International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 10, October-2018

# ON THE USE OF ELEPHANT HERDING OPTIMIZATION FOR MINING PROFITABLE PATTERNS

[1]Gomsi Yadav, [2]Akhilesh Tiwari

[1,2] *Department of CSE & IT Madhav Institute of Technology & Science Gwalior, M.P., India*

***Abstract— Profitable pattern mining explores the zone that highlights the motive of business. Every business aims the most conspicuous element i.e. Profit. To achieve the desired goal researchers determined the innovative way to generate the association rules based on profit.***

***This paper delineates a structure for mining profitable patterns using rough set theory followed by association rule mining algorithm and optimizing the results using Elephant Herding Optimization (EHO) algorithm. EHO is comparatively a new swarm intelligence metaheuristic algorithm that imitates the collective behavior of species from nature. The proposed approach generates profitable patterns (optimized) which will be advantageous for decision making in business. The obtained results are compared with the genetic algorithm which shows that EHO produces the most optimal (profitable) patterns as well as in a reduced time.***

***Keywords— Data mining, Association Rule Mining (ARM), Profitable Pattern Mining (PPM), Rough Set Theory (RST), Elephant Herding Optimization (EHO).***

## I. INTRODUCTION

Data mining (DM) [1] has pulled in a lot of consideration in the data business and generally in society because of the large data. Hence there is a fast requirement for turning such data into valuable information and knowledge which can be used for applications like market analysis, fraud detection, and customer preservation. This information are categorize with the help of DM techniques: classification, clustering, association rule mining, sequential pattern analysis, prediction, data visualization etc. Patterns that are hidden in the database or the information that users may miss are determined by Data mining (DM) Tools for example: MATLAB, WEKA etc.

Association rule mining (ARM) [2] is utilized to discover the association relationship from data recorded. It is a part of data mining, which is sourced to generate patterns that will predict the frequency of an item based on combinations of items that frequently co-occur in transactions. An association rule is such a rule, which includes a possible relation between sets of database attributes. The relationship between the attributes can be represented as

$$A \rightarrow B$$

where A and B are sets of items and $A \cup B = \phi$. Suppose the Transaction data set T is given, we are interested in finding all rules that satisfy the following constraints, *support* and *confidence*. For instance, let us consider a transactional database of a grocery store with only 100 transactions.

**Support:** It is defined as how frequently a rule is relevant to the given transaction records. For instance, the item Bread has occurred 80 times,

Support (Bread) = 80 / 100 x 100 = 80%

**Confidence**: It is defined as recurrence of an item when another item has been purchased before. If jam is bought along with bread in 60 transactions of 100, then

Confidence (Bread, Jam) = Support (Bread U Jam)/Support (Bread)
= 60 / 80 x 100 = 75%

Where MinSup = 50% and MinConf = 70%, are two thresholds. The above support and confidence of items lead to a decision in a form as Bread → Jam i.e., if bread is bought then jam is also bought along with it, the store brought up a conclusion to place jam near the bread, thereby increasing the sale of jam. But the patterns so obtained do not ensure the basic objective of any business organization i.e. profit. However a new approach to ARM is classified as **Profitable Pattern Mining (PPM)**.

Today, the major concern for mining profitable patterns is to determine whether an item is of interest to customer and also for an organization. For this a recommender recommends a pair of target items to the future customer whenever they buy non-target items [3]. Although this issue is defined in terms of retail sales, some techniques are implemented under the notion of profit. All this considerations lead to the evolution of profit mining.

The main motive of PPM is to build up a model which generates profitable rules, recommender rules that suggest target items and advancement procedure for future customer. It is a novel method of ARM which intends to discover those

patterns which contributes maximum profit for an organization. As the major obstacle in ARM application is the gap between the statistical based patterns extraction and valued based decision making, PPM helps in reducing those gaps. Fig 1 illustrates the hierarchy structure of PPM.

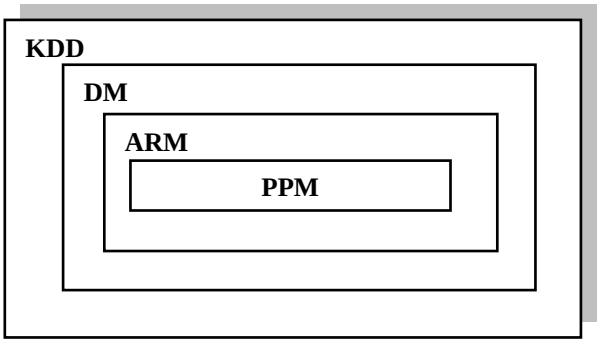



**Fig 1** Hierarchy of Profitable Pattern Mining

RST was introduced in 1982 by professor Pawlak, which is a mathematical tool for dealing imprecision, inconsistent, partial information & knowledge. It can find structured relationship of inaccurate data or noisy data, which is based on an equivalence class establishment with the given training data.

In this paper profitable rules are mined using the concept of RST accompanied by mining association rule (Apriori Algorithm). The rules so obtained are still huge, so the outcomes are optimized using EHO. Also, performance of genetic algorithm and EHO is compared based on accuracy to generate profitable rules.

The remaining paper has been organized in such a way; **Section 2** demonstrates the related work accompanied by EHO algorithm. **Section 3** introduce with proposed work and algorithm. Illustration and result along with the comparison of rules and accuracy of Genetic Algorithm with EHO were presented in **Section 4**. In the end, **Section 5** winds up the paper.

## II. RELATED WORK

**In 2002, Ke Wang et al**. [4] proposed the model for profit mining. They pointed out that the motive of profit mining is to construct a model to recommend suitable products and proper prices for customers in order to maximize net profit. Hence, this paper emphasizes value-based pattern evaluation in the process of knowledge discovery by using both objective and subjective measures to evaluate the interestingness and usefulness of association rules. The proposed approach can help analysts make decisions with discovered rules, which are profitable to enterprises. It exhibited an adaptable development of recommenders to address several important requirements in profit mining: pruning specific rules on a profit-sensitive basis, dealing with the behavior of shopping on unavailability, dealing with thinly dispersed and volatile search space, ensuring optimality and interpretability of recommenders. The goal of profit mining is to suggest "right" cost and "right" items.

**In 2004, Hong Yao et al.** [5] classify the utility into transaction utility and external utility. The transaction utility of an item describes the information that is gathered in the transaction, i.e, quantity. The external utility of an item is dependent on information that is not obtainable in the transaction, describing the user preferences i.e, profit to achieve the commercial goals.

**In 2007, Yaohua Chen et al**. [6] proposed a way to deal profit based business representation that calculates interestingness of rules. In this paper the relationships are investigated between discovered rules, marketing planning and profit growth. Or we can say the statistical factors of the rules are examined defining the norms, such that higher profit can be generated in related marketing strategies. Marketing strategies for raising the profit can be generally classified into two groups based on the two factors in the profit model:

- **Price-based strategy:** based on increasing the unit profit.
- **Volume-based strategy:** based on increasing the sales volume.

Relationship between unit profit and sales volume is clear by Graphical representation:-

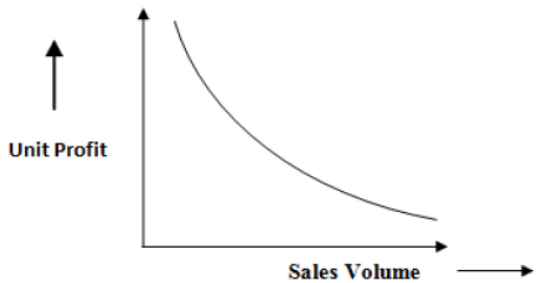



**Fig 2** Relationships between Unit Profit and Sales Volume

Final results for describing the notion of profit of the rules of type A → B (B is dependent on sales of A) can be concluded by following table:

**Table 1**

| Strategies / Item sets | Price Based Strategy | Volume Based Strategy |
|---|---|---|
| A | Profit Decreases | Profit Increases |
| B | Profit Increases | Profit Decreases |

**In 2009, Tzung-Pei Hong et al**. [7] proposed a two-phase mining algorithm along with a new mining measure called average utility to discover high average-utility itemsets. The proposed mining algorithm is divided into phases. In phase 1, this algorithm exaggerates the usefulness of items from the perspective of transactions; this process places the "Downward closure" property to capably clip impossible utility itemsets level by level. In phase-2, according to candidates obtained from Phase-1, only real high average utility items require scanning of records once. Assuming that the size of items is a significant reason to affect the utility projections of items, the measurement "average utility" is used to understand utility values.

**In 2010**, **Sandhu, P.S. et al.** [8] proposed a way to mine association rules dependent on utility and weight factor efficiently. Firstly the methodology utilizes conventional Apriori algorithm to acquire a set of association rules from a record. Weight age (W-gain) and utility (U-gain) are the two constraints that are employed over set of association rules. For every association rule mined, an integrated Utility Weighted Score (UW-Score) is calculated. They set a subset of valuable association rules based on the calculation of UW-Score. Consequently rules are formed which are a subset of traditional rules mined based on UW-score that can be advantageously applied for business development.

**In 2014, Sameer Kumar et al.** [9] proposed a model for profit pattern based on genetic algorithm. In this paper, the concept of profit pattern mining is applied with genetic algorithm to generate profit oriented pattern which helps out in future business expansion and fulfill the business targeted goal. Two kinds of profits are proposed i.e. value profit and percentage of profit along with two quality measures like completeness and Interestingness were proposed. They applied traditional ARM followed by genetic algorithm techniques. The implementation work depicts the efficiency of the proposed approach in terms of generating optimized profitable patterns that can thoroughly be applied in rise of business.

**In 2015, Satpal Singh et al.** [10] proposed a method that will modify classical support using formulas as classical support only finds statistical significance of items. This includes the profit factor. They proposed an approach of profit ARM that calculates profit significance based on count and profit values of items. They eliminated the fixed profit limitations by introducing new profit parameters, such as, actual profit of each item, its scaled value on the basis of its average profit, and then accumulating it with the classical support. Another factor that we consider is an inversely proportional parameter of inventory cost; since inventory cost has its negative effect on the profit, including inventory cost factor does ensure the absolute profit of any item.

**In 2016, Vivek Badhe et al.** [11] proposed a model for profit pattern mining based on genetic algorithm. In this paper they applied the theory of profit pattern mining with genetic algorithm to examine profit oriented pattern. Since genetic algorithms are capable to handle the problems related with the uncertainty, multidimensional, non-differential, non-continuous, and non-parametrical, non-linearity constraint and multi-objective optimization problems, it also enhanced the effectiveness and efficiency of ARM outcome. The generated profit oriented patterns helps out in future business expansion and fulfill the business motive.

### ○ ELEPHANT HERDING OPTIMIZATION(EHO)

Wang et al. [12] proposed a swarm-based metaheuristic optimization algorithm, EHO. This algorithm was inspired by herding behavior of elephants. It considered two operators, clan updating operation and the other is clan separation operation.
To solve all types of global adaptation problems in herding behavior of elephants, here preference would be given to simplify it in the following ideal rules:

- The elephant population is composed of some clans, and each clan has fixed number of elephants.
- A fixed number of male elephants will leave their family group and live solitarily far away from the main elephant group at each generation.
- The elephants in each clan live together under the leadership of a matriarch.

*Clan Updation Operation:*

The elephant population comprises of small group of clans which has finite number of elephants. The elephants in each clan stay together under the leadership of matriarch. Matriarch is considered as the fittest elephant in the clan. Its position is updated based on the position of the individuals in the clan.

*Clan Separation Operation:*

Once, the male elephants in the group grow up, they leave the group and go to live far away. This process gives the formation of separating operation when solving optimization problems.

*EHO Algorithm:*
**Step 1**: Initialize the population.
**Step 2**: Repeat
**Step 3**: Sort all the elephants according to their fitness value using Fitness Function.
**Step 4**: Apply clan updation operation. Each elephant j of clan i moves according matriarch where matriarch is the elephant $c_i$ with the best fitness value. The updated position is computed as

$$x_{new,ci,j} = x_{ci,j} + \alpha * (x_{best,ci} - x_{ci,j}) * r$$

Where $x_{new,ci,j}$ represent new position of elephant j in clan i, $x_{ci,j}$ represent old position of elephant j in clan i, $x_{best,ci}$ represent the position with best fit (matriarch), $\alpha \in [0,1]$ is algorithms parameter, $r \in [0,1]$ is the random number used to enhance the diversity of population in the upcoming stages of algorithm. Position update for the best fit in the clan $x_{best,ci}$ is computed as:

$$x_{new,ci} = \beta * x_{center,ci}$$

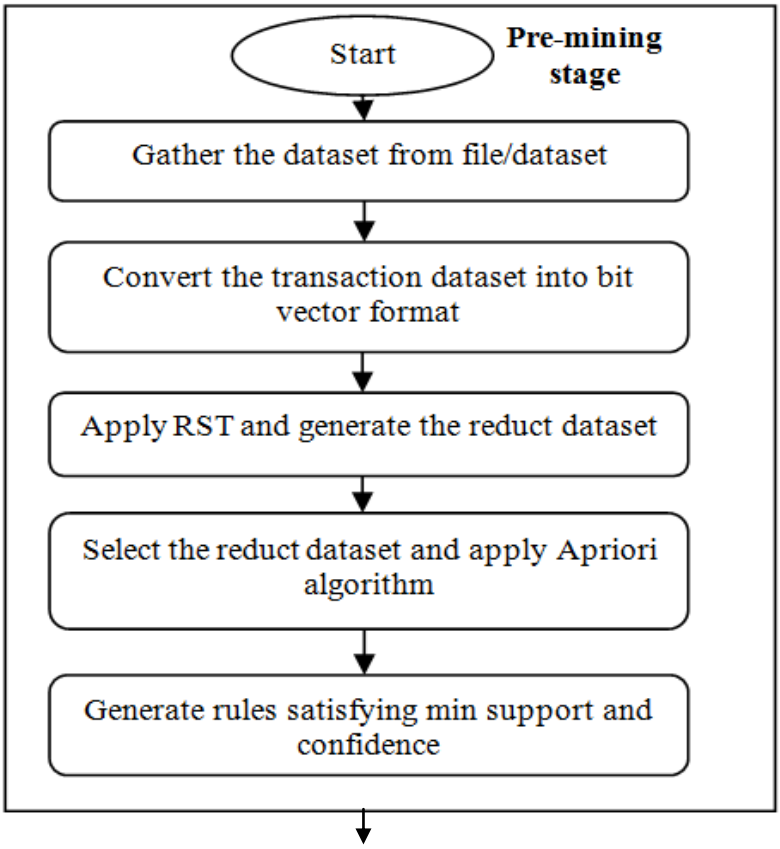
Where $x_{center,ci} = \sum_{j=1}^{n} \frac{x_{ci,j,d}}{n}$ , $\beta \in [0,1]$ is second parameter of the algorithm, where $x_{center,ci}$ represent the center of the clan $c_i$ and n is the number of elephants in clan $c_i$.
**Step 5**: Apply clan separation operation. In each clan i some number of the elephants with the worst value are moved to the new position according to following equation

$$x_{worst} = x_{min} + (x_{max} - x_{min} + 1) * rand$$

Where rand $\varepsilon$ [0,1], $x_{min}$ is the lower bound of the search space, $x_{max}$ is the upper bound of the search space.
**Step 6**: Evaluate population by the newly updated position.
**Step 7**: Until the stop criteria.
**Step 8**: Return the best solution.

## III. PROPOSED METHODOLOGY

The novel proposed structure and design is presented in Fig 2 for mining profitable patterns using Rough Set Theory (RST) along with Apriori Algorithm and Elephant Herding Optimization (EHO).



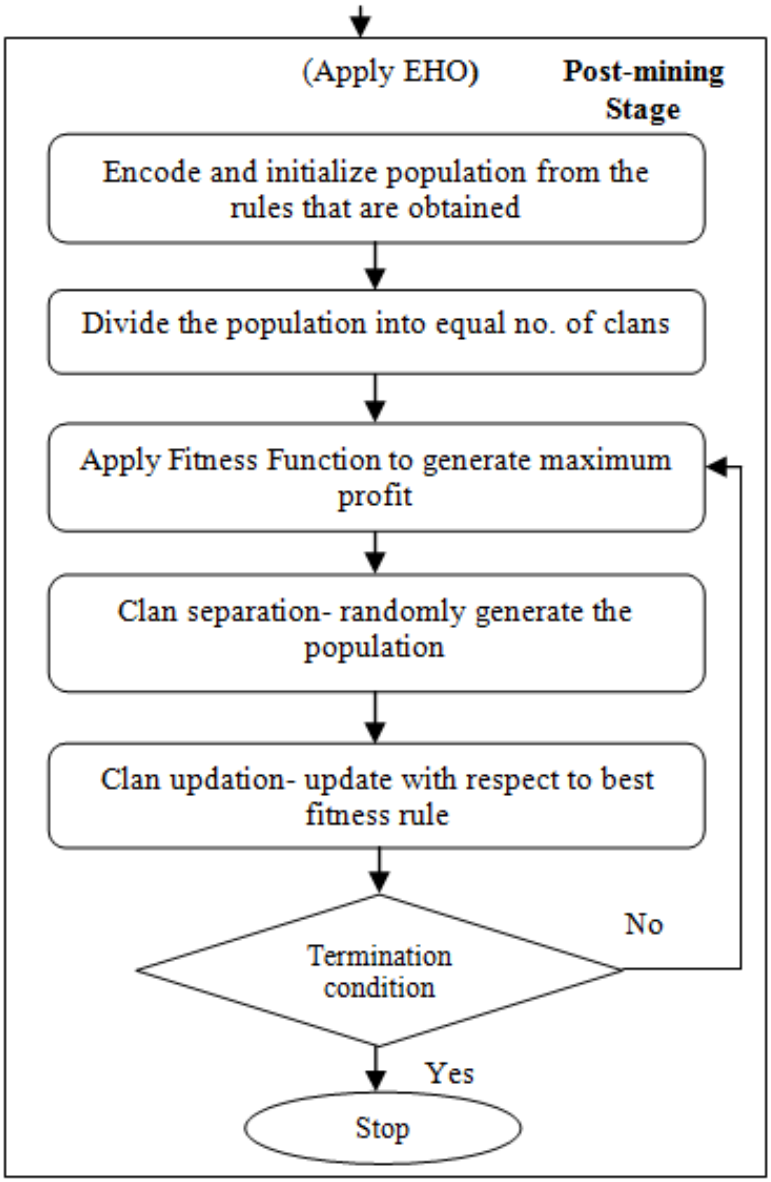



**Fig 3** Flowchart of Proposed Work

**Algorithm**: Profitable Pattern Mining

**Step 1**: Start
**Step 2**: Load the transaction dataset T and Cost-Quantity CO_QTY table.
**Step 3**: Transform the transaction dataset T into bit vector format.
**Step 4**: Apply RST and generate the reduct dataset.
**Step 5**: Apply Apriori algorithm to the reduct dataset with defined min support and confidence value for generating rules.
**Step 6**: Store the results in a rule set.
**Step 7**: Apply EHO on the rule set.

- Encoding : rules are represented by following chromosome schemes

| | | |
|---|---|---|
| Present & Antecedent | - | 11 |
| Present & Consequent | - | 10 |
| Absent | - | 00 |

For the rule A → C, encoding is done as follows. Since 'A' is antecedent and 'C' is consequent, rule is represented as 110010.

- Repeat for specified number of iteration

  Clan Separation: randomly generate the population into finite number of clans.

  Clan updation: following formula is used for updating the position of an individual in a clan

$$x_{new,\text{ci,j}} = x_{ci,j} \text{ OR } (x_{best,ci} \oplus x_{ci,j})$$

For the above rule A → C, the rule is XORed with the best fit rule in the clan B → A.

| | | |
|---|---|---|
| 110010 | -------- | A → C |
| | ⊕ | |
| 101100 | -------- | B → A |
| 011110 | OR | |
| 110010 | -------- | A → C |
| 111110 | -------- | A, B → C |

Therefore A, B → C is the new rule generated from above formula.

- Check termination condition

**Step 8**: Select and store the rules according to their fitness. Obtained results are optimized (profitable) rules.
**Step 9**: Mapped the result into a specified format.
**Step 10**: Stop.

As the key component for any business is Profit, the Fitness Function (FF) for EHO is characterized as the notion of profit which is defined as margin profit, weighted profit, and percentage profit. Although the notion of profit vary depending on type of business but generally, these three are considered.
For each rule A → B, the Fitness Function (FF) is defined as:-

$$FF = \frac{C * w1 + I * w2}{w1 + w2}$$

Where,
C- Completeness
I- Interestingness

The value of w1 and w2 is calculated as:-

$$W1 = \frac{Percentage\ Profit\ of\ A * Weighted\ Profit\ of\ A}{Percenatge\ Profit\ of\ B * Weighted\ Profit\ of\ B}$$

$$W2 = \frac{Margin\ Profit\ of\ A * Weighted\ Profit\ of\ A}{Margin\ Profit\ of\ B * Weighted\ Profit\ of\ B}$$

- *Completeness (C)*:- It is defined as:

$$C = \frac{|A \cup B|}{|B|}$$

- *Interestingness (I)*:- It is defined as:

$$I = \frac{Support(A \cup B)}{Support(A)} * \frac{Support(A \cup B)}{Support(B)} * \left(1 - \frac{Support(A \cup B)}{N}\right)$$

Where, N is the total no of transactions in dataset.

## IV. ILLUSTRATION AND RESULT

For the analysis of proposed methodology we are using MATLAB (R2013a) tool. Table 2 shows the synthetic Transaction dataset T that has been used for the above calculation:

**Table 2:** Transaction Dataset

| TID | Items |
|---|---|
| 1. | A, B, C |
| 2. | A, B, C, D, E, F |
| 3. | D |
| 4. | C |
| 5. | A, B, C |

| | |
|---|---|
| 6. | A, B, C, D, E |
| 7. | A, B, C, D, E |
| 8. | A, B, C, D |
| 9. | A, B, C |
| 10. | A, D |
| 11. | E |
| 12. | A, B, C, F |
| 13. | A, B, C, F |
| 14. | C |
| 15. | A, B, C |
| 16. | A |
| 17. | A |
| 18. | A, B, C, D |
| 19. | A, B, C, D, E, F |
| 20. | D, E |
| 21. | A, B, C |
| 22. | A, B, C, D, F |
| 23. | A, B, E, F |
| 24. | B |
| 25. | A, B, C, D, E |

**CO_QTY TABLE:**

Table 3 represents the CO_QTY table (Cost-Quantity table), which contains 6 tuples with 4 attributes, respectively, with items, cost price, sales price and weight (quantity).

**Table 3:** CO_QTY Table

| ITEMS | COST PRICE | SALES PRICE | WEIGHT(QUANTITY) |
|---|---|---|---|
| A | 10 | 15 | 2 |
| B | 08 | 16 | 1 |
| C | 13 | 16 | 3 |
| D | 12 | 14 | 1 |
| E | 10 | 13 | 2 |
| F | 10 | 11 | 1 |

Firstly the above two tables, table 2 and table 3 are loaded for further process. Table 2 is converted into Bit Vector Format as depicted in Fig 4.

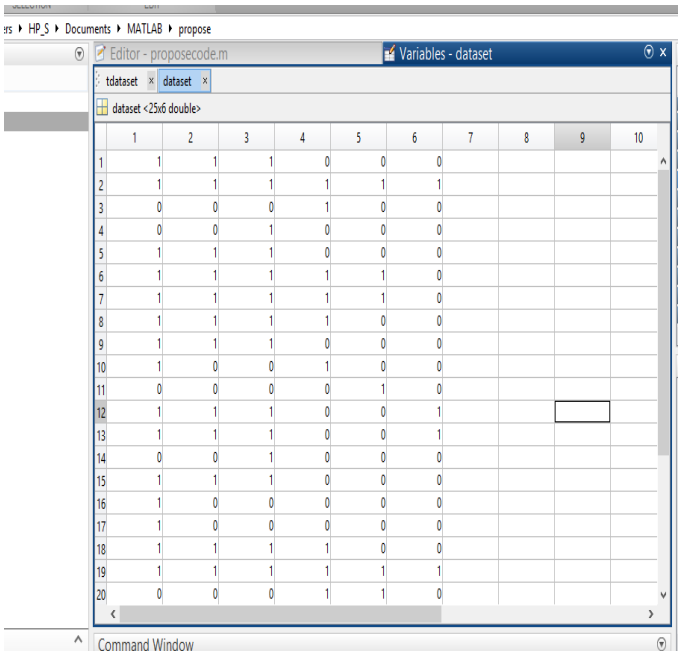



**Fig 4** Bit Vector Formatted Dataset

Apply RST over the above bit vector formatted dataset to generate the reduct dataset. Two reducts are generated after applying RST, one is **abdef** and the other is **acdef**. Any one reduct can be chosen for further calculation. Here **abdef** reduct is chosen for further process. The new reduct dataset is shown in fig 5.

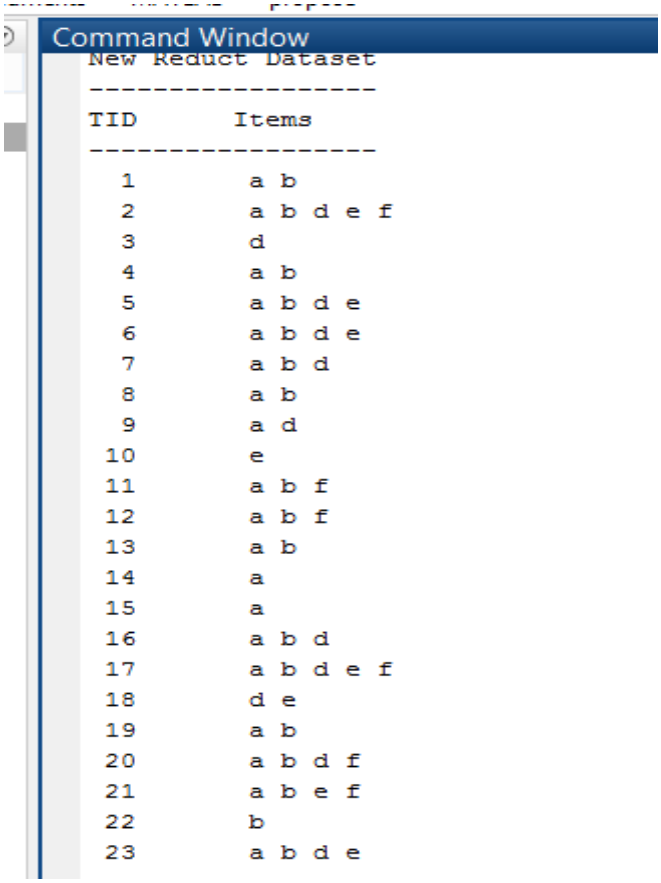



**Fig 5** Reduct dataset

Apply Apriori algorithm (min fitness = 3 and min confidence = 50%) to the reduct dataset and generate the rules as shown in fig 6.

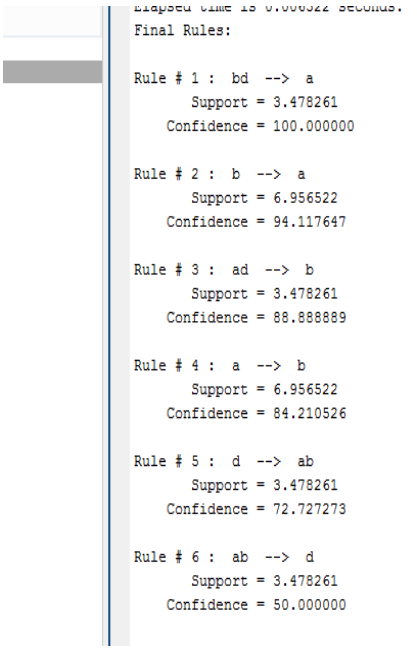



**Fig 6** Apriori generated rules

Now using the EHO, the above rules have been adapted for the production of new rules. Here minimum Fitness value is taken as 1.20. Finally we are acquired with new rules, fig 7.

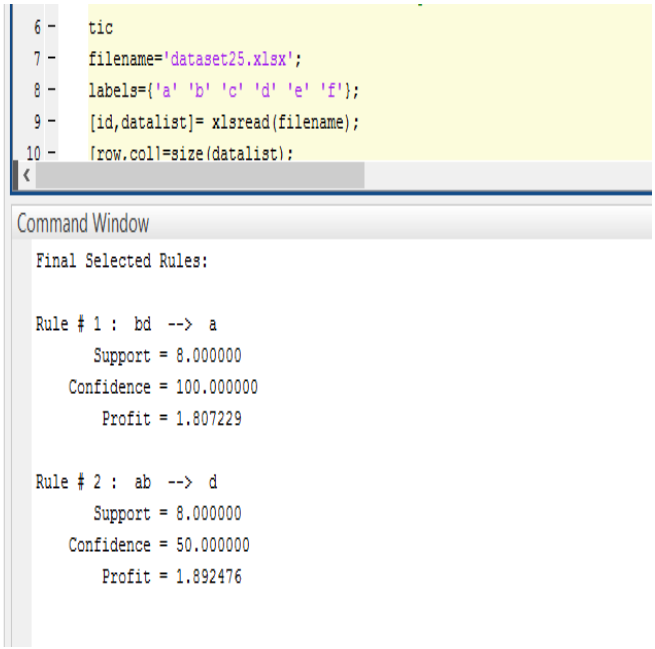



**Fig 7** optimized profitable patterns

Table 4 demonstrates the figured numbers of obtained results with Genetic Algorithm. Comparison chart, fig 8, describe the below table briefly.

**Table 4:** Result Analysis Table

| | GENETIC ALGORITHM | EHO |
|---|---|---|
| **No of patterns generated** | 3 | 2 |
| **Execution time (in seconds)** | 1.559662 | 0.150465 |

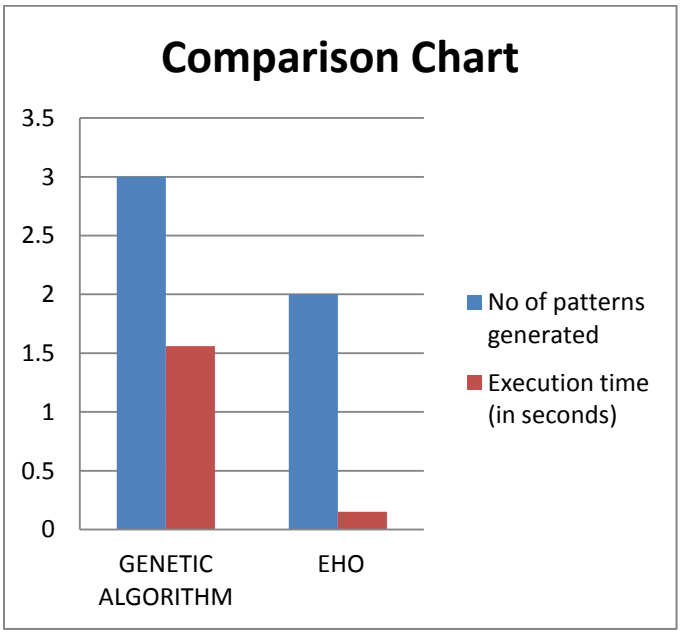



**Fig 8** Comparison Chart

The above graph clearly explains that EHO is better than genetic algorithm in terms of accuracy and optimized (profitable) patterns.

## V. CONCLUSION

Profitable Pattern Mining (PPM) is a new concept for providing decision making in business. The key commitment of this paper is to generate those patterns that are profit oriented. In our proposed work we have generated those patterns that are profit oriented. For achieving this, rough set theory (RST), followed by association algorithm (Apriori) has been used, and optimized the patterns by Elephant Herding Optimization (EHO) to get most Profitable patterns. As well as, number of patterns and execution time of EHO with genetic algorithm has been compared, which shows that EHO produces the most optimal (profitable) patterns, along with reduced time than the genetic algorithm. Since the major discussion for the above topic is profit-based patterns, researchers need to require more attention for defining the notion of a profit which is profitable for business growth and decision making.

## REFERENCES

[1] J. Han and M. Kamber, "Data Mining: Concepts and techniques", Morgan Kaufmann Publishers, Elsevier India, 2001.
[2] R Agrawal, T.Imielinski, and A.Swami,"Mining association rules between sets of items in large databases", in proceedings of the ACM SIGMOD Int'l Conf. on Management of data, pp. 207-216, 1993.
[3] Ke Wang, Senqiang Zhou, and Jiawei Han, "Profit Mining: From Patterns to Actions, C.S. Jensen et al. (Eds.): EDBT 2002, LNCS 2287, pp. 70–87, Springer-VerlagBerlin, 2002.
[4] Ke Wang, Senqiang Zhou, and Jiawei Han, "Profit Mining: From Patterns to Actions", Springer-Verlag Berlin Heidelberg, 2002.
[5] Hong Yao, Howard J. Hamilton, and Cory J. Butz, "A Foundational Approach to Mining Itemset Utilities from Databases", SIAM, 2004.
[6] Yaohua Chen, Yan Zhao and Yiyu Yao, "A Profit-based Business Model for Evaluating Rule Interestingness", Proceedings of the 20th Canadian Conference on Artificial Intelligence (CAI'07), 296-307, 2007.
[7] Tzung-Pei Hong, Cho-Han Lee & Shyue-Liang Wang "Mining High Average-Utility Itemsets", IEEE 2009.
[8] Parvinder S. Sandhu, Dalvinder S. Dhaliwal, S. N. Panda and Atul Bisht, "An Improvement in Apriori algorithm Using Profit And Quantity", IEEE 2010.
[9] Sameer Kumar Vishnoi, Vivek Badhe, "Association rule mining for profit pattern using Genetic algorithm", International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 5, May 2014.
[10] Satpal Singh, Vivek Badhe, "An exclusive survey on Profit Association Rules Mining", International Journal of Advanced and Innovative Research (2278-7844) / # 56 / Volume 4 issue 9.
[11] Vivek Badhe, R.S. Thakur, G.S. Thakur, "A model for profit pattern mining based on genetic algorithm", Volume: 05 Issue: 08 | Aug-2016, Available @ http://ijret.esatjournals.org.
[12] G.-G. Wang, S. Deb, X.-Z. Gao, and L. D. S. Coelho, "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour," International Journal of Bio-Inspired Computation, vol. 8, no. 6, pp. 394–409, 2016.