International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Data Analysis and Visualisation of Sales Data Using Artificial Intelligence

Swati Raut, *Tanmay Sakharkar , Ritik Zilpe, Akshay Shete, Nandini Agrawal , Akshay Kumar Talanje
Department of Computer Science, JD College of Engineering And Management, Nagpur, Maharashtra, India

Article Info
Volume 9, Issue 3
Page Number : 503-508

Publication Issue :
May-June-2022

Article History
Accepted : 10 June 2022
Published: 24 June 2022

ABSTRACT

This paper in brief describes the system of improvement of numerous case tasks via presenting a graduate degree path on Data Mining. It then outlines a specific case mission that describes the system of records extracting, records cleansing, records transfer, records warehouse layout and improvement. It additionally outlines the improvement of a records dice in addition to application to recognize enterprise intelligence. The outcomes may be useful to an trainer who desires to expand a realistic path or a practitioner venturing into the records warehousing and records mining region.

Keywords: Business intelligence, Data mining, Case, and Course.

I. INTRODUCTION

Because of the fierce opposition withinside the market, all people is busy with getting the most interest of people. For that manufacturer have to have merchandise which satisfies the wishes of clients. Huge scale studies goes on this field. In such situations, patron necessities are very essential. The cost of a manufacturing plan may be modeled as a characteristic that displays the communique of the business enterprise with specific agents, for instance, clients and competitors. The difficulty focused on this machine is to understand the manufacturing plan with the most application for a business enterprise, in which the application of a manufacturing plan is classified via way of means of anticipated variety of the clients for the selected merchandise withinside the plan. Consider a grocery store with a huge series of objects. Typical enterprise choices that the control of the grocery store has to make includes, what to place on sale, the way to layout coupons, the way to vicinity products on cabinets with a view to maximize the profit, etc. Analysis of beyond transaction records is a generally used technique with a view to enhance the exceptional of such choices. Until recently, but, handiest worldwide records approximately the cumulative income at some stage in a while duration (a day, a week, a month, etc.) turned into to be had at the computer. Progress in bar-code generation has made it viable to save the so known as basket records that shops objects bought on a per-transaction basis. Basket records kind transactions do now no longer always include objects sold collectively on the identical factor of time. It may also include objects sold via way of

means of a patron over a duration of time. Examples consist of month-to-month purchases via way of means of participants of a e-e book membership or a track membership. Several companies have accrued big quantities of such records. These records units are typically saved on tertiary garage and are very slowly migrating to database structures. One of the primary motives for the restrained achievement of database structures on this region is that cutting-edge database structures do now no longer offer essential capability for a consumer interested by taking benefit of this information. This mission introduces the hassle of mining" a huge series of basket records kind transactions for affiliation guidelines among units of objects with a few minimal designated self belief, and provides an green set of rules for this purpose. An instance of such an affiliation rule is the assertion that 90% of transactions that buy bread and butter additionally buy milk. The antecedent of this rule includes bread and butter and the ensuing includes milk alone. The variety 90% is the self belief thing of the rule. We Propose Apriori set of rules for locating the k-least merchandise which is likewise essential for manufacturing plan.

II. PROPOSED SYSTEM

In data mining, Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation).

The whole point of the algorithm (and data mining, in general) is to extract useful information from large amounts of data. For example, the information that a customer who purchases a keyboard also tends to buy a mouse at the same time is acquired from the association rule below:

Support: The percentage of task-relevant data transactions for which the pattern is true.

Support (Keyboard -> Mouse) =

$$\frac{\text{No. of transacti ons containing both Keyboard and Mouse}}{\text{No. of total transacti ons}}$$

Confidence: The measure of certainty or trustworthiness associated with each discovered pattern.

Confidence (Keyboard -> Mouse) =

$$\frac{\text{No. of transacti ons containing both Keyboard and Mouse}}{\text{No. of transacti ons containing (Keyboard)}}$$

The algorithm aims to find the rules which satisfy both a minimum support threshold and a minimum confidence threshold (Strong Rules).

Item: article in the basket.

Itemset: a group of items purchased together in a single transaction.

Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Other algorithms are designed for finding association rules in data having no transactions (Winepi and Minepi), or having no timestamps (DNA sequencing). Each transaction is seen as a set of items (an itemset). Given a threshold C, the Apriori algorithm identifies the item sets which are subsets of at least C transactions in the database.

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

The pseudo code for the algorithm is given below for a transaction database T, and a support threshold of ϵ. Usual set theoretic notation is employed; though note

that T is a multiset. Ck is the candidate set for level k. At each step, the algorithm is assumed to generate the candidate sets from the large item sets of the preceding level, heeding the downward closure lemma. Count[c] accesses a field of the data structure that represents candidate set C, which is initially assumed to be zero. Many details are omitted below, usually the most important part of the implementation is the data structure used for storing the candidate sets, and counting their frequencies.

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon) \\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\} \\
&\quad k \leftarrow 2 \\
&\quad \textbf{while } L_{k-1} \neq \emptyset \\
&\qquad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \nsubseteq L_{k-1}\} \\
&\qquad \textbf{for transactions } t \in T \\
&\qquad\quad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\} \\
&\qquad\quad \textbf{for candidates } c \in C_t \\
&\qquad\qquad count[c] \leftarrow count[c] + 1 \\
&\qquad L_k \leftarrow \{c \mid c \in C_k \wedge\ count[c] \geq \epsilon\} \\
&\qquad k \leftarrow k + 1 \\
&\quad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

III. SYSTEM DESIGN

A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of processes, or information about whether processes will operate in sequence or in parallel.

Fig 1. DFD Level-0

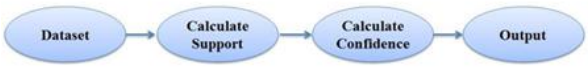

Fig 2. DFD Level-1

Fig 3. DFD Level-2

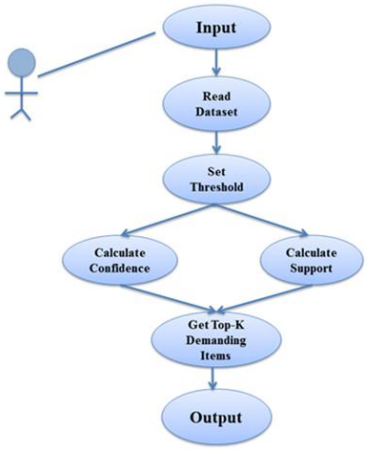

Fig 4. Use Case Diagram

IV. RESULT

Fig 5. Home Screen

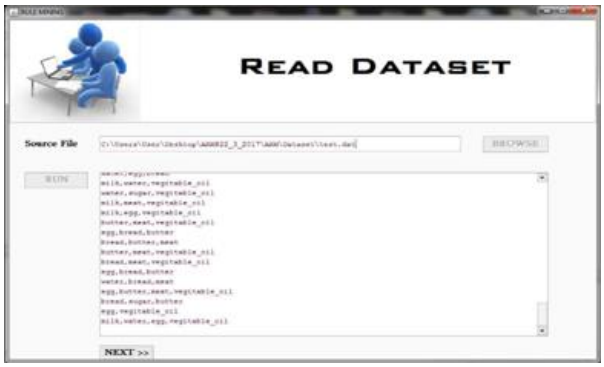

Fig 6. Read Dataset

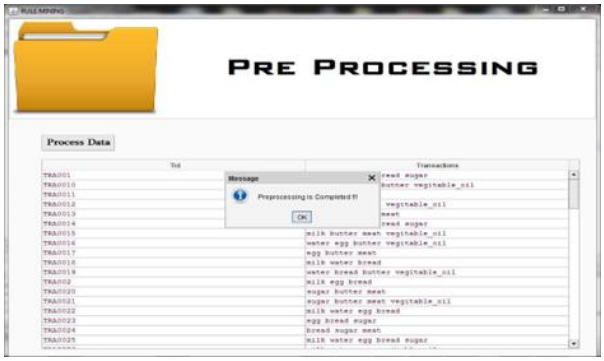

Fig 7. Data Pre-Processing

Fig 8. Input Thresholds

Fig 9. Unique Item Generation

Fig 10. Calculate Support

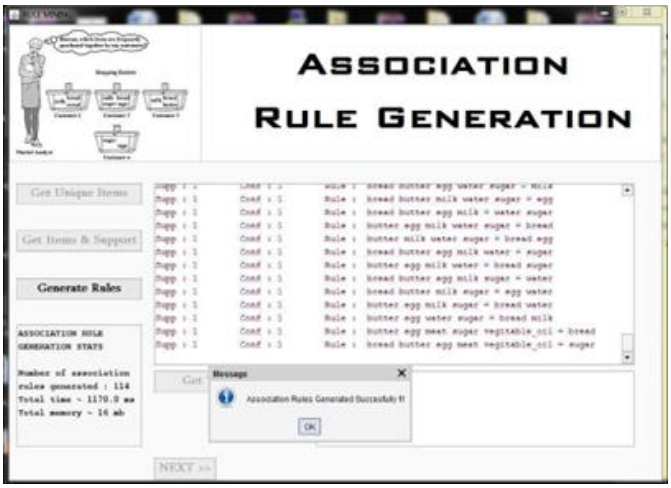

Fig 11. Generate Association Rule

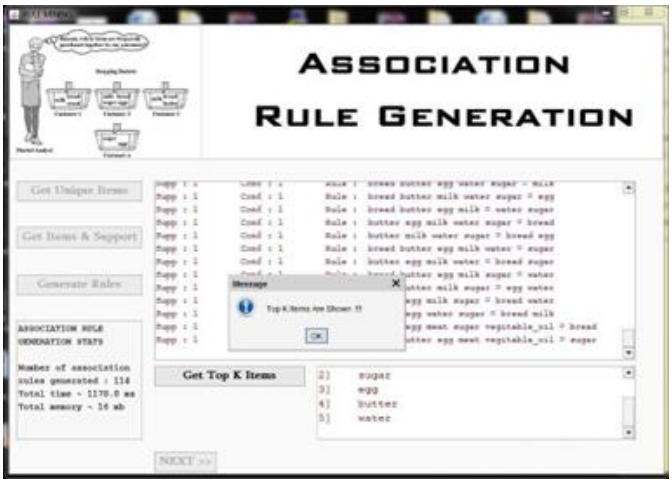

Fig 12. Generate Top-K Items

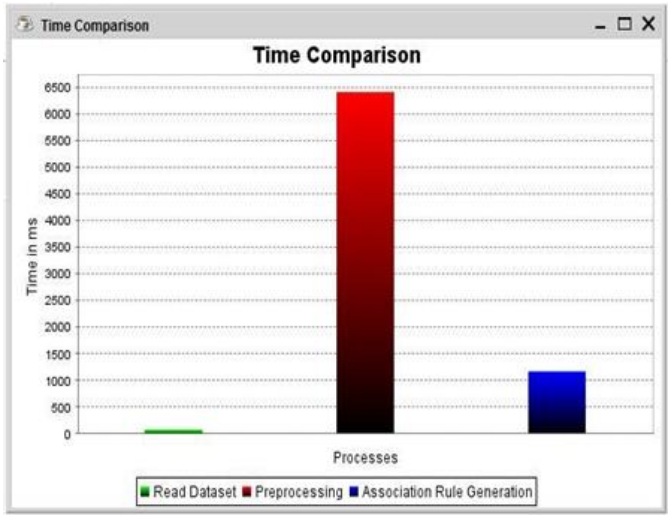

Fig 13. Time Comparison

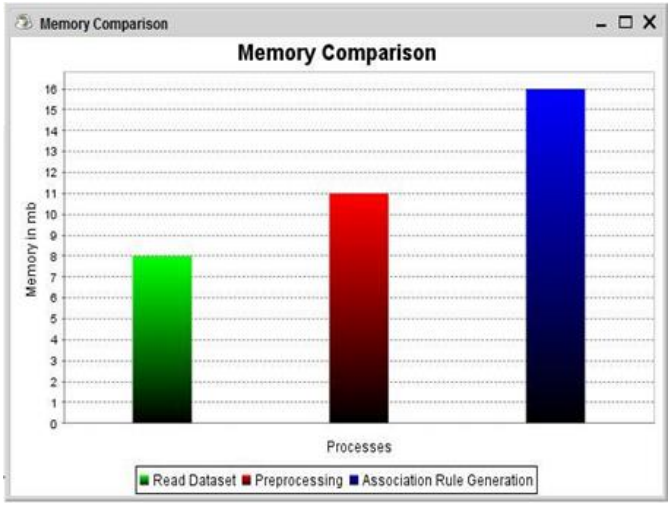

Fig 14. Memory Comparison

V. CONCLUSION

We applied Apriori set of rules for affiliation rule mining so that you can get the pinnacle ok disturbing gadgets from the transactional dataset. The Apriori precept can lessen the range of itemsets we want to examine. Put simply, the Apriori precept states that if an itemset is infrequent, then all its subsets ought to additionally be infrequent. This method that if become observed to be infrequent, we will expect to be similarly or maybe extra infrequent. So in consolidating the listing of famous itemsets, we want now no longer consider , nor every other itemset configuration that incorporates beer. Results display the implementation paintings and the outcomes generated in phrases of Time & Memory Consumption.

VI. REFERENCES

[1]. Agrawal, R., Imielinski, T., Swami, A., "Mining Association Rules between Sets of Items in Large Database", In: ACM SIGMOD International Conference on Management of Data (1993).

[2]. Yao, H., Hamilton, H.J., Buzz, C. J., "A Foundational Approach to Mining Itemset Utilities from Databases", In: 4th SIAM International Conference on Data Mining, Florida USA (2004).

[3]. Yao, H., Hamilton, H.J., "Mining itemset utilities from transaction databases", Data & Knowledge Engineering 59(3), 603–626 (2006).

[4]. Liu, Y., Liao, W.K., Choudhary, A., "A Fast High Utility Itemsets Mining Algorithm", In: 1st Workshop on Utility-Based Data Mining, Chicago Illinois (2005).

[5]. Erwin, A., Gopalan, R.P., N.R. Achuthan, "CTUMine: An Efficient High Utility Itemset Mining Algorithm Using the Pattern Growth Approach", In: IEEE CIT 2007.Aizu Wakamatsu, Japan (2007).

[6]. Han, J., Wang, J., Yin, Y., "Mining frequent patterns without candidate generation", In: ACM SIGMOD International Conference on Management of Data (2000).

[7]. Erwin, A., Gopalan, R.P., Achuthan, N.R, "A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets", In: International Workshop on Integrating AI and Data Mining. Gold Coast, Australia (2007).

[8]. CUCIS. Center for Ultra-scale Computing and Information Security, North-western University.

[9]. Yao, H., Hamilton, H.J., Geng, L., "A Unified Framework for Utility Based Measures for Mining Itemsets", In: ACM SIGKDD 2nd Workshop on Utility-Based Data Mining (2006).

[10]. Pei, J., "Pattern Growth Methods for Frequent Pattern Mining", Simon Fraser University (2002).S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[11]. Sucahyo, Y.G., Gopalan, R.P., CT-PRO: "A Bottom- Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure", In: IEEE ICDM Workshop on

Frequent Itemset Mining Implementation (FIMI), Brighton UK (2004).

[12]. G. Salton, Automatic Text Processing, Addison-Wesley Publishing, 1989.

[13]. J. Pei, J. Han, L.V.S. Lakshmanan, “Pushing convertible constraints in frequent itemset mining”, Data Mining and Knowledge Discovery 8 (3) (2004) 227–252.

[14]. Bin Chen, Peter Hass, Peter Scheuermann, “A New Two-Phase Sampling Based Algorithm for Discovering Association Rules”, SIGKDD '02 Edmonton, Albetra, Canada © 2002 ACM 1 58113 567 X/02/2007.

[15]. Ming-Yen lin, Tzer-Fu Tu, Sue-Chen Hsueh, “High utility pattern mining using the maximal itemset property and, lexicographic tree structures”, Information Science 215(2012) 1-14.

[16]. Sudip Bhattacharya, DeeptyDubey, “High utility itemset mining, International Journal of Emerging Technology and advanced Engineering”, ISSN 2250-2459, Volume 2, issue 8, August 2012.

[17]. Y.Liu, W.K. Liao and A. Choudhary, —A two phase algorithm for fast discovery of high utility itemset‖, Cheng, D. and Liu. H. PAKDD, LNCS. PP: 689-695, 2005.

[18]. J.Hu, A. Mojsilovic, —High utility pattern mining: A method for discovery of high utility itemssets‖, in: pattern recognition. PP: 3317-3324, 2007.

[19]. Y.-C. Li,j,-s. Yeh, and C.-C. Chang, —Isolated Items Discarding Strategy for Discovering High Utility Itemsets,‖ Data and Knowledge engg., pp: 198-217, 2008.

[20]. Liu Jian-Ping, Wang Ying Fan-Ding, ‖Incremental Mining algorithm Pre-FP in Association Rule Based on FP-tree‖, Networking and Distributed Computing, International Conference, pp: 199-203, 2010.

Cite this Article

Swati Raut, Tanmay Sakharkar , Ritik Zilpe, Akshay Shete, Nandini Agrawal , Akshay Kumar Talanje, "Data Analysis and Visualisation of Sales Data Using Artificial Intelligence", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 412-417, May-June 2022.

Journal URL : https://ijsrset.com/IJSRSET2293181