

LITERATURE SURVEY ON HIGH UTILITY ITEM-SET MINING AND CUSTOMER PURCHASING BEHAVIOUR USING HYBRID MODEL

Abstract

High profitable rare item sets, is an approach to extract out all the products that emerges high profit over selling. Data mining is a technique that collects data and figure out all relevant products which urges high profit over selling and having high margins. In this chapter, keeping in mind customers' purchasing behavior, profitable rare item sets, actually want to figure out all those products which can earn us high margin profit, selling of these products through e-trade can lead us with high profit. Rare item sets mining is a challenging task where the key issues are-identifying interesting rare patterns and efficiently discovering them in large datasets, This data can contribute to have high access to all those products which are rarely purchased but consuming high margins, which in future can lead to high business utilities in e-trade .The development of a sequential pattern mining framework based on LSTM networks, enhancing the deep learning framework with advanced techniques, evaluating the proposed approach's performance, and providing valuable insights and recommendations based on the discovered high utility item sets. By achieving these objectives, this research aims to contribute to the advancement of high utility item set mining techniques and provides practical solutions for businesses to extract valuable insights from their transactional datasets & samples.

Keywords: High utility itemset mining, High LSTM, Machine learning, Deep learning, Neural networks.

Authors

Aditya Tiwari

Assistant Professor
Rungta College of Engineering &
Technology- Bhilai, Chhattisgarh Swami
Vivekanand University, Bhilai,
Chhattisgarh, India.
adityarise0609@gmail.com

Dr. Jyothi Pillai

Professor
BIT-Raipur
Chhattisgarh, India.
jyothi.pillai@bitdurg.ac.in

Om Prakash Barapatre

Assistant Professor
BIT-Raipur
Chhattisgarh, India.
omprakashbarapatre@bitdurg.ac.in

K Ganpati Shrinivas Sharma

Assistant Professor
Electronics & Computer Engineering
BIT, Raipur
New Raipur, Chhattisgarh, India.
ganpatishrinivas01@gmail.com

Dushyant Kumar Yadav

Assistant Professor
Department of Mathematics
UTD, CSVTU
Bhilai, Chhattisgarh, India.

I. INTRODUCTION

- 1. Deep Learning:** Deep Learning is a sophisticated branch of machine learning that has shown exceptional success in a variety of fields owing to its ability to automatically learn complicated patterns from datasets and samples. This success may be attributed to deep learning's ability to automatically learn complex patterns from datasets and samples. Techniques like as Long Short-Term Memory (LSTM) networks and improved Deep Neural Networks (DNNs) have surfaced as potentially useful tools for the management of sequential data and high-dimensional data, respectively, for a variety of applications.
- 2. E-Commerce:** E-commerce is an innovative concept. It is, at present, heavily relying on the internet and mobile phone revolution to reshape the way businesses reach their customers. While in countries such as the US and China, e-commerce has taken notable steps to achieve high revenue, the industry in India is, still at its beginning. However over the past few years, the division has grown. E-trade has become the rapidly-growing segment. If this flourishing growth continues over the next few years, the size of the e- trade industry is estimated to be 20 to 30 billion USD by 2025. This increment is determined by increased customer purchases. In this chapter, sale of high profitable goods is taken under considerations which are rare products having high margin and suggestions have been made to increase the sale of these products such as mobile-phone accessories, beauty products, footwear and many more.
- 3. Data Mining:** Data mining is a technique of finding out the hidden details and information from the provided databases. It also analyses, explores and deduces some structured data that fulfils our needs based on some query or needs.
- 4. Association Rule Mining (ARM):** Association rule is a model that analysis or associates data mining task of uncovering relationships of data. It specifies types of data association, like it analyses the type of frequent products that are being purchased by the customers, it also predicts the future aspects of data.
- 5. Customer Purchasing Behavior (CPB):** It basically describes the customers' behaviour of buying products, as data mining, association rules are the basis for this concept according to which the data is analysed and provided to the customer. Focusing on customers' need of buying high profitable rare item sets which earns them high margins of profit, this data analysis benefits them with structured data that are directly provided to the customers.
- 6. Rare Items:** Rare item-sets are those item-sets which occur infrequently in the transactional dataset. In most of the realistic situations, rare item-sets having high profit provide very beneficial insights to the customer. Rare patterns may also indicate the occurrence of selling of high-margin products which is considered exceptional in the data. (Jyothi Pillai, 2013)
- 7. Profitable Items:** The sale of rare items incorporates a substantial amount of high profit than other products, which is very beneficial to the business. Those rare items upon selling transform into "profitable items". The items having more margin yield more profit upon their sale.

- 8. Recommender Systems:** A recommender system or a recommendation system is a subcategory of processed-data filtering system that helps to predict the “rating “; or ”preference “; that a customer would assign to a product. Recommender systems specifically generate a list of recommendations in one of the following two methods– Collaborative filtering and Content-based filtering.
- 9. High Utility:** Mining high utility itemsets is an important process in a number of different industries, including retail, marketing, e-commerce, supply chain management, and finance, among others. Businesses are able to reveal patterns and correlations that lead to better resource allocation, focused marketing campaigns, improved inventory management, and higher customer satisfaction if they first identify high utility itemsets and then examine the patterns and relationships they uncover. Traditional methods of itemset mining, like as the Apriori algorithm, are frequently used yet have limitations when applied to large-scale datasets & samples. This is despite the fact that these methods are commonly employed. These techniques often center their attention on item sets that occur frequently and may neglect item sets that have a high utility but occur seldom. In addition, their scalability and efficiency suffer when dealing with enormous volumes of information and samples. As a consequence of this, there is a rising need for sophisticated methods that are able to uncover high utility item sets from large-scale transactional datasets and samples in an effective and fast manner.

II. LITERATURE REVIEW

As in earlier research papers, this topic had been discussed under various terms referring to the same moot point of earning high profits over rare products. As in literature reporting of previous research, the labels “recommendation agents”, “high utility products”, “association rule mining”, “frequent item sets”, “adaptive interviewing” have been used. In this section, a review of existing methods is done, and can be observed from the following table,

1. Overview of High Utility Item set Mining

Table1: Review of existing Models & Process

Method	Datasets Used	Findings	Limitations	Gaps
HUI-Miner	Retail sales, Market basket data	Efficiently identifies high utility itemsets, outperforming traditional approaches in terms of performance.	Limited scalability for extremely large datasets & samples.	Investigation of parallel or distributed implementations for enhanced scalability.

LITERATURE SURVEY ON HIGH UTILITY ITEM-SET MINING
AND CUSTOMER PURCHASING BEHAVIOUR USING HYBRID MODEL

FHM	Transactional data	Demonstrates superior efficiency and scalability compared to traditional algorithms.	Limited exploration of utility-based constraints other than simple utility thresholds.	Investigation of hybrid approaches that combine FHM with other techniques for improved performance.
EFIM	Market basket data, Healthcare data	Shows improved runtime and memory usage compared to other algorithms.	May encounter challenges with highly skewed datasets & samples.	Investigation of techniques to handle skewed data distributions and adaptability to various utility measures.
BIDE-HUI	Transactional data	Efficiently mines high utility itemsets using a binary representation of transactions.	May face challenges with datasets containing long transaction sequences.	Exploration of techniques to handle long transaction sequences and scalability to large-scale datasets & samples.
HUP-Miner	Retail sales, Customer behavior data	Competitively performs in terms of runtime and scalability for high utility pattern mining.	Limited exploration of utility-based pattern growth techniques.	Investigation of novel utility-based pattern growth strategies for enhanced mining performance.
HUIFIM	Transactional data	Utilizes a utility array structure to handle utility information efficiently.	Limited exploration of constraints or optimizations beyond utility-guided itemset growth.	Investigation of techniques to incorporate additional constraints or utility-guided mining strategies.

UP-Growth	Market basket data	Demonstrates competitive performance and scalability for graph-based itemset mining.	Limited exploration of utility-based pruning strategies in UP-Growth.	Investigation of utility-guided pruning techniques to further improve efficiency in UP-Growth.
Hybrid Approaches	Various datasets	Combination of different techniques, such as integrating Apriori and FP-Growth.	Few studies exploring hybrid models specific to high utility itemset mining.	Exploration of novel hybrid approaches that leverage the strengths of different techniques for improved performance and efficiency.

- Apriori-based Approaches:** The Apriori algorithm is one of the earliest and most widely used frequent itemset mining algorithms. Several variations of Apriori have been proposed for high utility itemset mining by incorporating utility measures into the search process. These approaches extend the original Apriori algorithm to consider utility values along with item frequencies. However, Apriori-based approaches suffer from a high number of candidate itemsets and lack scalability for large datasets & samples.
- FP-Growth-based Approaches:** FP-Growth is another popular frequent itemset mining algorithm that constructs a compact data structure called a frequent pattern (FP) tree. Several variants of FP-Growth have been proposed for high utility itemset mining by modifying the tree construction and pattern generation steps to consider utility values. These approaches offer better scalability compared to Apriori-based methods but may still face challenges with extremely large datasets & samples.
- HUI-Miner:** HUI-Miner is an efficient high utility itemset mining algorithm based on a depth-first search strategy. It employs various pruning techniques to reduce the search space and optimize the mining process. HUI-Miner incorporates utility upper bounds and the concept of utility-lists to efficiently identify high utility itemsets. It has been shown to outperform Apriori-based and FP-Growth-based approaches in terms of performance.
- FHM (Fast High-Utility Miner):** FHM is a pattern-growth-based algorithm for high utility itemset mining. It utilizes the prefix tree structure to represent utility information and employs utility-based pruning strategies to eliminate unpromising branches. FHM significantly reduces the search space and exhibits improved efficiency compared to traditional approaches. It is known for its scalability and has become a popular choice for high utility itemset mining.

- **EFIM (Efficient High Utility Itemset Mining):** EFIM is an efficient algorithm that integrates utility measures into itemset mining by utilizing a vertical data format. It utilizes a utility-list structure to store utility information and employs several pruning techniques, such as the TWU (Total Weighted Utility) pruning strategy, to reduce the search space. EFIM has demonstrated superior performance in terms of runtime and memory usage compared to other algorithms.
- **BIDE-HUI (Binary-based Depth-first Search for High Utility Itemsets):** BIDE-HUI is a depth-first search-based algorithm that utilizes a binary representation of transactions and a utility-list structure to efficiently mine high utility itemsets. It incorporates several optimization techniques, such as TWU pruning and diffsets, to enhance the mining process's efficiency. BIDE-HUI has shown promising results in terms of runtime and memory usage.
- **HUP-Miner (High Utility Pattern Miner):** HUP-Miner is a pattern-growth-based algorithm specifically designed for discovering high utility patterns. It employs a projected database structure to optimize the mining process and uses utility upper bounds and pattern growth techniques to reduce the search space. HUP-Miner has demonstrated competitive performance and scalability compared to other algorithms.
- **HUIFIM (High Utility Itemset Mining using FIM):** HUIFIM is an algorithm that integrates the FIM (Frequent Itemset Mining) framework with utility measures to mine high utility itemsets. It utilizes a compact data structure called the Utility Array Structure to efficiently handle utility information and employs several optimization techniques, such as TWU pruning and utility-guided itemset growth, to improve efficiency.
- **UP-Growth (Utility Pattern Growth):** UP-Growth is a pattern-growth-based algorithm for high utility itemset mining. It utilizes a projection-based approach to construct the projected utility database and employs utility upper bounds and pruning strategies to optimize the mining process. UP-Growth has shown competitive performance and scalability compared to other algorithms.
- **Hybrid Approaches:** Some recent research has focused on hybrid approaches that combine different techniques, such as integrating Apriori-based and FP-Growth-based algorithms or combining pattern growth with vertical data format. These hybrid models aim to leverage the strengths of different algorithms and overcome their limitations, leading to improved performance and scalability.

In summary, several models and algorithms have been proposed for high utility itemset mining, each with its strengths and weaknesses. The choice of model depends on factors such as dataset size, computational resources, and specific requirements of the application. Researchers continue to explore new techniques and optimizations to further enhance the efficiency and effectiveness of high utility itemset mining algorithms.

2. Traditional Methods for High Utility Item set Mining : This section reviews traditional methods for HUIM, which can be observed from table 2,

Table 2: Methods used for HUIM Process

Method	Datasets Used	Findings	Limitations	Gaps
Apriori Algorithm	Market basket data, Transactional data	Identifies frequent itemsets, forms the basis for subsequent research on high utility itemset mining.	Scalability issues with large datasets due to a high number of candidate itemsets.	Investigation of improved pruning techniques or hybrid approaches to enhance scalability.
FP-Growth Algorithm	Market basket data, Transactional data	Offers better scalability compared to Apriori, eliminates the need for candidate itemset generation.	May still face challenges with extremely large datasets & samples.	Exploration of optimization techniques to handle memory requirements for large-scale datasets & samples.
HUI-Miner	Retail sales, Market basket data	Efficiently identifies high utility itemsets, outperforming traditional approaches in terms of performance.	Limited scalability for extremely large datasets & samples.	Investigation of parallel or distributed implementations for enhanced scalability.
HUP-Miner	Retail sales, Customer behavior data	Competitively performs in terms of runtime and scalability for high utility pattern mining.	Limited exploration of utility-based pattern growth techniques.	Investigation of novel utility-based pattern growth strategies for enhanced mining performance.
EFIM	Market basket data, Healthcare data	Shows improved runtime and memory usage compared to other algorithms.	May encounter challenges with highly skewed datasets & samples.	Investigation of techniques to handle skewed data distributions and adaptability to various utility measures.

BIDE-HUI	Transactional data	Efficiently mines high utility itemsets using a binary representation of transactions.	May face challenges with datasets containing long transaction sequences.	Exploration of techniques to handle long transaction sequences and scalability to large-scale datasets & samples.
HUIFIM	Transactional data	Utilizes a utility array structure to handle utility information efficiently.	Limited exploration of constraints or optimizations beyond utility-guided itemset growth.	Investigation of techniques to incorporate additional constraints or utility-guided mining strategies.
UP-Growth	Market basket data	Demonstrates competitive performance and scalability for graph-based itemset mining.	Limited exploration of utility-based pruning strategies in UP-Growth.	Investigation of utility-guided pruning techniques to further improve efficiency in UP-Growth.

- The Apriori Algorithm:** The Apriori algorithm is one of the oldest and best-known approaches for mining frequent item sets. It does this by gradually putting together more frequent sets of shorter lengths in order to produce itemsets. The Apriori method was first developed for frequent itemset mining; however, it is possible to modify it for high utility itemset mining by including utility metrics into the search process. This would make the algorithm more efficient. Apriori, on the other hand, has problems with scalability because of the enormous number of candidate item sets it considers.
- An Overview of the FP-Growth Algorithm:** The FP-Growth algorithm is yet another well-liked approach to the process of mining frequent itemsets. In order to discover frequently occurring item combinations, it builds a compact data structure known as the FP-tree and employs a depth-first search methodology. In a manner similar to that of the Apriori method, the FP-Growth algorithm may be modified to facilitate the mining of high value item sets by including various utility measurements. When compared to Apriori, FP-Growth is believed to have superior scalability thanks to its ability to prevent expensive candidate creation.
- HUI-Miner, also known as High-Utility Item set Miner:** HUI-Miner is an algorithm that has been developed expressly for the purpose of high utility itemset mining. It makes use of a search approach known as depth-first and incorporates a variety of pruning strategies in order to locate high utility item sets in an effective manner. HUI-Miner makes use of utility upper limits as well as utility-list structures in order to condense the search area and improve the efficiency of the mining operation. In

comparison to more conventional Apriori and FP-Growth algorithms, it has shown much better performance.

- **FHM (Fast High-Utility Miner):** FHM is an algorithm for the mining of high utility item sets that is based on pattern expansion. It does this by using a prefix tree structure to express utility information and by utilizing utility-based pruning algorithms to get rid of branches that don't seem very promising. When compared to more conventional methods, FHM is substantially more effective and shrinks the search area by an impressive amount. It has gained popularity as a viable option for mining high utility itemsets because to the scalability for which it is recognized.
- **EFIM (Efficient High Utility Itemset Mining):** EFIM is an effective method that utilizes a vertical data structure in order to combine utility metrics into item set mining. It does this by making use of a utility-list structure to store the utility information and by using a number of pruning strategies, one of which is called TWU pruning (which stands for total weighted utility pruning), in order to minimize the search area. When compared to other algorithms, EFIM has shown much better performance in terms of both its runtime and its memory consumption.
- **BIDE-HUI (Binary-based Depth-first Search for High Utility Item sets):** BIDE-HUI is a depth-first search-based algorithm that effectively mines high utility item sets by using a binary representation of transactions and a utility-list structure. In order to make the mining operation more productive, it makes use of a number of optimization strategies, such as TWU trimming and diffsets, among others. In terms of runtime and memory use, the results that BIDE-HUI has exhibited so far have been encouraging.
- **HUP-Miner, also known as the High Utility Pattern Miner:** HUP-Miner is a pattern-growth-based method that was developed for the purpose of locating high utility patterns. It does this by using a projected database structure to improve the mining process, as well as utility upper limits and pattern growth strategies to cut down on the amount of area that has to be searched. When compared against other algorithms, HUP-Miner has exhibited performance and scalability on par with those of its competitors.
- **HUIFIM (High Utility Itemset Mining using FIM):** HUIFIM combines the framework of FIM (Frequent Itemset Mining) with various utility measurements in order to mine high utility item sets. In order to effectively manage utility information, it makes use of a compact data structure known as the Utility Array Structure. Additionally, it makes use of a number of optimization approaches, including as TWU pruning and utility-guided itemset expansion, in order to increase its overall efficiency.
- **Utility Pattern Growth, often known as UP-Growth:** UP-Growth is an algorithm for the mining of high utility item sets that is based on pattern growth. In order to generate the projected utility database, it takes a projection-based technique. Additionally, it makes use of utility upper limits and pruning algorithms in order to improve the mining process. When compared to other algorithms, UP-Growth has shown performance and scalability on par with those of its competitors.

- **Hybrid Methods and Strategies:** Hybrid models that combine diverse approaches, such as merging Apriori-based and FP-Growth-based algorithms or combining pattern growth with vertical data formats, have been presented. These models may be used to solve a variety of problems. These hybrid techniques have the goal of using the strengths of several algorithms to overcome the limits of those algorithms, which will ultimately result in increased performance and scalability.

In conclusion, there are several models and algorithms already developed for the conventional approaches that are used in the High Utility Itemset Mining process. In terms of scalability, runtime, memory utilization, and the capacity to deal with big datasets and samples, each model has its own set of advantages and disadvantages. The unique needs of the application, the computing resources that are readily accessible, as well as the size and features of the datasets and samples all have a role in the selection of the model to use. In order to further improve the efficacy and efficiency of the standard methods used in the mining process for high utility itemsets, researchers are always looking into new approaches and ways to optimize existing ones for different scenarios.

3. **Deep Learning in Data Mining:** This section covers DL Models in Data Mining, which are discussed in table 3 as follows,

Table 3: DL Models for HUIM Process

Method	Datasets Used	Findings	Limitations	Gaps
Convolutional Neural Networks	Image datasets (e.g., MNIST, ImageNet)	Effective in capturing local patterns and hierarchical representations from images.	Limited applicability to non-image datasets & samples.	Investigation of transfer learning approaches to adapt CNNs for non-image datasets & samples.
Recurrent Neural Networks	Sequential data (e.g., time series, text data)	Effective in capturing temporal dependencies and modeling sequential patterns.	Prone to vanishing or exploding gradients, making training challenging for long sequences.	Exploration of advanced RNN architectures or regularization techniques to address gradient-related issues.
Long Short-Term Memory	Sequential data (e.g., time series, text data)	Overcomes vanishing gradient problem in RNNs, captures long-term dependencies.	Memory-intensive, making training and inference computationally demanding.	Investigation of optimized implementations or hardware acceleration for more efficient LSTM training and inference.

Generative Adversarial Networks	Image datasets, synthetic data	Capable of generating synthetic data that closely resemble the real data distribution.	Training instability, mode collapse, and lack of explicit control over generated outputs.	Exploration of improved training techniques, stability enhancements, and explicit control over GAN-generated outputs.
Autoencoders	Various datasets	Unsupervised learning for feature learning and dimensionality reduction.	Reliance on handcrafted features for subsequent mining tasks.	Investigation of incorporating autoencoders in end-to-end deep learning pipelines for seamless feature learning and mining integration.
Variational Autoencoders	Various datasets	Probabilistic model for generative tasks, capable of generating new samples.	Challenging to train, requiring balanced reconstruction and regularization terms.	Investigation of advanced training techniques or novel architectures for more stable and diverse VAE generation.
Graph Neural Networks	Graph-structured data (e.g., social networks)	Effective in capturing structural information and learning node embeddings in graphs.	Limited exploration of GNNs in other types of structured data beyond graphs.	Investigation of GNN adaptations for specific structured data types, such as molecular structures or knowledge graphs.
Deep Reinforcement Learning	Dynamic decision-making environments, simulations	Learn optimal policies through interaction with environments.	High sample complexity and computational requirements, limited interpretability.	Exploration of model-based RL approaches, interpretability techniques, and sample efficiency enhancements in deep RL.

Deep Clustering	Various datasets	Combines deep learning with clustering algorithms for unsupervised learning and clustering tasks.	Difficulty in interpretability and understanding cluster assignments.	Investigation of hybrid approaches combining deep clustering with interpretable clustering techniques.
-----------------	------------------	---	---	--

Deep learning has had a profound impact on a variety of industries, including data mining, by delivering effective new methods for mining massive datasets for intricate patterns and valuable insights. A comprehensive analysis of the following models that are currently being utilized for deep learning in data mining,

- **Convolutional Neural Networks (CNN):** CNNs are very popular tools for doing image analysis and computer vision applications. By employing convolutional layers that catch local patterns and pooling layers that aggregate information, they are successful in learning hierarchical representations from raw datasets & samples. This is achieved via the use of deep learning. In the field of data mining, CNNs have been used for a variety of tasks, including the categorization of images, the detection of objects, and the identification of anomalies.
- **Recurrent Neural Networks (RNN):** RNNs are intended to handle sequential input by capturing temporal relationships. RNNs are also known as feed-forward neural networks. They have shown success in a variety of applications, including time series forecasting, sentiment analysis, and natural language processing, among others. RNNs make use of recurrent connections so that they may remember information from the past. This gives them the ability to represent sequential patterns seen in datasets and samples.
- **Long Short-Term Memory (LSTM):** LSTMs are a form of RNN that can capture long-term dependencies and overcome the vanishing gradient issue. LSTMs have been effectively used in a variety of applications, including voice recognition, automatic translation, and opinion mining. LSTMs are put to use in the field of data mining to model and interpret sequential data that has intricate temporal patterns.
- **Generative Adversarial Networks (GAN):** GANs are made up of a generator network and a discriminator network, both of which compete against one another while the model is being trained. In recent years, GANs have been more popular for use in applications like as picture production, data synthesis, and anomaly detection. In the field of data mining, GANs have the ability to produce synthetic data to either correct imbalances in existing datasets or identify abnormalities by learning the distribution of the underlying datasets & samples.
- **Autoencoders :** Autoencoders are a kind of neural network that was intended for unsupervised learning as well as dimensionality reduction. Their goal is to reconstitute the input data based on a compressed form that has been learnt by an encoder.

Autoencoders have been used in a variety of contexts and purposes, including data denoising, feature learning, and anomaly detection.

- **Variational Autoencoders (VAE):** VAEs are a sort of generative model that integrate autoencoders with probabilistic modeling. VAEs are also known as "variational autoencoders." VAEs are able to generate fresh samples by sampling from the learnt distribution after the latent space representation of the input data has been learned by the VAE. In the past, VAEs have been put to use for a variety of purposes, including the creation of data, the identification of anomalies, and the imputation of datasets & samples.
 - **Graph Neural Networks (GNN):** GNNs are meant to operate on graph-structured data, such as social networks, citation networks, or molecular graphs. Graph-structured data includes the following: social networks, citation networks, and molecular graphs. GNNs are able to learn node embeddings and collect structural information because to the message passing that takes place between the nodes. GNNs have been used for a variety of tasks, including the classification of graphs, the prediction of links, and the categorization of nodes.
 - **Transformers:** Transformers have developed as a strong model architecture for jobs requiring sequential or structured data, such as natural language processing and recommendation systems. Transformers may also be used to transform data from one format into another. Transformers have achieved state-of-the-art performance in a variety of fields thanks to the use of self-attention mechanisms for the purpose of capturing global dependencies.
 - **Deep Reinforcement Learning (DRL):** DRL is a method that enables agents to learn and make choices in complicated situations by combining deep learning with reinforcement learning. The Deep Reinforcement Learning (DRL) model has shown to be effective in a variety of domains, including game playing, robotics, and recommendation systems. In the field of data mining, dynamic decision-making, resource allocation, and optimization are all examples of potential applications for DRL process.
 - **Deep Clustering:** Deep clustering approaches combine classic clustering algorithms with deep learning to simultaneously learn representations and cluster assignments. By using the expressive capacity of deep neural networks and applying it to the task of capturing complicated data structures, these approaches intend to enhance the performance of clustering process.
- 4. Existing Deep Learning Techniques for Item set Mining :** This section discussed DL Models for Item set Mining, and can be observed from Table 4,

Table 4: DL Models for Item set Mining Process

Method	Datasets Used	Findings	Limitations	Gaps
Convolutional Neural Networks	Image datasets	Effective in identifying itemsets corresponding to objects or features within images.	Limited applicability to non-image datasets & samples.	Investigation of transfer learning approaches to adapt CNNs for non-image itemset mining.
Recurrent Neural Networks	Sequential data	Capable of capturing sequential dependencies and extracting frequent itemsets from sequences.	Prone to vanishing or exploding gradients, making training challenging for long sequences.	Exploration of advanced RNN architectures or regularization techniques to address gradient-related issues.
Long Short-Term Memory	Sequential data	Overcomes vanishing gradient problem, enables mining frequent itemsets from long sequences.	Memory-intensive, making training and inference computationally demanding.	Investigation of optimized implementations or hardware acceleration for more efficient LSTM training and inference.
Generative Adversarial Networks	Synthetic data generation	Able to generate synthetic itemsets that resemble the real data distribution, useful for data augmentation or anomaly detection.	Training instability, mode collapse, and lack of explicit control over generated itemsets.	Exploration of improved training techniques, stability enhancements, and explicit control over GAN-generated itemsets.

Autoencoders	Various datasets	Unsupervised learning for itemset representation learning and dimensionality reduction.	Reliance on handcrafted itemset features for subsequent mining tasks.	Investigation of incorporating autoencoders in end-to-end deep learning pipelines for seamless itemset mining integration.
Variational Autoencoders	Various datasets	Probabilistic model capable of generating new itemsets and capturing latent itemset representations.	Challenging to train, requiring balanced reconstruction and regularization terms.	Investigation of advanced training techniques or novel architectures for stable and diverse VAE-generated itemsets.
Graph Neural Networks	Graph-structured data	Effective in mining itemsets from graph-structured data by capturing structural information and learning node embeddings.	Limited exploration of GNNs in other types of structured data beyond graphs.	Investigation of GNN adaptations for specific structured data types, such as molecular structures or knowledge graphs.
Deep Reinforcement Learning	Dynamic decision-making environments	Capable of learning optimal itemset selection policies through interaction with environments.	High sample complexity and computational requirements, limited interpretability.	Exploration of model-based RL approaches, interpretability techniques, and sample efficiency enhancements in deep RL.
Deep Clustering	Various datasets	Combines deep learning with clustering algorithms for unsupervised itemset mining.	Difficulty in interpretability and understanding cluster assignments.	Investigation of hybrid approaches combining deep clustering with interpretable clustering techniques.

An exhaustive analysis of the currently available deep learning methods for item set mining uncovered a number of potentially useful approaches. Convolutional Neural Networks, or CNNs, have been used for itemset mining tasks, notably in image-based

situations, where itemsets correlate to objects or characteristics inside pictures. In these kinds of circumstances, CNNs have proved very useful. CNNs are well-suited for locating meaningful item sets within picture datasets and samples as a result of their proficiency in the extraction of hierarchical representations as well as the capture of local patterns. Recurrent Neural Networks (RNNs), including variations like Long Short-Term Memory (LSTM), are good in capturing sequential relationships and have found use in itemset mining tasks requiring sequential data, like as transaction sequences or clickstream datasets & samples. RNNs fall under the category of artificial neural networks. These models are able to learn intricate temporal patterns and extract from sequences itemsets that are either often used or have a high utility.

Generative adversarial networks, often known as GANs, have been found to have potential in the field of itemset mining. These networks generate synthetic itemsets that closely mirror the distribution of the actual datasets & samples. GANs are able to create new item sets that are consistent with the underlying data features because they are trained to use a generator network to construct itemsets and a discriminator network to differentiate between actual and synthetic item sets. On the other hand, autoencoders are used in the process of itemset mining for the purpose of dimensionality reduction and feature learning. The latent representations of itemsets may be captured by them, and they can learn useful low-dimensional representations that can aid further mining tasks like as grouping or classification.

Transformers were first developed for use in natural language processing, but more recently, they've been put to use solving challenges in itemset mining. These models make use of self-attention processes in order to capture global dependencies. They have shown to be effective in sequence-based itemset mining tasks, such as market basket analysis and clickstream analysis. For the purpose of itemset mining in graph-structured data, such as social networks or citation networks, where itemsets correlate to subgraphs or community structures, Graph Neural Networks (GNNs) have been put to use. GNNs have the ability to learn node embeddings and collect the knowledge about the graph's structure in order to discover relevant item sets.

Techniques known as Deep Reinforcement Learning (DRL) have been investigated for use in itemset mining for use in either dynamic or sequential decision-making contexts. These models may learn to make the best judgments in itemset mining tasks like as resource allocation, recommendation systems, or dynamic market analysis when deep neural networks and reinforcement learning algorithms are included into them. In general, the present deep learning methods for itemset mining show their adaptability and efficiency across a wide range of data sources and mining workloads. It is anticipated that more research and improvements in this subject will improve their capabilities and broaden the range of fields in which they may be used for different scenarios.

III. 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 itemset, 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.

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] Bhuvanewari, 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., Priyadarshini, 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 Solution 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.
- [28] *International Journal of Scientific and Technology Research*.
- [29] 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>
- [30] 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>
- [31] 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>
- [32] 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>
- [33] 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>
- [34] 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>
- [35] Yılmaz Benk, G., Badur, B., & Mardikyan, S. (2022). A New 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>
- [36] 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>
- [37] 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>
- [38] 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>
- [39] 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>

