https://doi.org/10.62836/jcmea.v1i1.010102

Global Science Publishing

JCMEA

Journal of Computational Methods in Engineering Applications

https://ojs.sgsci.org/journals/jcmea

Article

Domain Adaptation-Based Machine Learning Framework for Customer Churn Prediction Across Varing Distributions

Diwei Zhu [1], Yunxiang Gan [2] and Xiaoyang Chen [3,*]

[1] *New York University, New York City, USA.*

[2] *Moloco, CA, USA.*

[3] *Radiawave Co., Ltd., Shen Zhen, China*

Abstract: In today's fiercely competitive business environment, the ability to accurately predict customer churn is essential for enhancing customer retention and reducing financial losses. Traditional statistical approaches, although beneficial, often struggle to perform effectively across diverse customer data domains due to variability in data distributions. This research introduces a refined method for predicting customer churn that utilizes domain adaptation techniques to overcome these challenges. Specifically, it employs the Correlation Alignment (CORAL) method to synchronize the feature distributions between the source and target datasets, significantly improving the logistic regression model's capacity to apply insights across various customer segments. The process involves segmenting the customer data into clearly defined clusters using the k-means algorithm, which helps pinpoint and adjust distributional discrepancies, thus boosting the model's accuracy. Early results indicate that incorporating domain adaptation not only bolsters the model's applicability across different domains but also drastically minimizes the covariance differences—from a substantial initial gap to nearly zero. This strategic approach demonstrates substantial potential to revolutionize how businesses anticipate and manage customer behavior, providing a more adaptable and effective framework for addressing the complex challenges of customer churn.

Keywords: customer churn predicition; business analytics; domain adaptation

1. Introduction

In the competitive landscape of modern business, customer retention is paramount for sustained success. A phenomenon often encountered by businesses across various sectors is customer churn, which refers to the loss of clients or customers who decide to stop using a service or product [1–3]. Understanding and predicting customer churn not only helps in retaining valuable customers but also in reducing the associated financial impact. The stakes are high; acquiring a new customer can be more expensive than retaining an existing one. Therefore, accurately predicting which customers are likely to churn becomes a critical strategic element.

As businesses accumulate vast amounts of data through customer interactions, traditional statistical methods for analyzing churn have been gradually overtaken by more sophisticated techniques. The advent of machine learning models has transformed the analytics landscape, offering new avenues for not only understanding but

Received: 15 September 2021; Accepted: 12 October 2021.

* Corresponding: Xiaoyang Chen (chenxiaoyang@radiawave.com)

also anticipating customer behaviors [4,5]. Machine learning models, such as decision trees, neural networks, and support vector machines, have been widely adopted due to their ability to learn complex patterns from data without explicit programming. For instance, Du et al. proposed an image recommendation algorithm based on deep neural networks for social networks, which addresses the inefficiencies of traditional text-based methods. This study sorts user interactions on social networks by timestamp and utilizes traditional feature algorithms (LBP, BGC3, RTU) along with CNN to create feature vectors [5]. Wang et al. developed a machine learning model to predict the mental health of medical workers, analyzing 32 factors from questionnaire responses of 5108 Chinese medical workers. The Self-reporting Inventory results were used to assess mental health status. The novel prediction model employs an optimization algorithm and neural network to identify and prioritize key factors impacting mental health [6].

Despite their effectiveness, these models often face challenges, particularly when dealing with heterogeneous datasets that include customers from different domains or demographics, exhibiting varied behaviors and characteristics. Such distribution differences can severely hamper the model's ability to generalize well across unseen data, leading to poor predictive performance. This is where the concept of domain adaptation becomes crucial. Domain adaptation techniques aim to bridge the gap between the source domain (data on which the model is trained) and the target domain (data on which the model is tested), thus enhancing the model's generalizability and robustness. Domain adaptation, a subfield of transfer learning [7,8], addresses the issue where the training data (source domain) and the testing data (target domain) have different distributions. Methods like Covariate Shift Adaptation [9], where the feature distribution differences are adjusted, and Concept Drift Adaptation [10], where changes in the data distribution over time are considered, are common approaches. However, another emerging technique, which is the focal point of this article, is the use of structural adaptations to align data distributions directly.

In this study, we propose a novel domain adaptation method tailored for customer churn prediction shown in Figure 1. Our approach begins with the segmentation of the dataset into distinct clusters with different distribution characteristics using the k-means clustering algorithm. This step helps in identifying inherent distributional differences within the data, which might correspond to various customer demographics or product usage types. Following the identification of these clusters, we employ the Correlation Alignment (CORAL) method to align the source domain distribution with the distributions of other clusters. CORAL minimizes domain shift by aligning the second-order statistics (covariances) of the source and target distributions [11], thereby making the features more comparable and transferrable. Finally, we test the efficacy of our domain adaptation strategy by employing the machine learning model to predict customer churn. We compare the performance of the model with and without domain adaptation techniques. The preliminary results indicate that incorporating domain adaptation significantly enhances the model's generalizability across different customer segments, thus proving the potential of domain adaptation in tackling the challenges posed by distribution differences in customer churn prediction.

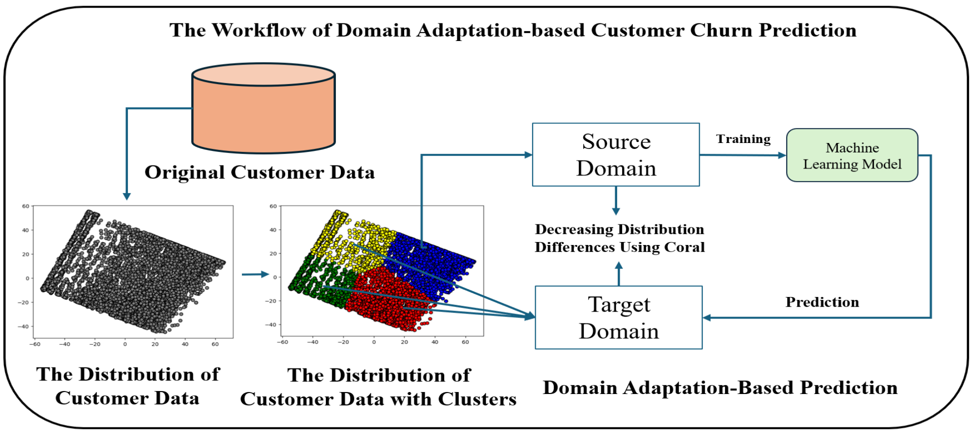


Figure 1. The workflow of the proposed domain adaptation-based method used for customer churn prediction.

2. Literature Review

2.1. Customer Churn Prediction

In customer churn prediction, two established algorithms—decision trees (DT) and logistic regression (LR)—are commonly employed due to their predictive accuracy and clarity [12]. Despite their popularity, these methods have notable drawbacks: decision trees struggle with linear relationships among variables, and logistic regression is less effective in managing interactions among variables. To overcome these challenges, the Logit Leaf Model (LLM) has been developed. This innovative algorithm enhances classification by applying models to data segments rather than the entire dataset, aiming to boost predictive performance while preserving the interpretability of the models. LLM involves a dual-phase approach: it first identifies distinct customer segments and then constructs a specialized model for each segment. Through testing and case studies, LLM has demonstrated several advantages over traditional decision trees and logistic regression in terms of both performance and usability [13].

Predicting customer churn is a key focus within data mining, offering a more effective approach than traditional survey methods [1]. Traditional surveys often face challenges such as high costs and restricted customer access. In contrast, data mining addresses these issues by drawing conclusions from the analysis of historical data. As a result, data mining has emerged as the predominant method for customer retention. It enables businesses to forecast customer churn and discern patterns based on historical customer data [14].

Various approaches have been employed to predict customer churn in the telecom industry, with most relying on data mining and machine learning techniques. While some studies have focused on a single data mining method to derive insights, others have compared multiple methods to enhance churn prediction accuracy [15]. For instance, Brandusoiu et al. introduced a contemporary data mining approach to predict prepaid customer churn, utilizing a dataset of 3333 customers with 21 features, and a binary churn variable indicating Yes/No [16]. Key features in this dataset include metrics such as the number of messages and voicemail usage. The study employed principal component analysis (PCA) to reduce the data dimensions, followed by tree-based machine learning algorithms like Bayes Networks and Neural Networks to predict churn. The model performance was evaluated using the AUC metric, with Bayes Networks achieving an AUC of 99.10% and Neural Networks reaching 99.55%. Notably, the dataset used was relatively small and contained no missing values.

Makhtar et al. developed a model for predicting telecom customer churn using rough set theory, noting that it outperformed traditional algorithms like Decision Trees (DT) and Linear Regression (LR) in terms of predictive accuracy [17]. However, while many models prioritize accuracy, few delve into the intuitiveness and understandability of the systems used to discern the underlying reasons for customer churn [18]. Addressing this gap, Idris et al. introduced an innovative churn prediction approach that utilizes the robust search capabilities of genetic programming (GP), enhanced by AdaBoost [19]. This method is particularly effective in identifying the specific factors that contribute to customer churn behavior in the telecom sector.

2.2. Domain Adaptation

Pan et al. have classified transfer learning into three primary categories based on variations in domains and tasks: inductive, transductive, and unsupervised transfer learning [20]. Inductive transfer learning is characterized by differing source and target tasks, regardless of whether the domains are the same. This form of transfer learning may involve annotated data in the source domain, but necessitates at least some labeled data in the target domain for training purposes. Transductive transfer learning, in contrast, maintains the same tasks across different domains, with labeled data only available in the source domain. During training, it also requires access to a subset of the target domain's unlabeled data to estimate its marginal probability distribution. Lastly, unsupervised transfer learning occurs when both tasks and domains differ, similar to inductive, but neither the source nor target domains provide labeled data, relying entirely on unlabeled datasets.

In terms of techniques, domain adaptation is generally divided into two main categories based on their underlying architectures: shallow and deep. Shallow domain adaptation methods, referenced in studies [21–23], primarily employ instance-based and feature-based strategies to align domain distributions. A common strategy within these approaches involves minimizing the distance between the domains using metrics such as maximum mean discrepancy (MMD) [24], Wasserstein metric, correlation alignment (CORAL) [11], Kullback-Leibler (KL) divergence [25], and contrastive domain discrepancy (CDD) [26]. In contrast, deep domain adaptation methods [27 – 29] leverage neural networks, often utilizing convolutional, autoencoder, or adversarial architectures to reduce the domain gap. These methods may also incorporate a distance metric at one or several layers of dual-network architectures, where one network processes source data and the other handles target data, to quantify discrepancies in feature representations across corresponding layers.

3. Method

3.1. Dataset Descrption and Preprocessing

The second dataset used in this study is also sourced from Kaggle and contains 7043 rows (observations) and 21 features (columns) [30]. This dataset provides detailed customer information related to their demographic attributes, services they subscribe to, and usage metrics. Some notable features include: (1) customer ID: Unique identifier for each customer. (2) gender: The gender of the customer (Male/Female). (3) Senior Citizen: Indicates whether the customer is a senior citizen (binary: 0 for No, 1 for Yes). (4) tenure: The number of months the customer has been with the company. (5) Monthly Charges: The monthly charges the customer incurs. (6) Total Charges: Total amount charged to the customer. The target variable for this dataset is Churn, which represents whether the customer has churned (Yes/No). This dataset offers more granular details compared to the previous one, and it includes both service usage data and customer demographic information, which can help in understanding customer behavior and predicting churn more effectively. The distribution of some features and the target variable are plotted in Figure 2.

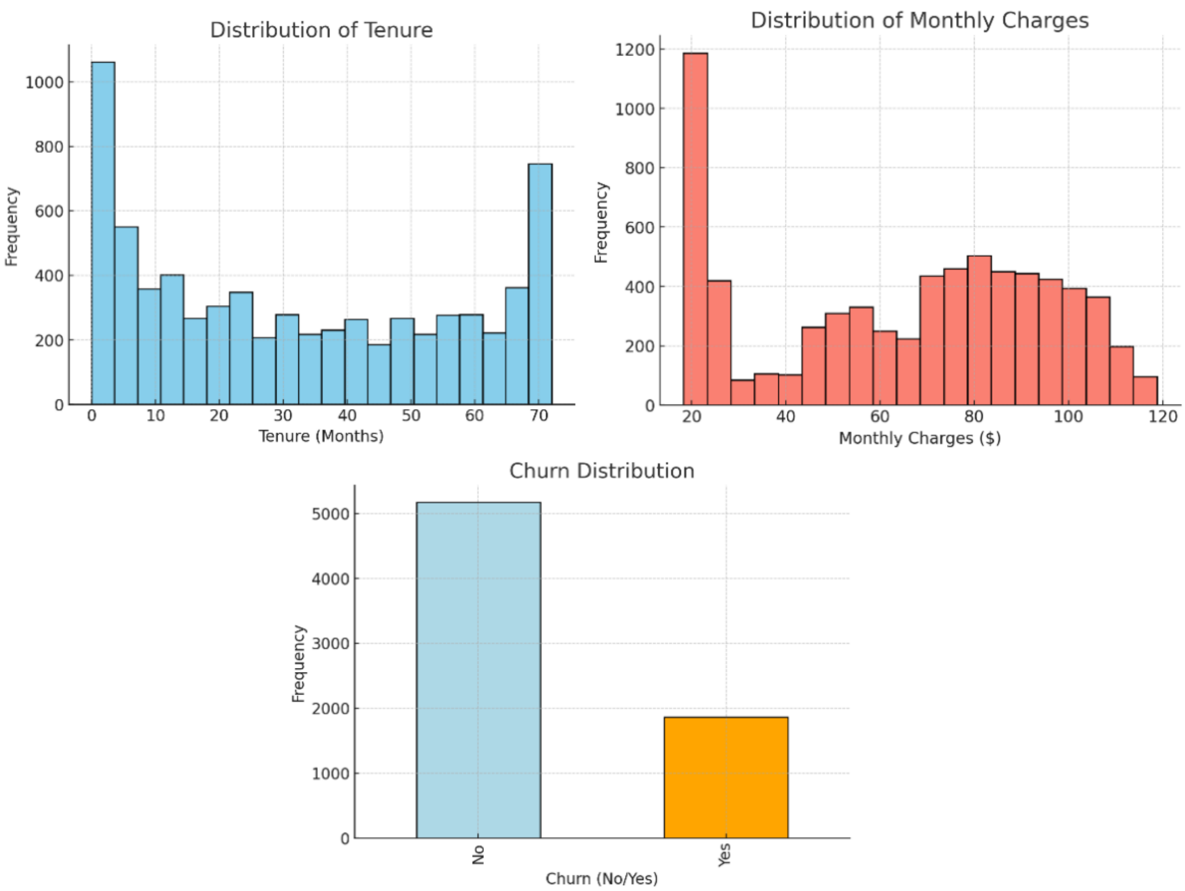


Figure 2. The distribution of some features and the target variable.

In the data preprocessing stage of this study, we first transformed the categorical features and the target variable into numerical values using one-hot encoding [31,32]. In terms of the target variable, ‘0’ represents no churn and ‘1’ represents churn. We then applied min-max normalization [33,34] to the features to ensure that all values fall within the same range, which enhances the performance and convergence of our machine learning model. To create a dataset with varying distributions of customer data for implementing domain adaptation algorithms, we used the K-means algorithm [35–37] to partition the data into clusters. K-means is a widely used unsupervised learning algorithm that divides data points into k clusters by minimizing the distance between points and their respective cluster centroids. To determine the optimal number of clusters (k), we employed the Silhouette score [38, 39], a metric that evaluates how well each data point fits within its assigned cluster compared to other clusters. A higher Silhouette score indicates a more clearly defined clustering. Our analysis found that the optimal clustering was achieved at k = 4. We designated the largest cluster, represented by blue points, as the source domain, and the smaller clusters, represented by yellow, red, and green points, as the target domains. Figures 3–5 in our study illustrate the PCA distribution [40] of the original data, the Silhouette score curve, and the PCA distribution after clustering, respectively.

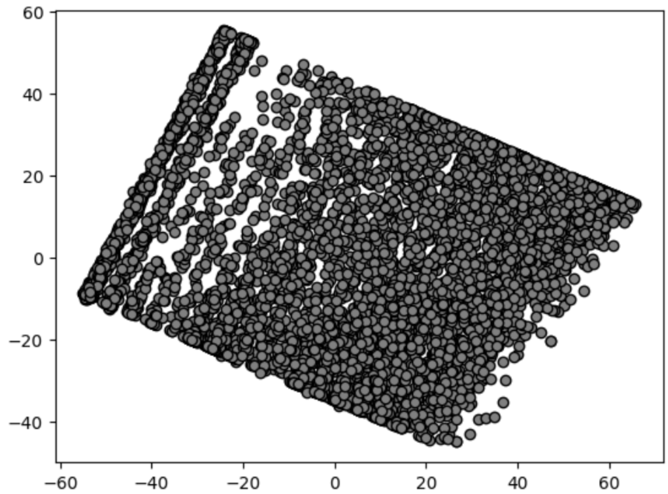


Figure 3. The PCA distribution of the original customer data.

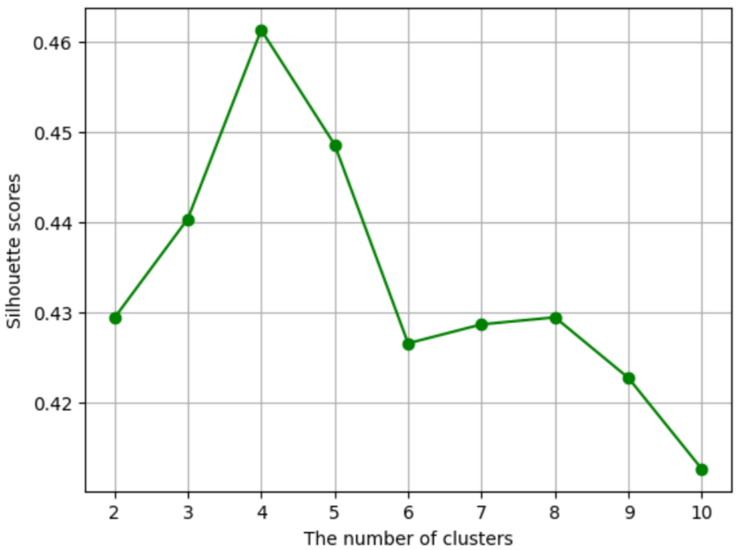


Figure 4. The curve of Silhouette scores used for determining k value.

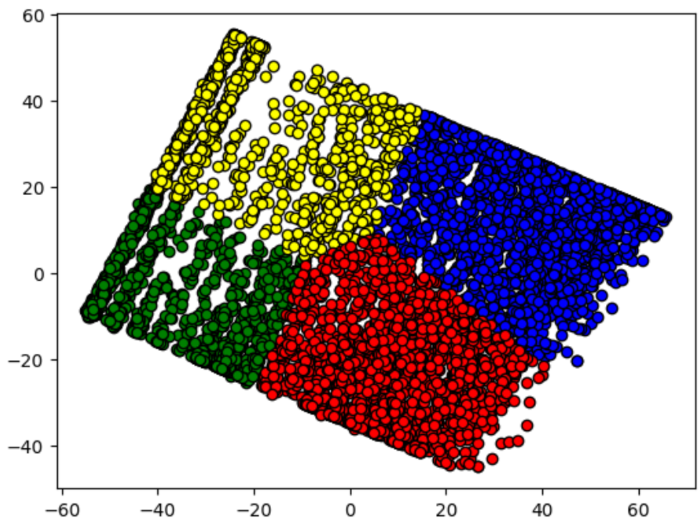


Figure 5. The PCA distribution of the data with identified clusters.

3.2. Domain Adaptation-Based Machine Learning Model

3.2.1. Logistic Regression

Logistic regression [41,42] was firstly considered in this study due to its wide application in the customer churn prediction. Logistic regression is a widely used statistical model that is particularly suited for binary classification problems, where the outcome variable is categorical with two possible values. In contrast to linear regression, which is used for predicting continuous outcomes, logistic regression models the probability that a given input point belongs to a particular class. This technique is commonly applied in fields such as finance, healthcare, and social sciences, where understanding the likelihood of an event occurring is essential. At the core of logistic regression is the logistic function, also known as the sigmoid function. This function takes any real-valued number and maps it to a value between 0 and 1, representing the probability of the dependent variable belonging to one of the two categories [43–49]. The logistic function is defined as:

$$sigmoid(z) = \frac{1}{1+e^{-z}} \tag{1}$$

where z is the linear combination of the input features. This transformation ensures that the model's predictions remain bounded between 0 and 1, making it ideal for probability estimation.

The logistic regression model predicts the probability of the target variable based on the input features. For instance, based on the dataset containing customer behavior information, logistic regression can help us estimate the probability that a customer will churn (leave the service) based on features like usage patterns, payment history, and customer demographics. The model computes the probability of an event happening, and a threshold (commonly 0.5) is applied to classify the event into one of the two categories.

Logistic regression is favored for its simplicity, ease of interpretation, and effectiveness in binary classification tasks. It provides insights into the relationship between features and the probability of an outcome, making it a powerful tool for making informed decisions. Despite being a linear model, logistic regression can be extended to more complex problems by introducing non-linear relationships through feature engineering or by combining it with other techniques such as polynomial regression. Its ability to handle categorical data, probabilistic outputs, and its relatively low computational cost make logistic regression a staple in the machine learning toolkit.

3.2.2. Domain Adaptation

CORAL, short for Correlation Alignment, is a domain adaptation technique used in machine learning, particularly in situations where the source domain (the data the model was trained on) and the target domain (the data on which the model will be tested or deployed) have different data distributions. The core idea behind CORAL is to align the feature distributions of the source and target domains by adjusting their correlations, allowing a model trained on the source domain to generalize better to the target domain.

The goal of CORAL is to minimize the discrepancy between the source and target domains by aligning their covariance matrices. Here's how the CORAL method works in a typical machine learning pipeline: (1) Covariance Calculation: First, CORAL computes the covariance matrices of the source and target domain feature representations. The covariance matrix captures how the different features in the dataset are related to one another. If the source and target domains have different feature correlations, the model's performance could suffer when applied to the target domain. (2) Whitening the Source Domain: CORAL applies a transformation to the source domain data to "whiten" it. Whitening is a process that removes any correlations between features by transforming the covariance matrix of the source domain to the identity matrix. (3) Recoloring: After whitening the source domain, CORAL transforms the source features so that their covariance matrix matches that of the target domain. This step ensures that the feature correlations in the source domain resemble those of the target domain, facilitating better alignment between the two domains. (4) Model Training: Once the source features have been transformed to align with the target domain's distribution, the model is trained on the aligned source data and applied to the target data.

Mathematically, CORAL minimizes the difference between the covariance matrices of the source and target domains using the following objective function:

$$L_{Coral} = \frac{1}{4d^2}|C_S - C_T|_F^2 \tag{2}$$

where, C_S and C_T are the covariance matrices of the source and target domain features, respectively. d is the dimensionality of the feature space. $|C_S - C_T|_F$ represents the Frobenius norm. The goal of this function is to reduce the difference between the covariance matrices, effectively aligning the two domains.

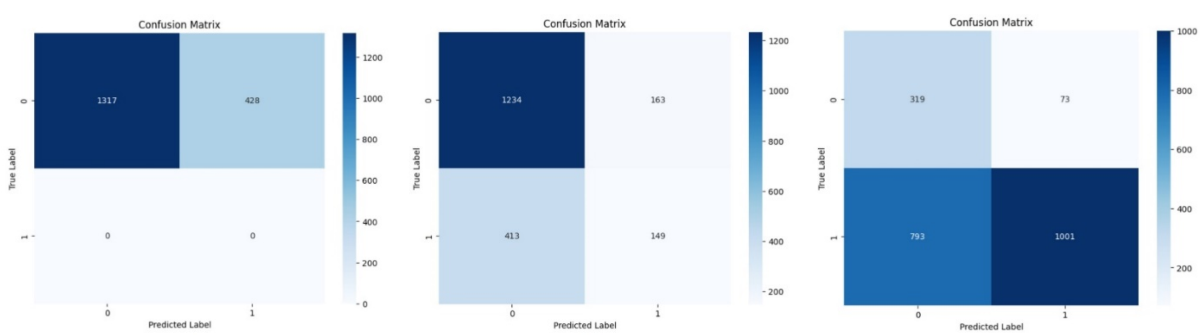
4. Results and Discussion

4.1. The Performance of the Model

Tables 1 and 2, Figures 6 and 7 provided showcase the performance of logistic regression in both the original prediction scenario and when domain adaptation techniques were applied across three target domains of customer data (green, yellow, and red). The metrics evaluated include Accuracy, F1 Score, Precision, and Recall, followed by confusion matrices that offer a detailed view of the classification outcomes for each domain.

Table 1. Original prediction performance using the logistic regression.

Metrics	Target Domain-1 (Green)	Target Domain-2 (Yellow)	Target Domain-3 (Red)
Accuracy	0.754	0.705	0.603
F1 score	0.430	0.575	0.561
Preicision	0.500	0.613	0.609
Recall	0.377	0.574	0.685



(a) **(b)** **(c)**

Figure 6. Original confusion matrix using the logistic regression.

Table 2. Domain adaptation-based prediction performance using the logistic regression.

Metrics	Target Domain-1 (Green)	Target Domain-2 (Yellow)	Target Domain-3 (Red)
Accuracy	0.755	0.761	0.627
F1 score	0.449	0.602	0.604
Preicision	0.507	0.616	0.631
Recall	0.628	0.594	0.674

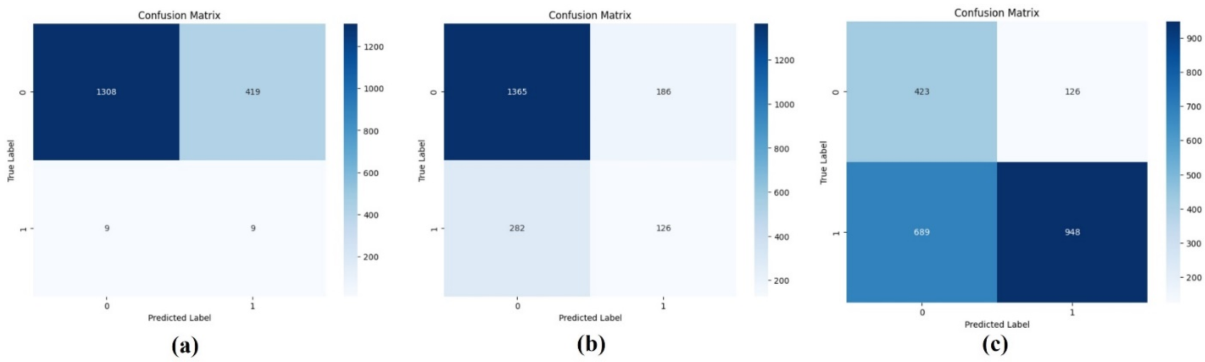


Figure 7. Domain-adaptation-based confusion matrix using the logistic regression.

The original prediction performance of the logistic regression model across the three target domains varies significantly. Starting with Target domain-1 (green), the model achieved an accuracy of 0.754, which suggests that the model performs relatively well in classifying data points in this domain. However, the F1 score of 0.430 and recall of 0.377 indicate that the model struggles with identifying the positive class (likely the churn class, given the application), implying a significant number of false negatives. The precision of 0.500 reveals that, when the model does predict the positive class, about half of those predictions are correct. In Target domain-2 (yellow), the accuracy is 0.705, slightly lower than in the first domain, but the recall of 0.574 shows an improvement, meaning the model is better at capturing true positives. However, the F1 score of 0.575 and precision of 0.613 reflect a better balance between precision and recall, which suggests improved prediction performance compared to Target domain-1. Target domain-3 (red), shows a lower accuracy of 0.603, the lowest among the three domains, signaling that the model struggles the most in this domain. However, the recall is the highest at 0.685, indicating the model's strength in identifying true positives. The F1 score of 0.561 and precision of 0.609 further confirm that while the model captures a larger number of true positives, it may also predict a fair number of false positives, leading to a relatively lower precision.

The confusion matrices in Figure 6a–c visually depict these results. In Figure 6a, for Target domain-1, the model predicts a high number of false negatives (428) compared to the true positives. In Figure 6b, Target domain-2 sees an improvement with a lower number of false negatives (413), while Figure 6c, for Target domain-3, shows a more balanced outcome, although there are still a considerable number of false negatives (793). Overall, the model's performance shows room for improvement, particularly in classifying the positive class in the three target domains.

The results after applying domain adaptation demonstrate improved performance across all metrics. In Target domain-1 (green), the accuracy remains relatively stable at 0.755, but there is a noticeable improvement in the recall, which increases to 0.628. This suggests that the model, after domain adaptation, has become much better at identifying the positive class (churn). The F1 score also increases slightly to 0.449, while precision remains similar at 0.507. These metrics indicate a better balance between precision and recall after domain adaptation, especially in minimizing false negatives. In Target domain-2 (yellow), the accuracy improves to 0.761, marking the highest accuracy among the three domains. The F1 score also increases to 0.602, and recall improves to 0.594, suggesting a more effective classification of the positive class. Precision remains consistent at 0.616, confirming that the model's improved recall did not come at the cost of a significant decrease in precision. Target domain-3 (red) sees a notable improvement in accuracy, increasing to 0.627. This is a substantial improvement over the original prediction results. The F1 score of 0.604 reflects a better balance between precision and recall, and the recall, while slightly lower than in the original model at 0.674, still demonstrates strong performance in identifying positive cases. Precision improves to 0.631, indicating fewer false positives and overall better classification of the positive class. The confusion matrices in Figure 7a–c show how domain adaptation improves prediction outcomes. In Figure 7a, for Target domain-1, the number of false negatives is reduced from 428 to 419, resulting in a more balanced classification. In Figure 7b, for Target domain-2, the false negatives drop from 413 to 282, further confirming the improved recall. In Figure 7c, for Target domain-3, the model significantly reduces false negatives from 793 to 689, marking a meaningful improvement in its ability to classify positive cases after domain adaptation.

It is evident that domain adaptation improves the model's ability to generalize across different target domains shown in Figure 8. Across all three domains, the recall improves after domain adaptation, which is critical for applications where correctly identifying the positive class (churn) is of high importance. The F1 scores also increase across all domains, indicating better overall performance. Accuracy remains relatively stable across the different domains, which suggests that while domain adaptation helps improve the model's performance on the positive class, it does not sacrifice the model's ability to correctly classify the negative class. Precision also sees some improvements, particularly in Target domain-3, showing that domain adaptation has helped balance the trade-off between precision and recall.

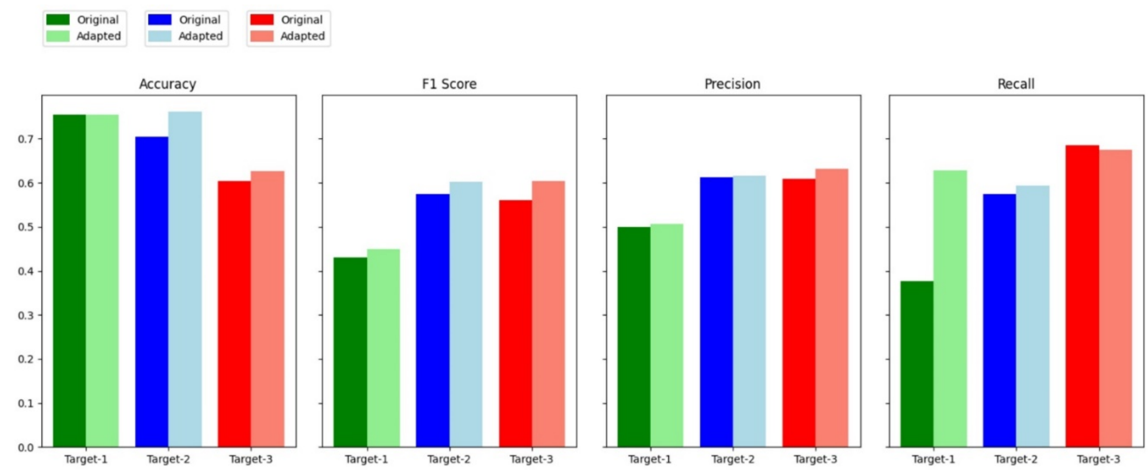


Figure 8. Comparison of Original and Domain Adaptation Performance Across Metrics.

The results depicted in Table 3 show a significant reduction in the covariance difference when comparing the previous to the new measurements after applying the CORAL technique. Specifically, the previous covariance difference was at a relatively high value of 48.453608, indicating a considerable disparity between the source and target domain data distributions. After the application of CORAL, this difference has been dramatically reduced to essentially zero (0.000002). This drastic reduction in covariance difference signifies that CORAL has effectively minimized the statistical discrepancies between the source and target domains by aligning their feature distributions. Such alignment is crucial for improving the performance of machine learning models when they are applied to new, unseen domains by ensuring that the model trained on the source domain can generalize effectively to the target domain. The near-zero new covariance difference suggests that the domains are now well-aligned, thereby enhancing the model's ability to make accurate predictions on the target domain data.

Table 3. The comparison of previous and new covariance difference comparison.

Previous Covariance Difference	New Covariance Difference
48.453608	0.000002

4.2. Discussion

While the results from the domain adaptation-based logistic regression model's application are promising, several limitations are identified: (1) Model Generalization: Despite observed improvements, the variation in performance metrics prior to adaptation suggests potential overfitting to specific domain characteristics. To counter this, future work could explore more robust regularization techniques to enhance model generalization capabilities. (2) Dependency on Domain Characteristics: The current reliance on domain adaptation for performance enhancement may indicate a dependency on specific domain traits. Further investigation into more intrinsic model resilience traits, such as feature selection and extraction, could help mitigate this dependency. (3) Scalability to Other Domains: The focus of the study on three specific domains raises questions about the adaptability and scalability of the proposed methods. Extending this approach to other domains or more varied data could help validate the adaptability and scalability of the methods used. In the future studies, several prospects can be considered: (1) Advanced Domain Adaptation Techniques: There is potential for exploring advanced techniques, such as Deep Correlation Alignment (Deep CORAL) and adversarial-based methods. These could provide deeper insights and potentially better alignment strategies. (2) Automated Feature Engineering: Leveraging machine learning to automate feature engineering could significantly enhance model adaptability and efficiency without extensive manual intervention. (3) Cross-Domain Model Validation: Implementing cross-validation techniques across more diverse domains could further substantiate the robustness of the domain adaptation strategies, ensuring the model's efficacy across various scenarios.

5. Conclusion

The implementation of a domain adaptation strategy in churn prediction, as explored in this study, indicates a substantial improvement in model performance across diverse customer datasets. The application of the CORAL method successfully reduced the covariance difference from a substantial disparity to nearly zero, underscoring its efficacy in aligning the source and target domain feature distributions. This alignment is pivotal for the logistic regression model to perform consistently and accurately across different domains, addressing the challenge of heterogeneous datasets that traditional models often struggle with. Despite the promising results, the model exhibits potential overfitting to specific domain characteristics, suggesting a need for more robust regularization techniques to enhance model generalization further. Additionally, while the study focused on three specific domains, expanding this approach to other domains or more varied data could provide deeper insights into its adaptability and scalability. Future research could explore advanced domain adaptation techniques, automate feature engineering, and implement cross-domain model validation to bolster the robustness of the predictive model. This study not only demonstrates the potential of domain adaptation in tackling the challenges posed by distribution differences in customer churn prediction but also sets the stage for further innovations in this critical area [50–57].

Funding

Not applicable.

Author Contributions

Conceptualization, D.Z., Y.G. and X.C.; writing—original draft preparation, D.Z., Y.G. and X.C. All of the authors read and agreed to the published the final manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Huang B, Kechadi MT, Buckley B. Customer Churn Prediction in Telecommunications. *Expert Systems with Applications* 2012; **39**(**1**): 1414–1425.

2 Çelik O, Osmanoglu UO. Comparing to Techniques Used in Customer Churn Analysis. *Journal of Multidisciplinary Developments* 2019; **4**(**1**): 30–38.

3 Qiu Y, Chen P, Lin Z, *et al*. Clustering Analysis for Silent Telecom Customers Based on k-Means++. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 1023–1027.

4 Xiong S, Yu L, Shen H, *et al*. Efficient Algorithms for Sensor Deployment and Routing in sensor Networks for Network-Structured Environment Monitoring. In Proceedings of the 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012; pp. 1008–1016.

5 Du S, Chen Z, Wu H, *et al.* Image Recommendation Algorithm Combined with Deep Neural Network Designed for Social Networks. *Complexity* 2021; **2021**(**1**): 5196190.

6 Wang X, Li H, Sun C, *et al.* Prediction of Mental Health in Medical Workers during COVID-19 Based on Machine Learning. *Frontiers in Public Health* 2021; **9**: 697850.

7 Farahani A, Voghoei S, Rasheed K, *et al*. A Brief Review of Domain Adaptation. In *Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020*; Springer: Cham, Switzerland, 2020; pp. 877–894.

8 Guan H, Liu M. Domain Adaptation for Medical Image Analysis: A Survey. *IEEE Transactions on Biomedical Engineering* 2021; **69**(**3**): 1173–1185.

9 Schneider S, Rusak E, Eck L, *et al*. Improving Robustness against Common Corruptions by Covariate Shift Adaptation. *Advances in Neural Information Processing Systems* 2020; **33**: 11539–11551.

10 Mehmood H, Kostakos P, Cortes M, *et al*. Concept Drift Adaptation Techniques in Distributed Environment for Real-World Data Streams. *Smart Cities* 2021; **4**(**1**): 349–371.

11 Sun B, Saenko K. Deep Coral: Correlation Alignment for Deep Domain Adaptation. In *Computer Vision–ECCV 2016 Workshops*; Springer International Publishing: Amsterdam, The Netherlands, 2016; pp. 443–450.

12 Verbeke W, Dejaeger K, Martens D, *et al.* New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach. *European Journal of Operational Research* 2012; **218**(**1**): 211–229.

13 De Caigny A, Coussement K, De Bock KW. A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees. *European Journal of Operational Research* 2018; **269**(**2**): 760–772.

14 Liu DS, Fan SJ. A Modified Decision Tree Algorithm Based on Genetic Algorithm for Mobile User Classification Problem. *The Scientific World Journal* 2014; **2014**(**1**): 468324.

15 Ahmad AK, Jafar A, Aljoumaa K. Customer Churn Prediction in Telecom Using Machine Learning in Big

Data Platform. *Journal of Big Data* 2019; **6**(**1**): 1–24.

16 Brânduşoiu I, Toderean G, Beleiu H. Methods for Churn Prediction in the Pre-Paid Mobile Telecommunications Industry. In Proceedings of the 2016 International Conference on Communications (COMM), Bucharest, Romania, 9–10 June 2016; pp. 97–100.

17 Makhtar M, Nafis S, Mohamed MA, *et al.* Churn Classification Model for Local Telecommunication Company Based on Rough Set Theory. *Journal of Fundamental and Applied Sciences* 2017; **9**(**6**): 854–868.

18 Bock K. W, Poel D. Reconciling Performance and Interpretability in Customer Churn Prediction Using Ensemble Learning Based on Generalized Additive Models. *Expert Systems with Applications* 2012; **39**: 6816–6826.

19 Idris A, Iftikhar A, Rehman Z.U. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO Undersampling. *Cluster Computing* 2017; **22**: 7241–7255.

20 Pan S.J, Yang Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 2010; **22**(**10**): 1345–1359.

21 Gopalan R, Li R, Chellappa R. Domain Adaptation for Object Recognition: An Unsupervised Approach. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 999–1006.

22 Jhuo I. H, Liu D, Lee D. T, *et al*. Robust Visual Domain Adaptation with Low-Rank Reconstruction. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2168–2175.

23 Pan SJ, Tsang I. W, Kwok JT, *et al*. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks* 2010; **22**(**2**): 199–210.

24 Gretton A, Borgwardt K, Rasch M, *et al*. A Kernel Method for the Two-Sample-Problem. *Advances in Neural Information Processing Systems 2006;* **19**.

25 Kullback S, Leibler RA. On Information and Sufficiency. *The Annals of Mathematical Statistics* 1951; **22**(**1**): 79–86.

26 Kang G, Jiang L, Yang Y, *et al*. Contrastive Adaptation Network for Unsupervised Domain Adaptation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4893–4902.

27 Ajakan H, Germain P, Larochelle H, *et al*. Domain-Adversarial Neural Networks. *arXiv* 2014, arXiv:1412. 4446.

28 Ganin Y, Lempitsky V. Unsupervised Domain Adaptation by Backpropagation. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 1180–1189.

29 Glorot X, Bordes A, Bengio Y. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Bellevue, WA, USA, 28 June–2 July 2011; pp. 513–520.

30 Kaggle. Telco Customer Churn, 2017. Available online: https://www. kaggle. com/datasets/blastchar/telco-customer-churn (accessed on 7 October 2024).

31 Rodríguez P, Bautista MA, Gonzalez J, *et al.* Beyond one-hot encoding: Lower Dimensional Target Embedding. *Image and Vision Computing* 2018; **75**: 21–31.

32 Okada S, Ohzeki M, Taguchi S. Efficient Partition of Integer Optimization Problems with One-Hot Encoding. *Scientific Reports* 2019; **9**(**1**): 13036.

33 Patro S. Normalization: A Preprocessing Stage. *arXiv* 2015, arXiv:1503.06462.

34 Gajera V, Gupta R, Jana PK. An Effective Multi-Objective Task Scheduling Algorithm Using Min-Max Normalization in Cloud Computing. In Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Bangalore, India, 21–23 July 2016; pp. 812–816.

35 Ahmed M, Seraj R, Islam SMS. The k-Means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics* 2020; **9**(**8**): 1295.

36 Sinaga KP, Yang MS. Unsupervised k-Means Clustering Algorithm. *IEEE Access* 2020; **8**: 80716–80727.

37 Bock HH. Clustering Methods: A History of k-Means Algorithms. In *Selected Contributions in Data Analysis and Classification*; Springer, Berlin/Heidelberg, Germany, 2007; pp. 161–172.

38 Shahapure K.R, Nicholas C. Cluster Quality Analysis Using Silhouette Score. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, NSW, Australia, 6–9 October 2020; pp. 747–748.

39 Ogbuabor G, Ugwoke FN. Clustering Algorithm for a Healthcare Dataset Using Silhouette Score Value. *International Journal of Computer Science & Information Technology* 2018; **10**(**2**): 27–37.

40 Maćkiewicz A, Ratajczak W. Principal Components Analysis (PCA). *Computers & Geosciences* 1993; **19**(**3**): 303–342.

41 LaValley MP. Logistic Regression. *Circulation* 2008; **117**(**18**): 2395–2399.

42 Hosmer DW, Jr., Lemeshow S, Sturdivant RX. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013.

43 Wang Z, Shafieezadeh A. Metamodel-Based Subset Simulation Adaptable to Target Computational Capacities: The Case for High-Dimensional and Rare Event Reliability Analysis. *Structural and Multidisciplinary Optimization* 2021; **64**: 649–675.

44 Zhang C, Wang Z, Shafieezadeh A. Error Quantification and Control for Adaptive Kriging-Based Reliability Updating with Equality Information. *Reliability Engineering & System Safety* 2021; **207**: 107323.

45 Wang Z, Shafieezadeh A. Highly Efficient Bayesian Updating Using Metamodels: An Adaptive Kriging-Based Approach. *Structural Safety* 2020; **84**: 101915.

46 Wang Z, Shafieezadeh A. On Confidence Intervals for Failure Probability Estimates in Kriging-based Reliability Analysis. *Reliability Engineering & System Safety* 2020; **196**: 106758.

47 Zhang C, Wang Z, Shafieezadeh A. Value of Information Analysis via Active Learning and Knowledge Sharing in Error-Controlled Adaptive Kriging. *IEEE Access* 2020; **8**: 51021–51034.

48 Rahimi M, Wang Z, Shafieezadeh A, *et al*. Exploring Passive and Active Metamodeling-Based Reliability Analysis Methods for Soil Slopes: A New Approach to Active Training. *International Journal of Geomechanics* 2020; **20**: 04020009.

49 Wang Z, Shafieezadeh A. Real-Time High-Fidelity Reliability Updating with Equality Information Using Adaptive Kriging. *Reliability Engineering & System Safety* 2020; **195**: 106735.

50 Mohammadi Darestani Y, Wang Z, Shafieezadeh A. Wind Reliability of Transmission Line Models Using Kriging-Based Methods. In Proceedings of the 3th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP13, Seoul, Korea, 26–30 May 2019.

51 Hur J, Wang Z, Shafieezadeh, A, *et al*. Seismic Reliability Analysis of NPP's Nonstructural Components Using Surrogate Models. Available online: https://s-space.snu.ac.kr/handle/10371/153449 (accessed on 7 October 2024).

52 Wang Z, Shafieezadeh A. A Parallel Learning Strategy for Adaptive Kriging-Based Reliability Analysis Methods. In Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13), Seoul, Korea, 26–30 May 2019; pp. 26–30.

53 Wang Z, Shafieezadeh, A. Reliability-Based Bayesian Updating Using Machine Learning. In Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13), Seoul, Korea, 26–30 May 2019; pp. 26–30.

54 Wang Z, Shafieezadeh A. ESC: An Efficient Error-Based Stopping Criterion for Kriging-Based Reliability Analysis Methods. *Structural and Multidisciplinary Optimization* 2019; **59**: 1621–1637. https://doi.org/10.1007/s00158-018-2150-9.

55 Rahimi M, Wang Z, Shafieezadeh A, *et al*. An Adaptive Kriging-Based Approach with weakly Stationary Random Fields for Soil Slope Reliability Analysis. In *Geo-Congress 2019: Soil Erosion, Underground Engineering, and Risk Assessment*; American Society of Civil Engineers: Reston, VA, USA, 2019; pp. 148–157.

56 Wang Z, Shafieezadeh A. REAK: Reliability Analysis through Error Rate-Based Adaptive Kriging. *Reliability Engineering & System Safety* 2019; **182**: 33–45.

57 Wang Z. Reliability Analysis and Updating with Meta-Models: An Adaptive Kriging-Based Approach. Ph. D. Thesis, The Ohio State University, 2019. Available online: https://rave.ohiolink.edu/etdc/view?acc_num=osu1574789534726544 (accessed on 7 October 2024).