International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Survey on Statistical Models for Estimating Churn Rate

Kapil[1] Prof. Pankaj Raghuwanshi[2]

ABSTRACT: Customer churn, the phenomenon of customers discontinuing their relationship with a business, poses a significant threat to the sustainability and profitability of companies across industries. As markets become increasingly competitive, retaining existing customers has proven to be more cost-effective than acquiring new ones. In this context, machine learning (ML) has emerged as a powerful tool for analyzing customer behavior and predicting churn with high accuracy. By leveraging vast datasets and sophisticated algorithms, businesses can proactively identify at-risk customers and take targeted actions to retain them. The success of churn prediction largely depends on the quality and relevance of input features. Important features include customer demographics, transaction frequency, service usage patterns, complaint records, and engagement metrics. Feature engineering, which involves creating new features or transforming existing ones, is a critical step in improving model performance This paper presents a comprehensive survey of statistical models for forecasting churn rates along with associated challenges that the sector faces.

Keywords: Churn Rate, Statistical Modelling, Machine Learning, Deep Learning, Regression Analysis.

I. Introduction

Customer churn can be voluntary, where customers choose to leave, or involuntary, due to factors like payment failures or service disruptions [1]. Understanding the reasons behind churn is crucial for developing effective mitigation strategies. Traditionally, companies relied on statistical analysis and manual insights to gauge churn, but these methods often lacked precision and scalability. With the advent of machine learning, it is now possible to analyze complex patterns in customer data that indicate a likelihood of churn [2]. There are several factors of customer churn which are:

Fig.1 Leading Causes of Churn

(Source: https://keywordseverywhere.com/blog/customer-retention-stats/)

Figure 1 depicts the top causes of customer churn. It is important to compute the customer churn rate (CRR). The Customer Retention Rate (CRR) is defined as [3]:

$$CRR = \frac{E-N}{S} * 100\% \quad (1)$$

Here,
S is number of customers at the beginning of a period.
E is the number of customers at the end of that period.
N is number of new customers you gained during that time.

Machine learning models can process and learn from large volumes of structured and unstructured data, including purchase history, customer support interactions, social media activity, and more. Algorithms like logistic regression, decision trees, support vector machines (SVM), and neural networks are commonly used to build churn prediction models. These models classify customers based on their likelihood of churning and generate actionable insights that guide retention strategies [4]. The success of churn prediction largely depends on the quality and relevance of input features. Important features include customer demographics, transaction frequency, service usage patterns, complaint records, and engagement metrics. Feature engineering, which involves creating new features or transforming existing ones, is a critical step in improving model performance. Data preprocessing techniques such as handling missing values, normalizing data, and dealing with class imbalance are also essential for building robust ML models [5].

II. Existing Statistical Models

The statistical machine learning models used for forecasting are presented in brevity in this section [6]:

Support Vector Machine (SVM):

Before the advent of deep learning, traditional machine learning models such as Support Vector Machines (SVM), Decision Trees, Random Forests, and K-Nearest Neighbors (KNN) were widely used for satellite object detection. These models typically relied on handcrafted features, such as texture, edges, and spectral indices, to distinguish between different objects [7].

The SVM classifies based on the hyperplane.
The selection of the hyperplane H is done on the basis of the maximum value or separation in the Euclidean distance d given by [8]:

$$d = \sqrt{x_1^2 + \cdots \ldots \ldots . x_n^2} \quad \textbf{(2)}$$

Here,
x represents the separation of a sample space variables or features of the data vector,
n is the total number of such variables
d is the Euclidean distance

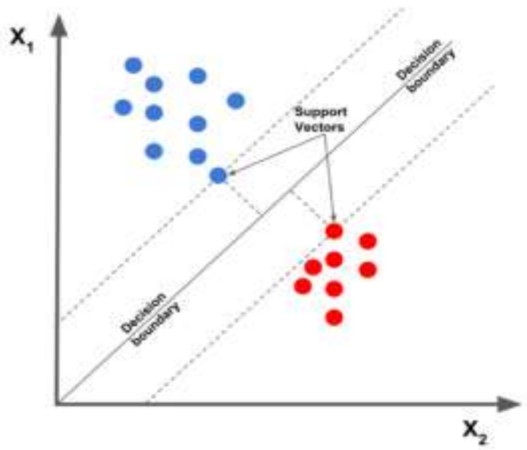

Fig.3 The SVM Model

Figure above depicts the SVM Model.

The (n-1) dimensional hyperplane classifies the data into categories based on the maximum separation. For a classification into one of **'m'** categories, the hyperplane lies at the maximum separation of the data vector **'X'.** The categorization of a new sample **'z'** is done based on the inequality [10]:

$$d_x^z = Min(d_{C1}^z, d_{C2}^z \ldots d_{C2=m}^z) \quad \textbf{(3)}$$

Here,
d_x^z is the minimum separation of a new data sample from **'m'** separate categories
$d_{C1}^z, d_{C2}^z \ldots d_{C2=m}^z$ are the Euclidean distances of the new data sample **'z'** from m separate data categories [11].

For instance, SVMs are effective for binary classification tasks, such as distinguishing between urban and rural areas, while Random Forests are used for multi-class classification problems, such as land cover mapping. However, these models struggle with complex patterns in high-resolution imagery and require extensive feature engineering, which limits their scalability and accuracy

ARIMA:
In an autoregressive integrated moving average model commonly known as the ARIMA model assumes that the future value of a variable can be linearly modelled as a function previous samples of the variables and errors of prediction [12].

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varphi_p y_{t-p} + \cdots \ldots \ldots \ldots . \theta_q \varepsilon_{t-q} \quad \textbf{(4)}$$

Here,
y_t is the value of the output variable at time 't'
ε is the prediction error
θ *and* φ are called the model parameters
p *and* q are called the orders of the model

One of ARIMA's key strengths lies in its ability to handle both stationary and non-stationary data. While the ARIMA model assumes the input time series is stationary (i.e., its statistical properties like mean and variance remain constant over time), it incorporates differencing techniques to convert non-stationary data into a stationary format. This makes it highly adaptable for real-world datasets that often exhibit trends or seasonality [13].

Neural Networks:
Owing to the need of non-linearity in the separation of data classes, one of the most powerful classifiers which have become popular is the artificial neural network (ANN). The neural networks are capable to implement non-linear classification along with steep learning rates. The neural network tries to emulate the human brain's functioning based on the fact that it can process parallel data streams and can learn and adapt as the data changes. This is done through the updates in the weights and activation functions [14].

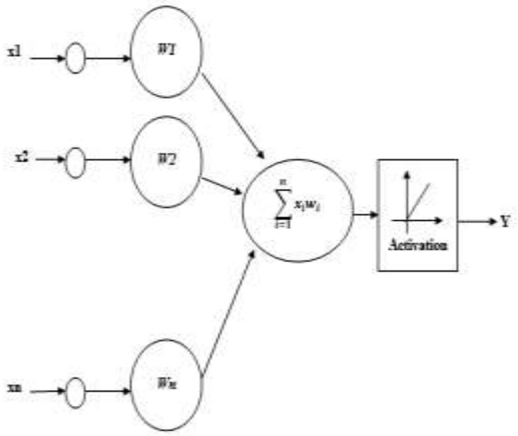

Fig.3 The ANN Model

Figure above depicts the ANN model.
The input-output relation of a CNN is given by [15]:

$$y = f(\sum_{i=1}^{n} x_i w_i + b) \quad \textbf{(5)}$$

Here,
x denote the parallel inputs
y represents the output
w represents the bias
f represents the activation function

The neural network is a connection of such artificial neurons which are connected or stacked with each other as layers. The neural networks can be used for both regression and classification problems based on the type of data that is fed to them [16]. Typically the neural networks have 3 major conceptual layers which are the input layer, hidden layer and output layer. The parallel inputs are fed to the input layer whose output is fed to the hidden layer. The hidden layer is responsible for analysing the data, and the output of the hidden layer goes to the output layer. The number of hidden layers depends on the nature of the dataset and problem under consideration. If the neural network has multiple hidden layers, then such a neural network is termed as a deep neural network. The training algorithm for such a deep neural network is often termed as deep learning which is a subset of machine learning. Typically, the multiple hidden layers are responsible for computation of different levels of features of the data [17].

Long Short Term Memory (LSTM):

The LSTM networks are a specialized type of recurrent neural network (RNN) designed to process and predict data sequences by learning long-term dependencies. Unlike traditional RNNs, which suffer from vanishing or exploding gradient problems during training, LSTMs incorporate a unique architecture with gates and memory cells that help retain important information over long periods [18].

The LSTM primarily has 3 gates:
1) Input gate: This gate collects the presents inputs and also considers the past outputs as the inputs.
2) Output gate: This gate combines all cell states and produces the output.
3) Forget gate: This is an extremely important feature of the LSTM which received a cell state value governing the amount of data to be remembered and forgotten.

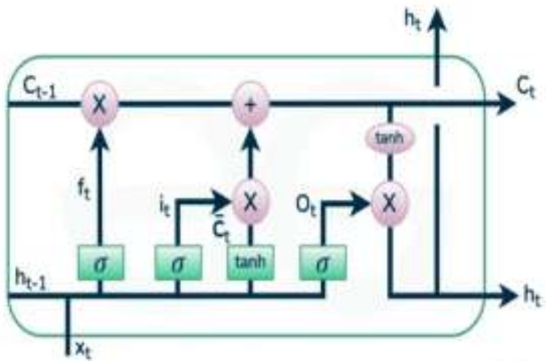

Fig.4 The LSTM Model

Figure above depicts the LSTM model.

The relation to forget by the forget gate is given by:

$$f = \sigma(W_f[h_{t-1}, x_t] + b_i) \quad \textbf{(6)}$$

Here,
f denotes forget gate activation
w_f are forget gate weights.
h_{t-1} Denotes Hidden state from the previous time step
x_t is present input.
b_i is the bias

The advantages of LSM are:
Capturing Long-Term Dependencies: LSTMs maintain long-term memory using the cell state, unlike traditional RNNs.
Mitigating Vanishing/Exploding Gradients: Gates help regulate gradient flow, enabling stable training over long sequences.
Versatility: Useful for several time series prediction problems.

However, the major challenge happens to be the problem of overfitting.

Convolutional Neural Networks (CNNs): The family of CNNs are the backbone of modern satellite object detection. CNNs automatically learn hierarchical features from raw images, eliminating the need for manual feature extraction. The Convolutional Neural Networks (CNNs) can automatically extract hierarchical characteristics from images, they have become the mainstay for image classification applications. These neural networks perform exceptionally well in applications like picture identification because they are specifically made for processing organised grid data [19].

Convolutional, pooling, and fully linked layers are among the layers that make up a CNN's architecture. Convolutional layers identify patterns in the input image by applying filters, hence identifying local features. By reducing spatial dimensions, pooling layers preserve significant information. High-level features are integrated for categorization in fully connected layers.

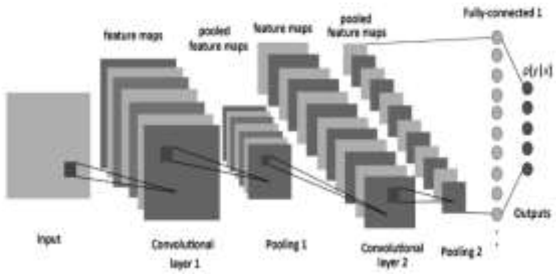

Fig.5 The CNN Model

Figure above depicts the CNN model.

The convolution operation is given by:

$$x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \qquad (7)$$

Here,
x(t) is the input
h(t) is the system under consideration.
y is the output
*is the convolution operation in continuous domain
For a discrete or digital counterpart of the data sequence, the convolution is computed using:

$$y(n) = \sum_{-\infty}^{\infty} x(k)h(n-k) \qquad (8)$$

Here
x(n) is the input
h(n) is the system under consideration.
y is the output
*is the convolution operation in discrete domain

In this approach, the back propagation based neural network model has been used. A backpropagation neural network for traffic speed forecasting typically consists of an input layer, one or more hidden layers, and an output layer. The number of nodes in the input layer corresponds to the features used for prediction, The hidden layers contain nodes that learn and capture the intricate patterns within the data, while the output layer provides the predicted traffic speed. The training of a backpropagation neural network involves the iterative application of the backpropagation algorithm. During the training process, historical data is used to feed the network, and the algorithm calculates the error between the predicted and actual energy demands. This error is then propagated backward through the network, adjusting the weights and biases of the connections to minimize the prediction error. This iterative process continues until the network converges to a state where the error is minimized. Successful backpropagation neural network models for traffic speed forecasting can be integrated into energy management systems [20].

III. Previous Work

A summary of noteworthy contribution in the domain is presented here:

Poudel et al. [21] proposed a gradient boosting machine learning (BGM) model to explain both local and global explanations of churn predictions. Various classification models, including the standout Gradient Boosting Machine (GBM), were used alongside visualization techniques like Shapley Additive Explanations plots and scatter plots for enhanced interpretability. The GBM model demonstrated superior performance with an 81% accuracy rate. A Wilcoxon signed rank test confirmed GBM's effectiveness over other models, with the value indicating significant performance differences. The study concludes that GBM is notably better for churn prediction, and the employed visualization techniques effectively elucidate key churn factors in the telecommunications sector.

Rao et al. [22] proposed churn prediction is crucial in a number of service- based businesses, including the telecom sector, life insurance, hospitality, banking, and gaming. A Prediction of consumer churn in the enterprises is crucial because it is one of the major risks to revenue loss and one of the top concerns for big enterprises. As per studies it is found that the expense of acquiring a new consumer is 10 times greater than the expense of an existing consumer. This issue has an impact on both the company's revenue and its ability to grow. One of the most important tasks for the business is to keep its current customers. As technology advances quickly, Deep Learning and Machine Learning techniques are evolved that businesses can employ to track customer churn behaviour. In this review we will cover, overview on Customer churn, Machine learning Classification and boosting Models as naive bayes, support vector machines, k-nearest neighbor, logistic regression, decision trees, Random Forest, Adaboost, XGBoost and artificial neural networks algorithms, as well as deep learning techniques.

Shobhana et al. [23] proposed that machine learning and data mining may be aided by examining this enormous quantity of data, analysing customer behaviour, and seeing potential attrition opportunities. The support vector machine is a popular supervised learning method in machine learning applications. Predictive analysis uses the hybrid classification approach to address the regression and classification issues. The process for forecasting E-Commerce customer attrition based on support vector machines is presented in this paper, along with a hybrid recommendation strategy for targeted retention initiatives. You may prevent future customer churn by suggesting reasonable offers or services. The empirical findings demonstrate a considerable increase in the coverage ratio, hit ratio, lift degree, precision rate, and other metrics using the integrated forecasting model. To effectively identify separate groups of lost customers and create a customer churn retention strategy, categorize the various lost customer types using the RFM principle.

Sobreiro et al. [24] proposed that customer churn rate prediction is a problem that is being addressed with machine learning algorithms; thus, appropriate approaches to address the dropout rate are needed. The selection of an algorithm to predict the dropout rate is only one problem to be addressed. Other aspects should also be considered, such as which features should be selected and how to measure accuracy while considering whether the features are appropriate according to the business context in which they are employed. To solve these questions, the goal of this paper is to develop a systematic literature review to evaluate the development of existing studies and to predict the dropout rate in contractual settings using machine learning to identify current trends and research opportunities. The results of this study identify trends in the use of machine learning algorithms in different business areas and in the adoption of machine learning algorithms, including which metrics are being adopted and what features are being applied. Finally, some research opportunities and gaps that could be explored in future research are presented.

Agarwal et al. [25] explained how to use machine learning algorithms to identify banking customers who may be considering switching financial institutions. This article demonstrates how machine learning models such as Logistic Regression (LR) and Naive Bayes' (NB) can effectively forecast which customers are most likely to leave the bank in the future by using data such as age, location, gender, credit card information, balance, etc. The article also uses data such as age, location, gender, credit card information, balance, etc. In addition, this article demonstrates the probabilistic predictions that may be generated using machine learning models such as Logistic Regression (LR) and Naive Bayes (NB). The findings of this research ultimately point to the conclusion that NB is superior to LR.

Typically, training algorithms try to attain low error rate metrics, which are defined next [16]:

The mean square error or mse given by:

$$mse = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad \textbf{(9)}$$

The final computation of the performance metric is the mean absolute percentage error given by:

$$MAPE = \frac{100}{M} \sum_{i=1}^{N} \frac{E - E_i|}{i} \quad \textbf{(10)}$$

The accuracy of prediction is computed as:

$$Ac = 100 - \frac{100}{M} \sum_{i=1}^{N} \frac{E - E_i|}{i} \% \quad \textbf{(11)}$$

Here,
n is the number of errors
i is the iteration number
E is the actual value
E_i is the predicted value

V. CONCLUSION

It can be concluded that customer retention is a critical component of sustainable growth. One of the most pressing challenges for companies, especially in

competitive markets, is customer churn—the rate at which customers discontinue their association with a business. Accurately forecasting churn is essential to proactively manage customer relationships and optimize revenue. In recent years, machine learning (ML) and deep learning (DL) have emerged as transformative technologies that significantly enhance the ability to predict churn with accuracy and scale. This paper presents the necessity to forecast customer churn rate along with existing baseline models which would allow future research in developing accurate forecasting models for the domain.

References:

[1] A. Sharma, D. Gupta, N. Nayak, D. Singh and A. Verma, "Prediction of Customer Retention Rate Employing Machine Learning Techniques," 2022 1st International Conference on Informatics (ICI), Noida, India, 2022, pp. 103-107.

[2] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," Algorithms, vol. 17, no. 6, p. 231, 2024.

[3] S. Wu, "Customer Churn Prediction in Telecom Based on Machine Learning," in Proc. 2nd Int. Conf. on Computer, Machine Learning and Artificial Intelligence (CMLAI 2024), Highlights in Science, Engineering and Technology, vol. 94, pp. 113–118, 2024

[4] R. Tyagi and K. Sindhu, "Customer Churn Analysis Using Machine Learning," in Proc. Int. Joint Conf. on Advances in Computational Intelligence, Algorithms for Intelligent Systems, Springer, Singapore, pp. 423–432, 2022.

[5] Y. Zhang, S. He, S. Li, and J. Chen, "Intra-Operator Customer Churn in Telecommunications: A Systematic Perspective," IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 948–957, Jan. 2020

[6] S. Wu, W.-C. Yau, T.-S. Ong, and S.-C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," IEEE Access, vol. 9, pp. 62118–62136, 2021

[7] P Sharmila, J Baskaran, C Nayanatara, "A hybrid technique of machine learning and data analytics for optimized distribution of renewable energy resources targeting smart energy management", Procedia in Computer Science, Elsevier 2019, vol.165, pp. 278-284.

[8] Y. Deng, D. Li, L. Yang, J. Tang, and J. Zhao, "Analysis and Prediction of Bank User Churn Based on Ensemble Learning Algorithm," in Proc. IEEE Int. Conf. on Power Electronics, Computer Applications (ICPECA), 2021, pp. 288–291.

[9] R. K. V. L. N. and P. Deeplakshmi, "Dynamic Churn Prediction Using Machine Learning Algorithms—Predict Your Customer Through Customer Behaviour," in Proc. IEEE Int. Conf. on Computer Communication and Informatics (ICCCI), 2021, pp. 1–6.

[10] R. Yahaya, O. A. Abisoye, and S. A. Bashir, "An Enhanced Bank Customers Churn Prediction Model Using a Hybrid Genetic Algorithm and K-Means Filter and Artificial Neural Network," in Proc. IEEE 2nd Int. Conf. on Cyberspace (CYBER NIGERIA), 2021, pp. 52–58.

[11] Y. E. Özköse, A. Haznedaroğlu, and L. M. Arslan, "Customer Churn Analysis with Deep Learning Methods on Unstructured Data," in Proc. Innovations in Intelligent Systems and Applications Conf. (ASYU), 2021, pp. 1–5.

[12] S. De, P. P., and J. Paulose, "Effective ML Techniques to Predict Customer Churn," in Proc. 3rd Int. Conf. on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 895–902.

[13] V. Geetha, A. Punitha, A. Nandhini, T. Nandhini, S. Shakila, and R. Sushmitha, "Customer Churn Prediction in Telecommunication Industry Using Random Forest Classifier," in Proc. Int. Conf. on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1–5.

[14] P. Bhuse, A. Gandhi, P. Meswani, R. Muni, and N. Katre, "Machine Learning Based Telecom-Customer Churn Prediction," in Proc. 3rd Int. Conf. on Intelligent Sustainable Systems (ICISS), 2020, pp. 1297–1301.

[15] I. Jahan and T. Farah Sanam, "An Improved Machine Learning Based Customer Churn Prediction for Insight and Recommendation in E-commerce," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 1-6.

[16] H. Zhao, X. Zuo and Y. Xie, "Customer Churn Prediction by Classification Models in Machine Learning," 2022 9th International Conference on Electrical and Electronics Engineering (ICEEE), Alanya, Turkey, 2022, pp. 399-407.

[17] R. K. Peddarapu, S. Ameena, S. Yashaswini, N. Shreshta and M. PurnaSahithi, "Customer Churn Prediction using Machine Learning," 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2022, pp. 1035-1040.

[18] S. R. Siva, S. N. Bushra, B. Maheswari and M. K. Prabukumar, "Performance Analysis of Different Machine Learning in Customer Prediction," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2022, pp. 1425-1430.

[19] F. Alhaqui, M. Elkhechafi and A. Elkhadimi, "Machine learning for telecoms: From churn prediction to customer relationship management," 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador, 2022, pp. 1-5.

[20] V Morozov, O Mezentseva, A Kolomiiets, “Predicting customer churn using machine learning in IT startups”, Lecture Notes in Computational Intelligence and Decision Making, ISDMCI, Springer 2021, pp. 645–664.

[21] SS Poudel, S Pokharel, M Timilsina, “Explaining customer churn prediction in telecom industry using tabular machine learning models”, Machine Learning with Applications, Elsevier 2024, vol.17, 100567.

[22] A. Raj and D. Vetrithangam, "Machine Learning and Deep Learning technique used in Customer Churn Prediction: - A Review," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 139-144.

[23] J Shobana, C Gangadhar, RK Arora, PN Renjith, “E-commerce customer churn prevention using machine learning-based business intelligence strategy”, Measurement: Sensors, Elsevier 2023, vol.27, 100728.

[24] P. Sobreiro, D. D. S. Martinho, J. G. Alonso and J. Berrocal, "A SLR on Customer Dropout Prediction," in IEEE Access, vol. 10, pp. 14529-14547, 2022.

[25] V. Agarwal, S. Taware, S. A. Yadav, D. Gangodkar, A. Rao and V. K. Srivastav, "Customer - Churn Prediction Using Machine Learning," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 893-899.

[26] N Alboukaey, A Joukhadar, N Ghneim,” Dynamic behavior based churn prediction in mobile telecom”, Expert Systems with Applications, Elsevier, 2020, vol.162, 113779.