IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PROACTIVE CHURN ANALYSIS FOR CUSTOMER RETENTION

Saranya M [1], Vinoth R [2], Karthiga J [3] ,Karthika M [4]

[1] Assistant Professor, [2] B.Tech-IT Student, [3]B.Tech-IT Student, [4]B.Tech-IT Student

[1,2,3,4] Department of Information Technology,

[1,2,3,4] paavai engineering college, Namakkal, Tamilnadu, India

Abstract: Customer retention is imperative for the banking sector because the cost of retaining existing customers far outweighs the cost of acquisition. In a highly competitive market, customer churn is one challenge that threatens profitability and growth. This project introduces Proactive Churn Analysis, the system that uses machine learning models and predictive analytics to pick up at-risk customers through early analysis of transaction history, product usage patterns, feedback, and engagement level. Unlike traditional reactive approaches, this system offers real-time monitoring and early warning alerts to enable banks to undertake timely and targeted retention strategies like offering personalized offers, improved support services, or financial incentives. Sentiment analysis detects dissatisfaction in customer interactions that could be deeper insights into concerns. User-friendly dashboards visualize data for effective decision-making. The scalable and adaptive design will address the complexity of customer behavior and banking services, which reduce churn rates and increase trust, loyalty, and long-term relationships. This data-driven, innovative approach improves customer satisfaction, drives revenue, and helps banks achieve sustainable growth in a competitive market.

Index Terms – Customer Retention, Churn Prediction, Data Visualization, Customer Engagement

I. INTRODUCTION

A Customer churn is one of the major issues in businesses because losing customers is not only a short-term loss of revenue but also affects the brand and its growth in the long term. For companies in industries like telecommunications, banking, and e-commerce, the cost of retaining customers often outweighs the costs of finding new ones because the latter can be substantially more expensive. The ability to predict churn allows businesses to implement preventive measures, reducing the likelihood of customers leaving and ensuring a more stable revenue stream. Churn prediction uses advanced analytics to identify patterns in customer behaviour that may indicate a risk of attrition. By studying historical data such as transaction patterns, product usage, interaction with customer service, and feedback, companies can gain valuable insights into the factors contributing to churn. This would, through predictive modelling techniques like supervised learning, identify the possibility of every customer leaving the service. Besides that, the use of sentiment analysis in customers' feedback identifies negative sentiments which might be early indicators of dissatisfaction. This project therefore seeks to develop a fully-fledged churn prediction system integrating machine learning algorithms towards classifying customers according to their possibilities of churning. Providing real-time monitoring, the system sends alerts to decision-makers about high-risk customers. Interventions could include personalized deals, better customer service responses, or proactive redressal of issues identified through sentiment analysis. The intent is not only to prevent churn but to create experiences that are personalized, leading to increased satisfaction, greater engagement, and long-term loyalty. The model is adaptive and scalable, and can work with multiple types of customer segments and data types. This allows the companies to have retention strategies tailor-made to different customer profiles, thereby making interventions more effective and resource-efficient. As more data is accumulated, the model can continue to evolve and improve in its accuracy and effectiveness at churn prediction. Ultimately, the proactive approach churning relationships through churn prediction will help organizations not only prevent customer loss but also maximize the value of existing relationships to achieve higher customer retention and ensure sustainable growth in a competitive market.

II.LITERATURE SURVEY

1. " A Review On Machine Learning Methods For Customer Churn Prediction And Recommentions For Business Practitioners:" by M . Atif Manzoor, M. Atif Qureshi, Etain Kidney , Luca Longo(2024):

In the Telecommunication Industry (TCI) customer churn is highly critical because the revenue of the service provider is very sensitive to the retention of existing customers. In this competitive market, it is essential for the service providers to figure out the concerns of their existing customers regarding their services because the cancellation of the services by the customers and switching to new service providers will not bring any good to the service provider. In the context of TCI, there are ample number of studies which has gone into prediction of customer churn though after performance analysis on these research it depicted there is more room to come up in it. Consequently, in this work, a novel architecture to predict the occurrence of customer churn, also known as Churn Net, in TCI.

2. "Deep Learning Enchanced Customer Churn Prediction In Telecommunication Industry:" by Somak Saha , Chamak Saha , Md. Mahidul Haque , Md . Golam Rabiul (2024):

The telecom sector generates a massive volume of data daily, as there is a vast client base. Decision makers and business analysts emphasized that acquiring new customers is more expensive than retaining the existing ones. For business analysts and CRM analyzers, it is important to understand the churn customers' reasons, as well as their behavior patterns from existing churn customers' data. This paper proposes a churn prediction model that uses classification, as well as clustering techniques to find out the churn customers and provides the factors behind the churning of customers in the telecom sector. Feature selection is done using information gain and correlation attribute ranking filter. The proposed model classifies churn customers' data using classification algorithms, in which the RF algorithm performed well by having correctly classified instances up to 88.63%. Effective retention policies are an essential task of the CRM to prevent churners. After classification, the proposed model divides the churning customer's data by classifying the churn customers into categories using cosine similarity for group-based retention offers.

3. "Predictability, and explainability of mobile telephony customer departure in telecommunication companies: " by David Freire, David Santos Mauricio Sanchez , Jose Lusis Castillo Sequera And Daniel Fiallo Moncayo (2021):

With significant growth since 2000, the telecommunication industry as at this date has 5.31 trillion users with revenues being by companies to tune up to $1.07 trillion in 2022. Competition among players drives the growth as several look for customers. The war sometimes leads to churn through telecom companies or changes of telecommunications operator through services received by its user. At present, algorithms are designed for churn prediction to help in initiating measures to prevent customer churn. However, despite the many factors influencing user retention or churn, there is no holistic list of churn factors and no list of algorithms describing the reasons for customer churn. This highlights the importance of our work in allowing us to know in advance which client will be a deserter (churn). This study aims to identify what factors intervene in the process of customer churn in telecommunication companies, outline existing techniques for prediction, and finally present advances in explainability through a systematic review of the literature from 2018 to August 2023 using the meta-search engines Scopus and Wos. In the case of churn-contributing factors, we determined 19 factors that were divided into 87 sub-factors and explained in 87 out of 112 articles reviewed. In the area of prediction, we were able to determine 26 unique algorithm techniques, and 16 combinations presented in 102 out of 112 articles reviewed.

4."Deep Learning Enhanced Customer Churn Prediction In Telecommunication Industry: " By Nagarjuna Jajam , Nagendra Panini Challa , Kamepalli S. L. Prasanna, And Ch Venkata Sasi Deepthi (2023):

To successfully operate online games, gaming companies are introducing the systematic customer relationship management model. Particularly, churn analysis is one of the most important issues, because preventing a customer from churning is often more cost-efficient than acquiring a new customer. Churn prediction models should, thus, consider maximizing not only accuracy but also the expected profit derived from the churn prevention. We, thus, propose a churn prediction method for optimizing profit consisting of two main steps: first, selecting prediction target, second, tuning threshold of the model. In online games, the distribution of a user's customer lifetime value is very biased that a few users contribute to most of the sales, and most of the churners are no-paying users. Consequently, it is cost-effective to focus on churn prediction to loyal customers who have sufficient benefits.

5."Sampaling Based Stack Frame Work For Imbalanced Learning In Churn Prediction:" By Some De ,P. Prabu (2002):

Churn prediction has recently gained a strong paradigm in the research community in support of data-driven operational decisions. Yet, datasets relevant to churn prediction are often imbalanced, with skewed class distributions presenting significant challenges. Traditional data-level solutions such as over-sampling and under-sampling have been widely used to alleviate these problems, but there is little coverage of case studies that develop these techniques into computationally advanced frameworks such as ensembles. Ensembles primarily rely on algorithmic diversity along with a fixed set of training instances for performance. This research introduces algorithmic diversity within the ensemble system by dynamically updating the training set using different sampling strategies, ultimately improving the performance of prediction related to imbalanced learning. Using data from the largest open hotel commerce platform, the study executes four experiments on the impact of sampling techniques and ensemble solutions. A new framework, "Stacking of Samplers for Imbalanced Learning," is proposed, aggregating the predictive power of sampling solutions to enhance the meta-features utilized by the ensemble.

III.EXISTING SYSTEM

The current system of response to customer churn uses simplistic post-event handling methods about managing customer issues and response adaptively. General comments are gathered from varied means such as online reviews and even social media, and inferences are drawn about principal trends and sentiments. Summation of customer responses is provided through this inferring process, and their behavior is a crucial variable for the churn prediction model. This approach gives insights into probable churn factors but relies more on retrospective data and lacks advanced mechanisms for predicting and addressing churn proactively

Disadvantages

• Retrospective Focus

The current system bases its operations on analyzing history, including customer complaints, reviews, and feedback, which are collected after an event has occurred. The data collected gives insights on why customers churn, but the system does not get to know the problems in real time or even predict potential churn risks. This delayed strategy hampers the system's ability to take proactive measures to retain customers. Due to this, organizations end up wasting precious time in fixing dissatisfaction, thus leading to higher churn rates.

• Lack Of Predictive Accuracy

The use of simple post-event handling methods in the existing system limits the capability of the system to predict churns with precision. These simple post-event handling methods may fail to capture the complex, multi-dimensional factors that influence customer behavior in terms of patterns in their transaction history, service usage, or interaction quality. Without the integration of advanced algorithms or machine learning techniques, the predictions may miss critical indicators of churn, leading to ineffective retention strategies. This identifies the necessity for more robust analytical models.

• Unstructured And Noisy Data

In the current system, use of simple methods of post-event handling leaves with the poor ability to predict churn events. It is so since most of the times such approaches fail to capture multi-dimensional factors, which would shape customer behavior, for instance, transaction history, patterns of service usage, and quality of interaction. Such predictions without integration of complex algorithms or machine learning can miss on some important indications of churn, thereby suggesting ineffective retention strategies. This gap indicates that more robust analytical models should be used.

• Scalability Issues

With growth in business, the volume of customer interactions and feedback grows exponentially. The existing system often depends on manual or semi-automated methods, which are not efficient enough for handling large datasets. It results in slow response times and misses opportunities to address potential churn. Besides, it can also impede the system's capability to process feedback in real-time when the organization has a significant customer base or diversified channels of communication.

These disadvantages point to the necessity of evolving the current churn prediction system into a more advanced, data-driven framework that can handle complexity, scale well, and provide for proactive decision-making.

IV. PROPOSED SYSTEM

The proposed system is designed for the customer churn prediction solution and addresses the limitation of the previous one through real-time processing, advanced predictive analytics, and scalable automation. It will therefore utilize real-time customer interaction data, transaction data, and social media feedback whereas the previous one was totally dependent on retrospective data. It will analyze complex factors that influence churn by using advanced machine learning techniques like Random Forest, XGBoost, and Neural Networks, which will improve the accuracy of the prediction. The system will also deal with issues of unstructured and noisy data by using Natural Language Processing to clean and structure feedback from sources like social media and online reviews. To ensure scalability, it is cloud-based and thus equipped to process large volumes of data in real-time. Such intuitive user interfaces will provide the actionable insights to decision-makers in much faster data-driven intervention modes. Moreover, continuous feedback loops for model improvement mean that the system continuously learns and adapts to new behaviors by customers, ensuring that proactive churn management happens alongside higher prediction accuracy, efficiently scalable.

Advantages

• Proactive Churn Management

The Proposed System helps businesses to identify churn risks in real-time based on customer interactions and feedback, which happen in real-time. Proactive intervention by the businesses before a customer decides to churn will allow them to engage with at-risk customers using targeted retention strategies that ultimately reduce the overall churn rate. Since the system ensures preventing churn early, businesses are able to retain their valuable customers, improving customer satisfaction and loyalty.

• Better Data Utilization

The proposed system enhances data processing by integrating Natural Language Processing (NLP) and advanced data-cleaning techniques for unstructured data sources such as social media, online reviews, and customer feedback. Extracting meaningful insights from such diverse and noisy data sources improves the overall accuracy of churn predictions. This capability ensures that all relevant customer data is used effectively, providing businesses with a comprehensive view of churn drivers and customer sentiment.

• Scalability And Efficiency

The proposed system can easily scale with business through cloud-based infrastructure. The high capacity of processing large amounts of data in real time keeps the system efficient and effective regardless of the size of customers that may be encountered along the way. It becomes possible to automate collection, analysis, and prediction processes and subsequently ensure that the system keeps increasing its capacity to process highly complex and voluminous data volumes without performance degradation. With such scalability, long-term sustainability and adaptability of a business system are assured.

V.METHODOLOGY

A method for predicting customer churn must therefore involve a rigorous approach toward obtaining the accuracy of actionable results. Starting from gathering data from transactions, profiles, feedbacks, and social media interaction, data preprocessing and cleaning handle missing values, removal of duplicate data, treating outliers, and encoding of categorical variables are involved, while feature engineering is the derivation of meaningful insights toward predictive enhancement. The prepared data was split into training and testing sets, and machine learning models such as Logistic Regression, Random Forest, and XGBoost were trained and optimized using techniques like cross-validation and hyperparameter tuning. To address class imbalances, resampling techniques such as over-sampling, under-sampling, or SMOTE were used. Finally, all the models are tested on an unseen set of data by computing measures of performance-accuracy, precision, recall, and AUC-values, for the best one to actually make predictions to predict who will churn next. With such predictions business can prepare proactive retention tactics and so enhance customer loyalty.

Data loading

The first step in the churn prediction process is data loading, where the relevant customer data is imported into the system. This data can be sourced from multiple channels such as CSV files, databases, or cloud storage. The dataset typically includes customer demographics, transaction histories, service usage, and customer feedback, which are crucial for identifying patterns related to churn. This data loading phase ensures access by the system to raw data that will be required for further processing and analysis.

Data cleaning and transformation

Following data loading, the next important step involves data cleaning and transformation. At this stage, a number of techniques are used to make the dataset machine learning friendly. Missing values are addressed using imputation methods while duplicates are removed to eliminate redundancy. Outliers are identified and either treated or excluded to prevent their distortion in the analysis. These categorical variables are encoded in numerical format using one-hot encoding or label encoding. Feature engineering is applied to derive meaningful new features from the available data. Finally, normalization or standardization is performed on the numerical features for consistent results and better performance of the model.

Model training

The most critical process after data loading is cleaning and transformation. This involves using different techniques to transform the dataset in preparation for machine learning. Missing values are dealt with using imputation methods, duplicates removed to remove redundancy, outliers detected and either treated or excluded in order not to disturb the analysis. Categorical variables are encoded as a numerical format, usually into one-hot encoding or even label encoding. Feature engineering is also used to engineer meaningful new features from existing data. Finally, normalization or even standardization is done in numerical features to make data consistent and to improve model performances.

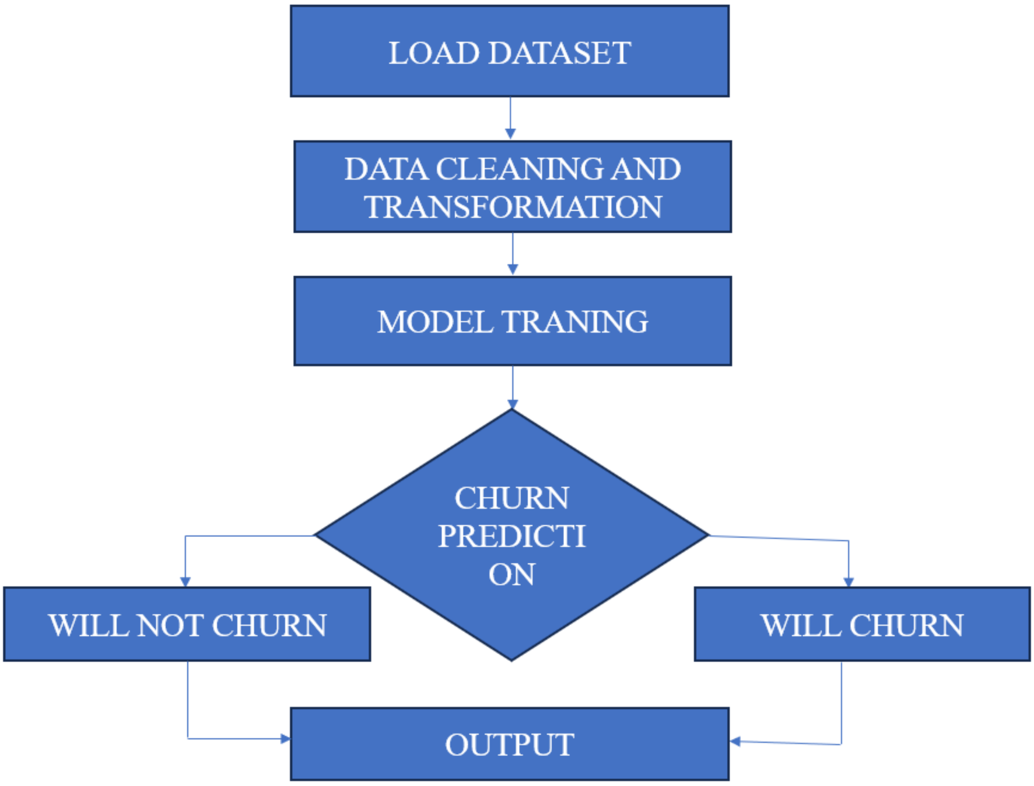
Predicting churn

The final phase involves using the trained model to predict churn in the testing dataset. From the patterns learned by the model during training, predictions are made on whether the customer will churn or be loyal. These predictions are analyzed to identify customers with high probabilities of leaving. This information can be used in proactive strategies such as individual retention offers, better support to customers, or relevant marketing campaigns to retain useful customers. This final step ensures that churn can indeed be predicted effectively, and in turn, businesses can implement timely actions to reduce the attrition of customers.

Output

The methodology of the system starts with loading the customer data from various sources like databases or files, which are key information sources, like demographics, transaction histories, and feedback. After loading, data cleaning and transformation processes follow, including handling missing values, duplicates, outliers, encoding categorical data, and scaling numerical variables. Feature engineering is done to extract extra insights and make sure that the data is ready for analysis. The cleaned data is divided into training and testing sets, where the machine learning models such as Logistic Regression, Random Forest, and XGBoost are trained with the training data. The phase also involves cross-validation techniques to optimize model performance and prevent overfitting. In the final step, the trained model is used to predict churn on the testing dataset by identifying customers at risk of leaving. These predictions allow organizations to strategically retain customers by using the system's output to enhance customer satisfaction and loyalty.

Flowchart



VI. CONCLUSION

Customer churn prediction offers a holistic framework for accurately identifying customers who are likely to leave. The system enhances the predictive power of machine learning models by addressing challenges such as data imbalances and leveraging advanced techniques like ensemble learning, sampling strategies, and stacked frameworks. The integration of multiple sampling methods ensures the generation of enriched meta-features, which in turn improves model performance with higher AUC and top-decile lift metrics. This approach provides businesses with actionable insights to help them develop targeted retention strategies and optimize efforts for customer engagement. It not only demonstrates superior accuracy over traditional frameworks but also represents a scalable and adaptive solution to changing customer behavior, thus ensuring sustained business growth and customer satisfaction.

VII. ACKNOWLEDGEMENTS

We would like to extend our gratitude to Mrs. M. Saranaya, Assistant Professor in Information Technology, for guiding us, encouraging us, and supporting this project at every stage. She has been instrumental in providing guidance over tough challenges and deepening our knowledge in this area of research. Finally, we must thank all those who have contributed directly and indirectly towards the successful completion of this project.

IX.REFERENCES

[1] Awais Manzoor, M. Atif Qureshi, Etain Kidney2, And Luca Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners" on IEEE, VOLUME 12, 2024.

[2] David Freire, David Santos Mauricio Sanchez, José Luis Castillo Sequera And Daniel Fiallo Moncayo, " Predictability, and Explainability of Mobile Telephony Customer Departure in Telecommunications Companies: A Systematic Review of the Literature on IEEE, "VOLUME 12, 2024.

[3] J. Kozak, K. Kania, P. Juszczuk, and M. Mitre¸ga, "Swarm intelligence goal-oriented approach to data-driven innovation in customer churn management," Int. J. Inf. Manage., vol. 60, Oct. 2021, Art. no. 102357.

[4] M. Sadeghi, M. N. Dehkordi, B. Barekatain, and N. Khani, "Improve customer churn prediction through the proposed PCA-PSO-K means algorithm in the communication industry," J. Supercomput., vol. 79, no. 6, pp. 6871–6888, Apr. 2023.

[5] Nagaraju Jajam, Nagendra Panini Challa 1, (Member, Ieee), Kamepalli S. L. Prasanna1,And C H Venkata Sasi Deepthi " Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction" on IEEE, VOLUME 11, 2023.

[6] P. P. simović, C. Y. T. Chen, and E. W. Sun, "Classifying the variety of customers' online engagement for churn prediction with a mixed penalty logistic regression," Comput. Econ., vol. 61, no. 1, pp. 451–485, Jan. 2023.

[7] Somak Saha, Chamak Saha, Md. Mahidul Haque, Md. Golam Rabiul Alam "Churn Net: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry" on IEEE, VOLUME 12, 2024.

[8] Seema and G. Gupta, "Development of fading channel patch based convolutional neural network models for customer churn prediction," Int. J. Syst. Assurance Eng. Manage., vol. 15, no. 1, pp. 391–411, Jan. 2024.

[9] S. Saha, C. Saha, M. Mahidul Haque, M. Golam Rabiul Alam, and A. Talukder, "ChurnNet: Deep learning enhanced customer churn prediction in telecommunication industry," IEEE Access, vol. 12, pp. 4471–4484, 2024.

[10] Y. Beeharry and R. Tsokizep Fokone, "Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry," Concurrency Computation, Pract. Exper., vol. 34, no. 4, Feb. 2022.