




HIGH UTILITY PATTERN MINING: A SURVEY ON CURRENT AND POSSIBLE AREAS OF APPLICATIONS

 **Abdullah Bokir^{1,2+}**
Vb Narasimha³

^{1,2}Department of Computer Science & Engineering, UCE, OU, Hyderabad, India.

^{1,2}Email: bokir102@gmail.com

³Email: vbnarasimha@gmail.com

²Hadhramout University, Mukalla, Yemen.



(+ Corresponding author)

ABSTRACT

Article History

Received: 18 February 2022

Revised: 8 November 2022

Accepted: 23 November 2022

Published: 15 December 2022

Keywords

Applications of high utility
Association rule mining
Customer behaviors analysis
Data mining
High utility pattern mining
Itemset mining
Market basket analysis
Pattern mining
Text mining applications.

High Utility Pattern Mining (HUPM) has a wide range of applications, including making recommendations, detecting outliers, analyzing customer behaviors, and solving a wide range of other problems. In fact, unlike other important data mining tasks such as outlier analysis and classification, high utility pattern mining can be used as an intermediary tool for providing pattern-centered insights for other data mining tasks. In this paper, we look at a wide range of different applications of high utility pattern mining that are available in the literature. We gathered the literature review papers from different resources based on the keywords related to the applications of high utility pattern mining. Then we classified the existing research papers into different categories according to the domain application, then we summarize the proposed method of each paper and the datasets used for the evaluation process. Rather than providing a full discussion of every possible application, the objective of this paper is to give readers an overview of what is feasible when high utility pattern mining approaches are used. This paper can help researchers to apply the techniques of high utility pattern mining to solve real-time problems and it can help to explore other possible domains to apply high utility pattern mining techniques by proposing new methods and evaluating them accordingly.

Contribution/Originality: An extensive literature review was done to analyze the existing research papers related to the application of high utility pattern mining. Summarize the existing literature related to the application of high utility pattern mining, their proposed methods, and the datasets used to evaluate them. Give a brief overview of the possible domains to apply the idea of high utility pattern mining to solve some real time problems.

1. INTRODUCTION

The idea of High Utility Pattern Mining (HUPM) was first introduced by Yao. Hong and his coauthors in Yao, et al. [1]. The primary motivation for this concept is to overcome the limitations of Frequent Itemset Mining (FIM) [2], which focuses exclusively on the frequency of items in a transactional database and ignores other factors such as the quantity of the product purchased and the profit earned from these products. On the other hand, High Utility Pattern Mining considers the number of items purchased and additional factors that reflect the product's usefulness; these additional factors are referred to as utility factors such as profit, cost, expiration date, storage safety, and any other factor that reflects the user's preferences.

For the last decade, researchers in the data mining community have attempted to develop various approaches for mining high utility patterns; these approaches can be classified as level-wise approaches such as Ho, et al. [3];

Chu, et al. [4], tree-based approaches such as Shie, et al. [5]; Ahmed, et al. [6]; Feng, et al. [7]; Kundu, et al. [8]; Kim and Yun [9]; Ryang and Yun [10]; Yun, et al. [11], list-based approaches such as Yen and Lee [12]; Yun, et al. [13]; Jaysawal and Huang [14]; Nam, et al. [15]; Kim, et al. [16]; Baek, et al. [17]; Bokir and Narasimha [18]. and approaches that are based on evolutionary techniques such as Lin, et al. [19]; Song and Huang [20]; Song and Huang [21].

The applications of high utility pattern mining are quite diverse, spanning a wide range of fields and incorporating data from a variety of distinct data domains. To address the unique challenges that are special to each domain, different types of variations of high utility pattern mining may be employed in conjunction with each other. For example, the kind of patterns that can be discovered in the context of geographic, multimedia, temporal, or biological data will be vastly different from one another. The following are some instances of the enormous number of areas and data domains where the idea of high utility pattern mining is applied:

- 1- **Affinity Analysis:** It's one of the data mining techniques for discovering meaningful relationships between different entities based on their co-occurrence in a dataset and selecting a set of entities that generate high benefits to the retail businesses. It is possible to gain important knowledge regarding unforeseen trends from practically any system or process by using affinity analysis techniques. In reality, affinity analysis makes use of the fact that it studies attributes that are related to one another, which aids in the discovery of hidden patterns in large amounts of data by developing association rules between the attributes.
- 2- **Customer Behaviour Analysis:** high utility pattern mining can be applied for analyzing customer behavior and finding a group of customers who contribute the most to the retail business's revenue. To put it another way, the concept is that regular connections between client purchasing behavior can be exploited to make valuable business decisions.
- 3- **Spatiotemporal data analysis:** Spatial data refers to the data where the items have both spatial and non-spatial characteristics such as the temperature readings on the sea surface. In such circumstances, high utility pattern mining can be used to define meaningful links between the spatial and non-spatial characteristics of the items. Spatiotemporal data, such as route data, may be studied using high utility pattern mining techniques. In particular, high utility pattern mining may be utilized to identify the critical parts of routes that are often utilized throughout the time.
- 4- **Text Mining Applications:** text mining is the process of transforming the text data from an unstructured format to a text with a structured format to identify meaningful patterns and new insights. High utility pattern mining has been used a lot in this domain to solve many problems.
- 5- **Healthcare and Biomedical applications Applications:** It is possible to forecast future trends, increase outreach, and even better regulate the spread of illnesses through the use of healthcare analytics. Insights may be gained at both the global and local levels in this diverse domain. It can lead to improvements in the quality of patient care, clinical information, diagnostics, and business development. High utility pattern mining can be used in the healthcare domain in many different ways for example it can be used for disease diagnostic by finding a set of symptoms that highly cause the respective disease.
- 6- **Web Mining Applications:** Web Mining is the process of looking for patterns in large amounts of unstructured data on the World Wide Web using data mining tools. It employs automated methods to obtain useful patterns from web pages, server logs, and link structures. High utility sequential pattern mining algorithms can be used to determine relevant traversal patterns from Weblogs. These traversal patterns can be utilized to construct and manage Web pages.
- 7- **Outliers Detection:** A measurement's variability, experimental mistakes, or novelty may be indicated by outliers, which are extreme observations that depart from the rest of the observations in the data. To put it another way, an outlier is an observation that deviates from the main trend in a collection of data. High

utility pattern mining can be utilized for outliers detection by extracting patterns with the least/ highest utility value.

A brief literature review of the research papers published previously in the above-mentioned application areas of high utility pattern mining will be presented in the next section. The number of potential applications for high utility pattern mining is diverse, and they can be found in a variety of disciplines. The primary purpose of this study is to present an overview of the scenario, with a particular emphasis on the important scenarios in which high utility pattern mining can be implemented. This will equip the readers with the tools necessary to comprehend how these strategies can be applied in a variety of situations.

2. LITERATURE REVIEW

In this section, the previous research papers published in the above-mentioned domain applications will be presented in detail.

Figure 1 depicts the number of papers that have been published in different domain applications as per our knowledge and our search results based on different keywords related to applications of high utility pattern mining.

Number of Research Papers Reviewed in Different Domain Applications

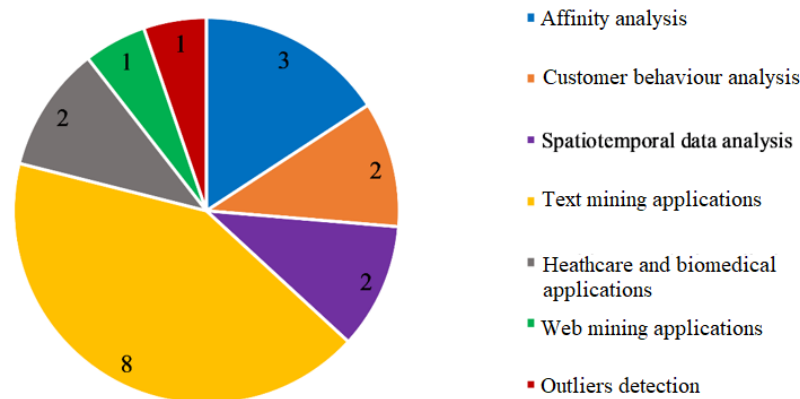


Figure 1. Ratio of papers published in different domain applications.

2.1. High Utility Pattern Mining for Affinity Analysis

Using a sliding-window-based technique of the data streams, the authors in [Amaranatha and Hazarath \[22\]](#) introduced a unique approach named “Extended Global Utility Item-sets Tree (EGUI-tree)” that can be used to retrieve high utility itemsets from a retail market data stream that contains both positive and negative profit items. Experimentation on real-world datasets reveals that the proposed EGUI-tree approach is both fast and more scalable than the existing standards.

A new strategy was proposed by [Gan, et al. \[23\]](#) for identifying non-redundant, correlated purchasing behavior by taking into account both the usefulness of the behavior and the degree of correlation. Patterns with high profit and strong correlation may be produced, which can lead to a stronger recall and disclose superior precision. The proposed technique is based on a database projection technique and uses different pruning strategies to reduce the search space. After conducting a thorough experiment, it was discovered that the proposed innovative non-redundant correlated high-utility pattern outperforms the existing algorithms in terms of efficacy. Furthermore, the suggested algorithm is both time and memory efficient when compared to the existing algorithms.

Applying the idea of high utility pattern mining for cross-selling strategy was done in [Padhye and Deshmukh \[24\]](#). The proposed method will first generate the rules associating the products, then by using objective function it

estimates the cross-selling profit of each rule. When selling to existing clients, it's said to be cross-selling them. For suggestion purposes, cross-selling makes use of products in the subsequent section of a rule, and with the implementation of a rule, it offers information on future profits. To optimize profit, administrators may utilize this knowledge on cross-selling margins, and the products that will be sold in the future are likely to also be high-utility products.

2.2. High Utility Pattern Mining for Customer Behaviour Analysis

To discover high-utility mobile sequential patterns, the researchers in [Shie, et al. \[25\]](#) integrated mobile data mining with utility mining techniques. The authors proposed two algorithms, one is based on a level-wise approach and the other is based on a tree-based approach. Through the use of experimental evaluations, a series of studies and comparisons of the performance of the two different types of algorithms are made. In particular, the findings demonstrate that the suggested algorithms surpass the current best mobile sequential patterns and that the tree-based algorithms outperform the level-wise algorithms under a variety of scenarios.

To increase the retailer's profit, the authors in [Mondal, et al. \[26\]](#) designed a new method named "Product Expiry-Aware and Revenue-conscious itemset placement scheme (PEAR)". The proposed method is a product placement method that is designed to increase store revenue by taking into consideration both product expiry and revenue considerations. The significant contributions of this work are divided into three categories. The challenge of managing retail itemset placement when the items might be connected with varied time periods of expiry is introduced first. Second, they suggest the expiry-aware PEAR system, which allows for the effective identification and placement of high-revenue itemsets, resulting in increased merchant income. Performance analysis was conducted using two real datasets and the results demonstrated that the proposed model is efficient in increasing retailer's revenue.

2.3. High Utility Pattern Mining for Spatiotemporal Data Analysis

The majority of earlier research concentrated on extracting high utility itemsets from transactional databases without taking into account the spatial and temporal features of items. To address this problem, the authors in [Bommisetty, et al. \[27\]](#) suggested a more flexible model of spatial high utility itemsets (SHUIs) that exist in spatiotemporal databases. Generally speaking, in a spatiotemporal database, an itemset is considered to be an SHUI if the utility of any of its items is more than a user-specified minimum utility threshold value, and the distance between any two of its items is greater than a user-specified maximum distance threshold value.

The study in [Kiran, et al. \[28\]](#) proposes a more flexible distributed technique to identify all needed itemsets from a database utilising the Spark in-memory computing architecture, which is built on the distributed computing paradigm. This article introduces many unique pruning ways to reduce the research scope. The strategy that is being presented incorporates a number of Spark's benefits, including decreased communication costs, fault tolerance, and excellent scalability.

2.4. High Utility Pattern Mining for Text Mining Applications

High utility itemset mining has been applied for academic literature recommendation in [Liang, et al. \[29\]](#). The authors of this work considered the references cited by a paper as an internal utility which can be either 1 or 0. The external utility in this work is a subjective value given by the user which represents the importance of that reference to the user. The recommendation in this approach is based on two steps. The first step is content-based filtering and the second one is high utility references recommendation. Experimental study using a real-world dataset have demonstrated the effectiveness of the proposed technique in producing customised suggestions without degrading their quality much.

For topics detection from microblog streams, the researchers in [Huang, et al. \[30\]](#) proposed a clustering framework that is based on high utility pattern mining. The proposed framework will identify a set of typical patterns from the microblog streams and then organize these patterns into subject clusters. Furthermore, the suggested framework is capable of detecting both previously identified topics and newly emerging topics at the same time. In-depth experiments study using Twitter and Sina Weibo streams demonstrate that the proposed method outperforms conventional topic detection algorithms.

The previous work for academic literature recommendation [\[29\]](#) was done based on static data. The authors in [Dhanda and Verma \[31\]](#) proposed a customized recommendation system that can recognize the ever-growing nature of research article repositories and adjust its recommendations accordingly. The results of the experimental study conducted by the authors demonstrate that the proposed method meets the researcher's specific criteria while also effectively dealing with the growing nature of the repository of the research papers.

Text clustering is a crucial topic in the field of text mining. The authors in [Tran, et al. \[32\]](#) designed a new technique for text clustering based on "frequent_weighted_utility_itemsets (FWUI)". The proposed technique calculates the frequency of each term in documents to create a weight matrix for all documents. After this, FWUI will be extracted from the number matrix and the weights of terms in documents. Finally, the documents will be clustered together based on frequent utility itemsets. The suggested method has been tested on three data sets, each of which had 1,600 documents spanning 16 different topics. The experimental results demonstrate that the proposed strategy enhances the accuracy of text clustering when compared to methods that make use of frequent itemset mining.

Emerging topic detection from Twitter was done by the authors in [Choi and Park \[33\]](#) using a high utility itemset mining technique that takes into account the frequency and utility of each pattern. For a set of tweets and by using a time-based windowing on the Twitter stream, the utility of words will be defined based on how often they appear and words with high frequency and high utility will be retrieved. The efficiency of the proposed method was shown by the experimental findings, which indicated greater performance and a shorter running time when compared to other examined topic detection methods. Specifically and in relation to the three datasets used for the experimental analysis, the suggested method demonstrated a 5 percent higher topic recall than the other evaluated methods.

In [Demir, et al. \[34\]](#) a new method that integrates high utility pattern mining and aspect-based sentiment analysis techniques to identify the group of features that have the potential to boost profits and those that must be enhanced in order to increase the level of user satisfaction. Testing the suggested method's pattern extraction against baseline methods reveals extremely profitable feature groups that can be discovered.

The majority of currently available techniques to news recommendation employ users' clicks as implicit feedback to better understand their behavior. "Clicks," on the other hand, may not be a reliable predictor of the users' real interests. A new and innovative news recommender system based on a news utility model is proposed in [Zihayat, et al. \[35\]](#). The proposed method performs two steps for news recommendation. Initial recommendations are generated at the article level using utility models, which are subsequently combined with probabilistic topic models to provide topic level recommendation rules by the proposed method. Using a big real-world dataset gathered from a prominent newspaper in Canada to evaluate the proposed method, the results show that the proposed method outperforms the existing approaches.

The High Average Utility Pattern Mining (HAUPM) technique is used by the authors in [Belhadi, et al. \[36\]](#) for hashtag recommendation. There are two major steps in the proposed method. Offline processing converts the Twitter corpus into a transactional database based on the timestamps of the tweets that have been tagged (tweets with hashtags). To get the top k high average utility patterns over time. Irrelevant tagged tweets and their ontology are also built offline. Online processing inputs the utility patterns, ontology, and irrelevant tagged tweets to extract the most relevant tags for a specific orphan tweet (tweet without hashtags). Comprehensive analyses

were carried out on a massive collection of tweets in order to better understand how they work. In terms of the quality of recommended hashtags as well as processing runtime, the proposed technique surpasses existing state-of-the-art hashtag recommender systems.

2.5. High Utility Pattern Mining for Healthcare and Biomedical Applications

Cerebellar Ataxia (CA) is a neurological condition characterized by poor coordination of movement and balance issues. It affects the cerebellum and affects the mobility and balance of the patient. The Heel-knee-shin test and the rapid alternating movements test are two main tests for evaluating CA. The authors in [Jin, et al. \[37\]](#), presented a non-contact approach for detecting CA and explores the practicality of using this method. Using wireless sensors operating in a frequency range of C-band, this body sensor network can collect data related to both types of tests without being noticed by the patient being tested. The results of the tests yield a wealth of helpful information on human health, which is extremely important to the participants who take part in the testing. The collected data from testing includes a lot of important and hidden information on human health that is extremely valuable to individuals. As a result, utility pattern mining (UPM) is employed for the goal of mining the activity patterns of individuals. The authors discover that the amplitude information included in the patients' activity patterns varies substantially amongst them. They were able to extract the amplitude information that was useful in assessing the test findings and determining whether the test was positive or negative. After that, they employed several classification techniques, such as "Support Vector Machines (SVM)", to classify the data samples.

A technique named "(Top-k Impactful Itemsets Miner)" is suggested in [Liu, et al. \[38\]](#) for determining the connections between the levels of gene expression. This method requires the user to specify the value for k, to be able to extract the most important top-k co-expressed genes. To assign various weights to different genes, a table with influence levels for each gene was generated based on the number of neighboring genes that are differentially expressed in the dataset inside gene regulatory networks for each gene. At last, the top-k significant itemsets were manually assessed using existing literature and examined using a Gene Ontology enrichment approach to determine which itemsets were the most significant. Two publicly accessible time course microarray datasets with two distinct experimental settings were used to assess the suggested approach. Co-expressed gene sets were discovered in both datasets and exhibited more accuracy for the proposed technique than the two comparable control approaches.

2.6. High Utility Pattern Mining for Web Mining Applications

For mining high utility web access sequences from static and incremental datasets, a new method was proposed in [Ahmed, et al. \[39\]](#). The proposed method is based on two novel tree structures and it doesn't scan the dataset multiple times. Intensive performance tests have revealed that the proposed technique is extremely efficient for mining high utility web access sequences from static and incremental datasets, demonstrating that it is both fast and accurate.

2.7. High Utility Pattern Mining for Outliers Detection

To detect outliers effectively from weighted stream data, the authors in [Cai, et al. \[40\]](#) designed a two-phase method called "Minimal Weighted Rare Pattern Mining approach (MWRPM-Outlier)". The first phase is called the pattern mining phase, where a method known as MWRPM is proposed to quickly mine the least weighted rare patterns, and then two deviation factors are created in the outlier detection phase (second phase) to quantify the anomalous degree of each transaction on the weight data stream.

Table 1. Summary of literature review.

[reference no] paper title (year of publication)	Domain application	Proposed method	Dataset for performance evaluation
[22] High Utility Item-set Mining from retail market data stream with various discount strategies using EGUI-tree (2021)	Affinity Analysis	A tree-based approach called “Extended Global Utility Item-sets Tree(EGUI-tree)”	Retail market from SPMF
[23] Extracting non-redundant correlated purchase behaviors by utility measure (2017)	Affinity Analysis	A projection-based algorithm named as non-redundant Correlated High-Utility Itemset Mining (CoHUIM)	1- Real-world datasets: i. Foodmart ii. Retail iii. Chess. iv. Mushroom. 2- Synthesis data i. T10I4D100K ii. T5I2N2KD100K
[24] A marketing solution for cross-selling by high utility itemset mining with dynamic transactional databases (2016).	Affinity Analysis	Tree-based approach.	-
[25] Efficient algorithms for discovering high utility user behavior patterns in mobile commerce environments (2012).	Customer Behaviour Analysis	Level-wise approach and Tree-based approach	Synthesis data
[26] PEAR: A Product Expiry-Aware and Revenue-Conscious Itemset Placement Scheme (2021).	Customer Behaviour Analysis	Product Expiry-Aware and Revenue-conscious itemset placement scheme (PEAR).	Retail and chainstore
[27] Discovering spatial high utility itemsets in spatiotemporal databases (2021).	Spatiotemporal Data Analysis	A fast single scan algorithm, called SHUI-Miner.	(Traffic) congestion database provided by JARTIC
[28] Distributed Mining of Spatial High Utility Itemsets in Very Large Spatiotemporal Databases using Spark In-Memory Computing Architecture (2020).	Spatiotemporal Data Analysis	Distributed Spatial High Utility Itemset-Miner (DSHUI-Miner).	1- Synthetic datasets: i- T10I4D100K ii- T10I4D100K.50X 2- Real- world datasets: i- Congestion, Kosarak ii- Connect
[29] A utility-based recommendation approach for academic literatures (2011).	Text Mining Applications	Two-phase algorithm	Real-world dataset, ACL Anthology
[30] Topic detection from large scale of microblog stream with high utility pattern clustering (2015).	Text Mining Applications	High Utility Pattern Clustering (HUPC) framework.	1- TwitterSet 2- SinaSet.
[31] Recommender System for Academic Literature with Incremental Dataset (2016).	Text Mining Applications	Modified Recommendation Systems that can comprehend the ever-expanding characteristics of the research article repository	ACL Anthology Network
[32] Text Clustering Using Frequent Weighted Utility Itemsets (2017).	Text Mining Applications	Modification Weighted Itemset Tidset- Frequent Weighted Utility Itemset (MWIT-FWUI).	downloaded from digital newspapers: 1- www.vnexpress.net 2- dantri.com.vn 3- thanhnien.vn
[33] Emerging topic detection in Twitter stream based on high utility pattern mining (2019).	Text Mining Applications	HUPM based method for detecting emerging topics from Twitter stream data.	Twitter data used for topic detection that include tweets about: 1- FA Cup Final. 2- Super Tuesday. 3- US Elections.
[34] Extracting Potentially High Profit Product Feature Groups by Using High Utility Pattern	Text Mining Applications	A new method that combines aspect-based sentiment analysis and high	Real-world datasets: 1- Mobile phones (Turkish language).

[reference no] paper title (year of publication)	Domain application	Proposed method	Dataset for performance evaluation
Mining and Aspect Based Sentiment Analysis (2019).		Utility pattern mining	2- Cell phones & Acc. (English language). 3- Musical instruments (English language).
[35] A utility-based news recommendation system (2019).	Text Mining Applications	Two-stage news recommendation framework	Globe and Mail dataset (Canadian news agency)
[36] A data-driven approach for twitter hashtag recommendation (2020).	Text Mining Applications	Pattern Mining for Hashtag Recommendation (PM-HRec).	Twitter data that contains tweets related to the following three topics: 1- health 2- cinema 3- sport
[37] Activity Pattern Mining for Healthcare (2020).	Healthcare and Biomedical applications	Utility pattern mining (UPM) is used for the purpose of mining subjects' activity pattern	Data have been collected using the proposed sensor network.
[38] Mining differential top-k co-expression patterns from time course comparative gene expression datasets (2013).	Healthcare and Biomedical applications	Top-k Impactful Itemsets Miner (TIIM)	Gene regulatory data from humans and mice were downloaded from the BioGRID and KEGG databases
[39] Mining high utility web access sequences in dynamic web log data (2010).	Web mining applications	Tree-based approach for mining web access sequences.	Synthetic datasets
[40] An efficient outlier detection approach on weighted data stream based on minimal rare pattern mining (2019).	Outlier Detection	Two-phase minimal weighted rare pattern mining-based outlier detection approach, called MWRPM-Outlier	Synthetic dataset

Table 1 presents a summary of the aforementioned research papers, it gives the domain application of the research paper, the proposed method, and datasets used to evaluate each paper.

3. POSSIBLE AREA FOR APPLYING HUPM

In the previous section for the literature review, we tried to give an overview of different domains where the idea of high utility pattern mining (HUPM) was applied to solve different problems. In this section, we will present our view on some of the possible areas to apply HUPM algorithm for discovering interesting and hidden patterns.

3.1. Online Advertisements

High utility pattern mining can be applied for online advertising that we usually see while browsing websites. It can help in selecting a set of ads that can be displayed on a web page by making sure that these ads are most clicked and generate high revenue when the users click them.

3.1.1. Example Data

As we can see from the example data given in Table 2 and Table 3, the historical data for each webpage and the set of ads in that webpage are given as a transaction data in Table 2. Here each ad is associated with a number of times that has been clicked by the visitors of the webpage. The amount of revenue generated every time a visitor clicks the respective ad is shown in Table 3. High utility pattern mining in this case can help in selecting a set of ads that can be shown on a webpage to increase the revenue from online advertisement.

Table 2. Transaction data for online advertisements.

Webpage	(Ad_ID, No_Clicks)
Page1	(Ad1, 5), (Ad2, 6), (Ad3, 2), (Ad5, 8)
Page2	(Ad2, 7), (Ad5, 9)
Page3	(Ad7, 4), (Ad4, 5), (Ad6, 9)

Table 3. utility table for advertisements.

Ad_ID	Revenu per click in \$
Ad1	5
Ad2	2
Ad3	6
Ad4	10
Ad5	7
Ad6	1
Ad7	5

3.2. Team Selection

If we want to select a team of members by considering that the members of the team have worked together before and they have high productivity, then we can apply the idea of high utility pattern mining.

3.2.1. Example Data

Table 4, shows the transaction data for team members where every row represents a team that was formed by n number of members. Table 5, shows the productivity rate (utility) of each member. Applying high utility pattern mining methods can help in selecting a team with high productivity. This idea can also be used in selecting a group of machines that work together. The utility can be chosen based on the domain application.

Table 4. Transaction data for team members.

Team	Members
Team1	M1, M3, M5, M6
Team2	M3, M4, M6
Team3	M1, M2, M5

Table 5. productivity (utility) data for every member.

Member	Productivity Rate
M1	8
M2	6
M3	10
M4	7
M5	2
M6	4

3.3. Recommendation Systems

High utility pattern mining can be used to build recommender systems for many domain applications such as products recommendation in e-commerce, news recommendations, movies or music recommendations, books recommendations, and many more. Although recommendation systems have been proposed in the literature, most of these systems assume the profit to be the only factor for utility and they ignore other important factors such as cost, safety to store, expiry time, ..., etc. There is a need for developing recommender systems that consider multiple factors for utility and balance these factors to measure the overall utility gain [41].

3.4. As a Tool for other Data Mining Techniques

High utility pattern mining can be used as a helping tool for other data mining techniques such as classification and clustering. For example in clustering, HUPM can help in clustering the data based on the utility factor.

3.5. Bug Detection in Software Programs

Identifying defects in software programs can be done using high utility pattern mining which can help in finding the most relevant trends in the underlying data.

4. CONCLUSION

In this article, a survey of the most significant applications of high utility pattern mining is presented for your perusal. In addition to customer behaviour analysis, affinity analysis, spatiotemporal data analysis, healthcare data analysis, and text or web mining applications, high utility pattern mining has been applied to a wide range of domain applications and different types of data. These applications include, but are not limited to, web mining and text mining. In this article, several potential domains are outlined, each of which is capable of making use of the methods of high utility pattern mining. There is a possibility that there are a great number of other sectors in which we may use high utility pattern mining in order to identify intriguing patterns in the underlying data.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: Both authors contributed equally to the conception and design of the study.

REFERENCES

- [1] H. Yao, H. J. Hamilton, and G. J. Butz, "A foundational approach to mining itemset utilities from databases," in *2004 SIAM Int. Conf. Data Min. Society for Industrial and Applied Mathem*, 2004, pp. 482–486.
- [2] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, 1994, pp. 487–499.
- [3] T. B. Ho, D. Cheung, H. Liu, Y. Liu, W.-K. Liao, and A. Choudhary, "A two-phase algorithm for fast discovery of high utility itemsets," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2005, pp. 689–695.
- [4] C.-J. Chu, V. S. Tseng, and T. Liang, "An efficient algorithm for mining temporal high utility itemsets from data streams," *Journal of Systems and Software*, vol. 81, pp. 1105–1117, 2008. Available at: <https://doi.org/10.1016/j.jss.2007.07.026>.
- [5] B.-E. Shie, S. Y. Philip, and V. S. Tseng, "Efficient algorithms for mining maximal high utility itemsets from data streams with different models," *Expert Systems with Applications*, vol. 39, pp. 12947–12960, 2012. Available at: <https://doi.org/10.1016/j.eswa.2012.05.035>.
- [6] C. F. Ahmed, S. K. Tanbeer, B.-S. Jeong, and H.-J. Choi, "Interactive mining of high utility patterns over data streams," *Expert Systems with Applications*, vol. 39, pp. 11979–11991, 2012. Available at: <https://doi.org/10.1016/j.eswa.2012.03.062>.
- [7] L. Feng, L. Wang, and B. Jin, "UT-Tree: Efficient mining of high utility itemsets from data streams," *Intelligent Data Analysis*, vol. 17, pp. 585–602, 2013. Available at: <https://doi.org/10.3233/ida-130595>.
- [8] M. K. Kundu, D. P. Mohapatra, A. Konar, and A. Chakraborty, "Time-fading based high utility pattern mining from uncertain data streams," *Smart Innovation, Systems and Technologies*, vol. 27, pp. 1–26, 2014. Available at: <https://doi.org/10.1007/978-3-319-07353-8>.
- [9] D. Kim and U. Yun, "Mining high utility itemsets based on the time decaying model," *Intelligent Data Analysis*, vol. 20, pp. 1157–1180, 2016. Available at: <https://doi.org/10.3233/ida-160861>.
- [10] H. Ryang and U. Yun, "High utility pattern mining over data streams with sliding window technique," *Expert Systems with Applications*, vol. 57, pp. 214–231, 2016. Available at: <https://doi.org/10.1016/j.eswa.2016.03.001>.

- [11] U. Yun, H. Nam, J. Kim, H. Kim, Y. Baek, J. Lee, E. Yoon, T. Truong, B. Vo, and W. Pedrycz, "Efficient transaction deleting approach of pre-large based high utility pattern mining in dynamic databases," *Future Generation Computer Systems*, vol. 103, pp. 58-78, 2020. Available at: <https://doi.org/10.1016/j.future.2019.09.024>.
- [12] S. J. Yen and Y. S. Lee, *An efficient approach for mining high utility itemsets over data streams. In data science and big data: An environment of computational intelligence*. Cham: Springer, 2017.
- [13] U. Yun, H. Ryang, G. Lee, and H. Fujita, "An efficient algorithm for mining high utility patterns from incremental databases with one database scan," *Knowledge-Based Systems*, vol. 124, pp. 188-206, 2017. Available at: <https://doi.org/10.1016/j.knsys.2017.03.016>.
- [14] B. P. Jaysawal and J. W. Huang, "Sohupds: A single-pass one-phase algorithm for mining high utility patterns over a data stream," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, 2020, pp. 490-497.
- [15] H. Nam, U. Yun, B. Vo, T. Truong, Z.-H. Deng, and E. Yoon, "Efficient approach for damped window-based high utility pattern mining with list structure," *IEEE Access*, vol. 8, pp. 50958-50968, 2020. Available at: <https://doi.org/10.1109/access.2020.2979289>.
- [16] J. Kim, U. Yun, E. Yoon, J. C.-W. Lin, and P. Fournier-Viger, "One scan based high average-utility pattern mining in static and dynamic databases," *Future Generation Computer Systems*, vol. 111, pp. 143-158, 2020. Available at: <https://doi.org/10.1016/j.future.2020.04.027>.
- [17] Y. Baek, U. Yun, H. Kim, H. Nam, H. Kim, J. C.-W. Lin, B. Vo, and W. Pedrycz, "Rhups: Mining recent high utility patterns with sliding window-based arrival time control over data streams," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 12, pp. 1-27, 2021. Available at: <https://doi.org/10.1145/3430767>.
- [18] A. Bokir and V. Narasimha, "High utility mining of streaming itemsets in data streams," in *Journal of Physics: Conference Series IOP Publishing*, 2021, p. 012027.
- [19] J. C.-W. Lin, L. Yang, P. Fournier-Viger, J. M.-T. Wu, T.-P. Hong, L. S.-L. Wang, and J. Zhan, "Mining high-utility itemsets based on particle swarm optimization," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 320-330, 2016.
- [20] W. Song and C. Huang, "Discovering high utility itemsets based on the artificial bee colony algorithm," *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10939 pp. 3-14, 2018. Available at: https://doi.org/10.1007/978-3-319-93040-4_1.
- [21] W. Song and C. Huang, "Mining high utility itemsets using bio-inspired algorithms: A diverse optimal value framework," *IEEE Access*, vol. 6, pp. 19568-19582, 2018. Available at: <https://doi.org/10.1109/access.2018.2819162>.
- [22] R. P. Amarathana and M. K. P. M. Hazarath, "High Utility Item-set Mining from retail market data stream with various discount strategies using EGUI-tree," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-12, 2021. Available at: <https://doi.org/10.1007/s12652-021-03341-3>.
- [23] W. Gan, J. C.-W. Lin, P. Fournier-Viger, H.-C. Chao, and H. Fujita, "Extracting non-redundant correlated purchase behaviors by utility measure," *Knowledge-Based Systems*, vol. 143, pp. 30-41, 2018. Available at: <https://doi.org/10.1016/j.knsys.2017.12.003>.
- [24] R. Padhye and R. J. Deshmukh, "A marketing solution for cross-selling by high utility itemset mining with dynamic transactional databases," in *2016 Int. Conf. Comput. Tech. Inf. Commun. Technol. ICCTICT*, 2016.
- [25] B.-E. Shie, H.-F. Hsiao, and V. S. Tseng, "Efficient algorithms for discovering high utility user behavior patterns in mobile commerce environments," *Knowledge and Information Systems*, vol. 37, pp. 363-387, 2013. Available at: <https://doi.org/10.1007/s10115-012-0483-z>.
- [26] A. Mondal, R. Mittal, V. Khandelwal, P. Chaudhary, and P. K. Reddy, "PEAR: A product expiry-aware and revenue-conscious itemset placement scheme," in *2021 IEEE 8th Int. Conf. Data Sci. Adv. Anal*, 2021, pp. 1-10.
- [27] S. C. Bommisetty, R. Penugonda, U. K. Rage, M. S. Dao, and K. Zettsu, "Discovering spatial high utility itemsets in spatiotemporal databases," *Lect. notes comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12798, pp. 53-65, 2021.

- [28] R. U. Kiran, S. Ito, M. S. Dao, K. Zettsu, C. W. Wu, Y. Watanobe, and T. C. Thang, "Distributed mining of spatial high utility itemsets in very large spatiotemporal databases using spark in-memory computing architecture," in *2020 IEEE International Conference on Big Data (Big Data) IEEE*, 2020, pp. 4724-4733.
- [29] S. Liang, Y. Liu, L. Jian, Y. Gao, and Z. Lin, "A utility-based recommendation approach for academic literatures," in *Proceedings - 2011 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT*, 2011, pp. 229-232.
- [30] J. Huang, M. Peng, and H. Wang, "Topic detection from large scale of microblog stream with high utility pattern clustering," in *PIKM 2015 - Proc. 8th Ph.D. Work. Inf. Knowl. Manag. co-located with CIKM 2015*, 2015, pp. 3-10.
- [31] M. Dhanda and V. Verma, "Recommender system for academic literature with incremental dataset," *Procedia Computer Science*, vol. 89, pp. 483-491, 2016. Available at: <https://doi.org/10.1016/j.procs.2016.06.109>.
- [32] T. Tran, B. Vo, T. T. N. Le, and N. T. Nguyen, "Text clustering using frequent weighted utility itemsets," *Cybernetics and Systems*, vol. 48, pp. 193-209, 2017. Available at: <https://doi.org/10.1080/01969722.2016.1276774>.
- [33] H.-J. Choi and C. H. Park, "Emerging topic detection in twitter stream based on high utility pattern mining," *Expert Systems with Applications*, vol. 115, pp. 27-36, 2019. Available at: Emerging topic detection in twitter stream based on high utility pattern mining.
- [34] S. Demir, O. Alkan, F. Cekineli, and P. Karagoz, *Extracting potentially high profit product feature groups by using high utility pattern mining and aspect based sentiment analysis. In High-Utility Pattern Mining*. Cham: Springer, 2019.
- [35] M. Zihayat, A. Ayanso, X. Zhao, H. Davoudi, and A. An, "A utility-based news recommendation system," *Decision Support Systems*, vol. 117, pp. 14-27, 2019. Available at: <https://doi.org/10.1016/j.dss.2018.12.001>.
- [36] A. Belhadi, Y. Djenouri, J. C. W. Lin, and A. Cano, "A data-driven approach for twitter hashtag recommendation," in *IEEE Access*, 2020, pp. 79182-79191.
- [37] J. Jin, W. Sun, F. Al-Turjman, M. B. Khan, and X. Yang, "Activity pattern mining for healthcare," *IEEE Access*, vol. 8, pp. 56730-56738, 2020. Available at: <https://doi.org/10.53733/jardcs/v11/20192570>.
- [38] Y.-C. Liu, C.-P. Cheng, and V. S. Tseng, "Mining differential top-k co-expression patterns from time course comparative gene expression datasets," *BMC Bioinformatics*, vol. 14, pp. 1-13, 2013. Available at: <https://doi.org/10.1186/1471-2105-14-230>.
- [39] Q. F. Ahmed, S. K. Tanbeer, and B. S. Jeong, "Mining high utility web access sequences in dynamic web log data," in *Proc. - 11th ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD2010*, 2010, pp. 76-81.
- [40] S. Cai, R. Sun, S. Hao, S. Li, and G. Yuan, "An efficient outlier detection approach on weighted data stream based on minimal rare pattern mining," *China Communications*, vol. 16, pp. 83-99, 2019. Available at: <https://doi.org/10.23919/jcc.2019.10.006>.
- [41] A. Muralidhar and P. Venkatasubbu, "HUPM-MUO: High utility pattern mining under multiple utility objectives," *International Journal of Computer Aided Engineering and Technology*, vol. 14, pp. 385-407, 2021. Available at: <https://doi.org/10.1504/ijcaet.2021.10030789>.

Views and opinions expressed in this article are the views and opinions of the author(s), Review of Information Engineering and Applications shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.