Grenze International Journal of Engineering and Technology, June Issue

ChurnShield: Machine-Learning-Driven Customer Churn Prevention System

Dhanshri Deshpande[1], Pralhad Deshpande[2], Tanvi Chopade[3], Aajit Aade[4], Nirbhay Chukekar[5] and Renuka Vaidya[6]
[1-6]Information Technology, Vishwakarma Institute of Technology, Pune, India
Email: dhanshri.deshpande24@vit.edu, pralhad.deshpande24@vit.edu, tanvi.chopade24@vit.edu, aajit.aade24@vit.edu, Nirbhay.chukekar24@vit.edu, renuka.vaidya@vit.edu

***Abstract*— An innovative research method involving a dataset with 21 columns and 7,044 cases is employed in an application that possesses a user-friendly interface based on Flask. The users can either input their data manually or upload their CSV files, thereby making the system an easy and effective method to analyze their data. Using the logistic regression and correlation exploration, the resulting model managed to attain high prediction accuracy and achieve a precision of 95.57% and overall accuracy of 92.08% when it makes customer churn predictions. Companies will henceforth regard this application because it provides abstraction of operational complexity with an ability to make predictions, enabling it to predict customer departures in order to offer enhanced customer retention strategies and decision-making. Further, since the research gave timely predictions customer relationship management processes can be enhanced, and assist companies in all industries reduce customer churn and enhance sustainable growth by means of higher operational efficiency and deliver business intelligence that is capable to enhance an organization's competitiveness. It employs continuous monitoring and constant model upgrades which allows the application to remain practical in assisting companies to ride through various environments and continue to deliver performance that will result to long-term impact.**

***Keywords*— Customer Churn Prediction, Logistic Regression, Ensemble Methods, Feature Engineering, Machine Learning
Deployment.**

I. INTRODUCTION

Today's competitive marketplace requires a company to focus on the retention of its customers, which has a direct impact on business growth and profit margins. There are significant challenges across industries as customers choose to no longer conduct business with companies. Companies need the right processes to find customers that have a high likelihood of attrition and act on them with some appropriate retention tactics to minimize revenue losses and maintain loyal subscribers. Manual review of customer attrition has many downsides since those methods cannot accurately predict attrition a timely manner. To mitigate this issue, we developed a machine learning algorithm that predicts attrition more effectively. The algorithm predicts customer attrition at a given time using previous customer data, including two stage support incidents, service subscriptions, and the length of the relationship.

Grenze ID: 01.GIJET.12.2.664_26
© Grenze Scientific Society, 2026

Businesses benefit from this forecasting method by correctly allocating their retention efforts, allocating resources, and increasing customer satisfaction. Additionally, businesses have dynamic access to churn forecasting tools that enable them to make more informed strategic decisions through proactive retention planning.

II. LITERATURE REVIEW

The Logit Leaf Model (LLM) is a novel approach that uses decision trees and logistic regression to generate predictions for customer churn, where data are first segmented using decision tree rules and then logistic regression models are developed within these segments. As shown in the study data below, the predictive ability of LLM is comparable to advanced methods such as random forests and logistic model trees, and outperforms standalone decision trees and logistic regression, while maintaining good interpretability [1].

A review of 212 published studies from 2015 to 2023 examined machine learning techniques that aimed to forecast customer attrition at every stage of the model prediction process. This study shows that combining demographic data with consumer behavior data from particular industries— gaming, finance, and telecommunications, in particular—is essential for effective churn prediction. To maximize customer retention and profit generation, this study identifies a dearth of research on profit-based evaluation metrics, which it suggests addressing through ensemble methods in conjunction with deep learning approaches and explainable techniques [2].

A supervised machine learning model was used to predict customer churn in Python. Results indicated that when evaluating labeled data, K-Nearest Neighbors (KNN) classified customers who churn with 2 percent greater accuracy than the Logistic Regression method. The confusion matrix indicated that KNN was better at forecasting customer attrition than relational regression.

A model using Random Forest (RF) and clustering algorithms was developed for the telecom industry to identify what customers would churn and better understand the factors affecting their decisions. When assessing churn events, the predictive model achieved an average accuracy range of 8863 percent. Since the prediction model was able to predict customer churn and identify customer characteristics, the model will help to improve CRM performance, and planned marketing programs, since the factors contributing to customer churn were identified. The model was assessed using a number of performance metrics including accuracy, precision, recall, f-measure, and ROC area [4].

This research identified a combined methods approach306, specifically support vector machines (SVM) and k-means clustering to formulate a B2C e-commerce customer churn prediction system. To optimize predictions and create initial consumer clusters, the model divides customers into three groups. Previous studies found that support vector machines (SVM) are more accurate than logistic regression when predicting customer churn rates [5].

This study investigates the ability of data mining techniques to forecast churn within the telecommunications industry. To locate the customers who are going to churn, a process of previously working with and processing data in addition to segmentation techniques is required. In this study several experiments were conducted in order to compare the effectiveness of different classification approaches in identifying customers that were about to churn. The outcome of the experiment demonstrates that accurate churn prediction supports organisations to stay in touch with their clients and, in this way, retain satisfied customers [6].

In their article, they employ a transactive learning paradigm with a data set from the banking industry and they create a customer churn prediction model using ANN (Artificial neural networks) for the purpose of prediction. They achieved 86% accuracy in customer attrition prediction using forward propagation and cross validation adjusting hyperparameters. The results show ANNs performed better than logistic regression on this dataset. The paper provides meaningful contributions into machine learning implementations to find ways to improve customer retention and to hold onto higher risk customers that can have some recommendations for proactive banking practice [7].

Artificial intelligence techniques related to direct marketing segmentation, customer profiling, and sales forecasting were all investigated. Customer segments: new, best, and sporadic. To determine the appropriate number of clusters, the K-means algorithm utilizes the elbow method, the gap statistics, and the silhouette coefficient [8].

Analyzing customer transaction data for this study, I attempted to build a customer segmentation and prediction model using a Markov model with K-means clusters. The Markov Model can make forecasts about future cluster movements by utilizing the transition matrix built from cluster groups. The model helps firms plan for marketing strategies and budgets, as it forecast the future value of a customer group. In addition, the model can be of great assistance to firms conducting large marketing campaigns in the retail, insurance, and e-commerce industries, as it allows them to segment their customers efficiently, profile them, and estimate their lifecycle value [9].

To create customer segments and predict consumer behavior, this ecommerce study uses Recency, Frequency, and Monetary (RFM) metrics. Instead of directly clustering the RFM table, this study used a separate one-dimensional cluster analysis. The company's customer retention and marketing plans are supported by machine learning algorithms and ensemble techniques that improve customer relationship management and segmentation [10].

This study explores how customer segmentation supports changes in consumer behavior using machine learning techniques. The study experiments with the conceptual and performance characteristics of popular machine learning algorithms, such as AdaBoost and Decision Trees, in addition to Logistic Regression. This study provides guidance for scientists working in related academic fields and outlines potential directions for future research on machine learningbased customer segmentation [11].

Effective customer retention and correlation management is enormously important for the success of any organization. This paper examines the effectiveness of customer retention and correlation management in an industry; banking and finance; while exploring these topics in telecommunications. Accurate customer churn prediction is very important to organizations to prevent losing customers and maintain competitive advantage. Decision trees and K-means clustering work together to accurately model churn prediction in the telecom space [12].

This research analyses the relationship between management and customer retention impacts on performance effects in the bank, financial and telecommunications sectors. To keep profitable relationships with clients and a position in the market in these sectors, correct assessment of customer churn is important. The three clustering techniques used for consumer segmentation are K-means, fuzzy C-means and probabilistic fuzzy C-means. The system applies decision trees to combine the supervised and unsupervised methodologies when training and conducting tests. Also, the author considers combining K means clustering with decision trees, to improve customer churn prediction outcomes in the telecommunications sector (Patil, 2018) [13].

This study examines the use of predictive neural networks to improve consumer segmentation in e-commerce. The study segments the target market using clustering algorithms and examines data components such as product evaluations, viewing and purchase activities, and time-based divisions. Network analyses determine consumer product preferences, but feature extraction requires the combination of unigram analysis, bigram analysis, and trigram examination. This study used classification and sentiment mining techniques to determine how consumers feel about certain brands and their offerings. It is possible to identify appropriate brands in the input dataset by evaluating prediction accuracy [14].

This paper discussed a comprehensive architecture for customer segmentation and attrition forecasting in telecommunications industrial contexts. Its six operational components entail factor analysis, data pre-processing, churn prediction, customer segmentation, and analytics on customer behaviours. Factor analysis was conducted using Bayesian Logistic Regression to establish the influential characteristics of churn customer segments. As part of this integrated system, K-means clustering was employed to complete the segmentation. By having an integrated system like this one, telecom firms can reduce customer attrition rates through targeted retention strategies[15].

The methodology of a new customer segmentation approach evaluates customer value change over time. Traditional segmentation methods calculate worth from a static customer evaluation perspective and do not include temporal value fluctuation analysis. This new approach uses time-based value changes instead of static-based segmented value identification analysis. Forecast predictions are more accurate when the model is trained using past customer behavior patterns. The present study used point of sale (POS) customer transaction records to apply this methodology [16].

III. RESEARCH GAP

Customer churn prediction has generated literature volume, which has included the use of statistical models, machine learning tools and data science techniques. Many studies have improved upon the methods of prediction, but they tended to focus on performance metrics such as accuracy or AUC. Critically, the studies have sacrificed interpretable prediction methods, which is important for business operations to know what drives customer churn and to develop targeted strategies to mitigate the churn in actionable ways.

A second notable gap in the literature is the limited features used. Many studies have focused heavily on demographic and transactional characteristics of the customer, while behavioral variables and service usage variables are not always present, or not sufficiently represented in existing literature beyond transactional use.

As Customer churn is a function of multiple factors, the selective usage of the features can limit the generalization and robustness of the customers' predictive models.
Furthermore, churn datasets are imbalanced; that is, there are generally far fewer churners relative to non-churners. Many of the studies failed to overcome this imbalance, resulting in biased models that tended to do well on the majority of cases (non-churners) while failing to do well on the minority churners. Handling this imbalance still remains an under-studied area of churn prediction research.
In addition, many studies test only one algorithm or a
limited number of models. This form of evaluation does not help to understand the benefits and tradeoffs between different classes of models in terms of interpretability and predictive power (e.g., models such as logistic regression, decision trees, ensembles, or hybrid models).
Finally, while churn prediction models are generally good at predicting customers that could be leaving, few studies go beyond prediction to provide actionable information. Without linkages to business strategies, the models become less useful.
All of these gaps suggest a need for a churn prediction framework that considers interpretability as well as predictive power, incorporates a wide variety of features, balances class imbalance, and can translate the model output into customer retention action steps.

IV. NOVEL CONTRIBUTIONS

- Thorough Feature Engineering - We build on the studies acknowledged in this research by using an assortment of customer demographics, service-usage, and behavioral features in a combination. Presumably the models will have enhanced predictive ability by including the behavioral factors and service usage.
- Comparison of Models - The extra machine learning algorithms (Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and Logit Leaf Model (LLM)) added depth of analysis, and provide further insight of explicability as opposed to purely predictive strength.
- Commitment to Interpretability - The literature often appears to combine predictive accuracy with ever-increasing complex digitally-derived models; this paper yielded accuracy, while primarily concentrating on explicable models (LLM and logistic linear regression) and even easier options to further improve performance (ie, simpler ensemble models). This gives us a better chance of practicability as output from this model can typically be defended and/or justified in practice.
- Tailored Accuracy Measure (Evaluation) - The models in this study were assessed with measures of accuracy, precision, recall, F1 score and AUC in addition to accuracy, to better understand their performance on churn datasets that had considerable class imbalances.

V. METHODOLOGY

The process kicks off by gathering datasets from various sources, or even creating new ones if necessary. The first step is to really understand the structure and content of these datasets. Once that initial analysis is done, it’s time to prepare the data for deeper analysis: this means filling in any missing values, identifying outliers, and crafting new features to extract more insights from the data. After preprocessing, the data is split into training and testing sets. A crucial part of the process is testing out different machine learning algorithms to find the one that fits the task best. Once we’ve selected a model, it gets trained on the training dataset, and we tweak the hyperparameters to boost its performance. Finally, the model is tested with the testing dataset to see how well it generalizes before it’s made available for users to interact with through an interface that allows for data input and CSV file uploads.
The process of adopting trained models for real-time implementation follows the model-evaluation stage. Either new system integration or a stand-alone program that uses the model is required for the implementation process. Depending on how well the user interface is created to encourage interaction, a model implementation will either succeed or fail.
Users can manually enter data or upload files via the system based on the needs of their application through an analysis selection process. As a crucial step, post-deployment monitoring of the model will assist in tracking its performance.
Changes in data patterns or business requirements necessitate periodic retraining of models. The predictive system is continuously improved using user feedback mechanisms, which provide the model with continuous improvement.

VI. RESULTS AND DISCUSSION

The website is designed as a flask-based website that helps users interact with the data and predict churn in the future. The main page of Fig 2 provides the user with two data entry options: manual entry through fig.3 or CSV file upload.

Fig. 1 shows a block diagram of the project workflow. The first stage of the system is the input phase, in which raw data are fed into the system. For this project, the raw data were customer-related data that contained contractual details, online security formats, and fee payment information. The preprocessing stage follows data input and includes the tasks of cleaning and normalizing data with feature extraction to arrive at an appropriate data format before model training can commence.

The model training procedure involved a correlation analysis using logistic regression algorithms. Once the model training is completed, the system has an easy offers two data entry methods because users can either enter data via a web based form or a mobile application.

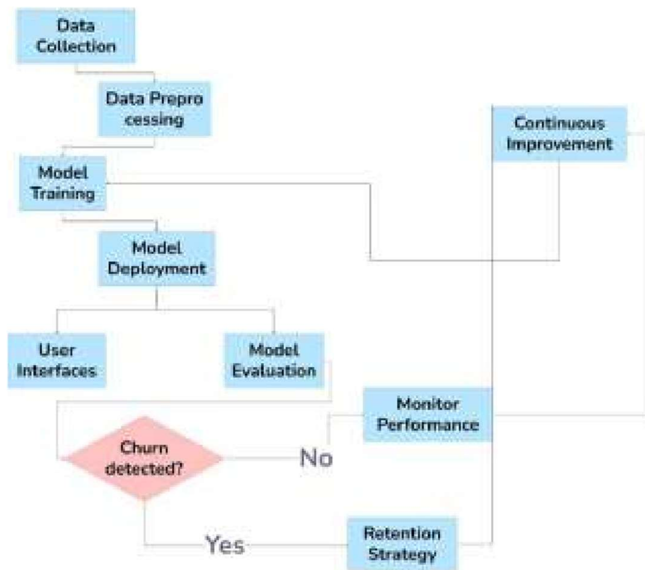

Fig 1. Block Diagram

A complete view of the project operation is provided by the block diagram, which shows how information moves through the model deployment stage to predict customer churn activity. An overview of the entire process, including its crucial phases, is provided to show the execution steps of the project. to-use front end that stakeholders can comfortably use. The system through fig.4. The manual entry form in Figure 3 allows a user to input key customer details, including contract type and online security status, alongside tenure and other relevant information.

The CSV file upload option in Fig.4 accelerates the prediction tasks by allowing easy handling of large datasets. The system offers two options to provide an enhanced pathway that accommodates all user preferences and increases the accessibility of the system for everyone.

Fig.2 Homepage

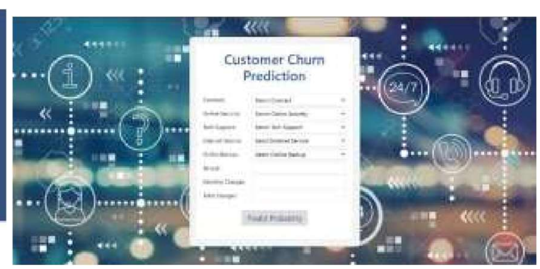

Fig. 3. Prediction via manual entry of data

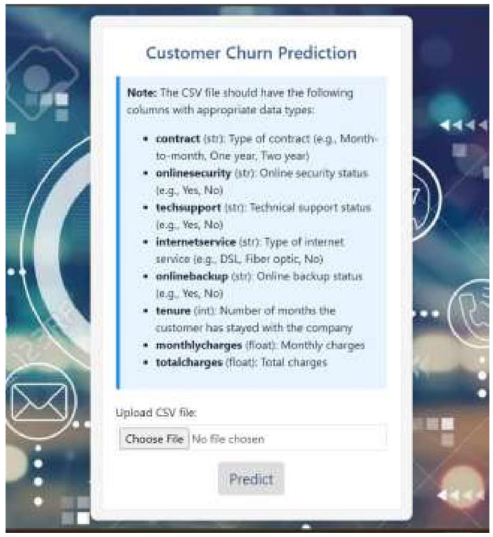

Fig. 4 Prediction via uploading csv file

As shown in Fig.5, in the interface screenshot, users input customer information manually in the form, and the system then immediately shows prediction results about churn probability on the webpage, so that users can get a fast understanding of churn probability by direct data entry of customer information. The streamlined process, along with direct prediction functionality, can provide a better user experience for businesses to make rapid decisions and allow users to gain a quick understanding of prediction results to set up proper measures for handling potential churn risks to improve customer retention.

Once a user uploads a CSV file with multiple data entries into the Flask powered website fig.6, churn prediction can be initiated by clicking the predict button. The churn prediction model trained with the uploaded dataset runs each entry through the model to evaluate the churn probabilities for each customer. The prediction system produces or allows for downloading a new CSV file that includes the original data along with a column containing churn status predictions for each record, as shown in fig.7. The system produces outputs in a format that allows for rapid evaluation of individual customer churn risks and customized retention plans. This seamless integration of the prediction output with practical decision components makes it possible for companies to enhance customer retention systems as well as business operations. The developed methods achieve high predictive accuracy for the machine learning component of the churn prediction system. A model trained using an extensive dataset with multiple characteristic fields is enabled by complex algorithms that allow the model to learn from a large number of customers to update its parameters and improve its performance. The churn prediction model has a high accuracy (92.08%) in distinguishing between customers who leave and those who stay.

The model was evaluated based on these three metrics: precision, recall, and F-measure. The precision score was 94.56% for identifying true cases of churn in the predicted positives, the recall score was 89.03% for identifying most of the real cases of customer churn in the dataset, and the Fmeasure metric provided a single assessment of 92.03% strength, showing the robustness of the machine learning approach to predict customer churn. By combining current practices with strict model assessment, organizations can identify patterns that may lead to meaningful insights that help prevent customer losses.

To really boost the accuracy and performance of our model, we need to pay attention to feature selection and ensemble algorithms. Plus, we can enhance things further by using hyperparameter optimization methods like Grid Search. It's also a good idea to apply L1 or L2 regularizers to logistic regressions and balance our datasets to get better accuracy. Data preparation is key here; we need to clean the data and encode categorical variables, and we can use techniques like KNN or MICE to tackle outliers and fill in missing values. Our dataset should provide detailed descriptions of the features and give us insight into how the classes are distributed. We can take performance up a notch by incorporating automated machine learning pipelines for technical model improvements and using cross-validation in our validation approaches. The accuracy, precision, recall, and F-measure of the churn prediction model indicate that it can play a crucial role in making significant business decisions and driving business growth.

VII. Comparative Analysis

Comparative Analysis of Machine Learning Classification Models

Model	Accuracy	Precision	Recall	F1Score	Interpretability
Logistic Regression	92.08%	95.57%	91 %	90 %	High (transparent coefficients & odds ratios)
Decision Tree	84%	76%	71 %	73 %	Moderate (rule based, easy to visualize)
Random Forest	88%	81%	77 %	79 %	Low (black box ensemble)
Gradient Boosting	90%	83%	79 %	81 %	Low (black box boosting)
Logit Leaf Model (LLM)	87%	80%	76 %	78 %	High (hybrid: interpretable + strong performance)

VIII. Future Scope

This study can be further extended in many ways. Below are possible directions for further exploration:
Advanced Models - Evaluate deep approaches (i.e., RNNs, transformers) that capture temporal and sequential customer behavior.
Explainability - Incorporate explainable AI solutions (i.e., SHAP, LIME) that compromise accuracy for interpretability. Real-Time Prediction - Incorporate streaming data, allowing for live churn detection and interventions. Domain-specific models - Create models that are unique to specific domains (i.e., telecom, bank, e-commerce). Prescriptive analytics - Predict customer retention and suggest recommendations based on prediction metrics for remaining customers at risk of churning.

IX. Conclusion

Better customer churn prediction technology is a first step for improving business strategies for customer retention practices and increasing profits. This platform utilizes machine learning algorithms to churn data to provide highly accurate predictive customer churn potential for organizations to reduce customer loss. The platform is presented in a way that is accessible to personnel with significantly different technical qualifications and in that, allows the opportunity for personnel to manually enter data or upload CSV files to access this tool. Predictive analysis will provide business executives the opportunity to allocate and direct organizational launching of resources around customer retention with the most success using the maximum return. The predictive customer attrition model described represents how insights from data can cause operational change for developing sustainable corporate growth. The predictive value of analytics to predict future customer behavior has been recognized by businesses and the predictive customer behaviour attrition model should be an essential tool to retain competitive advantage. Updates to the model to improve the version, will improve its predictive power of customer attrition potential to ensure predictable and lasting customer relationships for sustainable success.

References

[1] pp. 60134-60149, 2019, doi: 10.1109/ACCESS.2019.2914999

[2] Xiahou, Xiancheng, and Yoshio Harada. 2022. "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM" Journal of Theoretical and Applied Electronic Commerce Research 17, no. 2: 458-475. https://doi.org/10.3390/jtaer17020024.

[3] L. F. Khalid, A. Mohsin Abdulazeez, D. Q. Zeebaree, F. Y. H. Ahmed and D. A. Zebari, "Customer Churn Prediction in Telecommunications Industry Based on Data Mining," 2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA), Langkawi Island, Malaysia, 2021, pp. 1-6, doi: 10.1109/ISIEA51897.2021.9509988

[4] B. Baby, Z. Dawod, M. S. Sharif and W. Elmedani, "Customer Churn Prediction Model Using Artificial Neural Network: A Case Study in Banking," 2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakheer, Bahrain, 2023, pp. 154-161, doi: 10.1109/3ICT60104.2023.10391374.

[5] Kasem, M.S., Hamada, M. & Taj-Eddin, I. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Comput & Applic 36, 4995–5005 (2024). https://doi.org/10.1007/s00521-023-09339-6.

[6] A.S. Harish and C. Malathy, "Customer Segment Prediction on Retail Transactional Data Using K-Means and Markov Model," Intell. Automat. Soft Comput. , vol. 36, no. 1, pp. 589-600. 2023. https://doi.org/10.32604/iasc.2023.032030.

[7] A. Patra, R. Khan and S. Vijayalakshmi, "Customer Segmentation and Future Purchase Prediction using RFM measures," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 753-759, doi: 10.1109/ICAC3N56670.2022.10073993.

[8] Wang, Zhiyue. "Customer Segmentation Based on Machine Learning Methods." Highlights in Science, Engineering and Technology 92 (2024): 126-132.

[9] Sivasankar, E., and J. Vijaya. "Customer segmentation by various clustering approaches and building an effective hybrid learning system on churn prediction dataset." In Computational Intelligence in Data Mining: Proceedings of the International Conference on CIDM, 10-11 December 2016, pp. 181-191. Springer Singapore, 2017.
[10] Khalili-Damghani, Kaveh, Farshid Abdi, and Shaghayegh Abolmakarem. "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries." Applied Soft Computing 73 (2018): 816-828.
[11] Singh, Himanshu, and Shubhangi Neware. "Improving customer segmentation in e-commerce using predictive neural network." International Journal 9, no. 2 (2020).
[12] Wu, Shuli, Wei-Chuen Yau, Thian-Song Ong, and Siew-Chin Chong. "Integrated churn prediction and customer segmentation framework for telco business." Ieee Access 9 (2021): 62118-62136.