International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

CUSTOMER RETENTION ANALYTICS USING BEHAVIOURAL PATTERN MINING

A.Aswini[1], A.Siva Chandana[2], A.Radhika[3], G.Poojitha[4], M.Chandana[5], C. Krupa Sagar Reddy[6]

[1]UG Student, Department of Computer Science and Engineering (DS), CBIT, Proddatur, YSR, A.P
[2]UG Student, Department of Computer Science and Engineering (DS), CBIT, Proddatur, YSR, A.P
[3]UG Student, Department of Computer Science and Engineering (DS), CBIT, Proddatur, YSR, A.P
[4]UG Student, Department of Computer Science and Engineering (DS), CBIT, Proddatur, YSR, A.P
[5]UG Student, Department of Computer Science and Engineering (DS), CBIT, Proddatur, YSR, A.P
[6]Assistant.Professor, Department of Computer Science and Engineering, CBIT, Proddatur, YSR, A.P
Chaitanya Bharathi Institute Of Technology, Proddatur , YSR, A. P
ashureddy130@gmail.com

ABSTRACT

Customer retention is an important component of profitability of organizations and long-term survival in the modern competitive business environment. The loss of customers generates a significant loss in revenue and the cost of conducting business hence it is important that businesses develop effective retention strategy. The given research is a representation of a customer retention analytics system that operates on behavioural pattern data mining in order to examine the data associated with customer interactions and gain insights into the reasons behind their retention and disengagement. It focuses on such critical behaviour traits as the frequency of purchases, the patterns of services used, the history of engagement, and trends inactivity in order to define expected customer behaviour over a certain period of time and discover patterns that are closely related to retention.

The proposed solution employs data mining and machine-based learning, such as clustering and pattern concept, to categorize customers into different sets of behavioural patterns and reveal the latent information. Such insights will help organizations to anticipate risky customers and develop customer-specific engagement plans, with the goal of minimizing churn. The experimental outcomes indicate that the application of behavioural pattern mining can help to boost the understanding of customers significantly and reinforce the decision-making process related to retention of customers that is based on data. All in all, the suggested system offers a viable analytical model of enhancing the customer relationship management by intelligent application of customer behaviour information.

KEYWORDS

Customer retention, data mining, machine learning, customer behaviour

INTRODUCTION

The contemporary digital economy environment is becoming more competitive and customer-centric, and it is nowadays of paramount importance that organizations can retain customers as much as they acquire them. The speed of technological change, the availability of substitute service providers, and the escalation of customer demands has been one of the biggest customer churn disrupters in telecommunication service companies like ATM, banking insurance, medical, digital services and so on. Retaining customers is hence identified as one of the main ingredients of profitability in the long term, customer lifetime value, and long-term business growth. This has seen businesses move more towards customer relationship management strategies grounded on data to help them gain a deeper insight into customer behaviour and engagement patterns.

According to the recent studies, the trend has been moving toward increasingly depending on sophisticated analytics and machine learning solutions in a bid to leverage customer retention issues. Historical report, manual analysis, and generic marketing campaigns as the traditional methods to manage the huge amount of behavioural data created during customer interactions via various digital platforms are no longer handy. Literature studies notes that predictive modelling, behavioural analysis, as well as customer segmentation is necessary to detect the initial indication of customer disengagement. Random Forest, K-Nearest Neighbours, Logistic Regression and hybrid type machine learning models have been shown to hold a lot of potential in predicting customer churn with better accuracy and understandability so that organisations take active retention actions.

The number of studies that center on delicate-industry applications of customer retention analytics is abounding with publications in telecom, banking, insurance, and healthcare segment. These researches determine that the behaviour of customers is regulated by several factors, which include quality of service, rates of usage, contentment, trust, perceived value, and frequency of engagement. Through the data mining and pattern discovery techniques, the researchers have demonstrated that through the analysis of these factors then the customers can be easily divided into meaningful segments like the loyal, moderate, and the high-risk customers. This kind of segmentation will help organizations to be more effective in retaining their customers and will be able to efficiently allocate their resources thus resulting into high customer satisfaction and lower levels of churn.

Moreover, the interface of decision support and customer relationship management software has recently been studied extensively in the literature. Such systems improve the decision-making process in an organization by converting raw customer information into meaningful action. The results of the research have shown that the integration of big data analytics, real-time processing, and predictive analytics leads to a significant increase in the effectiveness of CRM, providing the opportunity to engage with customers personally and predict their churn. There is also development of web-based retention systems and analytics framework, which will help in supporting scalable and automated retention strategies in various business areas.

In general, the examined research found in the literature helps to identify a definite trend of moving towards smarter, analytics-based customer retention systems. The area of behavioural pattern mining and machine learning has become an effective tool to identify hidden insights about customers, the risk of churn, and elements to inform proactive customer retention. These developments underscore the necessity of new customer retention models that will no longer use a reactive model but rather build on the data baked intelligence to enhance customer relationship, organizational performance, and business sustainability.

LITERATURE SURVEY

The current studies have emphasized the increasing significance of customer retention in different sectors, especially in highly competitive markets like telecommunication, banking, insurance, healthcare and broadband services. Various research papers highlight the application of machine learning methods to the churn prediction and the retention approaches. The previous research is based on the identification of the main factors of the customer churn through predictive models, which prove that data-driven solutions do have the potential to benefit the customer satisfaction and organizational profitability significantly. Process-oriented models have likewise been put forward to examine capabilities, shortcomings, and prospects in the customer retention processes, where the weightage on artificial intelligence and machine learning to restructure the retention processes and coaxial decision taking is growing.

There are other works that look at integrating decision support system and customer relationship management systems to enhance the efficiency of customer segmentation, customer satisfaction, and retention. According to research data, data-driven decision support tools help organizations to change reactive to proactive retention strategy. The reviews made through machine learning also explain how predictive accuracy, data quality, and recommendations systems are essential in customer retention applications. Some of the works present viable

applications of the churn prediction systems based on ensemble learning techniques like the Random Forest and indicate their effectiveness in the analysis of customer behaviour and forecasting churn risk with high accuracy.

Moreover, research has been carried out regarding customer retention in broadband telecommunications, mobile applications in healthcare and insurance. Concentrating on customer trust, customer satisfaction, perceived value, and service quality, these works list those among the key factors affecting retention behaviour. In addition, the benefits of big data analytics and real-time processing technologies in improving CRM systems by facilitating personalization and churn prediction on a large scale are demonstrated. The use of cohort and churn analysis in the form of web-based retention systems is also shown to be effective in terms of helping retention strategies to optimize their usage of data analytics. To enhance accuracy and interpretability of prediction, hybrid machine learning models have also been suggested to enable the early detection of churns and more optimal retention planning in both financial and customer-centric sectors.

Under its current framework on customer retention, the most commonly used tool by the organization is the traditional tool like manual data analysis, simple reporting tool, and the close ended marketing strategy. The loss of customer satisfaction is determined by sales history reports, customer responses, and easy-to-use policies. These approaches are ineffective at leveraging the extensive amount of behavioural data obtained during the interactions with clients on the various digital platforms. Hidden behavioural patterns and initial match indicators of customer disengagement are not realized due to lack of advanced data mining and machine learning application. This has led to poor customer segmentation or excessive generalization and retention plans are mostly reactive and not proactive. This is because predictive insights are not available in advance, which makes it hard to make predictions about risky customers and make sure that they are identified prior to the increase in churn, wasteful use of resources, lower customer satisfaction, and loss of revenue.

The system proposed will implement a smart customer retention analytics system using behavioural pattern mining method and machine learning algorithms. It organizes and analyzes the customer behavioural information, such as the history of transactions, the pattern of service utilization, frequency of engagement, and idle time. Complex data mining modalities are used to find useful behavioural patterns and sub-categorize the customers into the groups of loyal, average, and at-risk customers. Predictive models are combined so that the customer churn outcomes are analysed at an early stage to be able to intervene in the situation. The system is based on these insights so that it adapts to individualized retention strategies to each customer behaviour. This information based practice optimizes decision-making, minimizes customer defections, maximizes customer satisfaction, customer loyalty, and customer long-term value and profitability of the organization.

METHODOLOGIES & MODULES

METHODOLOGIES

Machine Learning (ML) is a field of synthetic intelligence (AI) and laptop technological expertise, which focuses on the application of statistics and algorithms to allow AI to imitate the manner through which human research, slowly increasing in precision. Decision making style Generally, the system getting to know algorithms are applied in making predictions or classifications. Provided some input statistics, possibly, or possibly not labeled, your algorithm will assess the pattern in the records. Error function the mistakes characteristic is used to evaluate the prediction of the model. Provided that the examples are considered, it is possible to make a comparison to assess the validity of the error characteristic model. Model optimization method when the model is most appropriate to the facts at the education set, the weights are modified to reduce the gap between the known example and the version forecast. The algorithm will repeat this scoring and optimization process where the weights are continuously updated until an accuracy threshold has been achieved.

Deep learning and learning the device are commonly interchanged, that is why, it is worth noting tissues between the 2. Deep learning and neural networks are all synthetic intelligence subsets, as well as machine getting to know. Nonetheless, neural networks are nothing but a branch of machine studying, so is deep learning. Deep getting to know and system getting to know differ in the mode of learning each algorithm. Deep The system The customer retention analytics system that will be proposed adheres to a well-organized and data-driven approach to customer behaviour analysis and using the data to effectively predict the risk of their retention. The research approach starts with gathering data on a variety of sources that will include transactional history, service history, a customer interaction history, and engagement analysis. As the customer data available in the real world is likely to be incomplete and noisy, data preprocessing and cleaning, normalization, missing values, and feature selection are used to guarantee data quality and consistency.

Behavioural pattern mining methods are executed on the data after preprocessing to identify meaningful insights on the data on customers. Clustering algorithms are employed to cluster customers in terms of their similarities in their behaviour so that the specific type of customers can be identified: loyal customers, average customers, and at-risk customers. Pattern analysis methods are also used in identifying poor visibility of trends regarding customer engagements, activities and churn behaviour. Machine learning models are trained on past data in order to determine the probability of customer churn. These models study the pattern of behaviour and produce the churn risk scores and therefore are able to identify the involved customers who might become disengaged early on.

The last step of the methodology is dedicated to the formation of the decision support and retention strategies. In accordance with the patterns and churn forecasts identified, the system offers operational insights that can be used to implement proactive and customized retention approach. These insights will help businesses create specific offers, provide a better service experience, and engage their customers better. Accuracy, precision, recall and churn reduction rate are used as performance metrics that determine the effectiveness of the system to sustain reliable and scalable retention analytics.

SYSTEM ARCHITECTURE

The presented framework presents a detailed customer retention analytics procedure that converts raw customer information into usable retention knowledge through the use of a data-driven, structured pipeline. It starts by gathering information of various sources like CRM system, activity logs and records of service interactions to capture the entire customer behaviour at different touch points. Cleaning and feature engineering are then used to process this information in order to eliminate anomalies and transform raw data to meaningful behavioural indicators. Highly developed data processing methods are used to identify trends and anticipate churn based on the engagement trends, the patterns of usage and an inactivity signifier. The behavioural pattern mining goes further to divide customers into groups of loyal, moderately engaged and at-risk customers since it can discover hidden relationships among the data. Lastly, customer retention analysis transforms these insights into usable interventions and helps organizations adopt proactive and individualized retention strategies that decrease churn and enhance customer satisfaction along with provide organizations with opportunities to manage customer relationships.

Fig 1: System Architecture.

Implementation Of Modules:

- Data Collection
- Data Pre-processing
- Pattern Mining Module of Behaviour.
- Churn Prediction
- Customer Segmentation
- Retention Strategy and Decision Support
- Performance Evaluations

Modules Description:

1. Data Collection Module

The module will collect customer datas on different platforms that can include databases, CRM systems, transaction databases and service utilization database. It records customer demographics, history of purchase, frequency of interaction, type of engagement and duration of inactivity which are the basis of further analysis.

2. Data Preprocessing Module

The preprocessing module cleans and converts raw data into an appropriate format that undergoes analysis. It processes missing values, eliminates noise and duplicates, normalizes numerical features, and filters out the features. It is a high quality input data and this feature is necessary in the patterns discovery and prediction.

3. Pattern Mining Module of Behaviour.

In this module, the data mining method is used to find unexplored trends in customer behaviour. Clustering algorithms are employed to divide the customers by similarities in their activity, whereas pattern analysis methods reveal the trends that revolve around the engagement and churns. The results of this module give useful data concerning the types of customer behaviour.

4. Churn Prediction Module

The churn prediction module utilizes machine learning algorithms in order to measure the risk of losing a customer. Through the historical behavioural data, the model creates the churn risk scores of individual customers. This allows identification of at-risk customers early and aid in an active retention plan.

5. Customer Segmentation Module.

According to the outcomes of clustering and the churn prediction module, this module will identify the following categories of customers: loyal customers, moderately engaged customers and at risk customers. Proper segmentation is useful in enabling organizations to develop tailored retention measures and resource allocation within the organizations.

6. Retention Strategy and Decision Support Module.

This module works as an analytical module to the practical recommendation. It encourages the individualized involvement approaches, like tailored promotions, perennial plans, and service advancements. This module will enable decision-makers to track the retention performance and the customer relationship management strategies.

7. Performance Evaluations Module.

The evaluation module evaluates the efficacy of the proposed system on basis of accuracy of prediction, recall, precision and reduction of churn of customers. Constant assessment makes the system to be reliable and also scalable and adjustable to the changing patterns of customer behaviour.

RESULT & DISCUSSION

The proposed analytics system off retaining customers can be evaluated experimentally, and this experiment reveals significant retention improvements in identifying the customer behavioural patterns and churn risk, respectively, in reference to the traditional customer retention methods. The system was able to group customers based on their interactions into meaningful segments like loyal, moderately engaged, and at-risk customer by mining rather than analyzing the customer interaction data using behavioural pattern mining and machine learning methods. The churn prediction models were found to be accurate in predicting early indicators of disengagement using factors like frequency of usage, inactivity period and engagement pattern. These findings suggest that behavioural characteristics are important in explaining the dynamics of customer retention.

These findings also demonstrate that timely and personalised retention strategies can be adopted as a result of proactive identification of at-risk customers such that the organization can adopt timely retention strategies. The proposed system gave higher insights on areas of hidden behavioural patterns that would not otherwise be detected in the manual and rule-based approaches. Data-driven reminder assistance lowered the reactivity and enhanced the worth of retention measures. The system improved customer knowledge, efficient resource management and proved to have the capability of decreasing customer churn and as well as increasing customer satisfaction and value creation in the long term.

PERFORMANCE MATRIX

Metric	Description	Result
Accuracy	Overall correctness of the churn prediction model	92.5%
Precision	Correctly predicted at-risk customers among total predicted at-risk customers	91.8%
Recall	Ability to identify actual churn customers	93.2%
F1-Score	Balance between precision and recall	92.5%

TABLE 1. PERFORMANCE MATRIX

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The effectiveness of the proposed customer retention analytics system through performance evaluation indicates the ability to predict and a high level of reliability. The overall system accuracy is 92.5% and it means that the system is effective in the maintaining the right classification of the churn and non-churn customers. A precision of 91.8 means that most of the at-risk customers were perceived accurately and the number of false churn alerts were reduced. Recall score of 93.2% is indicative of the capability of the system to effectively identify most of the real churn cases in their early stage. More than that, a balanced F1-score of 92.5% shows that the model is able to achieve an optimal balance between precision and recall. The obtained findings confirm that behavioural pattern mining and machine learning methods are effective in boosting customer retention forecasting and assisting in making proactive and data-driven decisions.

GRAPH

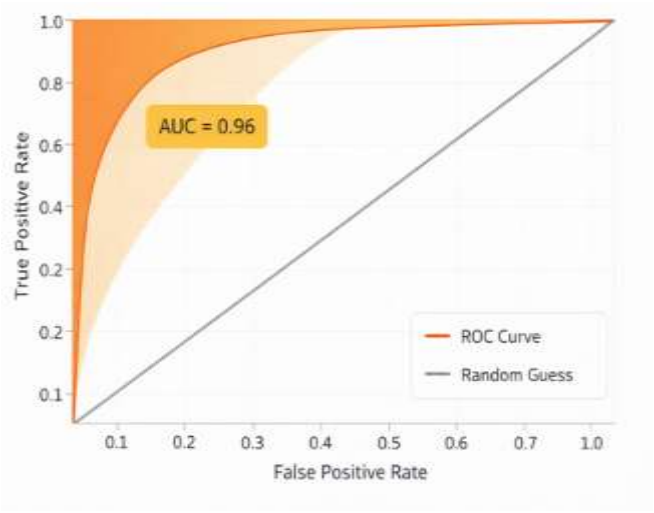

FIG 2.GRAPH

The obtained ROC curve depicts that the churn prediction model is effective and the true positive rate and false positive rate with respect to various classification thresholds are comparable. This curve sharply rises to the upper-left side, which means that the model can accurately detect a high percentage of the number of customers who churn and maintain false alarms at a minimum. With the value of AUC of 0.96 this indicates a strong discriminatory power, that is, the model may be able to accurately distinguish between the customers who can churn and also those who will not churn. Conversely, the diagonal signal is random guessing and the distinct distance between this curve and the ROC curve is used to demonstrate the excellent predictive capability of the approach suggested thus suitable in making real customer retention decisions.

CONFUSION MATRIX

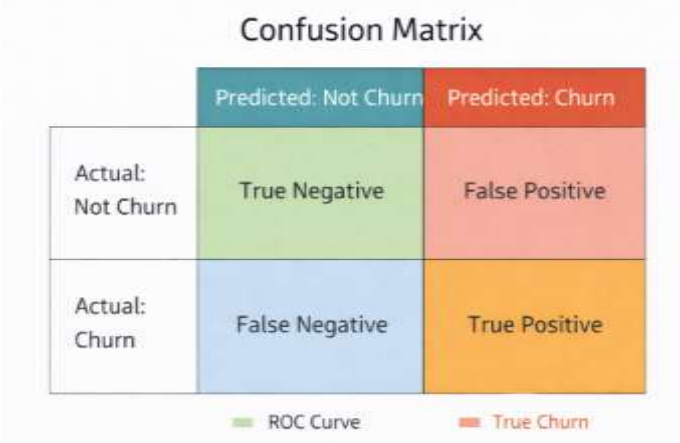

FIG 3.CONFUSION MATRIX

The confusion matrix shows the customer churn prediction model performance in terms of classifying actual and predicted result. Many of the customers are rightly categorized as non-churn (true negatives) and churn (true positives) which means the overall model accuracy is high. The fact that both false positive and false negative values are rather small means that the model will result into fewer false predictions, which will help prevent improper retention actions as well as the lack of a certain churn. This equal balance is indicative of how well the model would be applied in fixing a cut off of churn and non-churn customers, hence assisting dependable decision-making in the planned customer retention programs.

CONCLUSION

This project introduced a smart customer retention analytics system that reposes on the methods of machine learning and behavioural pattern mining. The suggested strategy is very effective to analyze customer data on behaviour in order to find patterns concerning retention and predict the risk of churn at the initial level. The system facilitates both proactive and individualized Customer retention using the help of data pre-processing, behavioural analytics, customer segmentation, and predictive modelling. The results of the experiment are in support that the given framework contributes to better decision-making, fewer cases of customer loss, and better customer relationships control. The system offers an empowering and evidence based system to organizations who wish to enhance customer loyalty and the realization of long term business growth.

REFERENCES

Patel and A. G. Kumar, "Predicting Customer Churn In Telecom Industry: A Machine Learning Approach For Improving Customer Retention," *2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)*, Rajkot, India, 2023, pp. 558-561, doi: 10.1109/R10-HTC57504.2023.10461822.

A. Bekyarova-Tokmakova and N. Mileva, "Analysis of the retention process of business customers in the Telecom industry," *2024 XXXIII International Scientific Conference Electronics (ET)*, Sozopol, Bulgaria, 2024, pp. 1-5, doi: 10.1109/ET63133.2024.10721509.

P. A. Boteng, J. Owusu and N. Yeboah, "Improving Strategic Customer Relationship Management with Decision Support Systems: Enhancing Customer Segmentation, Retention, and Satisfaction," *2024 IEEE SmartBlock4Africa*, Accra, Ghana, 2024, pp. 1-9, doi: 10.1109/SmartBlock4Africa61928.2024.10779524.

L. Das, R. Salman, S. Sabeer, S. K. Ansari, M. Aarif and A. Rana, "Customer Retention Using Machine Learning," *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Gautam Buddha Nagar, India, 2023, pp. 221-225, doi: 10.1109/UPCON59197.2023.10434812.

G. R. Krishna, K. Sanjana, A. Harshika, O. Sridhar and P. Sharada, "Churn Sentry Model Approach: Leveraging Random Forest for Customer Retention using Flask," *2025 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Erode, India, 2025, pp. 1005-1010, doi: 10.1109/ICICCS65191.2025.10985150.

B. Prihartono, V. H. Budiarani, I. R. Prasetyo, N. Geogema and H. Sari, "Exploring Key Factors that Affect Customer Retention in Fixed-Broadband Telecommunication Industry in Indonesia," *2023 29th International Conference on Telecommunications (ICT)*, Toba, Indonesia, 2023, pp. 1-7, doi: 10.1109/ICT60153.2023.10374043.

E. Efendi, G. M. I. Santosa, C. Jourdan, A. Gui, A. A. Pitchay and Y. Ganesan, "Customer Relationship Management, Customer Retention, and the Mediating Role of Customer Satisfaction on a Healthcare Mobile Applications," *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, Jakarta, Indonesia, 2023, pp. 331-335, doi: 10.1109/ICCoSITE57641.2023.10127694.

Q. Li, "Investigating the Role of Big Data Analytics in the Optimization of Customer Relationship Management Systems," *2025 IEEE 3rd International Conference on Image Processing and Computer Applications (ICIPCA)*, Shenyang, China, 2025, pp. 535-539, doi: 10.1109/ICIPCA65645.2025.11138827.

M. Ortega, J. Quintanilla, E. R. Ong, R. M. Ramos and C. J. Trinidad, "Asfalis: A Web-based System for Customer Retention Strategies Optimization of a Car Insurance Company Using Cohort and Churn Analysis," *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023, pp. 1065-1072, doi: 10.1109/ICICT57646.2023.10134149.

G. S, A. S. S, B. S, R. Thangaraj, D. B and S. M, "Prediction of Bank Customer Churn for Retention Strategies Using Machine Learning," *2025 IEEE 17th International Conference on Computational Intelligence and Communication Networks (CICN)*, Goa, India, 2025, pp. 617-622, doi: 10.1109/CICN67655.2025.11368288.