlines future work.

2. Methodology

The study is conducted in two main stages. The first stage consists of building the model and the platform, in which data is analyzed, and a predictive tool is developed based on the Random Forest algorithm, which is highly effective for predicting attrition [11]. The second stage consists of a pilot test to validate the platform's usability within a religious community. During this phase, qualitative feedback is collected from church leaders to evaluate the tool's usefulness and practical applicability, ensuring that the model is both accurate and functional within the community context.

2.1. Stage 1: Model Development and Web Platform

2.1.1. Data Collection and Analysis

In this phase, the religious community, its data sources, and the variables relevant to the study are selected. An Exploratory Data Analysis (EDA) is then performed to assess data quality, structure, and relevance. The process includes identifying missing values, inconsistencies, and duplicates to ensure data integrity [12], evaluating the distribution and importance of each variable, and selecting those most relevant for the predictive model while discarding redundant or uninformative features. Special attention is given to variables with potential predictive value, such as the date of entry into the community and other sociodemographic or participation characteristics identified during the analysis [13].

2.1.2. Web Platform Design

Following the data analysis, the web platform architecture and data flow diagrams are designed to support interaction with the predictive model and to define the system's components and data processing workflow [14, 15]. The data flow diagram, based on the Gane and Sarson notation, illustrates how data is captured, transformed, stored, and processed before implementation [15, 16]. Subsequently, the platform's functional and non-functional requirements are established [17]. Functional requirements define the system's core operations, such as user authentication, data management, and prediction visualization [18], while non-functional requirements specify quality attributes including security, scalability, performance, availability, and usability.

2.1.3. Model Development and Web Platform

Based on the exploratory analysis, a predictive model is developed to estimate the probability of member attrition in the religious community. Relevant features are selected, redundant variables are removed, and the dataset is split into training (80%) and testing (20%) sets to evaluate generalization performance [19]. The model is trained using the Random Forest algorithm, with hyperparameter tuning and cross-validation applied to improve robustness. The target performance is an accuracy and AUC above 90%, while maintaining strong precision, recall, and F1-score, particularly for the minority class (dropout). In parallel, a web platform is developed according to the defined functional and non-functional requirements, integrating the predictive model with interactive visualizations and a user-friendly interface that enables authorized users to analyze predictions without requiring advanced technical knowledge.

2.1.4. Unit and Functional Testing

Once the development of the web platform is complete, unit and functional tests are conducted to ensure its effectiveness and efficiency. Unit tests verify that the smallest components of the system function correctly when tested in isolation. These tests focus on analyzing each part of the code to ensure it fulfills its specific function [20]. On the other hand, functional tests validate that the system's features meet the defined requirements. These tests, typically black-box tests, are based on system requirements to ensure that the software's external behavior is as expected [21].

2.2. Stage 2: Pilot Test

In this stage, the web platform is validated through a pilot test with leaders of the selected religious community to evaluate both the predictive model's performance and its practical applicability [22]. Using real community data, participants assess whether the generated predictions align with their experience, providing expert validation of the model's relevance, reliability, and usefulness. User feedback is systematically collected through the SUS [23], focusing on prediction clarity, ease of use, and willingness to adopt the platform. The pilot also provides valuable input for refining the system and enhancing its potential impact before large-scale implementation [24].

3. Results

This section presents the research results, in accordance with the stages and activities defined in the methodology.

3.1. Stage 1: Model Development and Web Platform

3.1.1. Data Collection and Analysis

The study was conducted using membership records from a religious community in Medellín, Antioquia. Eleven CSV files were consolidated into a single dataset containing 234,361 records and 36 variables, with members classified as active or inactive. Privacy protocols were verified to ensure the absence of personally identifiable information. Although data quality could not be directly controlled, the dataset was considered reliable because it originated from the institution's official records [25]. After data cleaning, missing values, outliers, and inconsistencies were addressed, and the membership status was transformed into a binary target variable (1 = active, 0 = inactive). Finally, feature selection reduced the dataset from 36 to 11 variables, retaining only those most relevant for the predictive model, while discarding variables with limited relevance to the study.

Table 1
Variables Selected for the Prediction Model.

Name	Description
Age	Age of the religious community member
Gender	Gender of the religious community member (Male or Female)
Marital Status	Classifies the marital status of members as single, married, widowed, or divorced. It is a nominal categorical variable.
Occupation	Describes the occupation or profession of each member. This variable is also a nominal categorical variable.
Level of education	Indicates the level of formal education attained by each member. This is an ordinal categorical variable.
Date of Baptism	Records the year in which each person was baptized. It is a temporal variable.
Classification	This is a nominal categorical variable that indicates a member's frequency of church attendance within the community (e.g., "regular," "irregular," "to be transferred").
How did you learn about the religious community?	Explains how the member first encountered the religious community. It is a nominal categorical variable.
What was the deciding factor in your decision to be baptized?	Describes the main factor or reason that motivated the decision to be baptized. It is a nominal categorical variable.
Association	Indicates which regional association the church the member attends belong to. This is a nominal categorical variable.
Church Name	Identifies the specific congregation to which the member belongs. This is a nominal categorical variable.

3.1.2. Web Platform Design

After selecting the predictive variables, a three-layer web platform architecture was designed to separate the user interface, system logic, and data management. The front-end, developed with Next.js and React, provides user interaction, while the back end, implemented in Python with Flask, processes requests and hosts the Random Forest prediction model. The database stores user information and the data required to estimate member attrition probabilities. Figure 1 presents the resulting platform architecture.

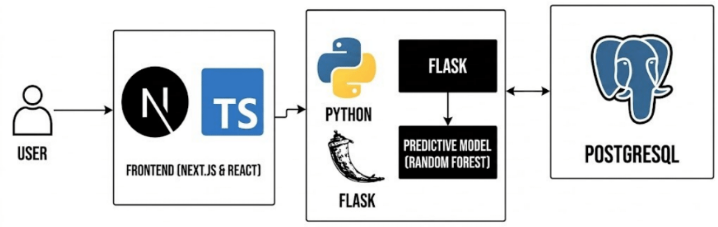

Figure 1: Architecture Diagram.

A data flow diagram was also designed using Gane & Sarson notation, which allows for the visualization of data movement through the application. This diagram, presented in Figure 2, illustrates how community member data is captured, processed, and transformed to make a prediction. The process begins when a user interacts with the website to upload member data. The data then undergoes a cleaning phase before being stored and, finally, used to train the predictive model.

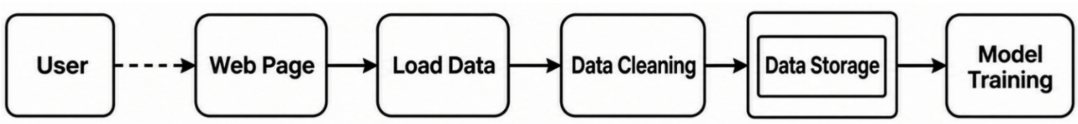

Figure 2: Data flow diagram of the predictive system.

In addition, the system's functional and non-functional requirements were defined based on the needs of the users (community leaders) and the project's objectives. Table 2 details the functional requirements, which describe the system's capabilities, and the non-functional requirements, which focus on performance and include usability, security, and code maintainability.

3.1.3. Development of the Model and Web Platform

After reviewing the variables, data cleaning was performed in Google Colab using Python libraries. Categorical values with spelling inconsistencies were standardized using the get_close_matches function, and categorical variables were encoded through Label Encoding. The dataset was then divided into 80% training and 20% testing subsets to ensure reliable model evaluation. The predictive model was developed using the Random Forest algorithm, selected for its effectiveness with datasets containing multiple, predominantly categorical variables [26].

As an ensemble learning method based on bagging, Random Forest builds multiple decision trees from random samples and combines their predictions, reducing overfitting while improving robustness, generalization, and feature importance estimation [27]. Feature importance analysis identified Baptism_Year (29%) as the most influential predictor, followed by Church Name (21%), Association (17%), Classification (13%), and Age (9%), whereas the remaining variables contributed only marginally to the prediction (Figure 3).

Furthermore, the predictive model developed using the Random Forest algorithm demonstrated outstanding performance in classifying community members. The evaluation was conducted on the test dataset, measuring overall and class-specific metrics. The model achieved an overall accuracy of

Table 2
System Requirements.

Requirement Type	Description
Functional	RF-01: The platform must implement a predictive model that allows determining the percentage of community members who will leave the community. RF-02: The platform must allow for the display of graphs showing community details, making it easier for leaders to access information. RF-03: The platform must have high-level security to protect the religious community's private data.
Non-Functional	RNF-01: The platform must be capable of supporting more than 500,000 community members without compromising performance. RNF-02: The user interface must be intuitive and easy for users to understand. RNF-03: The system must be compatible with major web browsers. RNF-04: The system must be compatible with the server infrastructure of the selected institution. RNF-05: The system must support the installation of environments and databases approved by the institution's Information System. RNF-06: The system must comply with established security regulations for the protection of religious community members' data.

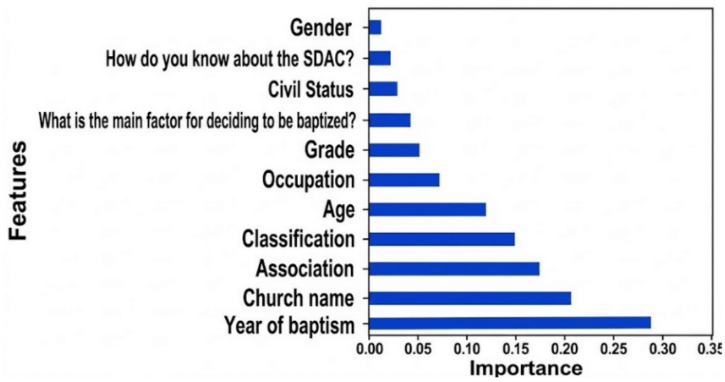

Figure 3: Prediction model for dropout, using the Random Forest algorithm.

91%, indicating a high hit rate. The model's discriminative power was confirmed by an AUC ROC value of 0.96, demonstrating its excellent ability to differentiate between members who remain active and those at risk of churn. Detailed performance by class is presented in Table 3.

Table 3
Model Classification Metrics Report.

Class	Precision	Recall	F1-score
0 (dropout)	0.82	0.79	0.80
1 (active)	0.94	0.95	0.94
Macro avg	0.88	0.87	0.87
Weighted average	0.91	0.91	0.91

As shown in Table 3, the model achieved a balanced performance across both classes. For Class 1 (active members), it obtained high precision (0.94), recall (0.95), and F1-score (0.94), demonstrating strong capability in identifying retained members. For Class 0 (churn), the model achieved a precision of 0.82, recall of 0.79, and F1-score of 0.80, indicating robust detection of attrition cases despite slightly favoring retention [28]. Figure 4 presents the confusion matrix, which shows that the model correctly classified 33,733 active members and 8,525 churn cases, while producing 1,895 false negatives and 2,307 false positives.

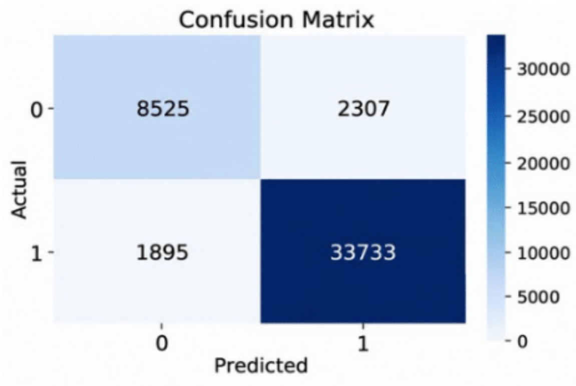

Figure 4: Confusion matrix.

The results indicate that the model is slightly more effective at identifying retained members than detecting attrition, while maintaining high overall performance. Its ROC AUC of 0.96 confirms excellent discriminatory capability between the two classes. In addition to the predictive model, a three-tier web platform was developed using Next.js and React for the front-end, Python with Flask for the back end, and PostgreSQL for data management.

The platform includes an authentication module to protect sensitive information, enables leaders to generate predictions for individual members, and provides interactive visualizations of community data. To mitigate the risk of identifying individual members through quasi-identifiers, the platform incorporates a secure authentication system; consequently, the raw dataset was not made publicly available. After authentication, users access a dashboard with general reports and usage indicators, including monthly and weekly login statistics (Figure 5).

Figure 5: Homepage with general reports.

Figure 6 presents the platform's Prediction Module, which enables community leaders to interact directly with the Random Forest model. The interface displays member records, including the variables used for training, such as church, age, gender, marital status, year of baptism, association, and status. By selecting the Predict button, the system processes the selected member's data and immediately displays the predicted status (active or inactive) together with the associated confidence level (probability).

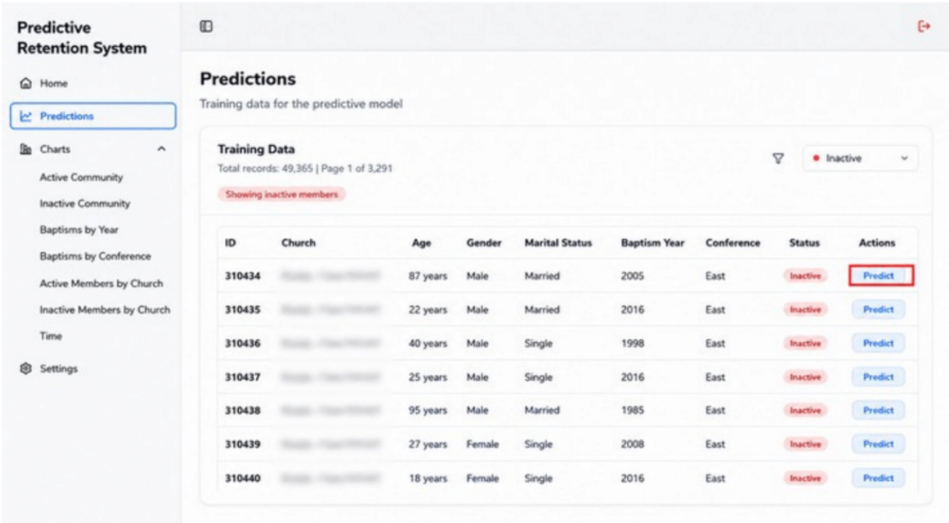

Figure 6: Prediction module.

Figure 7 illustrates the prediction results displayed after selecting the Predict button. The platform presents a results card showing the member's predicted status (Active or Inactive/At Risk of Churn) together with the probability of retention, which represents the model's confidence in the prediction. This interface provides community leaders with immediate and actionable information generated by the Random Forest model.

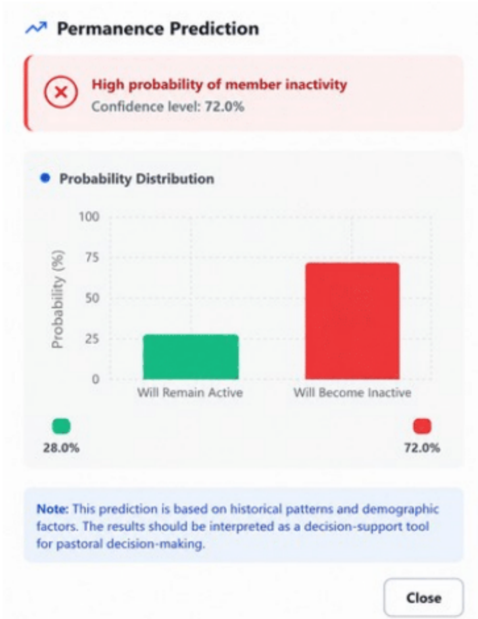

Figure 7: Prediction result.

The platform also includes a descriptive analytics module with interactive visualizations to support community analysis and decision-making. Figure 8 presents the distribution of active members by age group, showing that participation is highest among younger members, particularly those aged 10–19, followed by the 20–29 group. Participation gradually declines with increasing age, with the lowest representation observed in the 80–89 and 90+ age groups.

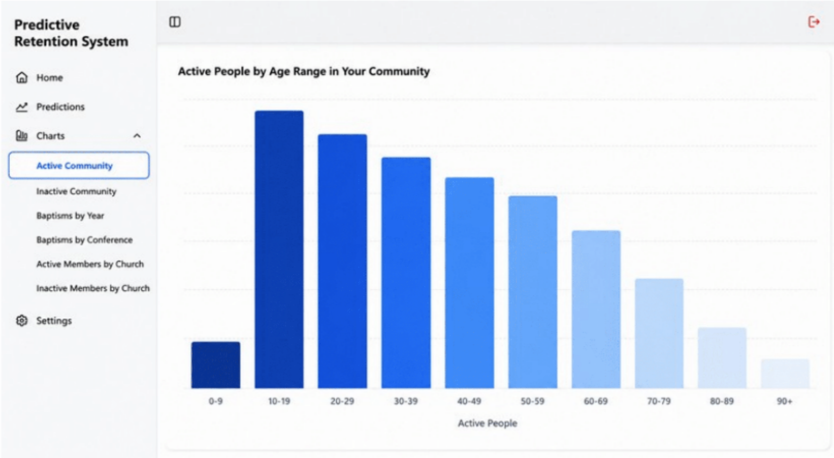

Figure 8: Active Members by Age.

Figure 9 presents the age distribution of inactive members, revealing that attrition is concentrated in middle and older adulthood. The highest number of inactive members is found in the 30–39 age group, followed by the 20–29 and 40–49 groups, indicating that working-age adults account for a large proportion of inactivity. As with the active community, the 0–9 and 90+ age groups have the lowest numbers of inactive members.

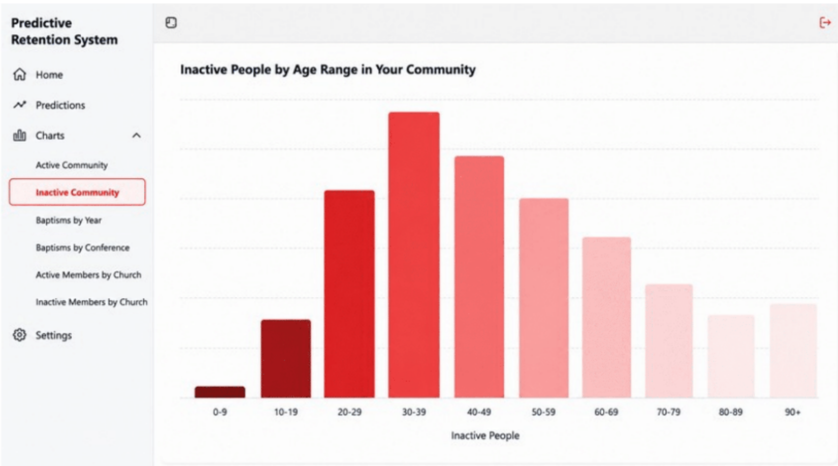

Figure 9: Inactive Population by Age.

Figure 10 illustrates the annual baptism trend, providing a historical view of community growth. Baptisms increased between 2015 and 2019, declined sharply in 2020, and then recovered steadily through 2021–2023, with 2023 recording the highest number of baptisms. Data for 2024 and 2025 show partial or lower figures compared to the 2023 peak. This analysis helps leaders identify periods of growth and years requiring further evaluation.

The platform also presents the distribution of members by congregation and administrative association, showing considerable differences in membership size across regions. Larger associations contain the highest numbers of active members, with some local congregations exceeding 1,400–1,800 active participants, providing leaders with valuable information for prioritizing retention efforts. The analysis of inactive members similarly identifies regions and congregations with the greatest concentrations of attrition, highlighting key areas for intervention. However, this descriptive analysis is limited by the quality and completeness of the available data, as some associations could not be included due to outdated membership records, preventing a comprehensive community-wide assessment.

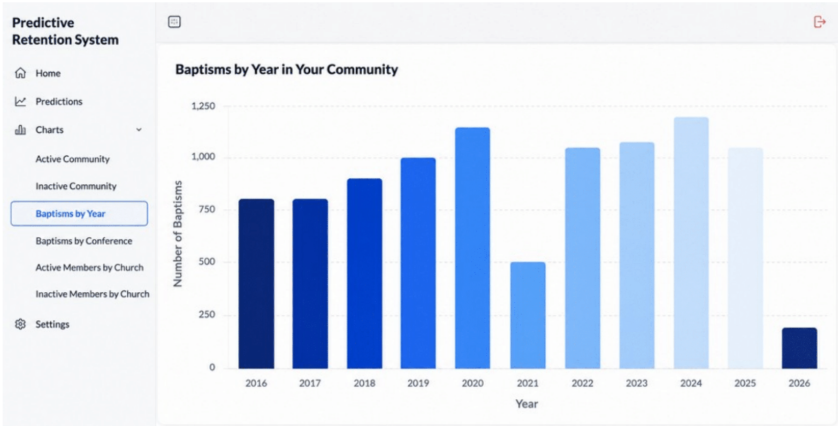

Figure 10: Baptisms by year.

3.1.4. Unit and functional testing

Unit testing verified that the platform's core components, including authentication, user registration, report generation, and prediction, operated correctly, allowing minor issues related to validation and exception handling to be resolved. Functional testing then confirmed that key processes, such as login, data queries, and database interactions, satisfied the defined requirements and behaved as specified from the user's perspective.

3.2. Stage 2: Pilot Test

The final phase evaluated the platform's usability and practical applicability with end users through a pilot study, combining qualitative feedback with the SUS. The usability evaluation was completed by a pilot sample of six participants, primarily serving as church pastors.

The SUS demonstrated good internal consistency, with a Cronbach's alpha of 0.84 [18]. Results indicated that participants generally perceived the platform as easy to use, consistent, and well-integrated, while opinions regarding the need for technical support were more varied. Overall, users considered the learning curve reasonable and the interface suitable for practical use, providing valuable feedback for further refinement. Figure 11 illustrate the SUS Scale Results.

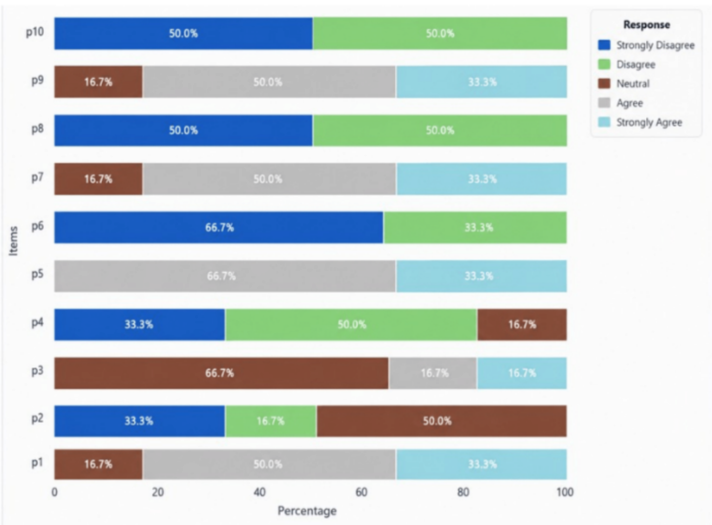

Figure 11: SUS Scale Results.

4. Discussion

The objective of this research was to develop a ML model to predict member retention in a religious community. The Random Forest model achieved high predictive performance (91% accuracy and AUC = 0.96), demonstrating that ML can effectively predict attrition in voluntary organizations, extending its application beyond traditional domains such as customer churn [9], employee attrition [7], and student dropout [29]. Feature importance analysis identified Year of Baptism, Church Name, and Association as the strongest predictors, reinforcing evidence that behavioral and engagement variables are more informative than static sociodemographic characteristics [4, 30].

These findings also confirm the suitability of Random Forest for datasets with predominantly categorical variables, consistent with previous studies highlighting its superior predictive performance. Beyond prediction, the platform provides descriptive analyses of age, baptism trends, and geographic distribution of active and inactive members, enabling leaders to identify high-risk groups and prioritize targeted interventions.

Finally, the usability evaluation showed positive perceptions of the platform's functional integration, supporting the transformation of predictive models into practical decision-support tools for non-technical users [31, 32]. The platform was designed to support twice the volume of data.

4.1. Limitations of the study

Despite the high predictive effectiveness achieved, it is imperative to acknowledge the inherent limitations of the study, particularly regarding the generalizability of the model and the depth of the predictive variables. First, the developed model is limited by its geographic and contextual scope to the specific community that served as the data source. Its direct transferability to other congregations is not guaranteed, as the factors of accuracy may vary substantially.

Second, the list of predictive variables is primarily external and observable. The limitation lies in the absence of intrinsic or psychosocial variables that directly measure the depth of an individual's faith or spiritual motivation. Faith, being an internal and fluctuating phenomenon, can lead to abandonment without external behavioral indicators reflecting it beforehand. Therefore, the model should be viewed as a high-performing proof of concept. Future research should explore incorporating qualitative data or indicators of internal commitment to enrich the dataset and create a model with greater explanatory power regarding long-term retention.

The pilot test is small because it is being conducted specifically with the leaders who handle the information. It could be extended to other congregations of the same denomination or other churches to obtain better results, but that is beyond the scope of this project. Thus, regarding the usability of the platform, it is restricted to a small sample of the benefits that the system provides.

5. Conclusion

This research addressed member attrition in religious communities by developing a Random Forest–based predictive model and integrating it into a web platform, followed by a pilot evaluation with community leaders in Medellín, Antioquia. The model achieved high predictive performance (91% accuracy, ROC AUC = 0.96, and F1-score = 0.80 for attrition), demonstrating the effectiveness of ML for predicting retention in a religious context. Feature importance analysis identified Year of Baptism, Church, and Association as the most influential predictors, highlighting the relevance of engagement-related variables over static sociodemographic characteristics.

The platform combined individualized predictions with descriptive analytics and demonstrated acceptable usability through the SUS evaluation, making it a practical decision-support tool for non-technical users. Overall, the study provides a data-driven approach to support retention strategies and community growth.

Future work should incorporate additional psychosocial variables, such as two or three relevant indicators, to further improve the model's predictive performance. The model should also be validated across different religious and geographic contexts, including other cities and countries, to assess its generalizability. Additionally, the pilot study should be expanded to include a larger sample size, providing more robust evidence of the platform's effectiveness. Future research should also examine the potential influence of baptism date and membership tenure on the model's predictions, considering whether the length of time since baptism may be associated with the risk of attrition. Finally, the platform could be extended with recommendation capabilities that automatically suggest targeted pastoral interventions for members identified as being at risk of attrition.

It is also recommended that a time split be made in the model training. Since dropout is an inherently time-dependent phenomenon, and the study itself shows changes in baptism trends over time, a random split could allow information from later periods to "leak through" into the training, producing an optimistic performance estimate relative to a realistic prospective prediction scenario.

Declaration on Generative AI

During the preparation of this paper, the first author used Gemini Notebook LM in order to verify information and enhance the text. After using this tool, the editing was done by the first author as needed, and he takes full responsibility for the final content.

References

[1] Oracle, ¿qué es la retención de empleados? importancia, beneficios y estrategia, 2024. URL: https://www.oracle.com/latam/human-capital-management/employee-retention/.
[2] H. J. Núñez Romo, Desarrollo de un plan estratégico para el crecimiento sano de una iglesia bíblica bautista, 2024.
[3] Alvis, Herrera, Pérez, Corporación universitaria adventista, 2007.
[4] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, A. Dinu, C. Stachl, Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics, Scientific Reports 13 (2023) 5705.
[5] R. Gupta, A. Sharma, A. Gupta, Predicting customer churn in telecom sector using ensemble machine learning techniques, 2020.
[6] M. S. Hossain, A. R. B. Ahlan, S. Hassan, A systematic literature review on customer churn prediction in telecommunication industry, 2021.
[7] A. Gonzalez-Nucamendi, J. Noguez, L. Neri, V. Robledo-Rella, R. M. G. García-Castelán, Predictive analytics study to determine undergraduate students at risk of dropout, in: Frontiers in Education, volume 8, Frontiers Media SA, 2023, p. 1244686.
[8] M. Afzal, S. Rahman, D. Singh, A. Imran, Cross-sector application of machine learning in telecommunications: enhancing customer retention through comparative analysis of ensemble methods, IEEE Access 12 (2024) 115256–115267.
[9] S. Arefin, R. Parvez, T. Ahmed, F. Jahin, F. Sumaiya, M. Ahsan, Enhancing bank consumer retention: a comprehensive analysis of churn prediction using machine-learning and deep learning techniques, in: International Conference on Advances in Information Communication Technology & Computing, Springer, 2024, pp. 265–281.
[10] D. O. Kireri, J. Gathogo, F. Mwaniki, Abandonment of church by youth-believers in the seventh-day adventist church in kisii county, kenya, 2022.
[11] X. Tan, S. Lim, T. Lim, C. Tan, N. Husaini, Leveraging random forest algorithm for enhanced lead conversion and customer retention, Int. J. Adv. Sci. Eng. Inf. Technol 14 (2024) 2141–2148.
[12] J. Han, M. Kamber, J. Pei, Data mining: Concepts and, Techniques, Waltham: Morgan Kaufmann Publishers 13 (2012).
[13] G. James, D. Witten, T. Hastie, R. Tibshirani, et al., An introduction to statistical learning: with applications in R, volume 2, Springer, 2021.

[14] E. Acosta Gonzaga, J. Álvarez Cedillo, A. Gordillo Mejía, Arquitecturas en n-capas: Un sistema adaptivo. polibits, 34, 34–37, 2006.

[15] L. Base, P. Clements, R. Kazman, Software architecture in practice, Boston, MA: Addison-Wesley, 2003.

[16] C. Gane, T. Sarson, Structured systems analysis: tools and techniques, McDonnell Douglas Systems Integration Company, 1977.

[17] Y. V. Regalado, La arquitectura de software como disciplina científica, Serie Científica de la Universidad de las Ciencias Informáticas 3 (2010) 1–4.

[18] L. J. Cronbach, Coefficient alpha and the internal structure of tests, psychometrika 16 (1951) 297–334.

[19] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, " O'Reilly Media, Inc.", 2022.

[20] S. V. Arias, T. M. Soria, P. N. Moya, P. M. Palma, Control de calidad del software mediante pruebas automatizadas de integración y pruebas unitarias, Ciencia Digital 2 (2018) 101–115.

[21] N. P. L. Larrea, M. V. R. Valencia, B. F. H. Ponce, A. E. S. Orozco, Análisis de la aplicación de pruebas funcionales y pruebas de usabilidad de software en el desarrollo de sistemas web, Ciencia Digital 3 (2019) 180–190.

[22] R. Hernández Sampieri, C. Fernández Collado, P. Baptista Lucio, et al., Metodología de la investigación, volume 6, Mcgraw-hill México, 2014.

[23] J. Brooke, System usability scale: A 'quick and dirty'usability scale, Usability evaluation in industry (1996).

[24] M. Q. Patton, Qualitative research & evaluation methods: Integrating theory and practice, (No Title) (2015).

[25] J. W. Creswell, J. D. Creswell, Research design: Qualitative, quantitative, and mixed methods approaches, Sage publications, 2017.

[26] R. Abuzayeda, I. Karamitsos, A. Kanavos, Advancing student guidance using classification data mining techniques, in: 2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), IEEE, 2024, pp. 1–6.

[27] T. Hastie, R. Tibshirani, J. H. Friedman, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, volume 2, Springer, 2009.

[28] T. Fawcett, An introduction to roc analysis, Pattern recognition letters 27 (2006) 861–874.

[29] S. Paigude, S. C. Pangarkar, S. Hundekari, M. Mali, K. Wanjale, Y. Dongre, Potential of artificial intelligence in boosting employee retention in the human resource industry, International Journal on Recent and Innovation Trends in Computing and Communication 11 (2023) 01–10.

[30] Y. A. Zelenkov, A. S. Suchkova, Predicting customer churn based on changes in their behavior patterns, 2023.

[31] T. Susnjak, Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and chatgpt, International Journal of Artificial Intelligence in Education 34 (2024) 452–482.

[32] K. Alalawi, R. Athauda, R. Chiong, An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention, International Journal of Artificial Intelligence in Education 35 (2025) 1239–1287.