STM JOURNALS

Journal of
Computer Technology & Applications

ISSN: 2229-6964 (Online)
ISSN: 2347-7229 (Print)
Volume 15, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoCTA

https://journals.stmjournals.com/jocta

Research JoCTA

Customer Churn Prediction Using ML Algorithms

Prajwal Waghole[1,*], Kalpna Saharan[2], Varad Wanwase[1], Shantanu Wasnik[1], Rajratan Gokhale[1]

Abstract

Comprehending customer churn is essential for businesses aiming to enhance and sustain customer relationships. This study introduces a machine learning approach aimed at forecasting customer churn by leveraging demographic and behavioral data. Our research involved developing predictive models using support vector machines (SVM), random forests, and decision trees, evaluating their efficacy using real-world data from the telecom industry. Our findings underscore that random forests consistently outperform SVM and decision trees in terms of accuracy, precision, recall, and F1-scores. This superiority indicates the robustness and adaptability of random forests across various data sets and conditions. The study emphasizes the potential of machine learning techniques in anticipating consumer behavior, offering valuable insights into customer churn dynamics. Furthermore, our model suggests several promising avenues for future research, such as exploring additional features or employing ensemble methods to further enhance predictive accuracy. By identifying at-risk customers early, businesses can proactively implement targeted retention strategies, thereby reducing churn rates and fostering long-term customer loyalty. In conclusion, our study provides actionable insights for businesses seeking to mitigate risks associated with customer churn. By leveraging machine learning, organizations can optimize customer retention efforts, ultimately fostering stronger, more sustainable customer relationships in competitive market environments.

Keywords: Machine learning, SVM, random forest, customer churn, decision tree

INTRODUCTION

Customer churn, defined as the phenomenon where customers discontinue their relationship with a company and opt for a competitor, presents a significant challenge in the business world. Retaining existing customers is generally more cost-effective than attracting new ones, making churn reduction a strategic priority for many businesses. Marketing initiatives aimed at attracting new customers require substantial investment, with the anticipation of long-term financial returns. Consequently, fostering a loyal customer base is crucial for sustained profitability.

In the telecommunications sector, the term "churners" refers to customers who switch from one provider to another. Predicting and mitigating churn is vital due to the high stakes involved; the industry faces potential churn rates exceeding 25% annually, which translates to substantial revenue losses. Numerous researchers have focused on developing predictive models to anticipate churn, aiming to enable proactive measures to retain customers. The cost of acquiring new customers is significantly higher, estimated at 15 times more, than the expense of retaining existing ones, underscoring the importance of effective customer retention strategies.

Effective churn prediction and management can lead to improved customer loyalty and stronger long-term relationships, providing businesses with a competitive edge. By leveraging machine learning

***Author for Correspondence**
Prajwal Waghole
E-mail: prajwalwaghole@gmail.com

[1]Student, Computer Engineering, RMD School College of Engineering, Pune, Maharashtra, India
[2]Assistant Professor, Computer Engineering, RMD School College of Engineering, Pune, Maharashtra, India

Received Date: May 28, 2024
Accepted Date: July 23, 2024
Published Date: August 01, 2024

Citation: Prajwal Waghole, Kalpna Saharan, Varad Wanwase, Shantanu Wasnik, Rajratan Gokhale. Customer Churn Prediction Using ML Algorithms. Journal of Computer Technology & Applications. 2024; 15(2): 70–75p.

techniques, companies can analyze demographic and behavioral data to forecast churn and implement targeted interventions, ultimately enhancing customer satisfaction and reducing churn rates.

LITERATURE SURVEY

Tien-Yu Tsai, Chin-Teng Lin, and Mukesh Prasad present a framework for predicting and responding to customer churn using intelligent techniques. It aims to enhance the accuracy of churn predictions and improve strategies for retaining customers through advanced data analysis and response mechanisms [1].

Bora Barduk discusses using modeling time statistics to improve predictions of customer churn. It likely focuses on how analyzing time-related data can enhance the accuracy of forecasts about when customers are likely to leave a service or product [2].

Laurie Butgereit explores how microservices can be utilized to construct a data pipeline specifically for machine learning applications, with a focus on a case study that demonstrates predicting customer churn. It highlights the benefits of microservices in managing and processing data efficiently to improve machine learning outcomes [3].

Xin Hu, Yanfei Yang, Lanhun Chen, and Siru Zhu present a hybrid model for predicting customer churn by combining decision tree and neural network techniques. This approach aims to enhance the accuracy of churn predictions by leveraging the strengths of both methods [4].

Pushkar Bhuse, Aayushi Gandhi, Parth Meswani, and Riya Muni focuses on using machine learning techniques to predict customer churn in the telecommunications industry. The study likely involves developing models to identify patterns and factors leading to customer attrition, helping telecom companies take proactive measures to retain customers [5].

Haotian Wu's article presents a customer churn prediction system that leverages self-attention mechanisms to enhance performance. The approach aims to improve the accuracy of predicting which customers are likely to leave a service or product, using advanced machine learning techniques to analyze customer data more effectively [6].

Asad Khattak, Zartashia Mehak, and Hussain Ahmad presents a method for predicting customer churn using a composite deep learning approach. This technique combines various deep learning models to enhance the accuracy of churn predictions, helping businesses to better understand and mitigate customer attrition [7].

Jitendra Maan and Harsh Maan presents a customer churn prediction model that utilizes explainable machine learning techniques. It focuses on enhancing the interpretability of the model's predictions, allowing businesses to understand and address the factors contributing to customer churn more effectively [8].

Shobhana J, Rakesh Kumar Arora, PN Ranjith, and J. Bamini discuss a strategy for preventing customer churn in e-commerce using machine learning. It highlights how business intelligence techniques can leverage machine learning to predict and mitigate the risk of customers leaving, aiming to enhance retention and improve overall business performance [9].

Youngjung Suh's article focuses on using machine learning techniques to predict customer churn in the home appliance rental sector. By applying these advanced methods, the study aims to enhance retention strategies and improve business outcomes in this industry [10].

PROPOSED SYSTEM

In today’s highly competitive business world, customer retention is essential to achieving sustainable growth and maintaining a competitive advantage. Customer engagement, which is when a customer’s

relationship with a company ends, is a major challenge for businesses across all industries. Customer influx not only affects revenue, but also damages brand reputation and market share. The ultimate goal is to use historical customer data to identify patterns and indicators before customers behave, so that businesses can take the first steps to retain customers and increase overall profitability and reduce customer flow. The ultimate goal is to use historical customer data to identify patterns and indicators before customers behave, so that businesses can take the first steps to retain customers and increase overall profitability.

Customer churn, also known as churn, is when a customer ends their relationship with a company or stops using the company's products or services. This is a challenge faced by businesses across all industries including communications, marketing, retail, and subscription services. Understanding and forecasting consumer competition is essential for businesses looking to build customer loyalty and maintain profitability in today's business environment.

The objective of this study is to develop and test machine learning-based approaches for predicting customer attendance in services, and system architecture is shown in Figure 1. The objective is to investigate the effectiveness of various machine learning algorithms and features in accurately predicting customers. In addition, this study aims to assess the feasibility of the model across different markets and customer segments. Additionally, it aims to provide recommendations and suggestions to businesses to support forecasting as a best practice to improve customer retention and improve business processes. Ultimately, this study aims to assess the potential return on investment (ROI) of the business impact of implementing churn forecasting.

This research is important for businesses because it can help customers develop effective strategies to reduce turnover, ultimately increasing profitability and sustainable growth. Support Vector Machine (SVM), using Random Forest (RF) and Decision Tree (DT) algorithms provides an efficient way to predict churn patterns and retain customers. By doing this, businesses can better distribute resources, improve efficiency, and increase overall customer satisfaction.

METHODOLOGY

We employed three machine learning algorithms to forecast customer churn:

1. Support Vector Machine (SVM).
2. The Random Forest.
3. The decision tree.

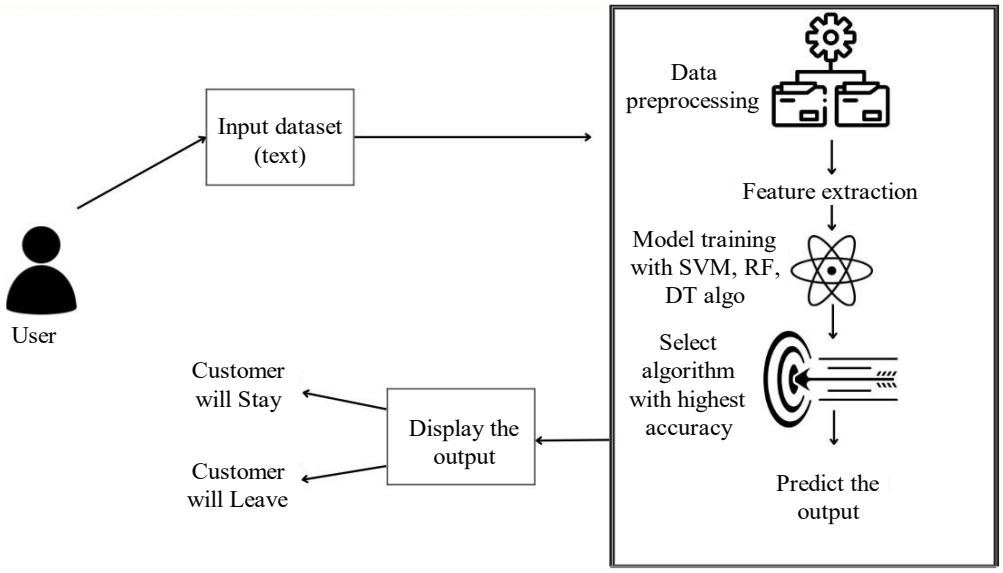

Figure 1. Customer attendance system architecture.

For each algorithm, we performed the following steps:

1. *Model Training:* Train the model using the training dataset.
2. *Hyperparameter Tuning:* Refine model hyperparameters using methods such as grid search or random search.
3. *Model Evaluation:* Assess the model's performance using metrics like accuracy, precision, recall, and F1-score.

RESULTS

After training and evaluating the models, we discovered that the random forest algorithm surpassed SVM and decision trees in prediction accuracy. The results are summarized below:

- *SVM:* Accuracy for SVM is 68%. It means our model had predicted churn of 68% customers from dataset.
- *Random Forest:* Accuracy for RF is 86%. It means our model had predicted churn of 86% customers from dataset.
- *Decision tree:* Accuracy for DT is 82%. It means our model had predicted churn of 82% customers from dataset
- *Additional terms*: Precision, recall, and F1-score offer additional insights into the model's performance:

Precision: Precision gauges the accuracy of the model's correct predictions. True precision is the ratio of the total number of good predictions made for the model.

Precision = true positive/(true positive + false positive).

In customer churn prediction, accuracy represents the proportion of correctly identified churn customers among all customers predicted to churn. High accuracy means that the model has few spurious items, i.e. it correctly identifies the sponsored customers without falsely listing too many non-churners.

Recall: Recall assesses the model's capability to identify all relevant instances within the dataset. It is the ratio of the total number of actual true predictions and actual true information in the data set.

Recall = True Positive/(True Positive + False Negative).

In customer churn forecasting, recall represents the proportion of correctly identified customers out of all actual churn customers. High recall means that the model captures a greater proportion of actual customer engagement.

F1-score: The F1-score is a harmonic mean of precision and recall. It strikes a balance between accuracy and recall, giving both metrics equal weight.

F1-score =2× (precision × residual)/(precision + residual).

DISCUSSION

After training and implementing all the algorithms we have compared the performances of each algorithm and found out that each algorithm gives different accuracy for the provided dataset. The SVM gives the accuracy of 68%, similarly, Random-forest gives accuracy of 86% and following with decision tree which gives accuracy of 82% as shown in Figure 2.

So, after comparing all the algorithms we have come up with result that:

RANDOM FOREST>DECISION TREE>SVM

Random forest gives highest accuracy followed by decision tree and then SVM. This discussion is totally based on results we achieved while testing the data with various inputs (Tables 1–3).

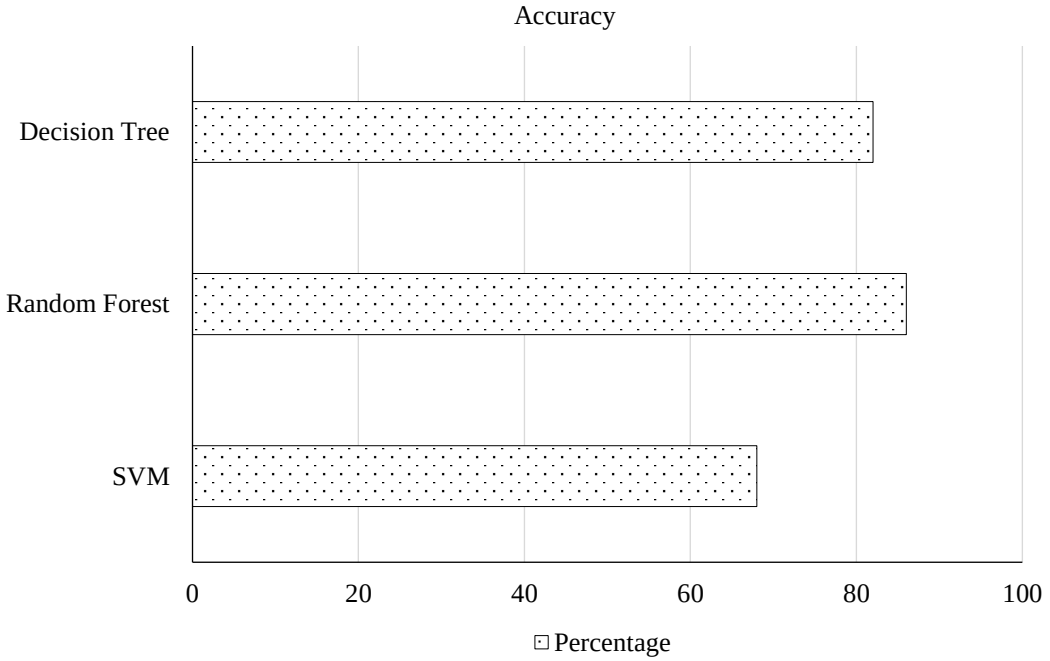

Figure 2. Accuracy Comparison Bar chart.

Table 1. SVM.

	Precision	Recall	F1-Score	Support
0	0.80	0.49	0.60	105
1	0.71	0.75	0.73	240
2	0.57	0.75	0.65	83

Table 2. Random Forest.

	Precision	Recall	F1-Score	Support
0	0.80	0.81	0.81	250
1	0.86	0.87	0.87	441
2	0.92	0.89	0.90	210

Table 3. Decision tree.

	Precision	Recall	F1-Score	Support
0	0.76	0.66	0.71	220
1	0.80	0.86	0.83	429
2	0.93	0.91	0.92	207

CONCLUSION

In this study, we developed and analyzed a machine learning model aimed at predicting customer churn within the telecom industry. Our objective was to offer actionable insights to aid the company in decreasing customer attrition and enhancing retention strategies. We evaluated the effectiveness of three well-known algorithms: Support Vector Machine (SVM), Random Forest, and Decision Tree. After rigorous testing and analysis, Random Forest was identified as the most effective algorithm, achieving a notable accuracy of 86%. It exceeded SVM and Decision Tree in accuracy, recall, and F1-score. The superior performance of Random Forest is due to its ensemble approach, which effectively captures complex data relationships and reduces overfitting. By combining predictions from multiple decision trees, Random Forest provides robust and reliable forecasts, making it a valuable tool for predicting customer churn in the telecom industry.

Future Scope

The future of machine learning for churn prediction has many promising avenues for research and application. The combination of advanced technologies such as deep learning and hybrid algorithms has the potential to increase the accuracy and power of predictions. Exploring the dynamics of time by analyzing time and individual patterns can lead to greater insight and retention strategies. Instant predictive and descriptive AI technology is effective and associated with trust. Integration of multiple sources provides greater understanding, while interdisciplinary and ethical consideration ensure broad application and responsible delivery. By addressing these challenges and opportunities, researchers and practitioners can spur innovation in many business areas, improve customer retention strategies, and promote ethical AI.

Acknowledgement

Grateful acknowledgement is made to Mrs. Kalpna Saharan who provided us with the suitable guidance, and Dr. Vina Lomte who gave us considerable help by means of suggestion, comments and criticism. Without their help, the completion of the study would be impossible. In addition, we deeply appreciate the various contributions made to this article.

REFERENCES

1. Tsai TY, Lin CT, Prasad M. An intelligent customer churn prediction and response framework. 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Dalian, China, 2019, p. 928–935. DOI: 10.1109/ISKE47853.2019.9170380.
2. Bardük B. Modelling time statistics for customer churn prediction. 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 2020, p. 1–4. DOI: 10.1109/SIU49456.2020.9302329.
3. Butgereit L. Work towards using micro-services to build a data pipeline for machine learning applications: A case study in predicting customer churn. 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE), Aswan, Egypt, 2020, p. 87–91. DOI: 10.1109/ITCE48509.2020.9047807.
4. Hu X, Yang Y, Chen L, Zhu S. Research on a customer churn combination prediction model based on decision tree and neural network. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 2020, p. 129–132. DOI: 10.1109/ICCCBDA49378.2020.9095611.
5. Bhuse P, Gandhi A, Meswani P, Muni R, Katre N. Machine learning based telecom-customer churn prediction. 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, p. 1297–1301. DOI: 10.1109/ICISS49785.2020.9315951.
6. Wu H. A high-performance customer churn prediction system based on self-attention. arXiv Preprint ArXiv:2206.01523. 2022 Jun 3.
7. Khattak A, Mehak Z, Ahmad H, Asghar MU, Asghar MZ, Khan A. Customer churn prediction using composite deep learning technique. Sci Rep. 2023;13:17294. DOI: 10.1038/s41598-023-44396-w.
8. Maan J, Maan H. Customer churn prediction model using explainable machine learning. arXiv Preprint ArXiv:2303.00960. 2023 Mar 2.
9. Shobana J, Gangadhar C, Arora RK, Renjith PN, Bamini J, Chincholkar YD. E-commerce customer churn prevention using machine learning-based business intelligence strategy. Measurement. 2023;27:100728.
10. Suh Y. Machine learning based customer churn prediction in home appliance rental business. J Big Data. 2023;10:41. DOI: 10.1186/s40537-023-00721-8.

