Dynamic Association Mining Techniques for the Faster Extraction of High Utility Itemsets from Incremental Databases

Subba Reddy Meruva
Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, India
subbareddy.meruva@gmail.com (corresponding author)

Bondu Venkateswarlu
Department of Computer Science and Engineering, School of Engineering, Dayananda Sagar University, Bangalore, India
bonduvenkat-cse@dsu.edu.in

Received: 16 October 2024 | Revised: 6 November 2024 | Accepted: 21 November 2024

 DOI: https://doi.org/10.48084/etasr.9295

ABSTRACT

Financial and market analysis applications require the mining of strong-utility itemsets. Finding frequent itemsets with high utility patterns is vital for such wide applications. Recent utility-based mining methods were successfully used in the current study to identify high value itemsets from static datasets. Stream databases or incremental databases update the itemsets at regular intervals (schedulers). Incremental Mining-based High Utility Itemset (IM-HUI) algorithms improve the methodologies based on High Utility Itemset (HUI) methods. The proposed technique refines the itemset values and updates the HUIs based on incremental schedulers. It reduces the time and space while mining HUIs over dynamic databases. The efficacy of the proposed work is compared experimentally to that of existing mining techniques on benchmark datasets.

Keywords-incremental mining; utility value; high utility itemset; association mining; schedulers

I. INTRODUCTION

Association mining [1] is a vital data mining functionality widely utilized in forming relationships among itemsets. Based on these relationships, frequent or customer-interesting itemsets can be defined that help improve marketing strategies. The frequency of itemsets is majorly considered during the mining process of itemsets in discovering strong association rules. For this mining process, traditional association techniques, such as FP-growth [2] and Apriori [3] are commonly used to filter out the demanded items based on customers' interests. Those items are commonly mentioned as frequent items or itemsets. Sometimes, metrics other than the frequency of itemsets may be adequate for marketing applications. The research on association mining has recently focused more on utility-based mining. Such techniques consider two essential factors, frequency and profit (or utility), while finding itemsets. These itemsets are usually called High Utility Itemsets (HUIs) [4]. Effective extraction of the HUIs based on a study of utility and frequency of itemsets can be mined with mining algorithms such as Utility Pattern Growth (UP-Growth) [5], HUI-Miner [6], and Fastest HUI Miner (FHM). Neither the HUI Miner nor the FHM use a support-confidence paradigm. They rather rely on a utility-confidence framework [7]. These techniques assess the set of correlated rules by their impact on firm profitability and itemsets' frequency. HUI Miner begins by computing utility lists, which compute the value of collections of items (or itemsets). Reducing the search space with HUI Miner's heuristic information can speed up the processing execution time. On the other hand, UP-Growth, initially computes Transaction Utility (TU) and Weighted Transaction Utility (TWU) and then constructs a tree based on Utility Patterns (UPs) to form a UP tree [9], which smoothly discovers HUIs. The concept of FHM [9] originated from implementing utility lists, which included a set of items with high utility. It is faster in extracting HUIs and can be effective for static databases. Real-time applications need the extraction of HUIs from dynamic databases. Thus, this paper proposes incremental mining-based HUI techniques for discovering HUIs over dynamic databases. The m ain contributions of the current study are:

- Utility-based itemsets are mined from the updated databases.
- A dynamic UP tree is established with the proposed work.
- Incremental mining techniques are designed for finding HUIs over dynamic databases.

- TU and TWU are updated based on defined regular intervals over dynamic databases.
- Performance study and empirical analysis are conducted to define the proposed method's efficiency.

II. BACKGROUND WORK OF MINING TECHNIQUES

The strong association rule and the framework of support-confidence are explained in [10] and [11], respectively. Association mining is the most useful [12], especially in the business and economic sectors. This system computes the degree to which different itemsets in a transactional database are associated or linked to one another. Association rule mining aims to discover the rules that will most stimulate the users' interest.

Equations (1) and (s) describe the support S and confidence C [13]:

$$S(X \rightarrow Y) = count\_value(XUY) / No\_of\_Transactions \quad (1)$$

$$C(X \rightarrow Y) = S\_count(XUY) / count\_value(X) \quad (2)$$

When evaluating the statistical information that is connected to the categories X and Y this set of values corresponds to what is put to use as part of the evaluation process. They do not supply the semantic measure that can be applied to these things. As a result, more may be needed to develop the utility-based mining system. So, to derive customer-interested rules, this framework searches for objects related to the utility. In situations like these, it is best to make use of yet another alternative framework, known as the utility-confidence framework [14, 15] in order to establish the norms of association based on utility [16]. A functionality-based association mining framework is expected to provide quantity, internal utility value (for example, the number of items sold), and benefits value while selling the products to others (for example, the profit of itemsets). The computation of the Transaction Utility (TU) and Weighted TU (TWU) is illustrated with a sample example in [17].

An improvised data mining method known as Weighted Association Rule Mining (WARM) [18] can assign weights to things based on importance or on high they are on a priority list. The WARM technique uses a linear model to automate the weight assignment process, nevertheless, it does not take into account the volumes involved in the trades. Authors in [19] take into account two primary values in their HUIM (HUI Mining) algorithm: the number of items involved in a given transaction, and the usefulness attribute of an item, which is defined by its utility and profitability. For active incremental association mining, IHUP is a technique introduced in [20] that makes use of the FP-growth notion and the HUIP tree as a representation. Keeping the utility value steady is essential for discovering HUIs. However, doing so at the expense of performance could be better when dealing with large numbers of candidate items or initial item generations. Another utility-based mining strategy that uses a tree-based approach is UP-growth [21], which uses a wide variety of techniques to assess the utility values of the itemset in order to boost mining efficiency. Using the utility-list data structure, HUI-Miner is an alternate mining approach [22] that helps conserve storage space in the event of a tree pruning by decreasing the volume of data, utilities, and heuristics that must be kept. It eliminates two potential sources of difficulty: the time and effort required to generate candidate items and the time and effort required to calculate the utility of sets of items. By skipping the phases of producing candidate itemsets in a depth-first search approach, high utility itemsets can be computed in a single phase, and the results can be extracted.

The HUI Miner approach, also reffered to as FHM [23], minimizes the amount of time and resources devoted to utility mining. It produces numerous item sets using EUCS (Estimated-Utility-Co-Occurrence-Structure), which makes processing more effective. The exploration of EUCS in large datasets requires a more significant amount of computational effort. The essential processes of FHM are [24, 25]: 1) The utility values of the itemsets are computed in terms of both their internal and external environments, and they are then recorded in their respective lists. 2) There are two forms of utility determined by this method: the raw utility of each transaction, or TU, and the weighted utility of each transaction, or TWU. 3) EUCS structures are derived from the utility values of the transactions. 4) Generations of user-interested, HUIs are established. While necessary, the joint operation can be expensive. Instead of the costly joint procedure, the FHM technique uses EUCS to solve this issue. As a result, compared to other methods, FHM significantly reduces the amount of memory used and the amount of time needed to complete a task. Databases in the "sparse" format are not suitable for mining HUIs. When dealing with sparse datasets, TIHUM [7] (Tree-integrated High Utility Miner) can overcome this difficulty. It enhances the mining algorithms for finding the HUIs dynamically for streaming databases. Its improvements are described with the proposed incremental mining-based HUI. The following section presents the IM-HUI algorithm for deriving the HUIs for updated dynamic databases.

III. THE PROPOSED IM-HUI

In order to mine HUIs from the static transactional databases, a number of HUIM techniques are employed. The concept of mining HUIs over incremental or dynamic databases is presented in this work. Algorithm 1 describes the proposed methodology of IM-HUI (Incremental Mining-based HUI) that dynamically determines the HUIs over the dynamic databases.

```
Algorithm 1 :     IM-HUI
Input       :
DTB   :     Dynamic Transactional Database
Mutil :     User Specified Minimum Utility
Threshold Value
Methodology :
Compute the Transaction Weighted Utility
of single items (or single item sets) from
DTB
Filter the item 'i' with satisfying the
condition TWU(i) ≥ Mutil and update
transactional item I and renamed as Inew
```

```
Apply the ordering in Inew based on the
TWU values of each item 'i' in Inew
Scan the DTB for building the utility list
of the item iєInew; construct the EUCS
structure as the algorithm of TIHUM in [
]
Construct_HUIs(Φ, Inew, Mutil, EUCS)

Algorithm 2:      Construct_HUIs
Output      :      results of HUIs
Methodology :
While (x є Inew) do
If add(x.utilitylist. iutils) ≥ Mutil then
Display x
If
add(x.utilitylist.iutils+add(x.utilitylist
.rutils) ≥ Mutil then
ExtensionsOfx←Φ
While (each y є Inew)
Such that every y>x do
If TWU(merged_items{x,y}) ≥ Mutil) then
Inewxy=Inewx U Inewy
Inewxy.utility_list←construct(Inew,
Inewx,Inewy)
Inewx←Inewx U Inewxy
End if
End while
Construct_HUIs( Inew,Inewx,Mutil)
End if
End while
```

Algorithm 1 shows the main procedure that can be used for deriving the high-utility itemsets. It scans the given database for computing the TWU of each item and filter outs the items in the itemset Inew based on a minimum utility threshold value (Mutil). Based on TWU values of items, the total ordering ">" is defined among the items in Inew. The EUCS structure is defined for the ordered itemsets Inew in a filtered transactional database per the procedure described in [7]. The EUCS structure stores TWU values for the itemsets (Inew), and the utility of the itemsets is not zero. DFS exploration of itemsets starts with the recursive calling Construct_HUIs procedure with the empty sets, Inew, Mutil, and the EUCS.

Algorithm 2 shows the recursive procedure of Construct_HUIs. Inewx shows the extensions of Inew with appending of x, Inewy shows the extensions of Inew items by appending y, and Inewxy shows the extensions of Inew by appending xy. Inewx, Inewy, and Inewxy are considered the high utility itemsets when their sum of "iutil" values of utility_list is above Mutil. It explores the updated Inewx, Inewy, and Inewxy for dynamically deriving the high-utility itemsets over incremental databases.

IV. EXPERIMENTAL WORK

The state-of-the-art HUIM methods, HUL_LIST_INS, HUIF_BA, HUIF_PSO [26-28] and the proposed IM_HUI were tested on two dynamic datasets, Retail and Food mart which can be both found in [29]. The number of transactions is dynamically added as follows: 1) In the Retail database, 100000 transactions are initially saved in the memory database. Each time 50000 transactions are added with a scheduler, the dynamic updates are considered in a transactional database for the experimental study of dynamic association mining. 2) In the Food mart database, dynamic updates are considered with an increased limit of 5000. Two performance parameters were considered: runtime and memory usage during the execution of existing and proposed methods. Table I shows the results.

TABLE I. PERFORMANCE ANALYSIS OF DYNAMIC ASSOCIATION MINING METHODS

HUIM Method	Transaction limit	Runtime (ms)	Memory (MB)	HUI count
		Retail		
IM_HUI	100000	5972	380.68115234375	22
HUL_LIST_INS	100000	5999	372.5	22
HUIF-BA	100000	151286	114.93023681640625	22
HUIF-PSO	100000	102027	127.58111572265625	22
IM_HUI	150000	5765	365.1206283569336	16
HUL_LIST_INS	150000	5982	354.001953125	16
HUIF-BA	150000	215837	107.4790267944336	16
HUIF-PSO	150000	149036	151.9121551513672	16
IM_HUI	200000	5525	361.501953125	9
HUL_LIST_INS	200000	5596	368.5	9
HUIF-BA	200000	251767	110.2098388671875	9
HUIF-PSO	200000	169309	108.16364288330078	9
IM_HUI	250000	5516	360.001953125	5
HUL_LIST_INS	250000	5531	358.0	5
HUIF-BA	250000	266831	240.22592163085938	5
HUIF-PSO	250000	191359	145.19060516357422	5
		Food mart		
HUIF-BA	10000	41431	251.4241714477539	428
HUIF-PSO	10000	19806	212.2222900390625	428
HUL_LIST_INS	10000	463	7.16595458984375	428
IM_HUI	10000	448	7.1724395751953125	428
HUIF-BA	15000	27479	164.43012237548828	89
HUIF-PSO	15000	13963	188.7872085571289	89
HUL_LIST_INS	15000	565	6.8936614990234375	89
IM_HUI	15000	381	6.898002624511719	89
HUIF-BA	16000	28103	236.42333984375	65
HUIF-PSO	16000	13655	205.257080078125	65
HUL_LIST_INS	16000	463	6.895133972167969	65
IM_HUI	16000	399	6.90338134765625	65
HUIF-BA	17000	28418	259.1466369628906	47
HUIF-PSO	17000	18304	64.50773620605469	47
HUL_LIST_INS	17000	479	6.8936614990234375	47
IM_HUI	17000	386	6.89996337890625	47

In Retail database, the difference in runtime is depicted in Figure 1, and the difference in memory usage is depicted in Figure 2. In all variant sizes of the database, the proposed IM-HUI outperforms the other incremental HUIM methods. In memory comparison analysis, IM-HUI did not show significant improvement compared to the existing methods. Thus, experimental demonstrations were conducted on another benchmarked incremental database, Food mart, to prove that the proposed method is efficient on both runtime and memory allocation. Figures 3 and 4 show the comparative analysis of runtime and memory allocation of existing and proposed techniques. It is observed that the proposed IM-HUI is more efficient regarding both time and memory in the Food mart database.

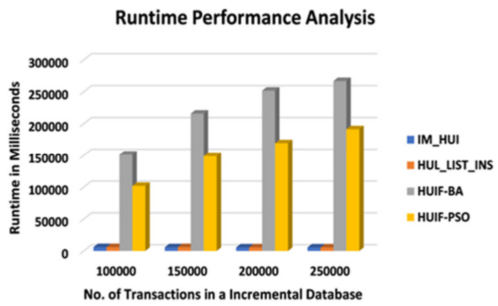

Fig. 1. Retail Dataset: Runtime Analysis.

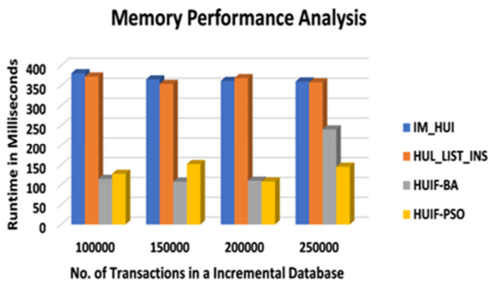

Fig. 2. Retail Dataset: Memory Usage Analysis.

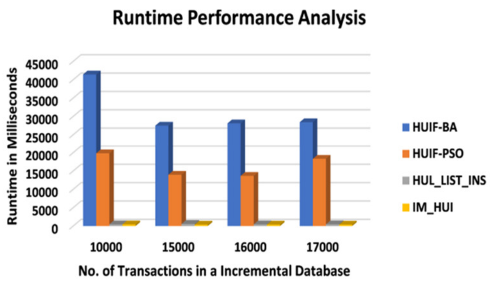

Fig. 3. Food Mart Dataset: Runtime Analysis.

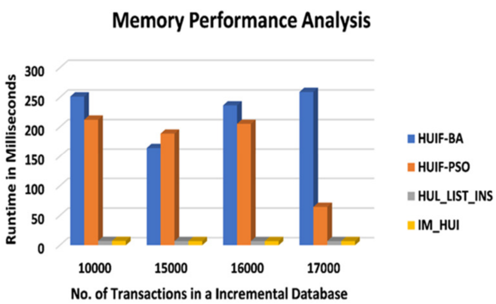

Fig. 4. Food Mart Dataset: Memory Usage Analysis.

V. CONCLUSION AND FUTURE SCOPE

Methods for mining highly useful itemsets are finding use in commercial and financial contexts. The utility-based framework is implemented in this paper in the proposed IM-HUI technique. Its framework considers the utility and frequency of items or itemsets in mining. Existing HUI mining techniques are unable to handle dynamic databases. State-of-the-art techniques handle dynamic databases useful for deriving frequent patterns incrementally. However, these techniques are unable to handle large transactional dynamic databases. The proposed IM-HUI delivered the dynamic association rule with the incremental construction of HUIs over the dynamic transactional databases. Time efficiency and memory usage were improved using the proposed IM-HUI with a rate of 30-45% compared to other incremental HUI mining methods on the Food mart database. A similar observation was made in the proposed IM-HUI in significant Retail data cases. The eventual goal of this effort is to incorporate closed set incremental mining into the IM-HUI for web video recommendation in search engines.

REFERENCES

[1] J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "A survey on the use of association rules mining techniques in textual social media," *Artificial Intelligence Review*, vol. 56, no. 2, pp. 1175–1200, Feb. 2023, https://doi.org/10.1007/s10462-022-10196-3.

[2] M. Shawkat, M. Badawi, S. El-ghamrawy, R. Arnous, and A. El-desoky, "An optimized FP-growth algorithm for discovery of association rules," *The Journal of Supercomputing*, vol. 78, no. 4, pp. 5479–5506, Mar. 2022, https://doi.org/10.1007/s11227-021-04066-y.

[3] Y. Xu, R. Zhan, G. Tan, L. Chen, and B. Tian, "An Improved Apriori Algorithm Research in Massive Data Environment," in *Cyber Security Intelligence and Analytics*, Z. Xu, K.-K. R. Choo, A. Dehghantanha, R. Parizi, and M. Hammoudeh, Eds. New York, NY, USA: Springer, 2020, pp. 843–851.

[4] W. Shen, C. Zhang, W. Fang, X. Zhang, Z.-H. Zhan, and J. C.-W. Lin, "Efficient High-utility Itemset Mining Based on a Novel Data Structure," in *IEEE International Smart Cities Conference*, Manchester, United Kingdom, Sep. 2021, pp. 1–6, https://doi.org/10.1109/ISC253183.2021.9562788.

[5] V. S. Tseng, B.-E. Shie, C.-W. Wu, and P. S. Yu, "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 8, pp. 1772–1786, Aug. 2013, https://doi.org/10.1109/TKDE.2012.59.

[6] S. R. Meruva, B. Venkateswarlu, "Tree Integrated High Utility Miner for Improving an Efficiency of Association Mining", *TEST Engineering & Management*, vol. 83, pp. 15938-15946, May-Jun. 2020.

[7] S. R. Meruva and V. Bondu, "Review of Association Mining Methods for the Extraction of Rules Based on the Frequency and Utility Factors," *International Journal of Information Technology Project Management*, vol. 12, no. 4, pp. 1–10, Oct. 2021, https://doi.org/10.4018/IJITPM.2021100101.

[8] Y. Liu, W. Liao, and A. Choudhary, "A fast high utility itemsets mining algorithm," in *1st International Workshop on Utility-Based Data Mining*, Chicago, IL, USA, Aug. 2005, pp. 90–99, https://doi.org/10.1145/1089827.1089839.

[9] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and Knowledge Discovery*, vol. 8, no. 1, pp. 53–87, Jan. 2004, https://doi.org/10.1023/B:DAMI.0000005258.31418.83.

[10] A. Inokuchi, T. Washio, and H. Motoda, "An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data," in *Principles of Data Mining and Knowledge Discovery*, New York, NY, USA: Springer, 2000, pp. 13–23.

[11] D. Enke and S. Thawornwong, "The use of data mining and neural networks for forecasting stock market returns," *Expert Systems with Applications*, vol. 29, no. 4, pp. 927–940, Nov. 2005, https://doi.org/10.1016/j.eswa.2005.06.024.

[12] M. Hamamoto and H. Kitagawa, "Ratio Rule Mining with Support and Confidence Factors," in *3rd International IEEE Conference Intelligent*

Systems, London, UK, Sep. 2006, pp. 500–505, https://doi.org/10.1109/IS.2006.348470.

[13] S. R. Meruva and B. Venkateswarlu, "A Fast and Effective Tree-based Mining Technique for Extraction of High Utility Itemsets," in *6th International Conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, Dec. 2022, pp. 1393–1399, https://doi.org/10.1109/ICECA55336.2022.10009213.

[14] M. S. Khan, M. Muyeba, and F. Coenen, "A Weighted Utility Framework for Mining Association Rules," in *Second UKSIM European Symposium on Computer Modeling and Simulation*, Liverpool, UK, Sep. 2008, pp. 87–92, https://doi.org/10.1109/EMS.2008.73.

[15] S.-J. Yen and Y.-S. Lee, "Mining High Utility Quantitative Association Rules," in *International Conference on Big Data Analytics and Knowledge Discovery*, Regensburg, Germany, Sep. 2007, pp. 283–292, https://doi.org/10.1007/978-3-540-74553-2_26.

[16] A. Erwin, R. P. Gopalan, and N. R. Achuthan, "Efficient Mining of High Utility Itemsets from Large Datasets," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Osaka, Japan, Dec. 2008, pp. 554–561, https://doi.org/10.1007/978-3-540-68125-0_50.

[17] S. Wadhwa and P. Gupta, "A review on Weighted Association Rule Mining (WARM) Using Python Programming Language," in *12th International Conference on Computing Communication and Networking Technologies*, Kharagpur, India, Jul. 2021, pp. 1–6, https://doi.org/10.1109/ICCCNT51525.2021.9579515.

[18] T. Wei, B. Wang, Y. Zhang, K. Hu, Y. Yao, and H. Liu, "FCHUIM: Efficient Frequent and Closed High-Utility Itemsets Mining," *IEEE Access*, vol. 8, pp. 109928–109939, Jan. 2020, https://doi.org/10.1109/ACCESS.2020.3001975.

[19] J. C.-W. Lin, M. Pirouz, Y. Djenouri, C.-F. Cheng, and U. Ahmed, "Incrementally updating the high average-utility patterns with pre-large concept," *Applied Intelligence*, vol. 50, no. 11, pp. 3788–3807, Nov. 2020, https://doi.org/10.1007/s10489-020-01743-y.

[20] W. Song, Y. Liu, and J. Li, "Mining high utility itemsets by dynamically pruning the tree structure," *Applied Intelligence*, vol. 40, no. 1, pp. 29–43, Jan. 2014, https://doi.org/10.1007/s10489-013-0443-7.

[21] M. Liu and J. Qu, "Mining high utility itemsets without candidate generation," in *21st ACM international conference on Information and knowledge management*, Maui, HI, USA, Nov. 2012, pp. 55–64, https://doi.org/10.1145/2396761.2396773.

[22] P. Fournier-Viger, J. C.-W. Lin, R. Nkambou, B. Vo, and V. S. Tseng, *High-Utility Pattern Mining: Theory, Algorithms and Applications*. New York, NY, USA: Springer, 2019.

[23] P. Fournier-Viger, J. C.-W. Lin, Q.-H. Duong, and T.-L. Dam, "FHM: Faster high-utility itemset mining using length upper-bound reduction," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Morioka, Japan, Aug. 2016, pp. 115–127, https://doi.org/10.1007/978-3-319-42007-3_11.

[24] P. Fournier-Viger, C.-W. Wu, S. Zida, and V. S. Tseng, "FHM: Faster High-Utility Itemset Mining Using Estimated Utility Co-occurrence Pruning," in *International Symposium on Methodologies for Intelligent Systems*, Roskilde, Denmark, Jun. 2014, pp. 83–92, https://doi.org/10.1007/978-3-319-08326-1_9.

[25] J. C.-W. Lin, W. Gan, T.-P. Hong, and J.-S. Pan, "Incrementally Updating High-Utility Itemsets with Transaction Insertion," in *International Conference on Advanced Data Mining and Applications*, Guilin, China, Dec. 2014, pp. 44–56, https://doi.org/10.1007/978-3-319-14717-8_4.

[26] 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, Jan. 2018, https://doi.org/10.1109/ACCESS.2018.2819162.

[27] C. Zhang, G. Almpanidis, W. Wang, and C. Liu, "An empirical evaluation of high utility itemset mining algorithms," *Expert Systems with Applications*, vol. 101, pp. 91–115, Jul. 2018, https://doi.org/10.1016/j.eswa.2018.02.008.

[28] S. R. Meruva and B. Venkateswarlu, "A Novel Data Stream High Utility Itemset Miner with the Batch Transaction Processing Model," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 21s, pp. 3858–3865, Mar. 2024.

[29] "Datasets," *SPMF*. https://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php.