Hybrid Ensemble Model for High-Utility Itemset Mining and Customer Purchase Behavior Analysis in Product Recommendation

Adity Tiwari[1,2,a], Jyothi Pillai[3,j],Rekhram Janghel[4,r]

[1] Department of Computer Science and Information technology Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India
[2] Department of Computer Science Engineering, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India

[3] Department of Computer Application, Bhilai Institute of Technology, Durg, Chhattisgarh, India
[4] Department of Information technology, National Institute of Technology, Raipur, Chhattisgarh,India
[a] adityarise0609@gmail.com, [j] jyothipillai71@gmail.com, [r] rrjanghel.it@nitrr.ac.in

ABSTRACT: In the modern era, high utility itemset (HUIS) mining has gained significant attention due to its wide-ranging applications in domains such as web services, retail marketing, and beyond. It enables the identification of itemsets within a transaction database that yield high utility, translating into greater profitability. Over the years, numerous methodologies have been introduced for HUIS mining and product recommendation. However, existing models often suffer from drawbacks such as lower prediction accuracy, high computational complexity, extended processing times, and costly implementation. To address these challenges, this study proposes an innovative system architecture that leverages a hybrid ensemble model for real-time HUIS mining and product recommendation based on customer purchasing behavior. The proposed framework has been trained and validated using well-established datasets, including Chainstore and Foodmart, achieving impressive performance metrics—98.94% accuracy, 97.59% precision, 97.99% recall, and an F1-score of 98.27%. While recent advancements have led to the development of various models for HUIS mining and customer behavior analytics, there remains ample scope for further research to create faster and more efficient models for analyzing purchasing trends.

KEYWORDS: Customer Behaviour, DNN, Ensemble Model, High Utility Itemset, Machine Learning, Product Recommendation.

1. INTRODUCTION

Data mining techniques play a crucial role in uncovering valuable hidden patterns within large datasets. Various data mining approaches, such as frequent itemset (FIS) mining and high utility itemset (HUIS) mining, have been explored to extract meaningful patterns for effective data analysis (Wu et al., 2021). FIS mining focuses on identifying itemsets that commonly appear together in a given dataset, referred to as frequent itemsets (Lin et al., 2020). This technique is widely used to determine correlations and associations between different items within a dataset. For example, in a retail store's transaction dataset, FIS mining can help identify items frequently purchased together, such as butter and bread (Sohrabi, 2020; Wu et al., 2020).

In contrast, high utility itemset (HUIS) mining focuses on identifying item sets that not only occur frequently but also generate higher utility values. The utility value of an itemset represents the associated profit, making HUIS mining particularly useful for detecting high-profit item combinations. For example, in a supermarket transaction dataset, HUIS mining can help identify frequently purchased item sets that yield higher profits, such as expensive wines and laptops (Luna et al., 2023; Yang et al., 2023; Yuan, 2022). Both FIS and HUIS mining are well-established data mining techniques and are applied across various domains, including market basket analysis, fraud detection, and recommendation systems (He et al., 2021; Magdy et al., 2023).

There are diverse gadgets that are being used by people in their day-to-day lives such as smartphones, tablets, laptops, computers, and many more. Such kinds of digital devices produce a massive amount of the dataset but such an unorganized dataset may not be useful until it is mined for extraction of desired data which might be utilized for diverse purposes (Agarwal et al., 2021; Bhuvaneswari et al., 2022; Hung et al., 2021). There is a communal method for data extraction by utilizing association rule mining. This is accomplished in two segments i.e., in frequent item set patterns are obtained in the beginning phase as well as solid rules amongst such patterns are produced in the later segment. Generation of the rules amongst frequent sets of items is very easy and due to that multifarious investigators have reported on FIS mining during the last two decades (Dao et al., 2023; K. Zhang et al., 2020).

Data mining is a procedure to uncover the diverse required patterns via a dataset collection that may be utilized in a large number of applications namely the manufacturing sector and many more. On the other hand, frequent itemset (FIS) mining is utilized for recognition of the frequent item grouping for facilitating effective choice-making which is one of the initial dataset mining jobs (Lakshmi & Krishnamurthy, 2022; Shayegan Fard & Namin, 2020). Though, the earlier discovered FIS protocols mainly concentrate only on the frequency occurrence of an item or a product to ignore the profit or value. In simple words, some of the frequent patterns are not much important or beneficial. For example, the frequent sets of items within the huge superstores e.g., apples and eggs may have very lower profits. Such is not the product choice-maker needs, even if they happen arbitrarily (Ma et al., 2020; Qiu et al., 2020; Sekhavat, 2020). The FIS job has been upgraded for the higher-utility sets of mining jobs which may recognize the product groups with higher benefits. However, the HUIS mining protocols concentrate not only on the product occurrence frequency but also on the quantity and profit margin (Cheng et al., 2023). Therefore, a large number of factors are being measured within the HUIS, the anti-monotonic assets have not been held within the mining procedure. Subsequently, the relation amongst the multiple utilities of a product as well as some extended set of items is intricate and increases the computational complexity of the mining process.

Ali et al. (2022) introduced a research study on an online product recommendation system based on consumer review analysis. Online consumer reviews serve as a valuable resource for both businesses and customers, enabling them to extract essential insights for informed decision-making. However, finding relevant and desired datasets while browsing the web remains a major challenge for consumers. To address this, the researchers proposed a novel recommendation model, “SmartTips,” which utilizes aspect-based evaluation to analyze consumer feedback, rate multiple products, and extract user preferences in real-time.

Similarly, Shahbazi et al. (2020) proposed a collaborative filtering-based approach for managing consumer big data, incorporating purchasing patterns and frequently bought products. Their method combines extreme gradient boosting (EGB) with the word2vec model to analyze consumer click behavior for improved product recommendations. Furthermore, Yılmaz Benk et al. (2022) developed a predictive framework for consumer lifetime value (CLV) in multi-class e-commerce platforms, employing a multiple-output deep neural network

(DNN) and explainable AI. With the rapid growth of online shopping due to convenience, extensive product availability, competitive pricing, and efficiency, CLV has become a key metric for retailers. Their study introduced two additional factors—drifts in the amount spent (DAS) and diverse product class (DPC)—to enhance long-term profit prediction by identifying high-value consumers in real-time.

The key contributions of the proposed research for HUIS mining and product recommendation based on customer purchasing behaviour analysis are described as follows.

- In this research work, the researchers implemented a hybrid ensemble model for HUIS mining and product recommendation based on customer behaviour analysis.
- Another aim of this research is to utilize multiple datasets for training and cross-validation of the suggested model for HUIS mining and customer behaviour analysis with improved accuracy.
- Our model has been developed using the modulated ECLAT, and rule-based ensemble model for effective analysis of the customer purchasing behaviour.
- Further, the proposed model uses a soft voting mechanism for accurate feature extraction and selection, thereby, provides optimal performance metrics in comparison to the existing approaches.
- In addition to this, suggested model effectively reduces the storage complexity of transaction dataset and consume very minimal time in execution, thereby enhance the performance.

Organization of the paper

In this research work, a novel hybrid ensemble model has been proposed for high-utility itemset mining and product recommendation based on customer purchasing behaviour analysis. This research article has been presented in a multifarious section which is described as follows. The introduction part is described in section 1. Related work has been described in section 2. The proposed methodology has been disclosed in section 3. The outcomes of the research study are presented in section 4. Lastly, the conclusion of the research study is given in section 5.

2. Important Definitions

Let us assume a transaction data set defined by K that is of a finite length and includes n number of utility, i.e., $K = \{K_1, K_2, K_3, K_3, K_4, K_5, K_6, \dots \dots K_n,\}$. TRD represents the considered dataset having the following $TD_1, TD_2, TD_3, TD_4, TD_5, \dots . TD_6$, transactions. All recorded transactions contain an identical identifier ID, which is recognized as a recorded transaction ID, where TRD∈K. The utility of K is associated to utility factor that is composed of revenue per unit, internal as well as external utility sets. The internal utility list is denoted by $R(K_i, TRD_n)$ and external selected utility list is denoted by $S(K_i)$. The opted datasets having seven transactions including the nine different item sets are depicted in Table 1. All utility transactions are denoted by total number of utilities and corresponding quantities. The utility of the transaction has been represented by KU. Furthermore, the utility revenue in datasets is depicted in table 2. The important definitions are given below.

Table 1: Illustrates considered transaction datasets example.

ID	Different Transactions	KU
K_1	(a:1), (b:4), (e:5)	32

K_2	(a:2), (b:1), (c:4), (d:5), (g:2)	38
K_3	(a:3), (e:1), (f:2)	55
K_4	(c:1), (d:2), (e:1), (f:4)	47
K_5	(a:1), (b:4), (d:7), (e:1), (f:3), (g:3)	65
K_6	(a:1), (b:4), (c:2)	68
K_7	(c:1), (d:2), (e:3), (f:4)	53
K_8	(a:3), (b:1)	46

Table 2: Illustrates the utility revenue.

a	6
b	3
c	5
d	2
e	3
f	7
g	9

Definition 1. "Calculation of itemset utility": A multiplication of itemset revenue and its frequency within any transaction may be defined as itemset utility. Simply, it is described as $U(K_i) = R(K_i, TRD_n) \times S(K_i)$.

Example. Calculation of item "a" utility in transaction "1" is $6\times 1 = 6$, as well as item "g" utility in transaction "5" is $9\times 3 = 27$.

Definition 2. Calculation of itemset 'K' utility: In a database transaction K_n, item K utility may be evaluated as sum of all utilities itemset, i.e., $U(R, TRD_n) = \sum U(K_i, TRD_n)$ for TRD_n ∈K.

Example. Calculation for the {a, e} item in 'T3' transaction is $6 \times 3 + 3 \times 1 = 21$.

Definition 3. Calculation of transactional utility: It is described as addition of all the utility item in a particular transaction *TRD.* It is represented by the *TU(TRDn).*

Example. For instance, in a transaction K_2 , $KU(K_2) = K(b) + K(c) + K(d) + K(g)$=1+4+5+2=12

Definition 4. Calculation of weighted utility: In any specific transactional item K, weighted utility (WU) is the addition of all utilities consisting K.

Definition 5. Calculation of item K utility in transaction database (TRD): Estimation of an item K utility considering overall database is equal to addition of total utility containing K from each recorded transaction. For instance, $\mathrm{U(K)} = \sum(K, TRD)$.

Example. In the recorded transaction, utility item {a, b} = U ({a, b}, K1) +U ({a, b}, K2).

3. RELATED WORK

Amel Hidouri et al. in (Hidouri et al., 2021) carried out research on mining closed HUIS rooted in propositional satisfiability. In this research, the author initially proposed to utilize emblematic artificial intelligence (AI) to compute the group of entire closed HUIS from the transactional dataset. This developed approach is rooted in the minimization of enumeration issues of propositional satisfiability. However, this development comprises miscellaneous downsides which are related to the lower accuracy as well as more complexity in the implementing phase and grabs the attention to developing the novel enhanced model. In (Verma et al., 2021), A. Verma et al. carried out a research study on high utility and distinct itemset mining. In this research, the authors extended the earlier research to develop a novel algorithm that is capable to compute distinct high-utility itemsets as well as different patterns in one segment. The findings of the research study depict this developed algorithm is more efficient in comparison to the earlier two-phase algorithm. However, this developed algorithm has a few confines namely the lower prediction accuracy in HUIS mining, challenges in real-time implementation, as well as more execution time.

P. Amaranatha Reddy et al. in (Amaranatha Reddy & Hazarath Murali Krishna Prasad, 2023) proposed a new model based on HUIS mining through retail marketing datasets stream along with multifarious concession schemes utilizing the extending global utility itemset (EGUI). The HUIS mining is a revolutionary remodel form of the frequent set of items mining. This article explores an advanced model namely the EGUI for extraction of HUIS from multiple retail datasets along with the positive as well as multiple negative yield itemset. This explored model has a few limits related to the scalability as well as the speed of the model in real time. R. Vasumathi et al. in (Vasumathi & Murugan, 2019), discussed research on HUIS mining for big data analytics. This article aims to explore a novel approach that provides more effective outcomes for big data analysis in real time. However, this developed method is not effective in terms of certain parameters namely less memory space as well as less time consuming for large datasets, which grabs huge attention towards new model development.

In (Logeswaran et al., 2020), K. Logeswaran et al. explored another modern evolutionary method for HUIS mining. In recent years, various investigators have utilized multiple meta-heuristic computational protocols in the area of HUIS mining for customer behaviour analysis. However, such approaches still have frequent flaws related to the model's scalability as well as error rate. In (Oguz, 2022), D. Oguz explored another approach for HUIS mining. The HUIS mining finds the itemset which is traded together as well as has higher utility standards than a specified least utility threshold. This work demonstrates an asymmetric scheme wherein the integral utility standards are overlooked when discovering higher-utility itemset along with higher external utility standards. However, this proposed scheme is having few drawbacks related to the higher model execution time, and more computing cost.

T. L. Dam et al. (Dam et al., 2019), discovered a novel CLS-minier protocol for efficient closed HUIS mining. HUIS mining is a well-recognized dataset mining process with applicability in multifarious arenas. Though, the conventional HUIS mining protocols originate a massive number of HUIS and due to that analysis of the HUIS becomes very complicated. However, this suggested protocol CLS-miner also uses utility-list architecture to straight computing of utility itemset deprived of creating the candidates. Hence, due to the aforementioned drawback, the CLS-miner protocol is less effective in terms of proper utilization of the memory as well as more time-consuming. T. Wei et al. in (Wei et al., 2020), discussed another protocol namely the frequently closed higher-utility itemset (FCHUIS) for high-utility sets of items from the transactional database. This proposed protocol is rooted in the overall summary list architecture to store and retrieve the list of utilities without repeated scanning of the transactional datasets. However, the FCHUIS protocol is inefficient in terms of higher memory utilization and more time ingesting which might result in massive challenges for users to evaluate accurate outcomes quickly.

In (Lin et al., 2021), J. C. W. Lin et al. discussed another larger-scale data fusion approach for the integration of revealed closed higher-utility patterns via multifarious distributed transactional datasets. This generic composite architecture has been utilized for clustering transactions about the pertinent correlation which may confirm the completeness as well as the correctness of the fusion prototypical. However, this approach is computationally more complex in terms of handling the larger-scale dataset for integration of revealed knowledge via diverse distributed datasets for HUIS mining. R. Saeed et al. in (Saeed et al., 2021), discussed another tree-rooted protocol namely effectively correlated higher utility patterns (ECHUP) mining for higher utility patterns mining which has a stronger correlation. Further, a novel dataset architecture rooted in the utility tree which is called CoUTlist has been explored for storing sufficient data to mine the required patterns in real-time. However, this proposed approach has some limitations which are related to the excess time taken in the mining procedure as well as higher memory consumption. T. D. D. Nguyen et al. in (Nguyen et al., 2021), proposed a novel protocol for mining the closed higher utility sets of the items in dynamic profit transactional datasets. In recent decades, there have been discovered multifarious key issues in HUIS mining in the huge transactional database are related to the existing model's memory consumption, prediction accuracy as well as execution time. This proposed algorithm also is limited to providing pragmatic accuracy during frequent itemset mining. However, in the case of HUIS mining from a larger transactional dataset, it consumes more time, and due to this the overall computing complexity of the system model increases.

4. PROPOSED METHODOLOGY

Effective mining and evaluation of the item sets which have higher utility values within a transactional dataset is a critical task. HUIS mining is a pragmatic data mining approach that identifies the item sets in a transaction dataset containing high utility values. The utility value of the item set is a degree of how much vital a particular item set is for the customers. Moreover, it's based on both items set frequency and associated monetary. The proposed HUIS mining process includes multifarious steps. Initially, the entire dataset is pre-processed for removing the redundant or irrelevant item sets. In the later step, the utility function requires to be described for measuring the item set utility. There are numerous benefits of HUIS mining particularly for the retail business in terms of identifying the most marginal item sets in the transactional database. In addition to this, the HUIS is more effective in optimization of the marketing approaches, price strategies for increasing revenue, as well as inventory management. In this research work, a novel hybrid ensemble model is proposed for high-utility item set mining and product recommendation to the customers based on the effective evaluation of their purchasing behaviour.

4.1. Dataset Utilized

HUIS mining is one of the most widely recognized approaches that is capable to extract the item sets which has higher utility values within a transactional database. For training and validation of the proposed model, two diverse datasets have been selected namely the Chainstore and Foodmart. The aim of selecting the diverse dataset was to test and train the proposed model on a multiple massive dataset for measuring the performance metrics accurately in order to validate the model.

4.1.1. Chainstore

This dataset is one of the high benchmark datasets which is frequently utilized in the arena of time series prediction. This dataset comprises a weekly sailing dataset of 52 retail stores of the United States rooted retail chain. This dataset comprises multifarious essential information regarding the number of stores, total departments in a particular store, the sales figures of all the departments of the store every week, etc. This dataset is free and accessible and can be easily downloaded from multifarious resources namely the Kaggle or UCI machine learning (ML) repository.

4.1.2. Foodmart

This dataset is mainly a sampling dataset which is developed by Arbor Software, a dataset warehousing corporation. The Foodmart dataset pretends a diversified fictional grocery chain comprising sailing datasets from multifarious stores as well as departments. This dataset is frequently utilized for performance evaluation of data warehousing as well as tools of data mining. Moreover, it aids in development of novel system models and applications. Foodmart dataset is freely accessible and may be downloaded easily from diverse resources such as the GitHub repository or Pentaho Community Wiki.

4.2. Sample

In this research work, we selected two diverse datasets namely the Chainstore and Foodmart for training and validation of the proposed hybrid ensemble model for determining the customer purchasing behaviour and recommending them the best suitable product following their needs. Table 3 illustrates the selected transactional dataset samples taken from Chainstore dataset for the proposed hybrid ensemble model training and validation. The samples were taken from the Chainstore datasets from 2019-2023, arbitrarily. From the year 2019 to 2023, the number of people who viewed, cart, and purchased different products are 500, 590, 600, 700, 450, 450, 480, 299, 500, 332, and, 300, 320, 250, 480, 240, respectively.

Table 3: Illustrate selected transactional dataset samples taken from Chainstore dataset for proposed hybrid ensemble model training and validation. The samples were taken from the Chainstore datasets from 2019-2023, arbitrarily.

Years	Number of people		
	View	Cart	Purchase
2019	500	450	300
2020	590	480	320
2021	600	299	250
2022	700	500	480
2023	450	332	240

Table 4 illustrates the selected transactional dataset samples taken from Foodmart dataset for the proposed hybrid ensemble model training and validation. The samples were taken from the Foodmart datasets from 2019-2023, arbitrarily. From the year 2019 to 2023, the number of people who viewed, carted, and purchased different products are 450, 580, 700, 675, 402, 412, 355, 400, 499, 322 and 290, 275, 270, 378, 212, respectively.

Table 4: Illustrate selected transactional dataset samples taken from Foodmart dataset for proposed hybrid ensemble model training and validation. The samples were taken from the Foodmart datasets from 2019-2023, arbitrarily.

Years	Number of people		
	View	Cart	Purchase
2019	450	412	290
2020	580	355	275
2021	700	400	270
2022	675	499	378
2023	402	322	212

4.3. System Configuration:

This proposed hybrid ensemble model for HUIS mining and product recommendation based on the customer purchasing behaviour analysis has been implemented on a personal computing machine which comprises the following described system arrangement: Intel, 11th Gen i7 processor, Graphic Card: NVIDIA, RTX 3060 Ti, DDR5 SDRAM: 16 GB, solid-state drive (SSD): 1TB, 64-bit operating system (OS) and Window 11. This model was implemented on the widely recognized tool Jupyter notebook (anaconda 3) in python 3.11.2 programming language. Furthermore, the scikit-learn and Matplot libraries have also been utilized in this research work for programming and data visualization in real-time.

4.4. System Architecture:

Nowadays, accurate and fast predictive analysis of customer behaviour is a top concern in the E-Commerce environment to recommend the best relevant product to the customers. Figure 1 illustrates the proposed predictive hybrid ensemble model architecture. The functionality of our proposed hybrid ensemble predictive model is described as follows. Initially, the collected dataset is pre-processed in the data pre-processing pipeline and multiple steps are performed such as data cleaning or filtering, data subset selection, data normalization, and transformation of data into structured data. Data pre-processing is a very important and required step in model development as it cleans and transforms the raw dataset. Further, the data analysis procedure is done by employing the modulated equivalence class clustering and bottom-up lattice traversal (ECLAT) algorithm. This modulated ECLAT algorithm deeply analyses the pre-processed datasets and accurately optimizes the required parameters for further evaluation such as search product name, product type, price of the product, data of searching, time of searching, and many others. The modulated ECLAT architecture performs improved pruning and handles streaming data effectively. It minimizes number of undesired computations in transactions search procedure and optimizes memory utilization, leading to the fast execution, significantly.

The modulated ECLAT algorithm initially generates all itemset possible list that occurs within the transactional database together. Further, this scans the entire database for determining which itemset are occurring together. The modulated ECLAT algorithm is capable to identify the complementary services or products which are rarely bought by the consumer in real-time. In summary, the modulated ECLAT comprises the following steps:

Step 1: In the initial step, the algorithm initializes by producing all possible itemset lists which are being occurred in the transactional dataset,

Step 2: The modulated ECLAT algorithm do scanning of the entire transactional database for determining which of these item sets are frequently occurring together.

Step 3: Further, the algorithm counts the occurrence frequency for an entire list of itemset within the transactional database.

Step 4: The modulated ECLAT algorithm, then trim the entire item sets list as well as remove which are not occurring together very frequently.

Step 5: In this step, the algorithm continuously repeats the entire procedure, as well as produces a novel possible item set list rooted on the trim list using the last iteration. Further, it also scans again the database for counting the occurrence frequency.

Step 6: This modulated ECLAT algorithm continues the same procedure, until the HUIS list is obtained, or till the pre-set stop rule is met.

Step 7: The modulated ECLAT protocol, then outputs the list of HUIS which are obtained in the process.

Further, a rule-based ensemble model is used which combines multiple algorithms such as transfer learning, XG-Boost and Random Forest to make predictions or decisions. In this rule-based mechanism, the outcomes of individual algorithms namely transfer learning, XG-Boost and Random Forest are combined for producing a final prediction. The rule-based ensemble model uses the weighted voting as an aggregation strategy. The data augmentation is done for artificially expansion of the datasets, reducing the overfitting and better generalization of the model. In the next step, the 10-Fold cross-validation operation is performed on the analysed dataset. This cross-validation method divides entire obtained optimized datasets into 10 parts, arbitrarily. By performing the 10-Fold cross-validation process the data is split into training data as well as testing data. For this proposed model performance computation, 80% dataset was used for training and the rest reserve 20% data is utilized for model testing. The 10-fold cross-validation is a method that is utilized for effective performance evaluation of the ML-based model. This technique partitions the entire data into 10-folds equally and iteratively tests and trains the model on multifarious fold combinations for accurate estimation of the model metrics in comparison to the normal train-test split approach.

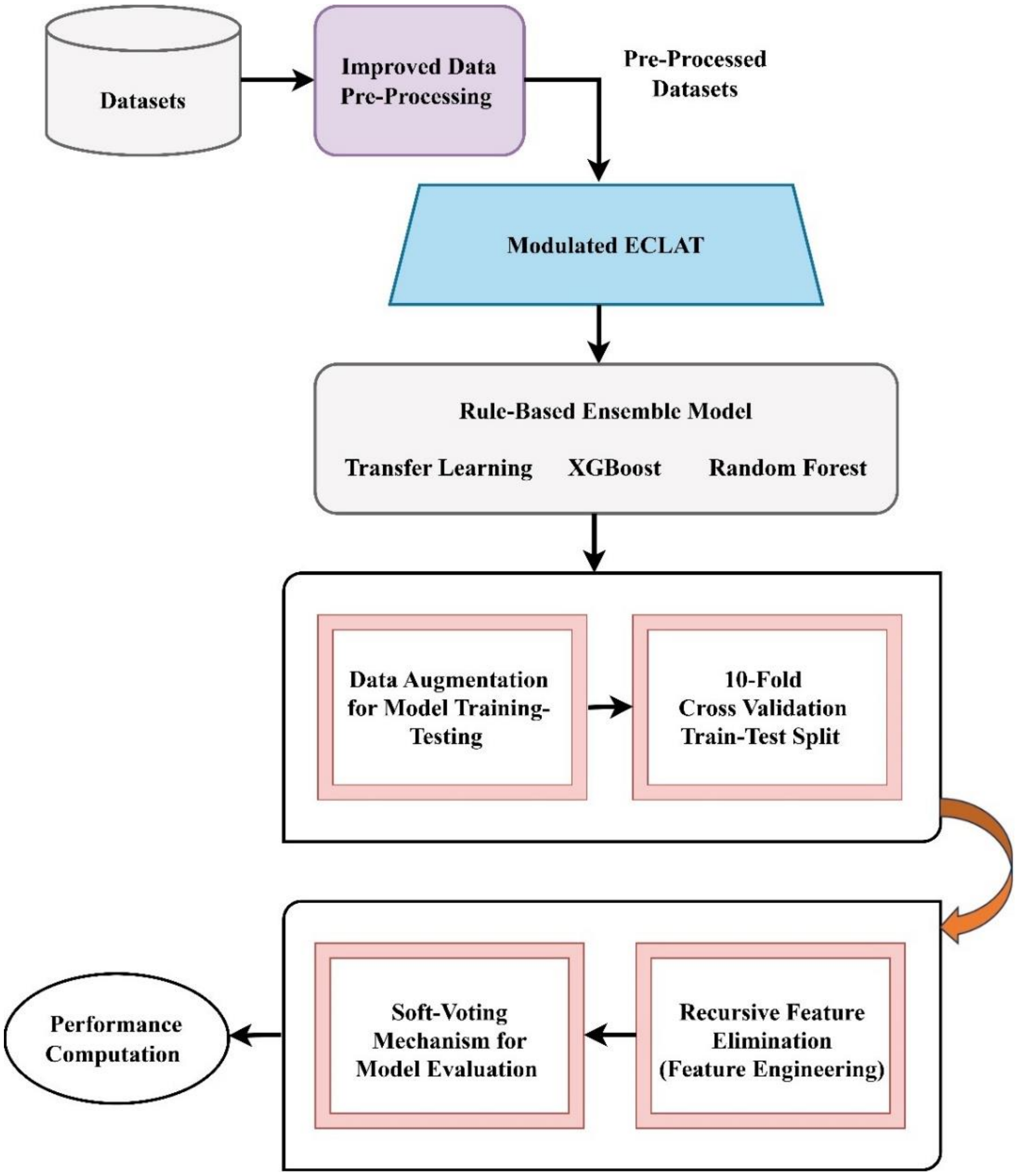

Figure 1: Illustrates proposed predictive hybrid ensemble model architecture.

The feature extraction is done utilizing the feature engineering approach namely recursive feature elimination. Furthermore, the features are selected and analysed using the soft-voting mechanism. Lastly, the model performance is computed on diverse epochs to validate this proposed model.

The process of the Recursive Feature Elimination algorithm is summarized in the following steps as follows:

Step 1: To originate the list of each possible itemset as well as their corresponding frequency within the defined transactional dataset.

Step 2: Eliminate the infrequent itemset from the obtained HUIS list rooted upon minimal support.

Step 3: Originate the list of frequent itemset of dimension two by joining the group of frequent items. For instance, if {X} and {Y} are considered frequent, in such case the {X, Y} is considered a candidate itemset.

Step 4: To, trim the candidate itemset list, which comprises an infrequent subset. For instance, if {X, Y} is considered a candidate itemset, while the {A} is found infrequent, in such a case, {X, Y} is to be pruned.

Step 5: To originate the frequent itemset list which has a dimension of three by combining the group of frequent itemset. For instance, if {X, Y} and {Y, Z} are considered frequent, in such a scenario the {A, B, C} is to be considered a candidate itemset.

Step 6: Repeat the procedure of combining as well as trimming until no more frequent HUIS is produced.

This Recursive Feature Elimination algorithm is based on multifarious association rules. These association rules mean here multiple statements used in the algorithm such as "if-then" which assists to illustrate the likelihood of relationship amongst multifarious data items and based upon that the predictive analytics is done for effective product recommendations. Thereby, the proposed hybrid ensemble model is an effective approach for customer behaviour analysis in real-time and based upon that recommends the HUIS to the customers in accurate and very less time.

Pseudo code: The Pseudo code for the proposed hybrid ensemble model for HUIS mining and product recommendation following the real-time analysis of the customer behaviour is described as follows.

Input: Datasets D1_Fc and D2_Fm i.e., Chainstore and Foodmart.

Output: All relevant HUIS products based on customer behaviour analysis.

Step 1: Scanning of D1_Fc and D2_Fm for receiving the united raw dataset i.e., UDs;

Step 2: product_List ← group of unique products in UDs saved on the support decreasing order;

Step 3: Run the algorithm for data pre-processing and obtains the datasets Xpp;

Step 4: Xpp ← parameter optimization using the modulated ECLAT algorithm to obtain Op;

Step 5: **if** product_List is not Null **do**

Step 6: Mx ← the end product in the product_List;

Step 7: through the top table of Xpp following the interconnection of nodes tagged Xpp to make the PoHUIS (Xpp);

Step 8: **if** p(Xpp) ≥ min_Us **then**

Step 9: PoHUIS (Xpp)← PoHUIS (Xpp) ∪ (Xpp)

Step 10: end if

Step 11: if p(Xpp) + tp(Xpp) ≥ min_Us **then**

Step 12: Call find (Xpp, PoHUIS (Xpp), min_Us, min_Cr)

Step 13: end if

4.5. Performance Evaluation Metrics

The evaluation of the proposed hybrid ensemble model has been carried out as means of multifarious performance metrics such as accuracy, precision, recall, and F1 score. All, this abovementioned evaluation matrix with the relevant mathematical equation is defined as follows.

Accuracy is a widely-recognized evaluation metric that is utilized to determine the performance of ML-based models. The accuracy metric is a degree of a proportion of instances that may be accurately categorized through the prototypical, out of entire instances. The accuracy of the proposed hybrid ensemble model may be evaluated using the following equation 1.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Recall is another performance metric used in ML-based models. It is a metric that measures the proportion of TP predictions amongst every actual positive occurrence. The Recall value may be evaluated for the proposed hybrid ensemble model by using the following equation 2.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

The precision metric is another important performance evaluation metric of the ML-rooted models. This metric is a measure of the proportions of TP predictions amongst every positive prediction. This evaluation metric is also recognized as a positive predictive value (PPV). The precision value for the proposed model may be evaluated by using the following equation 3.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

The F1 score is an evaluation metric that groups the recall and precision values within a single score. This is harmonic means of recall as well as precision and offers a composed measure of recall and precision metrics. The F1 score may be evaluated by using the following equation 4.

$$F1\ score = \frac{2 \times Recall \times Precision}{Recall+Precision} \tag{4}$$

5. IMPLEMENTATION RESULT AND DISCUSSION

The amount of data is constantly increasing all across the globe owing to the rapid development of e-commerce applications. The effective utilization of such data is very intricate for customer behaviour analytics, particularly recommending the users one of the pragmatic products following the consumer demands. However, finding a specified product that meets consumer requirements, is a very challenging job owing to the massive number of accessible products over the e-commerce platforms. There have been developed multifarious recommendation systems architecture and models utilizing state-of-the-art algorithms for improving the customer's experience or marketing outcomes through optimization of user choice-making process. In the modern era, forecasting seems to be the product recommendation premise, such that precise forecasting of the customer’s buying inclination can enhance the recommendation consequence and enhance the system. In the initial phase, owing to the immaturity of diverse accessible e-commerce applications and minimal datasets related to customer purchasing behaviour analysis, the investigation on its forecasting has been stagnant during the last decade. In recent years, owing to the fast progress of information technology (IT) infrastructure, a plethora of datasets on customer purchasing behaviour is collected utilizing the multifarious e-commerce applications that offer robust support for forecasting investigation.

During the last decade, diverse state-of-the-art methods have been developed and utilized by investigators for forecasting consumer behaviour following their product buying behaviour. However, such developed models for customer products purchasing behaviour analysis have miscellaneous limitations such as lower prediction accuracy of suitable products following the user requirement, implementation computational complexity in e-commerce applications, large time consumption in prediction, and many more. For, resolving these abovementioned key challenges, in this work, a novel hybrid ensemble model has been developed for HUIS mining from a transactional database for customer behaviour analysis.

Table 5: Illustrates the system environment and specifications.

System Component	Specifications
Processor	Intel, 11th Gen i7 processor
Graphic Card	NVIDIA, RTX 3060 Ti
RAM	DDR5 SDRAM: 16 GB
SSD	1 TB
Operating System	64-Bit
Installed Window	Window 11 version
Programming Language	Python 3.11.2
Programming Environment	Jupyter notebook (anaconda 3)
Used Libraries	Scikit-learn and matplotlib

The overall experimental settings are summarized in table 5. This experimental research work is carried out on a personal computing system which comprises the following described system arrangement: Intel, 11th Gen i7 processor, Graphic Card: NVIDIA, RTX 3060 Ti, DDR5 SDRAM: 16 GB, solid-state drive (SSD): 1TB, 64-bit operating system (OS) installed with a Window 11 environment. This proposed model has been implemented on the widely acknowledged tool Jupyter notebook (anaconda 3) in python 3.11.2 programming language. Moreover, the scikit-learn, as well as Matplot library, has also been utilized in this experimental research work for programming and data visualization in real-time.

Table 6: Illustrates the used dataset information in model training and testing.

Dataset Parameters	Information
Overall records	255,152
Overall purchased items	12762
Unique items	896
Unique products ID	1175
Unique customer numbers	48747
Used training datasets	80%

Used testing datasets	20%

Table 6 summarizes the used dataset information in model training as well as the testing procedure of the proposed hybrid ensemble model for HUIS mining and in accordance recommends a suitable product to the user in real-time searching for the desired item. During experimental work, we selected 255,152 overall products from the selected datasets i.e., Chainstore and Foodmart. Furthermore, we selected 12762 overall purchased items, 896 unique items, 1175 unique products ID, 48747 unique customer numbers, and training and testing datasets at 80% and 20%, respectively.

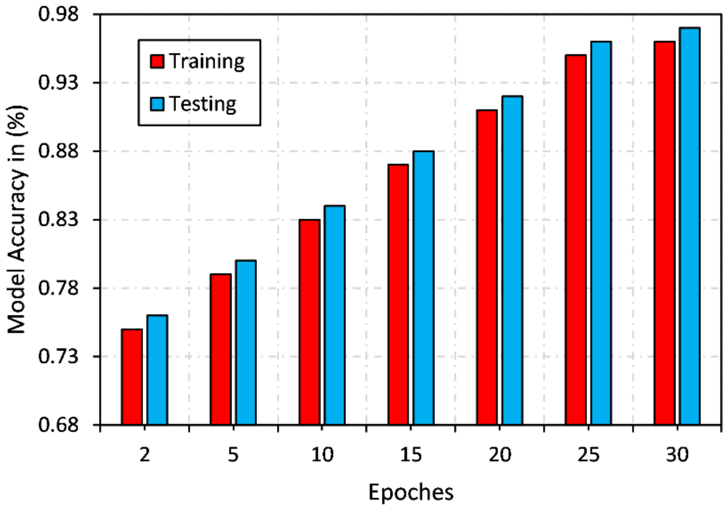

Figure 2: Demonstrates the training and testing accuracy of the proposed hybrid ensemble model.

Figure 2 demonstrates the training and testing accuracy of the proposed hybrid ensemble model for HUIS mining and product recommendation to the consumers based on their behavioural analytics. Our proposed hybrid ensemble model recorded training accuracy on the epochs 2, 5, 10, 15, 20, 25, and 30 is 0.75, 0.79, 0.83, 0.87, 0.91, 0.95, and 0.96, respectively. Further, the measured testing accuracy of the proposed hybrid ensemble model on the epochs 2, 5, 10, 15, 20, 25, and 30 is 0.76, 0.80, 0.84, 0.88, 0.92, 0.96, and 0.97, respectively. The accuracy finding of the proposed model on the chosen datasets is recorded and enhanced on the diverse epochs in the training and testing phase, which is more practical and optimal while comparing in real-time.

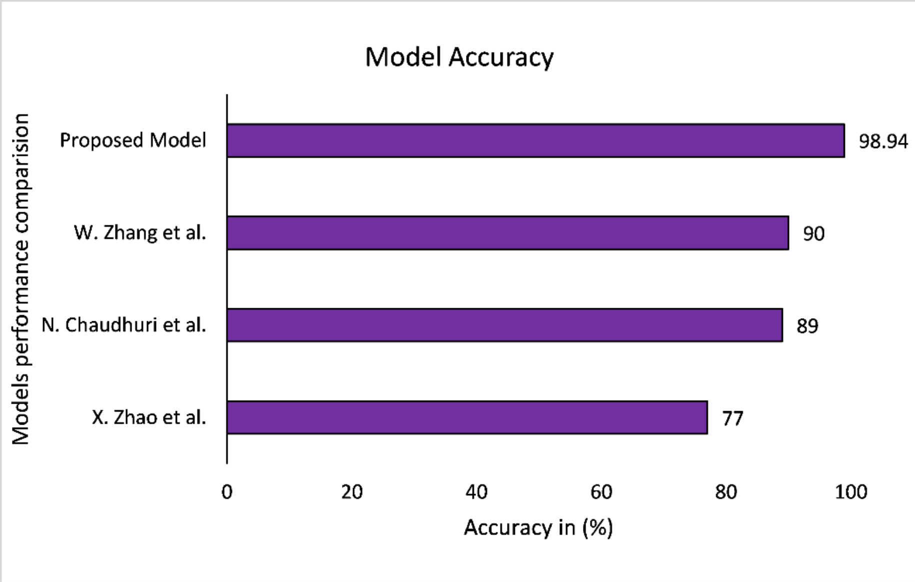

Figure 3: Illustrates the accuracy comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets (Chaudhuri et al., 2021; W. Zhang & Wang, 2021; Zhao & Keikhosrokiani, 2022)**.**

Figure 3 illustrates the accuracy comparison between the proposed hybrid ensemble model and previous methods, considering the Foodmart datasets. The measured accuracy of the existing models i.e., X. Zhao et al. (Zhao & Keikhosrokiani, 2022), N. Chaudhuri et al. (Chaudhuri et al., 2021), and W. Zhang et al. (W. Zhang & Wang, 2021) was 89 %, 90%, and 98.94 %, respectively. In contrast, the recorded accuracy of the proposed hybrid ensemble model in view of the HUIS mining and product recommendation to the consumer based on their behaviour analysis is recorded at 98.94%, on the Foodmart dataset. The accuracy comparative analysis of the suggested hybrid ensemble model and existing models shows that our models offer very optimal and improved accuracy in comparison to the previous work.

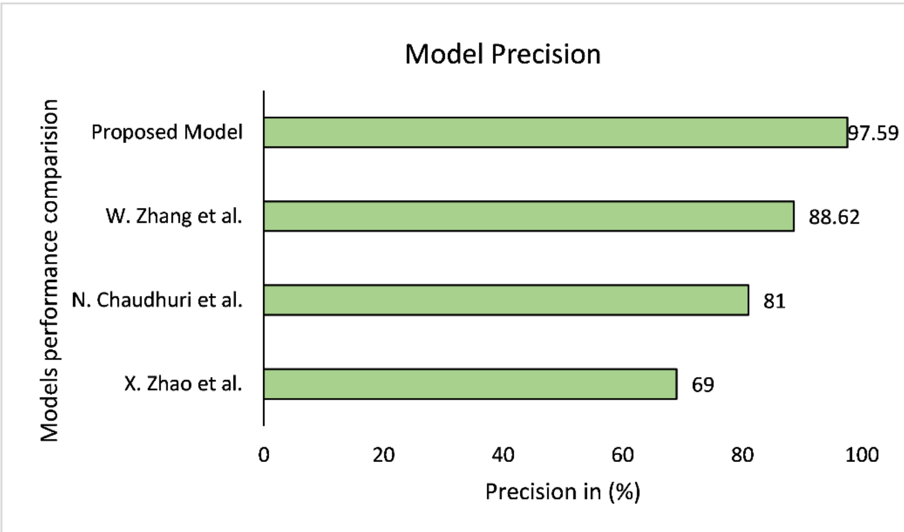

Figure 4: Illustrates the precision comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets (Chaudhuri et al., 2021; W. Zhang & Wang, 2021; Zhao & Keikhosrokiani, 2022)**.**

Figure 4 illustrates the precision comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets. The measured precision value of the existing models i.e., X. Zhao et al. (Zhao & Keikhosrokiani, 2022), N. Chaudhuri et al. (Chaudhuri et al., 2021), and W. Zhang et al. (W. Zhang & Wang, 2021) was 69%, 81%, and 88.62%, respectively. In contrast, the recorded precision value of the proposed hybrid ensemble model in view of the HUIS mining and product recommendation to the consumer based on their behaviour analysis is recorded at 97.59%, on the Foodmart dataset. The precision comparative analysis of the suggested hybrid ensemble model and existing models shows that the proposed hybrid ensemble models offer very pragmatic and enhanced precision in comparison to the previous models.

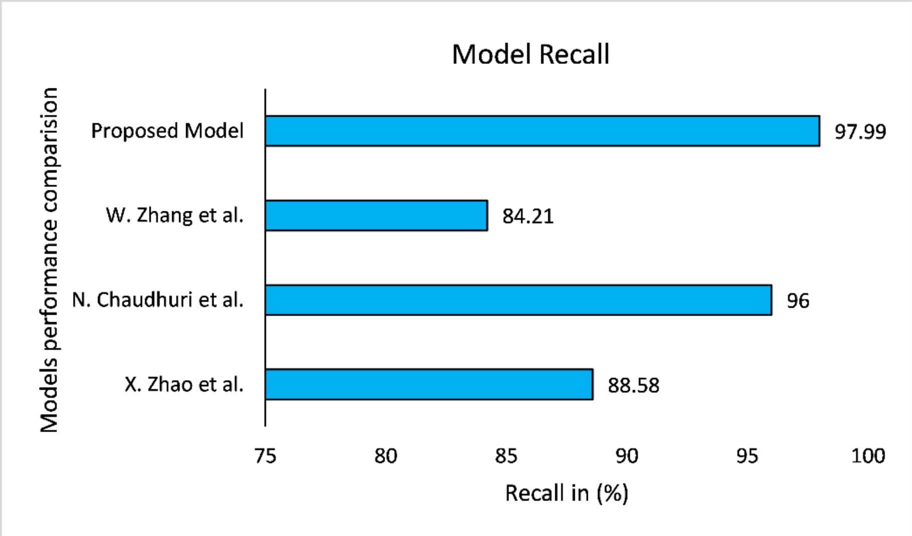

Figure 5: Illustrates the recall comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets (Chaudhuri et al., 2021; W. Zhang & Wang, 2021; Zhao & Keikhosrokiani, 2022)**.**

Figure 5 illustrates the recall value comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets. The measured recall value of the existing models i.e., X. Zhao et al. (Zhao & Keikhosrokiani, 2022), N. Chaudhuri et al. (Chaudhuri et al., 2021), and W. Zhang et al. (W. Zhang & Wang, 2021) was 88.58%, 96%, and 84.21%, respectively. In contrast, the recorded recall value of the proposed hybrid ensemble model in view of the HUIS mining and product recommendation to the consumer based on their behaviour analysis is recorded at 97.99%, on the Foodmart dataset. The recall comparative analysis of the suggested hybrid ensemble model and existing models shows that the proposed hybrid ensemble models provide more optimal and improvised precision in comparison to the previous models.

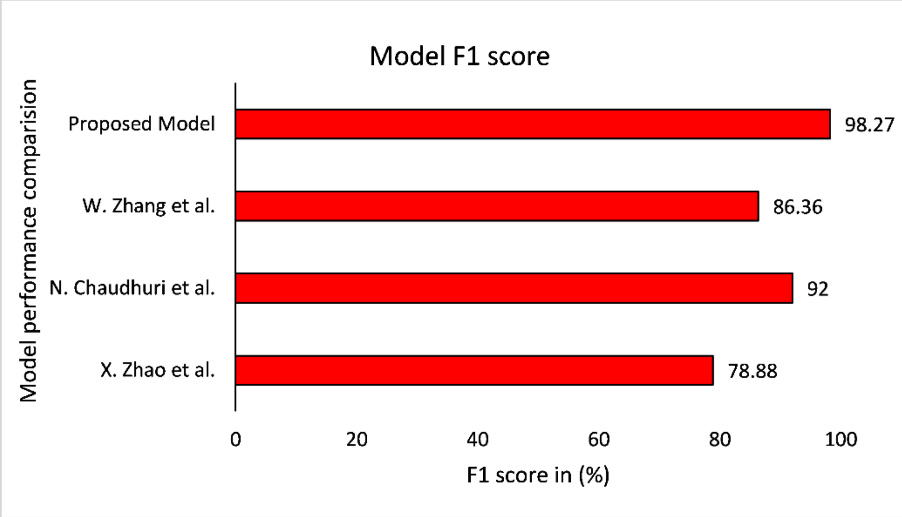

Figure 6: Illustrates the F1 score comparison between the proposed hybrid ensemble model and previous models, considering the Foodmart datasets (Chaudhuri et al., 2021; W. Zhang & Wang, 2021; Zhao & Keikhosrokiani, 2022)**.**

Figure 6 illustrates the measured F1 score value comparison between the suggested hybrid ensemble model as well as previous models, considering the Foodmart datasets. The measured F1 score value of the existing models i.e., X. Zhao et al. (Zhao & Keikhosrokiani, 2022), N. Chaudhuri et al. (Chaudhuri et al., 2021), and W. Zhang et al. (W. Zhang & Wang, 2021) was 78.88%, 92%, and 86.36%, respectively. In contrast, the recorded F1 score value of the proposed hybrid ensemble model in view of the HUIS mining and product recommendation to the consumer based on their behaviour analysis is recorded at 98.27%, on the Foodmart dataset. The F1 score comparative analysis of the proposed hybrid ensemble model and existing models shows that the proposed hybrid ensemble models provide a more optimal and enhanced F1 score in comparison to the previous works.

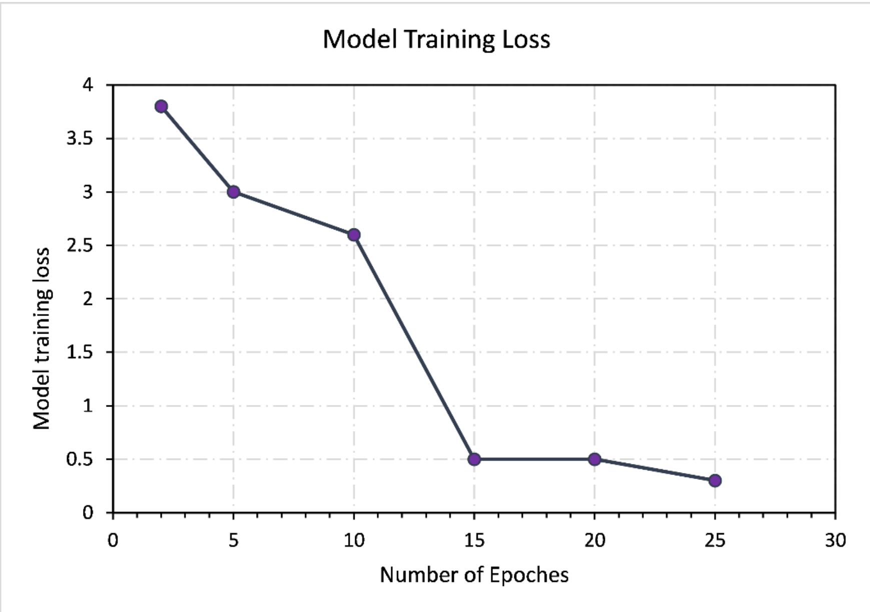

Figure 7: Illustrates the measured training losses of the proposed hybrid ensemble model.

Figure 7 illustrates the measured training losses of the proposed hybrid ensemble model for HUIS mining and product recommendation to the uses as per the real-time behaviour analysis. The measured training losses of the proposed hybrid ensemble model on the diverse epochs i.e., 0, 5, 10, 15, 20, 25, and 30 are 3.8%, 3%, 2.6%, 0.5%, 0.5%, and 0.3%, respectively. The findings demonstrate that our suggested hybrid ensemble models are more realistic and continuously decrease the losses during different epochs in the training phase.

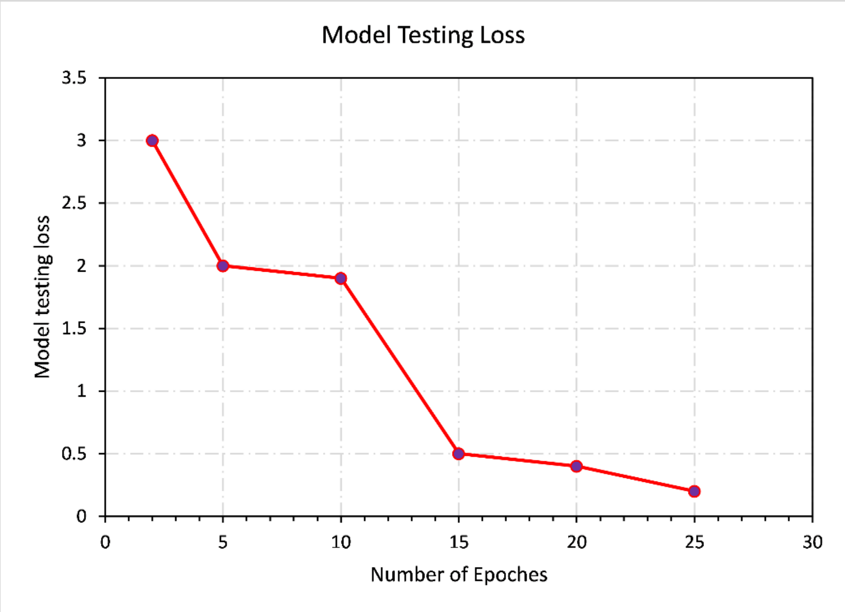

Figure 8: Illustrates the measured testing losses of the proposed hybrid ensemble model.

Figure 8 illustrates the measured testing losses of the proposed hybrid ensemble model for HUIS mining and product recommendation to the uses as per the real-time behaviour analysis. The measured testing losses of the proposed hybrid ensemble model on the diverse epochs i.e., 0, 5, 10, 15, 20, 25, and 30 are 3%, 2%, 1.9%, 0.5%, 0.4%, and 0.2%, respectively. The findings demonstrate that our suggested hybrid ensemble model is more practical and continuously minimizes the losses during different epochs in the testing phase.

Table 7: Illustrates the execution time comparison of the proposed hybrid ensemble model and existing models.

Different Models	Execution time (in Seconds)
R. Saeed et al. (Saeed et al., 2021)	100 seconds
B. Vo et al. (Vo et al., 2020)	60 seconds
Proposed hybrid ensemble model	14 seconds

Table 7 illustrates the execution time comparison of the proposed hybrid ensemble model and existing models. The existing models i.e., R. Saeed et al. (Saeed et al., 2021) and B. Vo et al. (Vo et al., 2020) consume 100 seconds and 60 seconds in the execution. In contrast, the proposed hybrid ensemble model for HUIS mining and product recommendation to the consumer based on their behaviour analysis takes only 14 seconds to the execution in comparison to the R. Saeed et al. (Saeed et al., 2021) and B. Vo et al. (Vo et al., 2020) models. The execution time plays a significant role in the model performance evaluation, the less time in execution, the better the model as it provides faster results with minimal effort. Thus, the proposed hybrid ensemble model for HUIS mining and product recommendations to the customers is more pragmatic and takes very minimal time the execution when compared to the existing models.

Table 8: Details of Dataset split ratio (Train-Test) for performance metrics evaluation.

Performance Metrics	**Dataset split ratio (Train-test)**			
	80:20(%)	70:30(%)	90:10(%)	60:40(%)
Accuracy	98.94%	97.22%	96.88%	95.22%
Recall	97.99%	96.77%	94.33%	94.36%
Precision	97.59%	92.88%	95.25%	93.56%
F1-Score	98.27%	93.71%	93.11%	92.35%

Table 8 provides the details of dataset split ratio (train-test) for performance metrics evaluation of proposed hybrid ensemble model for customer purchasing behaviour analysis and product recommendations. The optimized performance has been recorded for the portioned ratio 80:20 with predicted scores 98.94%, 97.99%, 97.59%, and 98.27%, for accuracy, recall, precision and F1-score, respectively.

6. CONCLUSION

HUIS mining is a data mining approach that is utilized for discovering the different patterns as well as a relationship amongst the products within the transactional database. One of the keys aims of the HUIS mining technique is to identify the item sets which recurrently happen together and have a higher influence over certain measures of interest namely the consumer satisfaction level or overall profit revenue. Likewise, conventional itemset mining concentrates on the identification of the recurrent set of items regardless of a distinct utility. HUIS mining considers the significance of every item set, considering their distinct value or weight. It allows the searching of the set of items that have a higher overall utility, even if certain sets of items seem rare or have a lower support. In the last decade, numerous distinct methods have been explored by the investigators for HUIS mining, however, such developed models do not analyse the behaviour of the customer in real-time in a more accurate way. Further, the existing system has some other limitations which include more computational time, intricate implementations etc. To solve, these aforementioned limitations, a novel hybrid ensemble model is presented for HUIS mining and product recommendation based on the customer purchasing behaviour analysis in real-time. The outcomes of the proposed hybrid ensemble model are very optimal and improved in terms of distinct performance metrics. The recorded accuracy, precision, recall, and F1 score of our proposed hybrid ensemble model is 98.94%, 97.59%, 97.99%, and F1 98.27%, respectively. Further, our novel hybrid ensemble model consumes only 14 seconds in the execution phase which is very much minimal in comparison to the previous system models. Although, during the last decade, several models have been explored for HUIS mining and customer behaviour analytics. However, still there is a mammoth scope for advanced investigation to build novel as well as more enhanced models for HUIS mining in a fast manner to analyse the customer’s product purchasing behaviour in real-time.

REFERENCES

1. Agarwal, R., Gautam, A., Saksena, A. K., Rai, A., & Karatangi, S. V. (2021). Method for Mining Frequent Item Sets Considering Average Utility. *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 275–278. https://doi.org/10.1109/ESCI50559.2021.9396947
2. Ali, N., Alshahrani, A., Alghamdi, A., & Novikov, B. (2022). SmartTips: Online Products Recommendations System Based on Analyzing Customers Reviews. *Applied Sciences*, *12*, 8823. https://doi.org/10.3390/app12178823
3. Amaranatha Reddy, P., & Hazarath Murali Krishna Prasad, M. (2023). High Utility Item-set Mining from retail market data stream with various discount strategies using EGUI-tree. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03341-3
4. Bhuvaneswari, M. S., Balaganesh, N., & Muneeswaran, K. (2022). H-Map-Based Technique for Mining High Average Utility Itemset. *IETE Journal of Research*, *0*(0), 1–13. https://doi.org/10.1080/03772063.2022.2075800
5. Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2021.113622
6. Cheng, Z., Shen, W., Fang, W., & Lin, J. C.-W. (2023). A Parallel High-Utility Itemset Mining Algorithm Based on Hadoop. *Complex System Modeling and Simulation*, *3*(1), 47–58. https://doi.org/10.23919/csms.2022.0023
7. Dam, T. L., Li, K., Fournier-Viger, P., & Duong, Q. H. (2019). CLS-Miner: efficient and effective closed high-utility itemset mining. *Frontiers of Computer Science*. https://doi.org/10.1007/s11704-016-6245-4
8. Dao, H. N., Ravikumar, P., Likhitha, P., Rage, U. K., Watanobe, Y., & Paik, I. (2023). Finding Stable Periodic-Frequent Itemsets in Big Columnar Databases. *IEEE Access*, *11*, 12504–12524. https://doi.org/10.1109/ACCESS.2023.3241313
9. He, R., Chen, J., Du, C., & Duan, Y. (2021). Stable Periodic Frequent Itemset Mining on Uncertain Datasets. *2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, 263–267. https://doi.org/10.1109/CCET52649.2021.9544352
10. Hidouri, A., Jabbour, S., Raddaoui, B., & Yaghlane, B. Ben. (2021). Mining Closed High Utility Itemsets based on Propositional Satisfiability. *Data & Knowledge Engineering*, *136*, 101927. https://doi.org/https://doi.org/10.1016/j.datak.2021.101927
11. Hung, N. M., NT, T., & Vo, B. (2021). A General Method for mining high-Utility itemsets with correlated measures. *Journal of Information and Telecommunication*. https://doi.org/10.1080/24751839.2021.1937465
12. Lakshmi, N., & Krishnamurthy, M. (2022). Frequent Itemset Generation Using Association Rule Mining Based on Hybrid Neural Network Based Billiard Inspired Optimization. *Journal of Circuits, Systems and Computers*. https://doi.org/10.1142/S0218126622501389
13. Lin, J. C. W., Djenouri, Y., & Srivastava, G. (2021). Efficient closed high-utility pattern fusion model in large-scale databases. *Information Fusion*. https://doi.org/10.1016/j.inffus.2021.05.011
14. Lin, J. C. W., Pirouz, M., Djenouri, Y., Cheng, C. F., & Ahmed, U. (2020). Incrementally updating the high average-utility patterns with pre-large concept. *Applied Intelligence*. https://doi.org/10.1007/s10489-020-01743-y
15. Logeswaran, K., Suresh, P., Ponselvakumar, A. P., Renuga, G., Priyadharshini, M., & Nivetha, R. (2020). Contemplate study of contemporary evolutionary approaches used for mining high utility item set. *International Journal of Advanced Science and Technology*.

16. Luna, J. M., Kiran, R. U., Fournier-Viger, P., & Ventura, S. (2023). Efficient mining of top-k high utility itemsets through genetic algorithms. *Information Sciences*, *624*, 529–553. https://doi.org/https://doi.org/10.1016/j.ins.2022.12.092
17. Ma, C., Wang, B., Jooste, K., Zhang, Z., & Ping, Y. (2020). Practical Privacy-Preserving Frequent Itemset Mining on Supermarket Transactions. *IEEE Systems Journal*. https://doi.org/10.1109/JSYST.2019.2922281
18. Magdy, M., Ghaleb, F. F. M., Mohamed, D. A. E. A., & Zakaria, W. (2023). CC-IFIM: an efficient approach for incremental frequent itemset mining based on closed candidates. *The Journal of Supercomputing*, *79*(7), 7877–7899. https://doi.org/10.1007/s11227-022-04976-5
19. Nguyen, T. D. D., Nguyen, L. T. T., Vu, L., Vo, B., & Pedrycz, W. (2021). Efficient algorithms for mining closed high utility itemsets in dynamic profit databases. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2021.115741
20. Oguz, D. (2022). Ignoring Internal Utilities in High-Utility Itemset Mining. *Symmetry*, *14*(11). https://doi.org/10.3390/sym14112339
21. Qiu, S., Wang, B., Li, M., Liu, J., & Shi, Y. (2020). Toward Practical Privacy-Preserving Frequent Itemset Mining on Encrypted Cloud Data. *IEEE Transactions on Cloud Computing*. https://doi.org/10.1109/TCC.2017.2739146
22. Saeed, R., Rauf, A., Quradaa, F. H., & Asim, S. M. (2021). Efficient Utility Tree-Based Algorithm to Mine High Utility Patterns Having Strong Correlation. *Complexity*. https://doi.org/10.1155/2021/7310137
23. Sekhavat, Y. A. (2020). CFM: Collusion-free model of privacy preserving frequent itemset mining. *International Journal of Information and Computer Security*. https://doi.org/10.1504/IJICS.2020.109476
24. Shahbazi, Z., Hazra, D., Park, S., & Byun, Y. C. (2020). Toward improving the prediction accuracy of product recommendation system using extreme gradient boosting and encoding approaches. *Symmetry*. https://doi.org/10.3390/SYM12091566
25. Shayegan Fard, M. J., & Namin, P. A. (2020). Review of Apriori based Frequent Itemset Mining Solutions on Big Data. *2020 6th International Conference on Web Research, ICWR 2020*. https://doi.org/10.1109/ICWR49608.2020.9122295
26. Sohrabi, M. K. (2020). An efficient projection-based method for high utility itemset mining using a novel pruning approach on the utility matrix. *Knowledge and Information Systems*. https://doi.org/10.1007/s10115-020-01485-w
27. Vasumathi, R., & Murugan, S. (2019). Efficient high average-utility pattern mining for big data. *International Journal of Scientific and Technology Research*.
28. Verma, A., Dawar, S., Kumar, R., Navathe, S., & Goyal, V. (2021). High-utility and diverse itemset mining. *Applied Intelligence*. https://doi.org/10.1007/s10489-020-02063-x
29. Vo, B., Nguyen, L. T. T., Bui, N., Nguyen, T. D. D., Huynh, V. N., & Hong, T. P. (2020). An Efficient Method for Mining Closed Potential High-Utility Itemsets. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2974104
30. Wei, T., Wang, B., Zhang, Y., Hu, K., Yao, Y., & Liu, H. (2020). FCHUIM: Efficient Frequent and Closed High-Utility Itemsets Mining. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3001975
31. Wu, J. M. T., Srivastava, G., Wei, M., Yun, U., & Lin, J. C. W. (2021). Fuzzy high-utility pattern mining in parallel and distributed Hadoop framework. *Information Sciences*. https://doi.org/10.1016/j.ins.2020.12.004
32. Wu, J. M. T., Teng, Q., Lin, J. C. W., & Cheng, C. F. (2020). Incrementally Updating the Discovered High Average-Utility Patterns with the Pre-Large Concept. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.2982415

33. Yang, Y., Ding, J., Wang, H., Xing, H., & Li, E. (2023). A High Utility Itemset Mining Algorithm Based on Particle Filter. *Mathematical Problems in Engineering*, *2023*, 7941673. https://doi.org/10.1155/2023/7941673
34. Yılmaz Benk, G., Badur, B., & Mardikyan, S. (2022). A New 360° Framework to Predict Customer Lifetime Value for Multi-Category E-Commerce Companies Using a Multi-Output Deep Neural Network and Explainable Artificial Intelligence. *Information*, *13*(8). https://doi.org/10.3390/info13080373
35. Yuan, S. (2022). Analysis of Consumer Behavior Data Based on Deep Neural Network Model. *Journal of Function Spaces*, *2022*, 4938278. https://doi.org/10.1155/2022/4938278
36. Zhang, K., Zhang, Y., & Wang, Z. (2020). Frequent Pattern Mining Based On Occupation and Correlation. *2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT)*, 161–166. https://doi.org/10.1109/ICEICT51264.2020.9334367
37. Zhang, W., & Wang, M. (2021). An improved deep forest model for prediction of e-commerce consumers’ repurchase behavior. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0255906
38. Zhao, X., & Keikhosrokiani, P. (2022). Sales prediction and product recommendation model through user behavior analytics. *Computers, Materials and Continua*. https://doi.org/10.32604/cmc.2022.019750