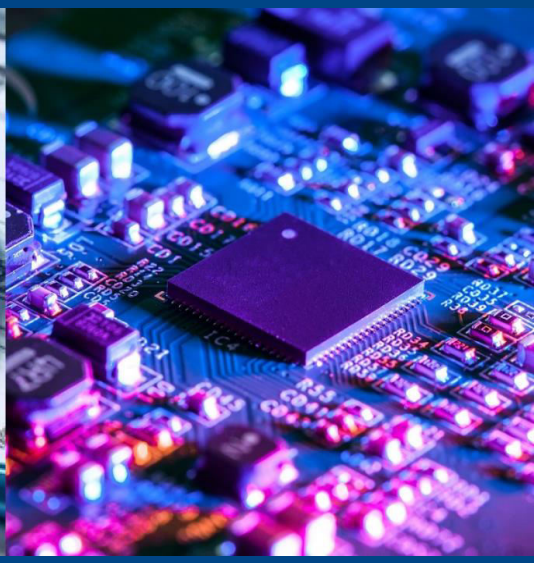
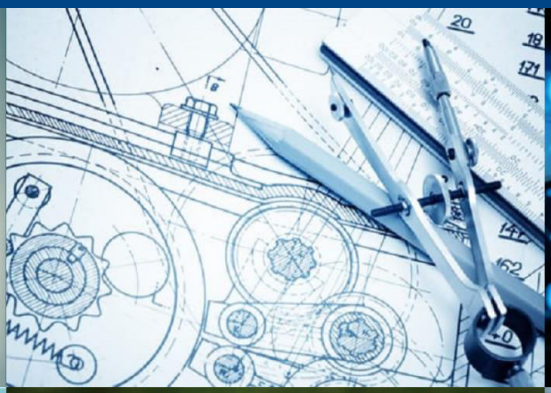
ISSN: 2582-7219

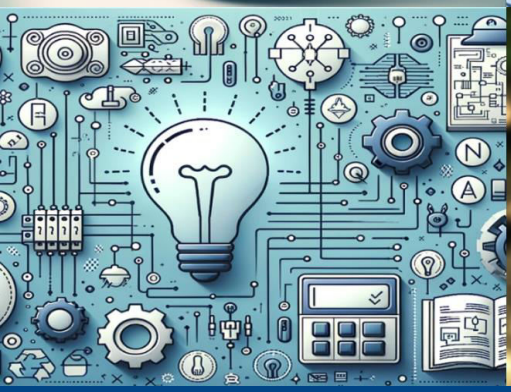




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.864

Volume 9, Issue 5, May 2026

© 2026 IJMRSET | Volume 9, Issue 5, May 2026| DOI:10.15680/IJMRSET.2026.0905033

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Adaptive Cross-Domain Customer Churn Prediction: An Automated Multi-Model Selection Framework

Mr. R.V. Dharmadhikari[1] , Dr. S.A. Shaikh[2] , Abhijeet Kadam[3], Chaitanya Gaikwad[4], Samarth Sonawane[5]

Professor, Department of Electronics and Computer Engineering, PREC, Loni, Maharashtra, India[1,2]

Student, Department of Electronics and Computer Engineering, PREC, Loni, Maharashtra, India[3, 4, 5]

ABSTRACT: Customer churn prediction is essential for improving retention and mitigating revenue loss, yet many existing machine learning solutions are domain-specific, require manual feature engineering, and rely on evaluation practices that are unreliable under class imbalance, limiting transferability across industries. This study aims to develop an adaptive cross-domain customer churn prediction framework that automates preprocessing, multi-model training, hyperparameter optimization, benchmarking, and best-model selection, and supports deployment for real-time decision support. The end-to-end pipeline ingests user-uploaded CSV datasets from heterogeneous domains (banking, gym membership services, and e-commerce), performs automated data cleaning, missing-value treatment, categorical encoding, scaling/normalization, validation, and stratified train–test splitting, and trains multiple binary classifiers (SVM, Random Forest, Gradient Boosting, and XGBoost) using stratified k-fold cross-validation. Models are evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC, with automated selection based on the highest mean ROC-AUC to ensure robust comparison in imbalanced settings. Experiments on a public bank churn dataset (≈10,002 records; churn rate 20.38%) showed that Random Forest achieved the best cross-validated performance (mean ROC-AUC 0.8379 ± 0.0081; mean accuracy 0.8478), outperforming XGBoost (mean ROC-AUC 0.8168 ± 0.0088; mean accuracy 0.8391) and demonstrating stable generalization across folds. Ensemble methods consistently exhibited superior behavior due to their ability to capture nonlinear patterns and handle mixed-type tabular data while reducing overfitting. The selected model was deployed via a Flask web application with secure authentication, role-based access control, prediction-history storage (PostgreSQL), and outputs including churn probability, binary churn classification, and actionable recommendations. Overall, the framework provides a scalable, deployment-ready approach for cross-domain churn prediction, while highlighting future needs for interpretability, temporal modeling, transfer learning, and streaming-data integration..

I. INTRODUCTION

Customer retention has turned into a major issue for firms in very competitive markets. Churn refers to the situation when clients stop utilizing a company's goods or services. Churn results in huge losses due to customer acquisition costs being significantly higher than retention. According to the findings, the price of finding a new client can be up to five times greater than the cost of maintaining the current customers.

With the increase in data availability regarding customers' behavior patterns, machine learning methods have been applied to predict customers who would most likely leave. However, most churn prediction systems tend to operate only within their domain and use pre-trained machine learning models that cannot be adapted to novel datasets. In addition, these systems usually require much feature engineering and complicated architecture to build.

Existing works mainly deal with single-domain data sets, for example, telecommunications or banking. Thereby, there is no system that could handle multiple heterogeneous datasets and be evaluated and deployed properly.

In order to overcome the described limitations, this paper introduces a system for cross-domain adaptive customer churn prediction with automatic multi-model selection. The framework allows uploading different datasets on such topics as banking, gym services, and online purchases. Data processing and classifier training and parameter tuning are

© 2026 IJMRSET | Volume 9, Issue 5, May 2026| DOI:10.15680/IJMRSET.2026.0905033

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

performed automatically. Classification algorithms include Support Vector Machine, Random Forest, and XGBoost models. The optimal model is selected based on cross-validation ROC-AUC mean score. Finally, a web-based system has been developed using Flask framework. The interface allows performing client-side user authentication and makes predictions possible in real time. Besides, prediction history could be saved in the database.

Main contributions of this project are listed below:

1. Cross-domain adaptive churn prediction framework development.
2. Multi-classifier learning and evaluation via cross-validation.
3. Best model selection based on optimization of mean ROC-AUC score.
4. Web-based deployment with user authentication, real-time prediction capabilities.

II. LITERATURE REVIEW

Many studies on predicting customer churn have been conducted by employing statistical and machine learning methodologies. The early approaches used statistical models like logistic regression and survival analysis to predict churn rates. While providing good explanations of the findings, these approaches were inadequate in capturing non-linear relations among variables in customer datasets. In recent years, machine learning models, including Decision Trees, SVM, Random Forest, and Gradient Boost, have shown improved performances in churn prediction. For instance, Caigny et al. introduced a hybrid classification approach that combines logistic regression and decision trees to enhance churn detection in structured data. Further, Lalwani et al. applied different machine learning algorithms in the prediction task and achieved higher accuracy with ensemble models. Recently, research in customer churn prediction is shifting towards data mining and ensemble learning methods. Tangyuan and Moro examined the progress in churn prediction literature and observed that there is an increasing trend towards using tree-based ensembles due to robustness and the capability to work with diverse datasets. Alboukaey et al. developed dynamic behavior-based churn prediction models for telecom customers, focusing on the evolution in customers' behaviors. Although there has been substantial improvement in performance, many solutions are still domain-specific, being trained on single datasets. Besides, numerous researchers have not considered automated model selection, cross-domain generalization ability, and the development of deployable systems in their churn models. Little emphasis has been placed on developing frameworks that can handle heterogeneous datasets from different sectors such as finance, retail, or health care. Moreover, some studies have measured the performance of churn prediction models using the accuracy metric only, which may not be appropriate when dealing with imbalanced churn data. Other metrics like ROC-AUC and F1 score give a better understanding of the robustness of classification but have rarely been incorporated in automated model optimization within churn prediction systems. To overcome the shortcomings mentioned above, this paper suggests developing a cross-domain adaptive framework for churn prediction using automated multi-model benchmarking and model selection using cross-validated ROC-AUC. Contrary to traditional domain-specific solutions, the proposed system will allow dataset uploading, automated data preprocessing, hyperparameter tuning, and model deployment via web-based interface.

III. METHODOLOGY OF PROPOSED SURVEY

A. System Overview

The proposed framework is organized into five major modules, as shown in Figure 1: data acquisition, data preprocessing, model training, model evaluation and selection, and deployment. The complete architecture of the system is illustrated in Figure 1.

The framework allows users to upload customer datasets in CSV format from different business domains, including banking, gym membership services, and e-commerce platforms. This cross-domain capability improves the adaptability and practical usability of the system across multiple industries.

B. Data Acquisition and Preprocessing

After uploading, the dataset undergoes an automated preprocessing stage to maintain data quality and ensure compatibility with the system. The preprocessing process also supports time-independent analysis. Let the dataset be represented as:

$$D = \{(x_i, y_i)\}_{i=1}^{n}$$

© 2026 IJMRSET | Volume 9, Issue 5, May 2026| DOI:10.15680/IJMRSET.2026.0905033

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

where $x_i$represents the feature vector and $y_i$denotes the churn label, where 0 indicates non-churn and 1 indicates churn.
The preprocessing pipeline includes the following steps:

- Handling missing values
- Encoding categorical variables
- Feature scaling and normalization
- Data cleaning and validation
- Stratified train-test splitting

These operations transform raw and unstructured data into a structured feature matrix suitable for machine learning classification tasks.

C. Model Training

Customer churn prediction is formulated as a binary classification problem:

$$\hat{y} = f(x)$$

where $f$represents the classification model and $\hat{y}$is the predicted churn output.
The following machine learning algorithms were implemented in the framework:

1. Support Vector Machine (SVM)
2. Random Forest
3. XGBoost

Algorithm	Key Strength	Limitation in Churn Prediction
Support Vector Machine (SVM)	Handles high-dimensional data effectively and provides clear decision boundaries	Performance may decline on very large datasets and requires careful kernel selection
Random Forest	Achieves strong accuracy and minimizes overfitting using ensemble learning	Computational cost increases with a large number of trees
XGBoost	Provides scalable boosting with excellent predictive capability	Requires extensive parameter tuning and increases training complexity

Table 1: Comparative Analysis of Machine Learning Algorithms

Each classifier was trained using stratified k-fold cross-validation to improve robustness and reduce variance in performance estimation.

D. Hyperparameter Optimization

To improve predictive accuracy, hyperparameter tuning was conducted using cross-validation-based search methods.
The objective of optimization was to enhance generalization performance while reducing overfitting.
The optimization objective can be represented as:

$$\theta^* = \arg\max_{\theta} \text{ROC-AUC}(f_\theta)$$

ROC-AUC was selected as the primary evaluation metric because churn datasets are generally imbalanced, and ROC-AUC provides threshold-independent performance measurement.

E. Automated Model Selection

After training, all models were evaluated using multiple performance metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC.
The model achieving the highest average ROC-AUC score was automatically selected as the final classifier. This automated benchmarking mechanism removes the need for manual model selection and improves adaptability across heterogeneous datasets.

F. Deployment and Prediction Module

The selected model was deployed using a Flask-based web application to support real-time churn prediction. The system provides:

- Churn probability estimation

© 2026 IJMRSET | Volume 9, Issue 5, May 2026| DOI:10.15680/IJMRSET.2026.0905033

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Binary prediction output (Churn / No Churn)
- Business-oriented recommendations for customer retention

In addition, prediction records are stored in a database, and role-based access control mechanisms (Normal User, Premium User, and Admin) are implemented to ensure secure and reliable system operation.

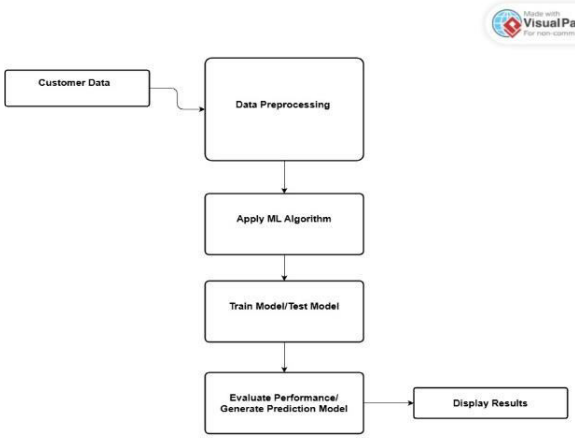


Figure 1: Architecture of the Proposed Customer Churn Prediction Framework

IV. CONCLUSION AND FUTURE WORK

This study introduced an adaptive cross-domain customer churn prediction framework that combines automated model benchmarking and intelligent model selection within a unified machine learning pipeline. The framework integrates data preprocessing, stratified cross-validation, hyperparameter tuning, and ensemble learning techniques to improve prediction accuracy and system reliability. Unlike conventional churn prediction systems that are limited to a single domain, the proposed approach supports heterogeneous datasets collected from different industries such as banking, gym membership services, and e-commerce platforms.

The experimental results showed that ensemble-based algorithms, particularly Random Forest and XGBoost, delivered strong predictive performance. Among all evaluated models, Random Forest achieved the highest cross-validated ROC-AUC score on the banking churn dataset. The adoption of stratified k-fold cross-validation improved the robustness and consistency of the evaluation process, especially for imbalanced datasets. In addition, the automated model selection strategy based on average ROC-AUC scores reduced manual effort and enhanced the adaptability of the framework across different datasets.

The selected model was successfully deployed through a Flask-based web application capable of performing real-time churn prediction. The system supports secure dataset uploads, churn probability estimation, and storage of prediction history. Furthermore, the framework generates actionable insights that can assist organizations in customer retention and business decision-making processes.

Although the proposed framework achieved effective results, certain limitations remain. The overall prediction accuracy is highly dependent on dataset quality and the relevance of selected features. Hyperparameter optimization on large-scale datasets may also require substantial computational resources. Moreover, the current implementation focuses primarily on structured tabular datasets and does not consider temporal or sequential customer behavior patterns.

Future enhancements of the framework may include:
Integration of Explainable Artificial Intelligence (XAI) methods such as SHAP and LIME to improve model interpretability and transparency.
Implementation of deep learning approaches for handling large and complex datasets.
Application of cross-domain transfer learning techniques to improve generalization capability.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 9.864| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Integration of real-time streaming data for continuous and dynamic churn prediction.
Development of hybrid ensemble architectures that combine multiple high-performing classifiers for enhanced prediction accuracy.

REFERENCES

[1] W. M. Van der Aalst, "Process modeling and analysis," in Process Mining: Data Science in Action, 2nd ed. Berlin, Germany: Springer,2016, ch. 3, pp. 55–88.
[2] S. Sakr, Z. Maamar, A. Awad, B. Benatallah, and W. M. P. van der Aalst, "Business process analytics and big data systems: A roadmap to bridge the gap," IEEE Access, vol. 6, pp. 77308–77320, 2018.
[3] Z. Tangyuan and S. Moro, "Research trends in customer churn prediction: A data mining approach," in Proc. World Conf. Inf. Syst.Technol., 2021, pp. 227–237.
[4] A. D. Caigny, K. Coussement, and K. W. D.Bock, "A new hybrid classifica-tion algorithm for customer churn prediction based on logistic regression anddecision trees," Eur. J. Oper. Res., vol. 269, pp. 760–772, Sep. 2018.
[5] N. Alboukaey, A. Joukhadar, and N. Ghneim,"Dynamic behavior based churn prediction in mobile telecom," Expert Syst. Appl., vol. 162,Dec. 2020, Art. no. 113779
[6] K. Viol, H. Scho¨ller, A. Kaiser, C. Fartacek, W. Aichhorn, and G. Schiepek, "Detecting pattern transitions in psychological time series—A validation study on the pattern transition detection algorithm (PTDA)," PLoS ONE, vol. 17, no. 3, Mar. 2022, Art. no. e0265335.
[7] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn predic- tion system: A machine learning approach," Computing, vol. 104, no. 2, pp. 271–294, Feb. 2022.
[8] G. Mohammadi, R. Tavakkoli-Moghaddam,and M. Mohammadi, "Hierar- chical neural regression models for customer churnprediction," J. Eng., vol. 2013, pp. 1–9, Feb.2013.
[9] Z. Chen, S. Zhang, S. McClean, B. Allan, and I.Kegel, "Sequence mining TV viewing data using embedded Markov modelling," in Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv.Trusted Comput., Scalable Comput. Com mun.,Internet People Smart City Innov. (SmartWorld/ SCALCOM/ UIC/ATC/IOP/SCI),Oct. 2021, pp. 665–670

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com