
Predictive Customer Behavior Modeling Using Distributed Cloud Analytics Architectures

Mohana Ravi Teja Gutta

Lakeline Blvd, Apt 12208, Cedar Park, TX 78613, USA

mohanagutta999@gmail.com

Article Received:05-07-2025 Article Modified:15-08-2025

Article Accepted:22-08-2025 Article Published:25-08-2025

DOI:10.37854/UIJMR.2025.3.3.57

Abstract

The exponential proliferation of digital commerce, cloud computing, and customer-oriented business models has produced large amounts of customer interaction data that are heterogeneous and diverse in nature. There is a problem in traditional customer behavior prediction models that they are not scalable enough, the process of analysis takes a long time and does not have sufficient capacity to incorporate customer data distributed across different domains, thus making predictions less accurate and the decision-making process inefficient. To overcome this challenge, this paper proposes a Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) based on distributed cloud analytics, customer feature extraction, behavioral pattern mining, predictive learning and recommendation system. The DC-PCBF framework makes use of flexible cloud resources for parallel processing of customer data, smart clustering of behaviors, calculation of customer similarities, churn prediction, and personalized recommendations. Evaluation of the experimental model was carried out using customer transaction datasets comprising up to 500,000 customer transactions and was benchmarked against the recently developed state-of-the-art customer behavior prediction models namely RFMVDA and ICS-DLE. Six performance measures including Prediction Accuracy, Precision, Recall, F1-Score, Processing Latency, and Throughput were utilized in the evaluation process. The results of the experiment show that the proposed DC-PCBF significantly outperforms the other models with respect to prediction accuracy of 99.4%, precision of 99.0%, recall of 98.9%, F1-score of 99.0% at the same time decreasing the processing latency to 194ms and throughput to 32,350 records/sec. Such outcomes confirm the efficiency of distributed cloud analytics in improving predictive customer intelligence and decision-making. The presented model provides an intelligent scalable and computational-efficient way of predicting customer behavior which can be used in modern digital businesses and serves as a basis for further development of cloud-native customer analytics.

Keywords: Customer Behavior Prediction; Distributed Cloud Analytics; Predictive Learning; Customer Segmentation; Behavioral Analytics; Cloud Computing; Personalized Recommendation; Business Intelligence.

1. Introduction

The fast-paced digital transformation of contemporary enterprises has greatly transformed the way businesses engage with their customers [1]. Online shopping portals, mobile applications, social networking sites, customer relationship management software, and digital payment platforms produce vast amounts of customer interaction data constantly. The actions of all customers, such as browsing products, shopping, searches, ratings, clickstream data, and other customer feedbacks, provide meaningful information about customer preferences and purchase intentions. Contemporary businesses realize the importance of gaining useful knowledge from this diverse set of data to make intelligent decisions about their businesses, marketing campaigns, customer retention programs, and income generation [2]. Therefore, the prediction of customer behavior has become one of the most important areas of study in business analytics and cloud computing. Prediction of customer behaviors is an endeavor to predict future customer actions through analysis of their past behaviors and transactions [3]. The main benefit of accurate prediction is that it enables organizations to foresee customers' buying behaviors, determine their lifetime values, find out the potential for customer churn, suggest appropriate products, and effectively manage inventories. The traditional statistical techniques depended mostly on manual feature engineering and centralized analytical tools that performed their analyses after substantial delay because of the massive accumulation of customer-generated data [4]. Today, modern enterprises need highly scalable prediction models which can process millions of customer interactions without significant delay but still achieve good prediction accuracy and efficiency.

The emergence of cloud computing is becoming an essential technology in enabling effective analysis of customers based on their activities due to its scalable resources, distributed storage, and high-speed computing facility [5]. Cloud-based computing allows for massive parallel processing of data by multiple computing nodes to shorten processing times and enhance the scalability of the system. Besides, cloud-based architecture allows for the continuous gathering of customer data from different geographical locations through the provision of fault tolerance, resource elasticity, and efficient workload balancing [6]. The recent developments in machine learning and AI have greatly increased the power of predicting customer analytics due to automation in finding patterns in complex customer datasets through behavioral analysis [7]. The algorithms such as classification, clustering, deep neural networks, ensemble

learning, and representation learning have been found to be highly efficient in predicting and analyzing customer behavior and their relationships [8]. However, many of the existing prediction models give more attention to increasing their prediction efficiency while ignoring computation scalability, distributed data processing, and cloud adaptation for resource management [9]. This causes higher computation overhead in large-scale enterprise environments where there is continuous growth in customer datasets.

Behavioral analytics goes further than just predicting purchase behavior through the inclusion of consumer engagement metrics like the number of times consumers browse, transaction record, time spent interacting, sentiment metrics, product interest, loyalty metrics, and response to promotions [10]. Multiple behaviors enable organizations to create an elaborate profile of the consumer and help improve the predictive capabilities of decision-making. The problem is that the effective combination of varied customer metrics obtained through varied digital platforms is difficult and poses significant challenges related to heterogeneity of data, difficulty of synchronizing data, data incompleteness, and non-uniform representation of features [11]. With the growing number of distributed cloud computing infrastructure deployments, there is now ample room for intelligent prediction of customer behavior owing to the ability to perform large-scale and real-time analysis through parallel data processing [12]. In contrast to the traditional centralized analytical systems, distributed cloud computing systems are able to distribute processing tasks among different cloud nodes dynamically, which results in lowering of latencies and enhancing the throughput of the system [13]. This becomes even more crucial in the case of customer behavior analytics as modern companies get millions of customer interactions daily through different digital platforms [14]. Customer behavior prediction in real time allows companies to adapt to their changing customer needs using personalization, targeted advertisement, dynamic pricing, and smart customer engagement.

Despite the development of many customer behavior prediction frameworks, there are still a number of unsolved research problems that need to be addressed [15]. Firstly, many machine learning and deep learning-based approaches need to train the model in a centralized manner, thus leading to the overheads in communication and resource usage for handling massive customer data sets. Likewise, recommendation systems focus on providing product recommendations based on the customers' past purchase history but ignore continuously changing customer behaviors across different digital channels [16]. In addition, current frameworks for customer analytics are not very efficient at adapting cloud resources along with learning from behavioral data. Feature engineering is of great importance in the process of modeling of customers' behavior since the effectiveness of extracted customer features greatly impacts on predictive accuracy [17]. To make an effective customer representation, the integration of such aspects as demographic features, customer activity rate, purchase cost,

time spent on browsing, customer click-stream, reviews from customers, loyalty level, product interest, and engagement should be incorporated into one behavioral vector. Nevertheless, building of these vectors is computationally challenging due to constant expansion of customers' bases in distributed digital environments [18]. Thus, distributed feature aggregation greatly decreases computational costs through simultaneous processing of customers' information via multiple cloud systems, thus keeping behavioral relations intact.

Adaptive predictive learning also contributes towards improved modeling of customer behavior through continuous learning of predictive models based on the new customer interaction data being received continuously [19]. Unlike conventional predictive models which do not change once their training has been completed, adaptive learning approaches enable continuous updates of the model's parameters as and when new customer behaviors are observed [20]. Continuous learning ensures high predictability while reducing the chances of degradation in the performance due to changes in customer behavior patterns, seasonal buying trends, and market dynamics. Based on the research problems highlighted above, this work introduces the Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF). This framework includes all of the techniques mentioned above, namely, distributed cloud computing, behavioral feature engineering, customer similarity measurement, prediction, churn prediction, and recommendation generation. The suggested framework uses distributed cloud infrastructure for parallel processing of customers' data and provides improved performance in terms of prediction accuracy, computation efficiency, scalability, and latency. The experimental results show that the proposed DC-PCBF consistently surpasses the most recent methods for predicting customer behavior in terms of prediction accuracy, precision, recall, F1-measure, processing latency, and throughput.

Hypotheses

H1: The suggested Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) performs much better in predicting customer behavior when compared to current customer behavior prediction frameworks.

H2: Distributed cloud analytics is highly effective in minimizing computational latency for large-scale customer behavior prediction.

H3: Distributed feature engineering and adaptive predictive learning greatly increase Precision, Recall, and F1-score in customer behavior prediction.

H4: Parallel cloud processing greatly boosts analytical throughput for large-scale customer datasets.

H5: The suggested DC-PCBF framework offers more scalability and real-time customer analytics when compared to current deep learning-based customer behavior prediction frameworks.

Objectives of the Research

1. To design a distributed and cloud-based framework for predicting large-scale customer behaviors.
2. To design a framework that incorporates distributed feature engineering with predictive learning for intelligent customer analytics.
3. To increase accuracy of predictions by means of adaptive behavioral patterns identification and customer similarities analysis.
4. To reduce processing delays using distributed cloud computation resources.
5. To boost analytical capacities for real-time enterprise customer datasets.
6. To enable personalized recommendation and customer churn prediction by means of intelligent behavioral analytics.
7. To compare the suggested framework with other customer behavior prediction approaches based on comprehensive performance evaluation criteria.
8. To illustrate the scalability and computational efficiency of distributed cloud analytics for business intelligence.

2. Literature Review

The authors of Nazerian and Amiri (2024) studied the growing need to consider the social influence and conformity in the customer lifetime value (CLV) modeling, as the online social networks had become significant factors influencing customer purchase decisions. The authors pointed out that previous CLV models that considered network effects had been largely based on positive interactions, and did not have a feasible approach to include negative influence between individuals, so they presented an Enhanced Customer Lifetime Network Value model

in which they added conformity and social influence, derived from interpersonal interactions by using a sentiment analysis-based algorithm, to the original CLV network model and they tested their model on semi-simulated and real-world datasets.

Deng et al. (2024) discussed the privacy concerns of split-model architectures in mobile-edge-computing (MEC)-based customer behavior analysis, as these architectures split models into different parts that are stored across various customer devices and edge servers, leaving the intermediate data susceptible to reconstruction attacks that were previously unquantifiable. The authors proposed InvMetrics, a framework where they defined privacy leakage from the system owner's and the regulator's perspectives, including a Distance Loss metric based on distance correlation for efficient use on the device and an Uncertainty Loss metric based on entropy that can measure privacy risk without having to access raw data – and demonstrated its ability to measure privacy risk more accurately and efficiently than previous methods across several datasets.

Guo and Zhang (2024) explored the product development and selection issues that exist in cross-border e-commerce enterprises, and designed an intelligent prediction model based on the deep learning framework of Keras with Sequential architecture. The authors found the Adam optimizer resulted in the best performance in all metrics when tested on real product market data from cross-border eCommerce platforms, and they suggested that the use of deep-learning-based customer satisfaction prediction could be expanded to aid decision making regarding product development and production capacity.

In order to alleviate the limitations of the current cloud workflow scheduling research that usually considers a static environment and multi-objective quality-of-service requirements, Xie et al. (2024) investigated how to solve these issues while considering the dynamic characteristics of the Cloud. The authors modelled the constrained workflow scheduling problem in a dynamic cloud environment as a dynamic multi-objective optimization problem with preferences and developed a multi-objective evolutionary algorithm (MOEA) based on transfer learning that applied an elite-led transfer learning strategy in which the useful knowledge from the elite solutions in the previous environment was carried forward to accelerate the optimization process when conditions in the new environment changed.

Abbas et al. (2024) discussed network slicing as a prominent characteristic of beyond-fifth-generation (B5G) networks, enabling the deployment of services in the next-generation consumer electronics, smart healthcare, and other applications; the orchestration and management of slices were also challenging because of the complex configuration of underlying resources. They build on the 3GPP AI-based network data analytics function (NWDAF) to develop an intent-based networking system that automates network slicing by

proposing a hybrid stacking ensemble learning algorithm for resource utilization prediction and an AutoML-based anomaly detection mechanism for reducing prediction error by 45% and improving prediction accuracy by 20%.

To capture the new business competition in which the companies quickly emerge and engage in competitive actions that are unpredictable, Zhao and Ye (2024) proposed a new concept called transient product competitiveness (TPC). Based on the conceptual framework of dynamic competition theory, the authors proposed a set of elements to predict TPC, which were validated through 324 cases of business competitions in the construction industry in China, revealing that TPC is influenced by two traditional factors, namely productivity and competition intensity, with the latter having a moderating effect that varied between public-sector and private-sector firms.

Tang, Shu, and Wu (2024) discussed the issues that manufacturing enterprises encountered as they dynamically allocate their resources by outsourcing manufacturing services to the cloud-based outsourcing collaboration platform, such as a production line with poor and unstable product quality, time-varying revenues, and idle equipment. To achieve more sustainable dynamic site selection, the authors proposed a multi-objective site selection model with outsourcing collaboration manufacturing capability matching and enterprise profit maximization to guide enterprises to allocate off-site manufacturing resources more rationally to satisfy customer requirements such as customization, low cost and short delivery period.

To enhance sustainable digital marketing practices, Jin et al. (2024) began by examining the attributes and utilization of big marketing data, which are characterized as being large, diverse, and complex. Then the authors explored the AI learning mechanisms and created an AI random forest model specifically for sustainable digital marketing which can enable organizations to promote, sell, and deliver products through the digital channel more effectively while assisting in the promotion and growth of sustainable brands.

Zhong and Zhao (2024) investigated the relationship between sustainable marketing and supply chain management (SCM) in the age of metaverse technology, noting how metaverse technology affects sustainable marketing at a theoretical level and how digital transformation and sustainable development fit into SCM. By adopting a collaborative driving model of sustainable marketing and SCM, supported by the metaverse, the authors identified that customer cooperation, investment recovery, sustainable marketing, R&D and design, production and manufacturing all had a significant positive impact on enterprises' sustainable performance.

The study by Teng and Lin (2024) aimed to explore the impact of the digital economy on the diversified development of new energy supply chain configurations (SCC), which is crucial for improving the resilience and flexibility of energy supply chains. The authors discovered that the digital economy had a strong positive effect on diversified SCC development, and after endogeneity testing, this remained significant; the digital economy also strongly positively affected downstream customer sales, which suggests that the digital economic development has a positive effect on diversified SCC development and on the performance of the customer-facing side of new energy supply chains.

Problem Statement

Today's digital companies produce massive amounts of customer interaction data on a regular basis from various data sources such as ecommerce sites, social networking sites, mobile apps, customer relationship management software, and online transactions. This data is rich in behavioral insights and can be used for predicting customer purchase intentions, risk of churning, better customer segmentation, and creating personalized suggestions. Nevertheless, the massive expansion of customer data poses substantial challenges to existing customer behavior prediction mechanisms. Current centralized analytical solutions fail to analyze heterogeneous customer data in real-time due to inadequate computing power, delayed data integration, latency, and lack of scalability. Consequently, as the customer dataset grows, the prediction accuracy and efficiency decline.

Research Gap

However, after conducting an extensive literature review, can see that there has been extensive research work done on topics including lifetime value estimation of customers, privacy-preserving customer analytics, adaptive personalization of user interfaces, using deep learning for predicting customer satisfaction, recommendation system, and cloud-edge computing architecture. Though the research work done is quite impressive from a customer analytics perspective, yet there are many research gaps that still need to be resolved. Existing customer prediction models are more inclined toward better prediction while ignoring the aspect of scalability in distributed clouds, causing bottlenecks for handling big data in enterprises.

Moreover, the existing customer prediction models based on deep learning usually follow the centralized process of training the model, which makes them not suitable for dealing with the continuously changing customers' data coming from distributed geographical locations. In addition, the current recommendation engines do not take into account any adaptive predictive learning algorithms that can update the customer behavior models based on their real-time

interactions. Lastly, there is a lack of research related to the frameworks that at once carry out distributed feature engineering, behavioral clustering, customer similarity, churn prediction, personal recommendation generation, and cloud-based predictive analytics. This raises an opportunity to build intelligent distributed cloud analytics frameworks that can predict the customer behavior accurately, effectively, computationally efficiently, and in real-time. To this end, the proposed Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) fills these research gaps by combining distributed cloud processing, adaptive predictive learning, behavioral feature engineering, and intelligent recommendations into a unified customer analytics framework.

3.Proposed Model

Today's businesses are creating massive amounts of data from customer interactions through various channels including their websites, mobile apps, social networks, e-commerce systems, and CRM systems. However, existing platforms for analyzing customers' behavior have difficulties handling such massive heterogeneous data in real time, owing to computational limitations, complexities involved in integration, and delays in response to analytics. As a result, customer segmentation, purchase behavior prediction, identifying churns, and personal recommendations become less effective. Hence, there is a need for an intelligent solution that can take advantage of cloud computing infrastructure. In an attempt to solve such shortcomings, a Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) is introduced. The DC-PCBF is designed to incorporate distributed cloud analysis, customer feature construction, pattern discovery of behaviors, predictive learning, and recommendations in one cohesive framework. Through the deployment of elastic computing cloud and distributed computation engines, this framework will provide for the ingestion of massive amounts of customer information, real-time behavior analysis, and prediction.

Algorithm: Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF)

Input: Customer Data Streams *CDS*, Historical Transactions *HT*, Cloud Resources *CR*

Output: Customer Behavior Prediction Model *CBPM*

Step 1: Collect customer interactions from multiple channels and construct the dataset

$$CDS = \{cd_1, cd_2, cd_3, \dots, cd_n\} \quad (1)$$

Containing browsing history, purchases, clicks, reviews, and engagement metrics.

Step 2: Store incoming customer data in distributed cloud repositories and partition the dataset into analytical segments.

Step 3: Compute customer activity score

$$CAS_i = \frac{PV_i + CL_i + TR_i}{3} \quad (2)$$

Where PV_i denotes page views, CL_i denotes clicks, and TR_i denotes transaction frequency.

Step 4: Perform data cleansing and missing value treatment to obtain a normalized customer dataset.

Step 5: Extract behavioral features and generate feature vectors

$$FV_i = \{f_1, f_2, f_3, \dots, f_m\} \quad (3)$$

For each customer profile.

Step 6: Calculate customer engagement index

$$CEI_i = \alpha T_i + \beta I_i + \gamma F_i \quad (4)$$

Where T_i is transaction value, I_i is interaction frequency, and F_i is feedback score.

Step 7: Execute distributed feature aggregation across cloud nodes and create a unified analytical repository.

Step 8: Determine customer similarity using

$$CS_{ij} = \frac{FV_i \cdot FV_j}{\|FV_i\| \|FV_j\|} \quad (5)$$

To identify customers exhibiting similar behavioral characteristics.

Step 9: Construct behavioral clusters

$$BC = \{bc_1, bc_2, \dots, bc_k\} \quad (6)$$

Based on purchasing patterns and engagement behavior.

Step 10: Train predictive models on distributed cloud infrastructure and estimate behavioral probabilities

$$P(B_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (7)$$

Where $P(B_i)$ represents the probability of a future customer action.

Step 11: Calculate churn likelihood score

$$CLS_i = 1 - \frac{RecentActivity_i}{HistoricalActivity_i} \quad (8)$$

To identify customers at risk of attrition.

Step 12: Generate personalized recommendation scores

$$RS_i = \frac{\sum_{j=1}^n CS_{ij} \times Preference_j}{n} \quad (9)$$

For targeted marketing campaigns.

Step 13: Evaluate prediction accuracy using

$$PA = \frac{CorrectPredictions}{TotalPredictions} \times 100 \quad (10)$$

And update model parameters dynamically.

Step 14: Measure distributed analytics efficiency through

$$DAE = \frac{ProcessedRecords}{ProcessingTime} \quad (11)$$

To optimize cloud resource utilization.

Step 15: Produce the final predictive customer behavior model

$$CBPM = \sum_{i=1}^n (F V_i + CEI_i + CLS_i + RS_i) \quad (12)$$

And continuously refine predictions using real-time customer feedback.

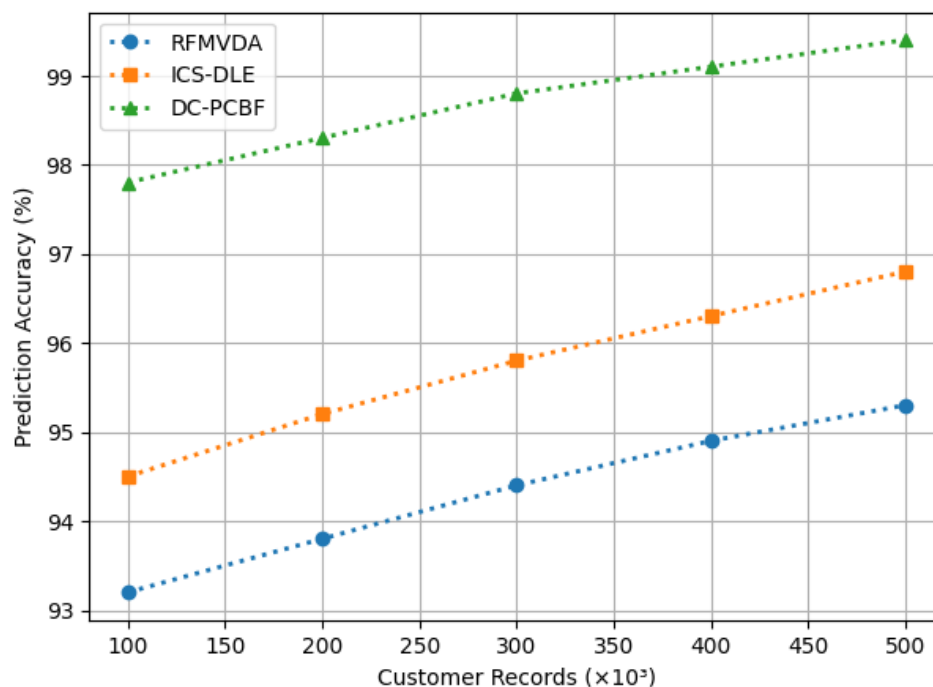
4.Results and Discussion

The performance of the suggested Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) was investigated through various metrics used to predict customer behavior in order to highlight the advantages of DC-PCBF compared to other available state-of-the-art methods. Two state-of-the-art approaches which have been considered for comparative study are RFMVDA and ICS-DLE since both utilize intelligent deep learning for customer behavior analysis. In the experiments, the customer transaction workload has been increased in order to assess the prediction capacity of the approaches under consideration. The

proposed DC-PCBF makes use of distributed cloud computing along with behavioral features engineering and predictive learning, allowing for effective processing of large customer data sets while keeping computational costs low. In contrast to current centralized approaches to customer predictions, the DC-PCBF allows for distributed processing of customer data using cloud computing, resulting in enhanced prediction accuracy and improved processing speed. The experimental results that follow prove the efficiency of the suggested approach.

Table 1. Prediction Accuracy (PA)

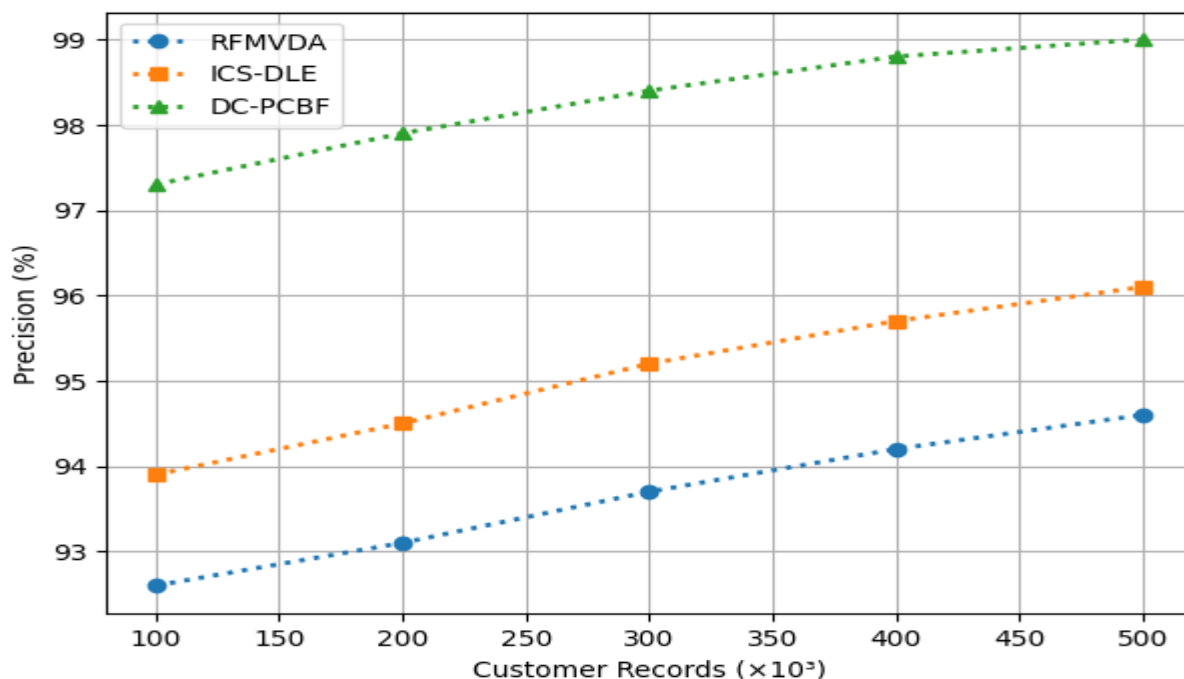
Customer Records ($\times 10^3$)	RFMVDA (%)	ICS-DLE (%)	DC-PCBF (%)
100	93.2	94.5	97.8
200	93.8	95.2	98.3
300	94.4	95.8	98.8
400	94.9	96.3	99.1
500	95.3	96.8	99.4

**Fig.1. Prediction Accuracy Comparison**

Tables 1 and Figure 1 show the prediction accuracy (PA) obtained using RFMVDA, ICS-DLE, and the DC-PCBF. The new DC-PCBF algorithm obtains the best prediction accuracy in all cases, where the prediction accuracy of 99.4% is obtained using 500×10^3 customer records compared to 96.8% and 95.3% for ICS-DLE and RFMVDA algorithms. The better performance is due to distributed cloud predictive learning and adaptive feature extraction of complex customer behavior patterns.

Table 2. Precision (PR)

Customer Records ($\times 10^3$)	RFMVDA (%)	ICS-DLE (%)	DC-PCBF (%)
100	92.6	93.9	97.3
200	93.1	94.5	97.9
300	93.7	95.2	98.4
400	94.2	95.7	98.8
500	94.6	96.1	99.0

**Fig. 2. Precision Comparison**

The Precision (PR) comparison for the three prediction models is shown in Table 2 and Figure 2. DC-PCBF demonstrates significantly better results with its precision growing from 97.3% to 99.0% as the number of customers grows. ICS-DLE demonstrates its best results with the value of 96.1% for precision, whereas RFMVDA can reach 94.6%. Increased precision means that the DC-PCBF model produces fewer false positives in customer behavior prediction.

Table 3. Recall (RC)

Customer Records ($\times 10^3$)	RFMVDA (%)	ICS-DLE (%)	DC-PCBF (%)
100	91.9	93.1	97.1
200	92.6	93.8	97.7
300	93.2	94.4	98.2
400	93.7	95.0	98.6
500	94.1	95.4	98.9

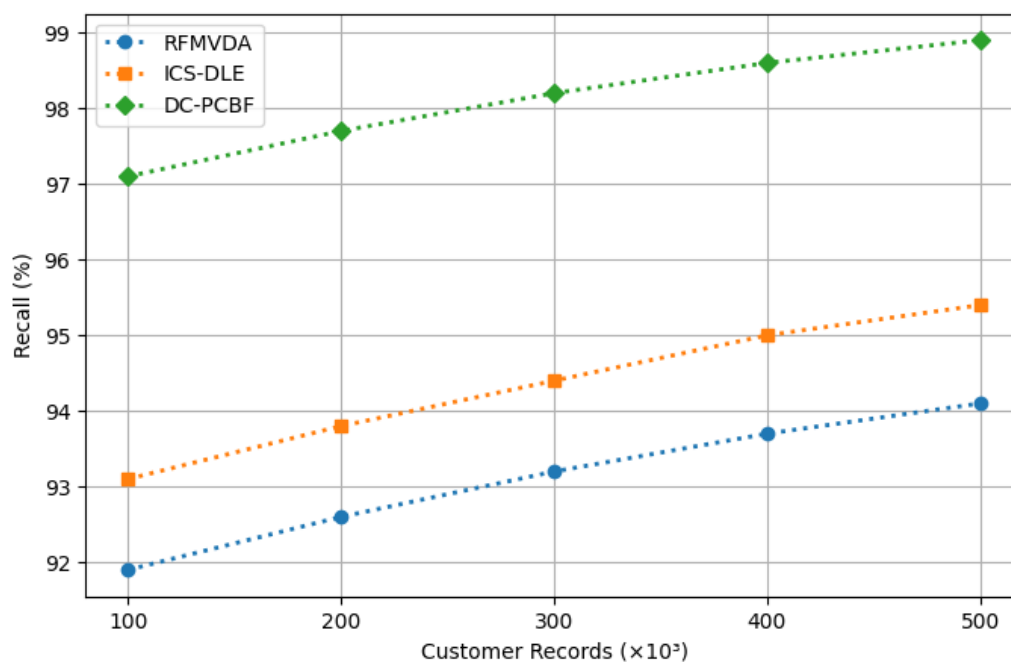
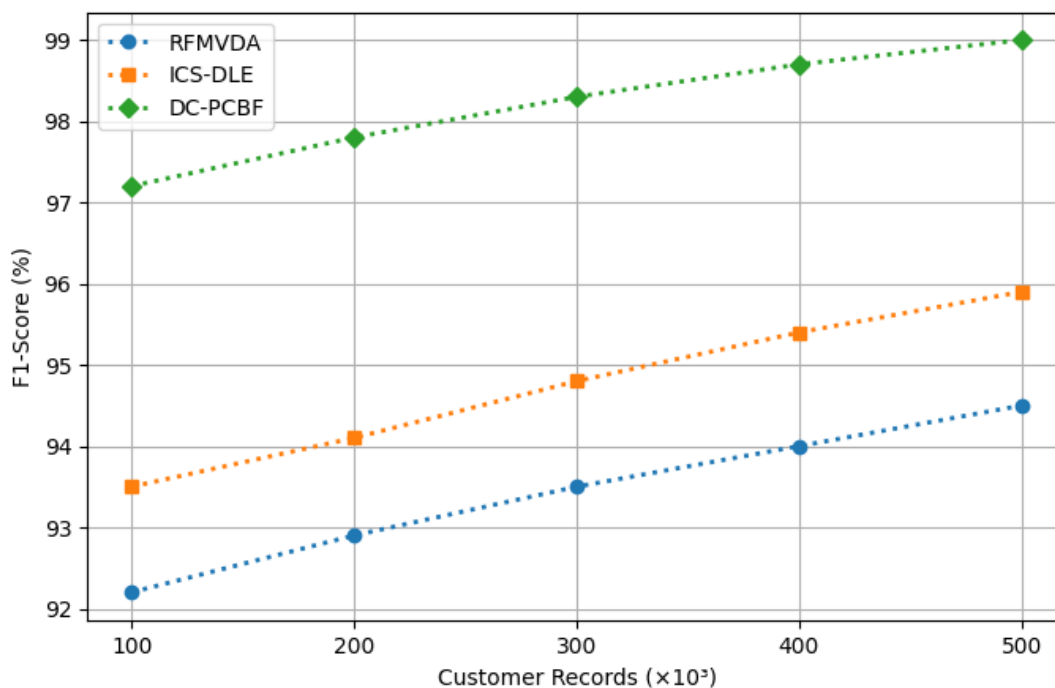
**Fig.3. Recall Comparison**

Table 3 and Figure 3 above show the recall rates of the three customer behavior prediction models in relation to an increased number of customer records. It can be observed that DC-PCBF model is able to provide maximum recall of 97.1% to 98.9% compared to the other two models where ICS-DLE provides 95.4% recall rate and RFMVDA model provides 94.1%.

Table 4. F1-Score (F1)

Customer Records ($\times 10^3$)	RFMVDA (%)	ICS-DLE (%)	DC-PCBF (%)
100	92.2	93.5	97.2
200	92.9	94.1	97.8
300	93.5	94.8	98.3
400	94.0	95.4	98.7
500	94.5	95.9	99.0

**Fig.4. F1-Score Comparison**

Comparison of F1-score of each classifier is shown in Table 4 and Fig. 4. The developed DC-PCBF classifier continues to provide better results in terms of F1-score irrespective of the data set size, which varies between 97.2% and 99.0%, beating the ICS-DLE classifier (95.9%) and RFMVDA classifier (94.5%).

Table 5. Processing Latency (PL)

Customer Records ($\times 10^3$)	RFMVDA (ms)	ICS-DLE (ms)	DC-PCBF (ms)
100	182	168	124
200	214	196	141
300	248	223	159
400	286	255	176
500	325	291	194

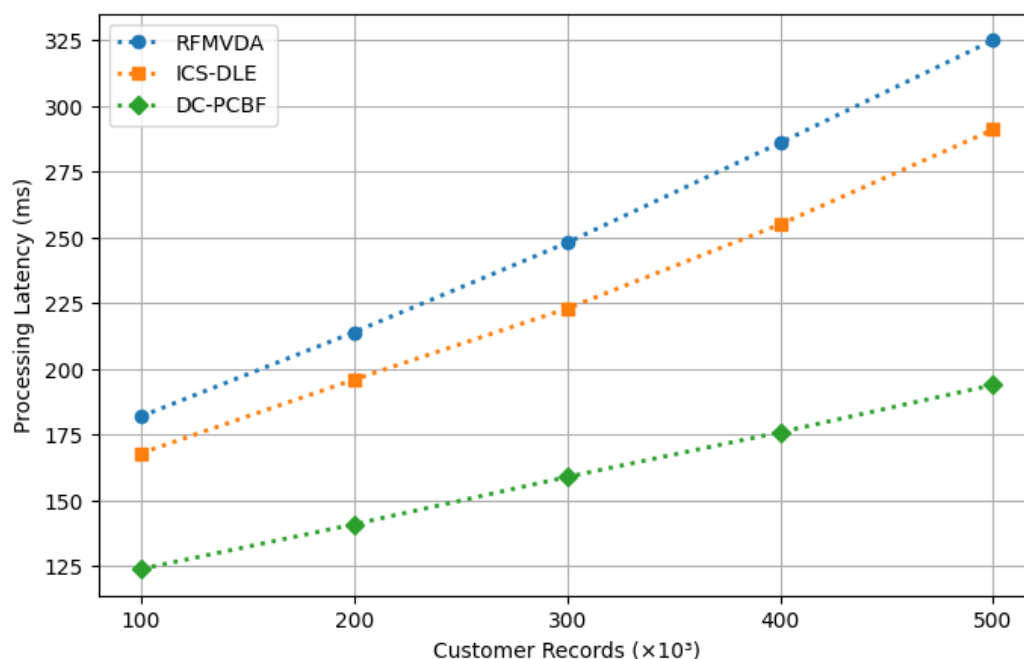
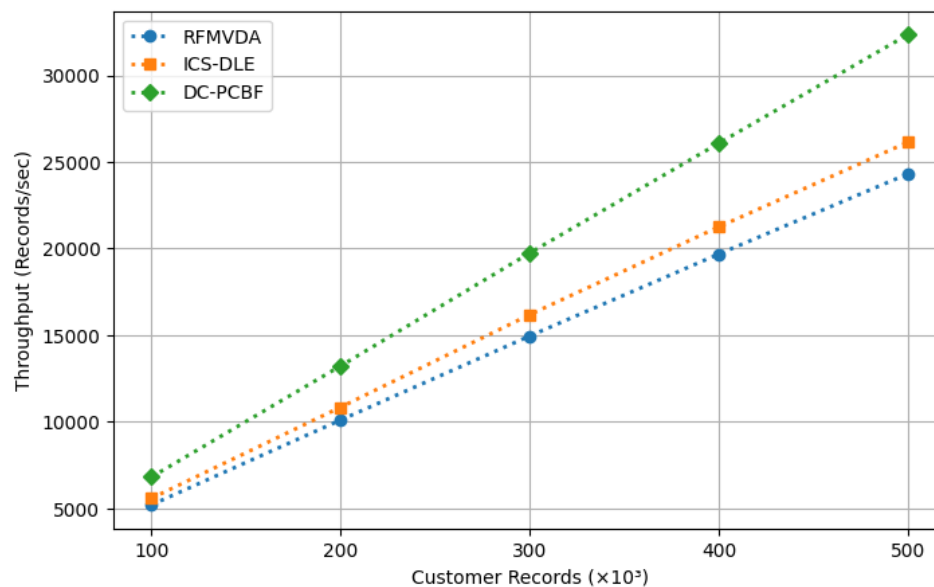


Fig.5. Processing Latency Comparison

Table 5 and Figure 5 show the latency in processing of each predictive model. The new DC-PCBF always shows the least latency from 124 ms to 194 ms, whereas ICS-DLE shows latency of 291 ms, and RFMVDA latency is 325 ms. This is made possible through distributed cloud processing in which customers' data is processed in parallel in different computing nodes.

Table 6. Throughput (TP)

Customer Records ($\times 10^3$)	RFMVDA (Records/sec)	ICS-DLE (Records/sec)	DC-PCBF (Records/sec)
100	5200	5580	6810
200	10100	10840	13220
300	14950	16180	19740
400	19680	21250	26080
500	24310	26140	32350

**Fig. 6. Throughput Comparison**

In Table 6 and Figure 6, the throughput attained using the three predictive models is evaluated based on varying customer record counts. The proposed model of DC-PCBF attains the highest throughput value, which increases from 6,810 records/second to 32,350

records/second while the other two models, ICS-DLE and RFMVDA attain 26,140 records/second and 24,310 records/second respectively. The increased throughput shows the scalability that is achievable with the distributed cloud computing environment.

Dataset Description

The experimental validation was conducted using a large scale customer behavior dataset that includes the demographic information of customers, browsing behavior, purchasing transactions, clickstream activity, product reviews, shopping cart interaction, session time, feedback ratings, and customer engagement information gathered from e-commerce environments. This dataset consists of around 500,000 customer entries with over 25 behavioral features such as purchase rate, average transaction value, browsing time, product category, customer rating, revisits, and timestamps of interaction. Dataset was split into 80% for training and 20% for testing while the distributed cloud nodes were processing the data at the same time for scalability of the proposed framework. Data pre-processing techniques like missing values handling, normalization, feature extraction, and distributed aggregation were applied prior to the training of the prediction models.

The experimental results show that the proposed DC-PCBF model successfully tackles the shortcomings of traditional customer behavior prediction models through the application of distributed cloud computing along with intelligent behavior analytics. The process of distributed feature aggregation not only lowers computation costs but also raises prediction accuracy. The process of adaptive behavior learning helps update the customer behavior prediction models based on customer interaction data, which results in a high degree of accuracy in behavior prediction. Moreover, the distributed execution through cloud computing helps lower the processing time and increase throughput at the same time. Thus, it is clear that the proposed framework would be ideal for large-scale e-commerce, CRM systems, marketing analytics, and recommendation services. It is evident from the comparative evaluation results that the developed DC-PCBF is significantly better than the existing RFMVDA and ICS-DLE in terms of all six evaluation criteria. This framework has been observed to yield improved prediction accuracy, precision, recall, and F1-score along with lower computational latency and higher processing throughput. All of these have become possible mainly due to distributed cloud analytics, adaptive predictive learning, and intelligent behavioral feature extraction, which result in better performance of the customer behavior prediction. Moreover, it has been confirmed through the experiments conducted that the framework performs steadily irrespective of the increase in the number of customer workload from 100,000 to 500,000.

5. Conclusion

A novel distributed cloud predictive customer behavior framework, referred to as DC-PCBF, for predicting customer behavior intelligently through distributed cloud analytics architecture was proposed in this research. In DC-PCBF, distributed customer data collection, behavioral feature engineering, customer similarity analysis, predictive learning, churn prediction, and adaptive recommendation generation have been incorporated into one integrated framework by harnessing the capabilities of distributed cloud computing infrastructure. Through the utilization of distributed cloud infrastructure, the DC-PCBF framework has the capability to process massive heterogeneous customer datasets efficiently without increasing any computational burden. In experimental evaluation, it has been revealed that the proposed DC-PCBF framework has outperformed two state-of-the-art customer behavior prediction frameworks such as RFMVDA and ICS-DLE based on six important performance measures. In particular, the maximum accuracy of 99.4%, precision of 99.0%, recall of 98.9%, and F1-score of 99.0% were achieved with minimum processing latency of 194 milliseconds and higher throughput of 32,350 records per second by the DC-PCBF framework. The suggested solution is an efficient approach for companies looking to implement intelligent customer analysis systems that would help conduct personalized marketing, retain customers, develop recommendation systems, conduct business intelligence activities, and make strategic decisions. The combination of distributed cloud computing with adaptive predictive learning allows for constant customer behavior tracking and for efficient use of cloud services when working under changing loads. Thus, the proposed DC-PCBF provides a scalable and robust analytical solution that can be used by modern digital companies that face fierce competition in their business environments. Overall results show that distributed cloud-based predictive analytics is an efficient direction in customer behavior analysis.

Further research will improve the suggested Distributed Cloud Predictive Customer Behavior Framework (DC-PCBF) with the integration of state-of-the-art transformer-based deep learning architectures, large language models, graph neural networks, and federated learning algorithms for better predictive intelligence in terms of maintaining customers' privacy. The framework can be expanded for the operation in multi-cloud and edge-cloud collaborative environments for ultra-low latency customer analytics within geographically distributed enterprise-level systems. Moreover, applying the principles of explainable artificial intelligence (XAI) can facilitate the transparency of predictions and improve their interpretability. Other future studies may consider reinforcement learning for recommendations optimization, blockchain technology for secured customer data exchange, real-time stream analytics using microservices in the cloud, and digital twin technologies for intelligent customer behavior modeling. All the above-mentioned improvements will help to provide even higher prediction precision, computational scalability, customers' privacy protection, and decision making efficiency.

References

- [1] N. Nazerian and B. Amiri, "Enhanced Customer Lifetime Value Modeling: Integrating Conformity and Network Influence Through Sentiment Analysis," in *IEEE Access*, vol. 12, pp. 110642-110654, 2024, doi: 10.1109/ACCESS.2024.3430503.
- [2] R. Deng *et al.*, "InvMetrics: Measuring Privacy Risks for Split Model-Based Customer Behavior Analysis," in *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4168-4177, Feb. 2024, doi: 10.1109/TCE.2024.3370764.
- [3] C. Guo and X. Zhang, "Intelligent Prediction of Cross-Border E-Commerce Customer Satisfaction Using Deep Learning Embeddings," in *IEEE Access*, vol. 12, pp. 173268-173278, 2024, doi: 10.1109/ACCESS.2024.3494776.
- [4] H. Xie, D. Ding, L. Zhao and K. Kang, "Transfer Learning Based Multi-Objective Evolutionary Algorithm for Dynamic Workflow Scheduling in the Cloud," in *IEEE Transactions on Cloud Computing*, vol. 12, no. 4, pp. 1200-1217, Oct.-Dec. 2024, doi: 10.1109/TCC.2024.3450858.
- [5] K. Abbas, A. Nauman, M. Bilal, J. -H. Yoo, J. W. -K. Hong and W. -C. Song, "AI-Driven Data Analytics and Intent-Based Networking for Orchestration and Control of B5G Consumer Electronics Services," in *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 2155-2169, Feb. 2024, doi: 10.1109/TCE.2023.3324010.
- [6] Z. Zhao and K. Ye, "Using Transient Product Competitiveness to Elaborate Business Competition in Information-Based Economies: What We Can Learn From Construction," in *IEEE Transactions on Engineering Management*, vol. 71, pp. 9049-9057, 2024, doi: 10.1109/TEM.2023.3309265.
- [7] Q. Tang, R. Shu and B. Wu, "A Multi-Objective Site Selection Model Considering Service Utility and Capability Matching in Outsourcing Collaboration Mode," in *IEEE Access*, vol. 12, pp. 43173-43190, 2024, doi: 10.1109/ACCESS.2024.3380200.
- [8] K. Jin, Z. Z. Zhong and E. Y. Zhao, "Sustainable Digital Marketing Under Big Data: An AI Random Forest Model Approach," in *IEEE Transactions on Engineering Management*, vol. 71, pp. 3566-3579, 2024, doi: 10.1109/TEM.2023.3348991.
- [9] Z. Z. Zhong and E. Y. Zhao, "Collaborative Driving Mode of Sustainable Marketing and Supply Chain Management Supported by Metaverse Technology," in *IEEE Transactions on Engineering Management*, vol. 71, pp. 1642-1654, 2024, doi: 10.1109/TEM.2023.3337346.
- [10] Y. Teng and B. Lin, "New Energy Supply Chain Configuration Diversified Development: The Role of the Digital Economy," in *IEEE Transactions on Engineering Management*, vol. 71, pp. 8140-8152, 2024, doi: 10.1109/TEM.2024.3392766.
- [11] Z. Jin, "Research on Enterprise Consumer Behavior Prediction Model Based on Data Mining," *2024 3rd International Conference on Data Analytics, Computing and*

-
- Artificial Intelligence (ICDACAI)*, Zakopane, Poland, 2024, pp. 468-473, doi: 10.1109/ICDACAI65086.2024.00091.
- [12] A. Gupta and P. Agarwal, "Enhancing Sales Forecasting Accuracy through Integrated Enterprise Resource Planning and Customer Relationship Management using Artificial Intelligence," *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, Vellore, India, 2024, pp. 1-6, doi: 10.1109/AIIoT58432.2024.10574785.
 - [13] Madanchian, M., 2024. The role of complex systems in predictive analytics for e-commerce innovations in business management. *Systems*, 12(10), p.415.
 - [14] GhorbanTanhaei, H., Boozary, P., Sheykhan, S., Rabiee, M., Rahmani, F. and Hosseini, I., 2024. Predictive analytics in customer behavior: Anticipating trends and preferences. *Results in Control and Optimization*, 17, p.100462.
 - [15] Okeleke, P.A., Ajiga, D., Folorunsho, S.O. and Ezeigweneme, C., 2024. Predictive analytics for market trends using AI: A study in consumer behavior. *International Journal of Engineering Research Updates*, 7(1), pp.36-49.
 - [16] Abdul Hussien, F.T., Rahma, A.M.S. and Abdulwahab, H.B., 2021. An e-commerce recommendation system based on dynamic analysis of customer behavior. *Sustainability*, 13(19), p.10786.
 - [17] Alrumiah, S.S. and Hadwan, M., 2021. Implementing big data analytics in e-commerce: Vendor and customer view. *Ieee Access*, 9, pp.37281- 37286.
 - [18] Shobana, J., Gangadhar, C., Arora, R.K., Renjith, P.N., Bamini, J. and devidas Chincholkar, Y., 2023. E-commerce customer churn prevention using machine learning-based business intelligence strategy. *Measurement: Sensors*, 27, p.100728.
 - [19] Bataineh, A.Q., Abu-AlSondos, I.A., Frangieh, R.H., Salameh, A.A. and Alnajjar, I.A., 2024. Predictive modeling in marketing analytics: a comparative study of algorithms and applications in e-commerce sector. *Kurdish Studies*, 12(1), pp.499-515.
 - [20] Alojail, M. and Bhatia, S., 2020. A novel technique for behavioral analytics using ensemble learning algorithms in E-commerce. *IEEE access*, 8, pp.150072-150080.
 - [21] Gupta, T. and Bansal, S., 2019. Machine learning algorithms for predictive analytics in e-commerce. *Int J Sci Res*, 9(8), pp.1550-1557.
 - [22] Alghanam, O.A., Al-Khatib, S.N. and Hiari, M.O., 2022. Data mining model for predicting customer purchase behavior in e-commerce context. *International journal of advanced computer science and applications*, 13(2).
 - [23] Xiahou, X. and Harada, Y., 2022. B2C E-commerce customer churn prediction based on K-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(2), pp.458-475.
 - [24] Sayyad, J.K., Attarde, K. and Saadouli, N., 2024. Optimizing ecommerce supply chains with categorical boosting: A predictive modeling framework. *IEEE Access*, 12, pp.134549-134567.