Building a Predictive Churn model By Applying Machine Learning Techniques To Customers Behaviour Data

[1]Mrs. A. Beneta Mary. M.E, [2]E. Aishwarya, [3]P. Swarna Rubi, [4]P. Tamilazhagi

[1]Assistant Professor, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – beneta1610@gmail.com

[2]Student, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – aiswariyasuji@gamil.com

[3]Student, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – swranarubi777@gmail.com

[4]Student, Computer Science and Engineering, St. Joseph College of Engineering, Chennai-602117, Tamil Nadu,

Email Id – ptamilazhagi07@gmail.com

Abstract – This study focus on customer churn prediction is a data driven approach used to categorize customers into distinct groups based on shared attributes such as purchasing behavior, spending patterns, demographics, and preferences. This segmentation enables businesses to tailor their marketing strategies, enhance customer engagement, and optimize resource allocation. However, existing systems often rely on static segmentation models, which fail to account for dynamic customer behaviors or provide insights into potential customer churn. These limitations result in generalized marketing efforts, reduced customer retention, and an inability to predict future customer behavior effectively. To address these drawbacks, the proposed project integrates dynamic customer segmentation and churn prediction into a unified framework. The system begins by segmenting customers into three categories—Gold, Silver, and Normal—using advanced clustering techniques based on Regency, Frequency, and Monetary (RFM) metrics. This ensures precise categorization of high-value, moderately active, and low-engagement customers. Building upon this segmentation, the project employs Logistic Regression, a robust supervised learning algorithm, to predict customer churn. Key features such as RFM scores, behavioral patterns, and demographic data are utilized to identify customers at risk of leaving.

I.INTRODUCTION

The customer's concentration on the providers has prompted many new telecom associations to emerge. These new firms usually specialize in providing a specific service or product that the customer cannot find from the incumbent providers. These new firms can provide this service or product at a lower price than the incumbent providers, allowing them to capture a larger market share. The incumbent providers, however, can retain most of the market by pricing their products higher than the new firms.

This competition between the incumbent providers and the new firms has caused the rates that the associations charge to change. The associations' rates are often determined by the amount of competition in the market. The more competition, the higher the rates the associations can charge. In markets with a low level of competition, the associations can charge rates lower than in markets with a high level of competition. The rates that the associations' charge is also affected by the number of services that the associations can offer.

The more services the associations can offer, the higher the rates that the associations can charge. Churning, in marketing terms, refers to the number of customers who stopped using a particular product. Always the churn rate must be low. Customer churning is common with any product when there are multiple options for a single problem. Usually, customers will churn when they face any difficulties or disappointments in the services rendered by the product. The churn rate is usually measured for a specific time.

Any organization's primary motive should be satisfying customers and retaining existing customers. Retaining existing customers is equally important as gathering new customers. Customer churn prediction is the most important issue in adopting an industry's product. Managing customer churn is one major challenge companies face, especially those offering subscription-based services. Customer churn also called customer attrition is the loss of customers, and it is caused by a change in taste, lack of proper customer relationship strategy, change of residence, and several other reasons.

If businesses can effectively predict customer attrition, they can segment those customers that are highly likely to churn and provide better services to them. Hence, a churn prediction model is a mandate needed in today's digitized economy. An organization can achieve a high customer retention rate and maximize its revenue.

II. BACKGROUND AND MOTVATION

Overview

Customer churn, or customer attrition, occurs when customers stop using a product or service. It is a major challenge for businesses, especially those operating in subscription-based industries like telecom, banking, e-commerce, and streaming services. High churn rates lead to loss of revenue, increased customer acquisition costs, and reduced market share. Identifying customers likely to churn and implementing strategies to retain them is crucial for business sustainability.

Machine learning (ML) techniques provide an advanced and data-driven approach to predicting churn by analyzing customer behaviour patterns, ML models help businesses anticipate which customers are likely to leave, enabling them to take proactive measures to retain high-value customers. Traditionally, churn prediction relied on statistical methods and rule-based systems, which were limited in their ability to detect complex patterns in customer behaviour . With the increasing availability of customer data such as transaction history, engagement metrics, complaints, and service usage machine learning offers a more effective approach.

Motivation of this research

The primary motivation for building a predictive churn model is to help businesses improve customer retention and profitability. Studies show that acquiring a new customer costs five times more than retaining an existing one. By predicting churn, businesses can take timely actions to reduce customer loss and maintain a strong customer base. Churn prediction also enables companies to personalize their retention strategies. Some commonly used ML techniques for churn prediction include.

Data Collection is gathered from various sources, including customer interactions, transaction logs, support tickets, and demographic information.

Data preprocessing involves cleaning, handling missing values, and normalizing variables to ensure consistency.

Feature Engineering Identifying key features that influence churn, such as, Customer demographics: Age, location, income, Engagement metrics: Website visits, time spent on platform, login frequency, Financial transactions.

Logistic Regression: A simple and interpretable model that estimates the probability of churn.

Decision Trees: Identify key factors influencing churn through hierarchical decision-making.

Random Forest: An ensemble model that improves accuracy by combining multiple decision trees.

Neural Networks: Deep learning models that capture complex relationships in large datasets.

The dataset is split into training and testing sets to ensure model generalization. Performance is evaluated using metrics such as Accuracy, Precision & Recall. balanced metric considering both. Measures the model’s ability to distinguish between churners and non-churners.

III. ROLE AND POTENTIAL FOR CUSTOMER CHURN PREDICTION MODEL

Role:

The primary role of a customer churn prediction model is to identify customers who are likely to

stop using a company's products or services in the near future. This is a crucial function for businesses because Retention is cheaper than acquisition Acquiring new customers is often significantly more expensive than retaining existing ones Lost revenue Churn directly impacts a company's revenue stream. Predicting and preventing churn helps maintain and grow revenue.
Customer lifetime value (CLTV) Retaining customers for longer increases their overall value to the business. Informed decision-making: The insights from a churn prediction model can inform various business strategies related to customer engagement, marketing, product development, and customer service.
Proactively identify at-risk customers by analyzing historical data and current behavior, the model flags customers with a high probability of churning. Understand the reasons for churn by examining the features that contribute to the prediction, businesses can gain insights into why customers are leaving. Targeted interventions the model enables businesses to focus their retention efforts on the customers who are most likely to churn, optimizing resource allocation.
Personalized retention strategies Understanding the factors driving churn for specific customer segments allows for the development of tailored retention offers and engagement strategies. Measure the impact of retention efforts by tracking the churn rate of customers who received retention interventions compared to those who did not, businesses can assess the effectiveness of their strategies.

Potential:

A. Reduced Churn Rate :The most direct potential is a significant reduction in the number of customers who leave. by proactively addressing the needs and concerns of at-risk customers, businesses can improve retention rates.

B. Increased Revenue Lower: Churn directly translates to higher revenue retention and potentially increased revenue through extended customer lifespans. Improved Customer Loyalty: Proactive engagement and personalized retention efforts can foster stronger relationships with customers, leading to increased loyalty and advocacy.

C. Optimized Marketing Spend: Instead of broad marketing campaigns, resources can be focused on retaining valuable customers identified by the model, leading to a higher return on investment (ROI) for marketing efforts.

D. Improved Forecasting: By understanding churn patterns, businesses can improve their revenue and customer base forecasts, leading to better strategic planning.

IV. FUTURE RESEARCH DIRECTIONS FOR CUSTOMER CHURN PREDICTION MODEL

Future research directions for customer churn prediction models are abundant and span several key areas, aiming to create more accurate, interpretable, and actionable models. Here are some potential directions:

Advanced Machine Learning and Deep Learning Techniques:

Hybrid Models: Exploring combinations of different machine learning algorithms (e.g., ensemble methods with deep learning) to leverage the strengths of each and improve prediction accuracy and robustness. For instance, combining the interpretability of tree-based models with the feature learning capabilities of neural networks.

Temporal and Sequence Modeling: Utilizing Recurrent Neural Networks (RNNs), LSTMs, and Transformers to better capture the sequential nature of customer interactions and predict churn based on evolving behaviour over time

Graph Neural Networks (GNNs): Applying GNNs to model customer relationships and network effects on churn. Customers within a social network or those interacting with each other might exhibit correlated churn behaviour.

Automated Machine Learning (AutoML): Further research into AutoML techniques to automate feature engineering, model selection, hyperparameter tuning, and deployment for churn prediction, making it more accessible and efficient.

Enhanced Feature Engineering and Data Integration:

Incorporating Diverse Data Sources: Integrating data from various touchpoints beyond traditional CRM data, such as social media activity, IoT device data, website clickstreams, and customer service interactions (using text and sentiment analysis).

Behavioural Feature Engineering: Developing more sophisticated features that capture the nuances of customer behaviour, such as the sequence of feature usage, the time elapsed between actions, and the context of interactions.

Feature Importance: Researching methods to understand how the importance of different features changes over time and across different customer segments. Privacy-Preserving Feature Engineering: Exploring techniques to extract meaningful features for sensitive data while adhering to privacy regulations (e.g., federated learning, differential privacy).

Improved Model Interpretability and Explainability:

Explainable AI (XAI) for Churn: Focusing on making churn prediction models more transparent and understandable to business users. This includes using techniques like SHAP values, LIME, and attention mechanisms in deep learning to identify the key drivers of churn for individual customers and segments.

Causal Inference: Moving beyond correlation to understand the causal factors that lead to churn. This can help businesses implement more effective intervention strategies by addressing the root causes.

Interactive Visualization Tools: Developing user-friendly tools that allow business users to explore churn predictions, understand the contributing factors, and simulate the impact of different retention strategies.

VI. CONCLUSION

Building a customer churn prediction model using machine learning on customer behaviour data holds immense potential for businesses to proactively identify at-risk customers and implement targeted retention strategies. However, the journey is fraught with significant challenges spanning data quality and management, the complexities of machine learning modeling, and the practicalities of implementation and integration.

Success hinges on addressing these hurdles through meticulous data preparation, thoughtful feature engineering, careful algorithm selection and tuning, robust deployment strategies, and seamless integration with existing business processes. Ultimately, a well-designed and maintained churn prediction model, coupled with effective actioning of its insights, can translate into significant improvements in customer retention, reduced costs, and enhanced business value. Continuous monitoring and adaptation of the model are crucial for sustained effectiveness in the face of evolving customer behaviour.

A well-built customer churn prediction model, leveraging machine learning on customer behaviour data, empowers businesses to proactively identify and retain at-risk customers. While its development presents challenges across data, Modelling, and implementation, successfully overcoming these hurdles can lead to significant improvements in customer retention, cost reduction, and overall business value. Continuous monitoring and adaptation of the model are crucial for sustained effectiveness in the face of evolving customer behaviour.

VI. REFERENCE

[1] Omar Adwan, Hossam Faris, Khalid Jaradat, Osama Harfoushi, Nazeeh Ghatasheh, Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis, Life Sci. J. 11 (3) (2014) 75–81.

[2] Mohammad Ridwan Ismail, Mohd Khalid Awang, M. Nordin A. Rahman, Mokhairi Makhtar, A multi-layer perceptron approach for customer churn prediction, International Journal of Multimedia and Ubiquitous Engineering 10(7) (2015) 213–222.

[3] Farhad Shaikh, Brinda Pardeshi, Ajay Jachak, Akash Bendale, Nandakishor Sonune, Mangal Katkar, Customer churn prediction using nlp and machine learning: an overview, International Journal Of Advance Scientific Research And Engineering Trends 6 (2) (2021) 40–45.

[4] Anuj Sharma, DrPrabin Kumar Panigrahi, A neural network based approach for predicting

customer churn in cellular network services, Int. J. Comput. Appl. 27 (11) (2011) 26–31.

[5] Ammara Ahmed, D. Maheswari Linen, A review and analysis of churn prediction methods for customer retention in telecom industries, in: 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2017, pp. 1–7.

[6] S. Babu, N.R. Ananthanarayanan, V. Ramesh, A study on efficiency of decision tree and multi layer perceptron to predict the customer churn in telecommunication using WEKA, Int. J. Comput. Appl. 140 (4) (2016) 26–30.

[7] Fatih Kayaalp, Review of customer churn analysis studies in telecommunications industry, Karaelmas Science & Engineering Journal 7 (2) (2017).

[8] Davoud Gholamiangonabadi, Jamal Shahrabi, Seyed Mohamad Hosseinioun, Sanaz Nakhodchi, Soma Gholamveisy, Customer churn prediction using a new criterion and data mining; A case study of Iranian banking industry, in: Proceedings of the International Conference on Industrial Engineering and Operations Management, 2019, pp. 5–7.

[9] Kamorudeen A. Amuda, Adesesan B. Adeyemo, Customers Churn Prediction in Financial Institution Using Artificial Neural Network, 2019, 11346 arXiv preprint arXiv:1912.

[10] Anam Bansal, Churn prediction techniques in telecom industry for customer retention: a survey, J. Eng. Sci. 11 (4) (2020) 871–881

[11] Tianyuan Zhang, S´ergio Moro, Ricardo F. Ramos, A data-driven approach to improve customer churn prediction based on telecom customer segmentation, Future Internet 14 (3) (2022) 94.