OPTIMISED ASSOCIATION RULE MINING FOR HEALTH DATA

Purnima Das[1*], John F. Roddick[1], Patricia A. H. Williams[1] and Mehwish Nasim[1,2]

[1]College of Science and Engineering, Flinders University, Tonsley, SA 5042
[2]The University of Western Australia, Perth, WA 6009

ABSTRACT

Association Rule Mining (ARM) has been recognised as a valuable and easy-to-interpret data mining technique in response to the exponential growth of big data. However, research on ARM techniques has mainly focused on enhancing computational efficiency while neglecting the automatic determination of threshold values for measuring the "interestingness" of items. Selecting appropriate threshold values (such as support, confidence, etc.) significantly affects the quality of the association rule mining outcomes. This study proposes an algorithm that utilises Particle Swarm Optimization (PSO) and ARM techniques to determine optimised threshold values in the health domain automatically. The algorithm was evaluated using the UCI machine learning medical database for heart disease. Results show that the proposed algorithm is capable of generating frequent itemsets and rules in an efficient manner and can detect the optimum threshold values. This research has practical implications for the health domains, as it can extract valuable results.

KEYWORDS

Association Rule Mining; Particle Swarm Optimisation; Health Data; Heart Disease; Optimised Frequent Itemsets.

1. INTRODUCTION

With the advent of information technology, various organisations and researchers have been able to release data publicly across several domains, including medical data, health data, scientific data, financial data, network data, and transaction-based marketing data [1]. The ability to effectively analyse and utilise this data to uncover critical hidden knowledge and relationships depends on the quality and accuracy of the data, as well as the ease with which researchers can deploy analytical tools.

Data mining (DM) tools have been widely used by researchers in various domains over the past few decades to find patterns in data [2, 3]. DM has demonstrated excellent performance in multiple areas, including medical, business, and scientific analysis. The growth of DM technology has also significantly improved its accuracy and computational efficiency.

Among the different data mining approaches, Association Rules Mining (ARM) is a widely used and easily understandable technique [4]. The Apriori algorithm, as the earliest and most paradigmatic algorithm among the variations of ARM techniques, has been refined multiple

David C. Wyld et al. (Eds): CMLA, GRAPH-HOC, CIoT, DMSE, ArIT, WiMoNe, CSEIT – 2023
pp. 83-93, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131007

times to enhance its efficiency and accuracy. However, users need to provide threshold values for identifying frequent itemsets and relationship rules, which limits the objectiveness and efficiency of the algorithm's results. Additionally, there has been little work done to apply automatic methods in the medical domain, which is crucial for achieving higher accuracy, objectivity, and acceptance.

Particle Swarm Optimisation (PSO) is an effective and fast computational technique that does not require substantial domain knowledge to achieve optimisation for a particular problem [6]. The health domain generates complex data, making it challenging to understand the relationships that may exist within it, despite extensive analysis. To ensure high accuracy even without domain knowledge, this study proposes a PSO-based ARM technique to achieve optimised thresholding automatically. The paper is organised as follows: Section 2 explains the concepts and background of the ARM and PSO algorithms, Section 3 outlines the methodology and implementation of the proposed model, Section 4 discusses the results and findings of the proposed model, and Section 5 highlights the significance of the work and provides suggestions for future research.

2. BACKGROUND

This section provides an overview of previous research conducted in the field, covering a range of ARM and PSO algorithms.

2.1. Association Mining

ARM is a well-known and extensively utilised technique in the field of data mining and knowledge discovery technology due to its ease of implementation and interpretation [2]. One of the most popular ARM techniques is Apriori [5], which was the earliest version of the algorithm. Although it has undergone significant improvements, Apriori still effectively represents the concepts underlying ARM. Apriori was used to analyse transaction data from market baskets to identify frequent patterns of items, assisting business experts in comprehending customer purchasing behaviour. It is expected that readers are generally familiar with ARM algorithms [4].

2.1.1. Definition of Association Rule

In the context of market basket applications, an association rule is denoted by $A \rightarrow B$, where A and B represent Itemsets (I), a set of products. An itemset is a group of items, $\{i_1, i_2, ..., i_n\}$, selected from the complete set of possible items. The source dataset $D = \{t_1, t_2, ..., t_n\}$ consists of transactions that contain items from these itemsets and is utilized to generate frequent itemset rules.

2.1.2. Extensions of ARM

The Apriori algorithm is computationally intensive, and therefore, most ARM proposals aim to enhance its computational efficiency. Numerous enhancements have been suggested to improve the algorithm's efficiency; however, only a small number of the algorithms mentioned below eliminate the requirement for users to provide threshold values. Zhang *et al.* [7] present a generic framework based on Spark for evolutionary-based ARM (P-EAARM) to reduce computational cost. Kuo *et al.* [2] employed PSO in numerical ARM for multi-objective purposes in terms of confidence, comprehensibility and interestingness, which can be obtained by the equations (1), (2) and (3).

$$Confidence(X \rightarrow Y) = \frac{\delta(X \cap Y)}{\delta(X)} \quad (1)$$

Where δ(X ∩Y) is the total number of transactions that contain both X and Y, and δ(X) is the total number of transactions that contain X only.

$$Comprehensibility = \frac{\log(1+|C|)}{\log(1+|A \cap C|)}. \quad (2)$$

$$Interestingness = \left[\frac{SUP(A \cup C)}{SUP(A)}\right] \times \left[\frac{SUP(A \cup C)}{SUP(C)}\right] \times \left(1 - \frac{SUP(A \cup C)}{SUP(D)}\right). \quad (3)$$

Unlike other numerical algorithms, this method does not require discretization of numerical data. Instead, this study employs an adaptive grid concept, where the objective space is divided into several hypercubes.

Martínez-Ballesteros et al. [8] incorporated an evolutionary-based algorithm into the mining of quantitative multi-objective association rules. The authors proposed an improved version of the genetic algorithm for identifying high-utility itemsets in a quantitative ARM domain space. Martín et al. [9] devised a parallel-based framework to accelerate the evolutionary-based Quantitative ARM process. However, neither of these methods can handle complex structured data.

Sahota and Verma [10] substituted the user-defined minimum support of the classical ARM technique with average support to enhance the traditional Apriori method. Although this method can decrease the number of generated rules, it may discard useful information. Djenouri et al. [11] suggested the integration of cluster-based and GPU parallel computing approaches with natureinspired metaheuristic-based optimization algorithms to reduce computation time in solving ARM problems for high-dimensional data. In this approach, each solution in the evolutionary process is effectively presented. The fitness function, which is a combination of support and confidence, was used to evaluate the interestingness of the rules.

2.2. Particle Swarm Optimisation Algorithm

Particle Swarm Optimization (PSO) was first proposed by Eberhart and Kennedy [6] as a nature inspired meta heuristic-based algorithm for optimising non-linear functions. PSO has since become a widely used and efficient evolutionary computational technique in various domain spaces [12].

PSO employs the concept of birds flocking, where each "bird" in the PSO search space represents a single solution or particle. To optimise a problem, a fitness function is assigned to evaluate the fitness of each particle in the search space. Each particle's flight is directed by its velocity, and all particles in the problem space follow the current optimum particles. Initially, a group of random particles or solutions with random positions and velocities are assigned to generate the initial population. Then, with each iteration or update generation, the particles search for the optimal solution.

During the iteration process, the position and velocity of each particle are updated with the help of the two "best" values: "Pbest" or local best and "Gbest" or global best. "Pbest" represents the locally found best solution or fitness against a specific particle so far, while "Gbest" represents the globally found best solution or fitness achieved by any particle in the population so far. After obtaining the "Pbest" and "Gbest" values, the particles update their position and velocity using the following equations (4) and (5) in each iteration [13].

$$v_{id}^{n+1} = v_{id}^{n} + c_1 rand()(Pbest - x_{id}) + c_2 rand()(Gbest - x_{id}). \tag{4}$$

$$x_{idn+1} = x_{idn} + v_{idn+1}. \tag{5}$$

The definition of the variables is as follows: v_{id}, x_{id} is the particle velocity and particle position of the id^{th}particle; dimension of the searching space is d; c_1 is the solitary coefficient factor, and $c_2$is the societal factor; rand() is a random number range between 0 and 1. Normally, the value of c_1 and c_2 are given as 2 [14]. Moreover, the fitness of each particle is evaluated using a fitness function. To prevent particle velocities from exceeding the maximum velocity value, vmax, in each dimension, a user-specified parameter is defined and the sum of accelerations in each dimension is checked. If the sum exceeds vmax, the velocity of that dimension is limited by vmax. This approach is known as the "vmax method" [6]. Shi and Eberhart [15] introduced another technique for PSO, known as the "inertia weight method," which incorporates weight with velocities. The updated velocities and positions are measured in this method by the equations (6), (5)

$$v_{id}^{n+1} = w v_{id}^{n} + c_1 rand()(Pbest - x_{id}) + c_2 rand()(Gbest - x_{id}). \tag{6}$$

Here this added weight, *w,* balances the global and local search, and the value of *w* can be a positive constant, a positive linear or nonlinear function of time. Another discrepancy of PSO proposed by Clerc [16] is known as the “constriction factor method”. This method updates particle velocities and positions using equations (7) and (8).

$$v_{id}^{n+1} = k[v_{id}^{n} + c_1 rand()(Pbest - x_{id}) + c_2 rand()(Gbest - x_{id})]. \tag{7}$$

$$k = \frac{2}{|2-\delta^2-\sqrt{\delta^2-4\delta}|}. \tag{8}$$

Where $\delta = c_1 + c_2$, $\delta > 4$

Enhancing social interaction among swarms is a crucial factor in improving the performance of the PSO algorithm. To achieve this, Zhao et al. [17] proposed a variation of PSO that extends the original algorithm by effectively influencing the global optimal solutions through enhanced social interaction among swarms.

3. METHODOLOGY

This section presents a new approach that combines Particle Swarm Optimisation and Association Rule Mining, building upon the background knowledge discussed earlier. The detailed process is depicted in Figure 1.

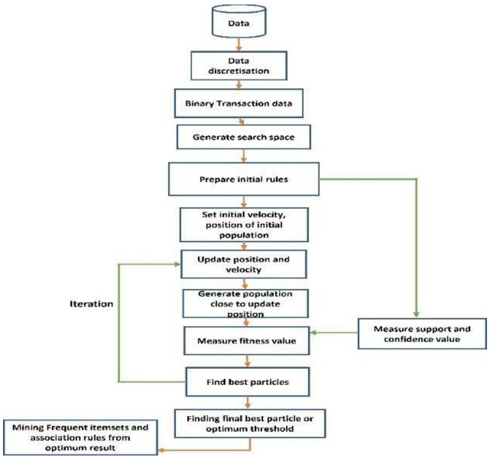

Figure 1: The proposed PSO-based ARM algorithm

3.1. The Proposed Algorithm

The proposed algorithm consists of two main modules: pre-processing and rule generation. The pre-processing module is responsible for data preparation and fitness evaluation of individuals, while the rule generation module finds the optimal threshold value for a given database.

Firstly, the data is discretised and converted into a binary transaction format to facilitate the frequent itemset scanning. Then, the particle swarm optimisation technique updates the population based on the quality of each particle, and the PSO iteration continues until either all individuals achieve the same result or a predefined iteration limit is reached.

3.2. Pre-Processing of Algorithm

The ARM requires categorical data, but real-world data is often not in a categorical form and therefore needs to be transformed. There are several techniques available to transform data into a discretised form, including unsupervised methods such as equal width, equal frequency, k-means, and supervised methods such as decision trees. In this study, the equal-width technique was used to maintain consistency with the item values.

To speed up the scanning process and facilitate the calculation of support and confidence values, Wur and Leu's binary transformation technique [18] was employed to transform the transaction data into a binary format. Figure 2 illustrates the transformation process for five distinct items labelled Item 1 through Item 5, where each transaction record T1, . . . T5 contains one or more items. In the binary transformed data, each column represents a distinct item, and each row represents a transaction.

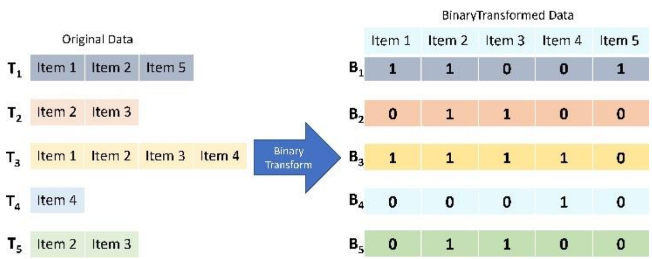

Figure 2: Transaction data transformation

3.3. PSO based ARM Algorithm

To ensure all important and semi-important rules are included in the results, this study considers them in a way that small threshold values are defined to generate a search space for implementing the evolutionary process. The fitness value of each particle is determined using Kung's fitness function [19]. The function is shown in equation (9):

$$Fitness(R) = confidence(R) \times \log(support(R) \times length(R) + 1. \quad (9)$$

The equation above defines the confidence(R), support(R), fitness value, and length of rule type R. Apriori definitions were utilised to measure the support and confidence values. The PSO evolutionary process begins with generating an initial population that will be updated by the local and global best performance of each particle. In this study, the initial population was randomly selected from predefined rules. The global best value is the particle with the maximum fitness value. The local best value of each particle is modified according to equations (4) and (5). The PSO evolution process measures the updated position to select the iterative population closest to the best value quickly. The termination criteria for this PSO evolution are when all particles achieve the same best results in terms of fitness value or position.

4. MODEL EVALUATION, RESULTS AND DISCUSSION

In this study, we utilised the heart disease datasets [20] from the UCI machine learning database to explore the performance of the proposed PSO-based ARM algorithm. In these datasets, the heart disease dataset contains 11025 records with 14 columns. This data is real and of categorical type, and the columns are age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal and target. The sample heart disease data is shown in Table 1.

Table 1: Sample Heart disease dataset

age: Age of the patient in years; cp:chest pain type; sex: Gender of the patient; chol: serum cholesterol in mg/dl; trestbps :resting blood pressure; restecg: resting electrocardiographic results; fbs: fasting blood sugar; exang: exercise induced angina; thalach: maximum heart rate achieved; slope: the slope of the peak exercise; oldpeak: depression induced by exercise relative to rest; target:outcome of the disease; ca: number of major vessels (0-3) colored by fluoroscopy;

4.1. Mining Results with PSO-based ARM

For our research, we extracted age, sex, cp, trestbps, chol, fbs, restecg, thalach, oldpeak, thal, and target column values from the heart disease dataset. We transformed the real-type column values of age, trestbps, chol, and thalach into discretized form to derive meaningful rules. The remaining categorical columns were renamed.

To obtain initial rules, we used the FPGrowth [21] algorithm, followed by applying the PSO technique on those rules. The fitness of each rule was measured by calculating its support and confidence values. In this experiment, equation (9) was used to evaluate the fitness of each particle.

Figure 3 illustrates the number of generated frequent itemsets across different medical datasets with varying support values. Increasing support values indicate stronger itemsets being generated.

The relationship between the number of generated rules and the variation of confidence values across different datasets is illustrated in Figure 4. Higher confidence values indicate stronger rules. To generate these confidence values, we used the proposed PSO-Based ARM method with various population sizes.

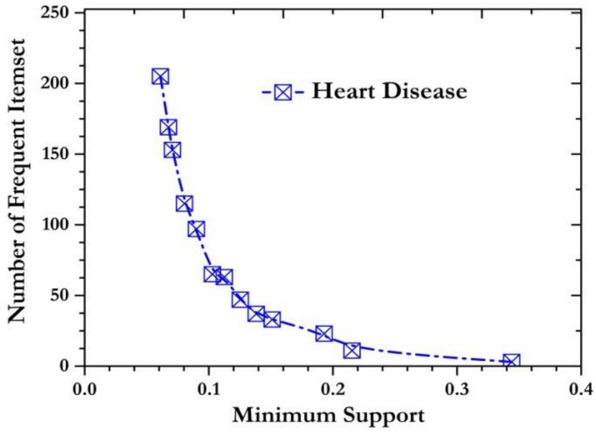

Figure 3: Relationship between minimum support and number of frequent Itemsets

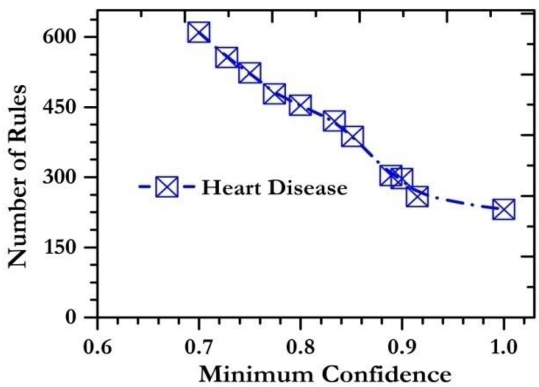

Figure 4: Relationship between minimum confidence and number of rules

To examine the relationship among data attributes, we categorised the heart disease dataset into four groups: patients diagnosed with positive heart disease, patients diagnosed with negative results, heart disease cases with female gender, and heart disease cases with male gender. Table 3 displays the strongest frequent itemsets, and Table 4 shows the strongest rules with the best support and confidence values measured by our proposed algorithm for heart disease data with a target value of disease type.

Overall, our proposed algorithms demonstrate excellent performance in terms of execution time and mining strong frequent itemsets and rules with high support and confidence values.
For patients diagnosed with negative heart disease, the most robust frequent itemsets and rules are presented in Table 5 and 6, respectively.

Table 3: Strongest frequent itemsets from “target=disease” type filtered data.

Support	Itemsets
0.26616	(target→disease)
0.226236	(fbs→0)
0.201521	(thal→fixed defect)
0.226236	(target→disease, fbs→0)
0.201521	(target→disease, thal→fixed defect)

Table 4: Strongest rules from “target→disease” type filtered data

Antecedents	Consequents	Antecedent support	Consequent support	Support	Confidence
(fbs→0)	(target→disease)	0.226236	0.26616	0.226236	1
(thal→fixed defect)	(target→disease)	0.201521	0.26616	0.201521	1

Table 5: Strongest frequent itemsets from “target→no-disease” type filtered data

Support	Itemsets
0.675351	(target→no-disease)
0.547094	(fbs→0)
0.54509	(chest-pain-type→0)
0.543086	(sex→male)
0.547094	(target→no-disease, fbs→0)
0.54509	(chest-pain-type→0, target→no-disease)
0.543086	(target→no-disease, sex→male)

Table 6: Strongest rules from “target→no-disease” type filtered data **t**

Antecedents	Consequents	Antecedent support	Consequen support	Support	Confidence
(fbs→0)	(target→no-disease)	0.547094	0.675351	0.547094	1
(chest-pain-type→0)	(target→no-disease)	0.54509	0.675351	0.54509	1
(sex→male)	(target→no-disease)	0.543086	0.675351	0.543086	1

4.2. Evaluation of Performance

In this section, a comparison is made between our proposed algorithm, PSO-based ARM, and PSO-GES [22], regarding variations in computational time with respect to population size and iterations. The dataset used for this analysis is the heart disease dataset consisting of 499 transaction records with 104 items. To prepare the summary matrix of PSO-GES, two transactions were used in each partition. The computational time was measured for one iteration with varying population size and for particle size 5 with varying iterations.

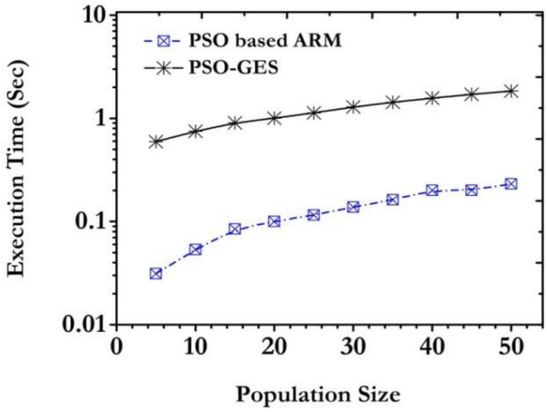

Figure 5: Relationship between population size and execution time.

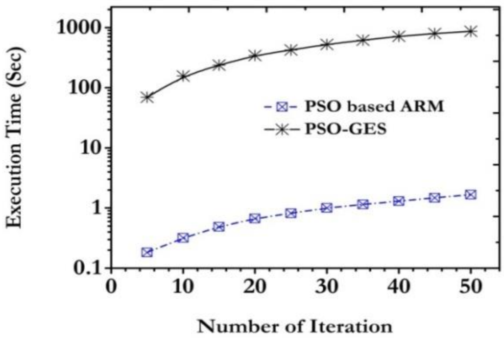

Figure 6: Relationship between number of iterations and execution time.

Figures 5 and 6 depict the comparison between the proposed algorithm and PSO-GES with respect to population size and the number of iterations, using a logarithmic y-axis. Both figures demonstrate that the proposed algorithm performs better than the PSO-GES algorithm. Furthermore, it is observed that when the number of items in transaction data is 300, PSO-GES exhibits a higher computational time.

5. CONCLUSIONS

The discovery of significant and useful knowledge is crucial when dealing with large datasets. Therefore, it is essential that the tools used to extract knowledge from these datasets produce reliable and understandable results. Association Rule Mining (ARM) is one such technique, but it is limited by the difficulty in selecting optimal thresholds for support and confidence. Automating the selection of these thresholds would make the ARM technique more practical and widely accepted.

In summary, this study proposed a PSO-based optimisation technique that integrates with the ARM process to efficiently select threshold values for support and confidence and extract the most important knowledge from the data. The proposed algorithm was tested on various medical datasets, and the results demonstrate its ability to identify data relationships and diagnose diseases. Moreover, the algorithm's performance in terms of execution time was evaluated and found to be satisfactory with varying population size and iterations.

ACKNOWLEDGEMENTS

Purnima Das gratefully acknowledges the financial support of the Flinders University through Flinders University Research Scholarship. The authors thank to the College of Science and Engineering, Finders University for providing the facility to carry-out the research.

REFERENCES

[1] UCI, “Machine learning repository.” [Online]. Available: https://archive.ics.uci. edu/ml/index.php
[2] R. Kuo, M. Gosumolo, and F. E. Zulvia, “Multi-objective particle swarm optimization algorithm using adaptive archive grid for numerical association rule mining,” *Neural Computing and Applications*, vol. 31, no. 8, pp. 3559–3572, 2019.
[3] S. Sharmila and S. Vijayarani, “Association rule mining using fuzzy logic and whale optimization algorithm,” *Soft Computing*, vol. 25, no. 2, pp. 1431–1446, 2021.
[4] A. Ceglar and J. F. Roddick, “Association mining,” *ACM Computing Surveys*, vol. 38, no. 2, 2006.
[5] R. Agrawal, T. Imieli´nski, and A. Swami, “Mining association rules between sets of items in large databases,” in Proceedings of the ACM SIGMOD international conference on Management of data, Conference Proceedings, 1993, pp. 207–216.
[6] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. IEEE, Conference Proceedings, pp. 39–43.
[7] T. Zhang, M. Shi, J. Wang, and G. Yang, “P-eaarm: A generic framework based on spark for eas-based association rule mining,” in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, Conference Proceedings, pp. 99–104.
[8] M. Mart´ınez-Ballesteros, A. Troncoso, F. Mart´ınez-Alvarez, and J. C. Riquelme, “Improving a multi-objective evolutionary algorithm to discover quantitative association rules,” *Knowledge and Information Systems*, vol. 49, no. 2, pp. 481–509, 2016.
[9] D. Mart´ın, M. Mart´ınez-Ballesteros, D. Garc´ıa-Gil, J. Alcal´a-Fdez, F. Herrera, and J. Riquelme-Santos, “Mrqar: A generic mapreduce framework to discover quantitative association rules in big data problems,” *Knowledge-Based Systems*, vol. 153, pp. 176–192, 2018.
[10] S. Sahota and P. Verma, “Improved association rule mining based on abc,” *International Journal of Computer Applications*, vol. 135, no. 10, pp. 6–10, 2016.
[11] Y. Djenouri, D. Djenouri, Z. Habbas, and A. Belhadi, “How to exploit high performance computing in population-based metaheuristics for solving association rule mining problem,” *Distributed and Parallel Databases*, vol. 36, no. 2, pp. 369–397, 2018.
[12] R. J. Kuo, C. M. Chao, and Y. Chiu, “Application of particle swarm optimization to association rule mining,” *Applied Soft Computing*, vol. 11, no. 1, pp. 326–336, 2011.
[13] “Pso tutorial.” [Online]. Available: http://www.swarmintelligence.org/tutorials.php
[14] M.-P. Song and G.-C. Gu, “Research on particle swarm optimization: a review,” in Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826), vol. 4. IEEE, Conference Proceedings, pp. 2236–2241.
[15] Y. Shi and R. Eberhart, “A modified particle swarm optimizer,” in 1998 IEEE international conference on evolutionary computation proceedings. IEEE world congress on computational intelligence (Cat. No. 98TH8360). IEEE, Conference Proceedings, pp. 69–73.
[16] M. Clerc, “The swarm and the queen: towards a deterministic and adaptive particle swarm optimization,” in Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol. 3. IEEE, Conference Proceedings, pp. 1951–1957.
[17] B. Zhao, C. Guo, B. Bai, and Y. Cao, “An improved particle swarm optimization algorithm for unit commitment,” *International Journal of Electrical Power and Energy Systems*, vol. 28, no. 7, pp. 482–490, 2006.
[18] S.-Y. Wur and Y. Leu, “An effective boolean algorithm for mining association rules in large databases,” in Proceedings. 6th International Conference on Advanced Systems for Advanced Applications. IEEE, Conference Proceedings, pp. 179–186.
[19] S.H.Kung, “Applying genetic algorithm and weight item to association rule,” Thesis, 2002.
[20] J. Andras, P. William, Steinbrunn, P. Matthias, and D. Robert, “Heart disease databases,” 1988. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[21] 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, 2004.

[22] G. B. Bar´o, J. F. Mart´ınez-Trinidad, R. M. V. Rosas, J. A. C. Ochoa, A. Y. R. Gonz´alez, and M. S. L. Cort´es, "A pso-based algorithm for mining association rules using a guided exploration strategy," *Pattern Recognition Letters*, vol. 138, pp. 8–15, 2020.

AUTHORS

Purnima Das is a Research Scholar, Casual Academic, and Final Year PhD Candidate at the College of Science and Engineering, Flinders University, Australia. Before joining Flinders University, she worked as a Data Scientist in the School of Information Technology and Mathematical Sciences at the University of South Australia, Adelaide, Australia. Purnima has a Master of Science and a Bachelor of Science in Computer Engineering. Her research interests include Machine Learning and Data Mining for developing state-of-the-art algorithms to solve challenging problems in various industrial sectors, including medical and health. Additionally, she is a Computer Science and Information Technology Lecturer at The University of Adelaide College, Melbourne campus, Australia.

John Roddick is a Professor at the College of Science and Engineering, Flinders University, Australia. He joined Flinders in April 2000 after 15 years at the Universities of Tasmania and South Australia. His time as an academic followed 10 years' experience in the computing industry. He was Dean of the School of Computer Science, Engineering and Mathematics between January 2008 until July 2017, Chair of Academic Senate between 2017 and 2021 and more recently, until the end of 2022, the Dean of Education in the College of Science and Engineering. Professor Roddick is a Fellow of the Institute of Engineers, Australia (FIEAust) and a Fellow of the Australian Computing Society (FACS).

Trish Williamsis a leader in research and innovation in digital health. Trish is Cisco Chair and Professor of Digital Health Systems at Flinders University, Director of Flinders Digital Health Research Centre, and Director of Cisco-Flinders Digital Health Design Lab and Digital Health IOT Laboratory. Internationally recognised in her field, Trish applies 40 years of experience in healthcare computing to research and practical outcomes in cybersecurity, health IoT, mobile health, medical devices, governance, patient safety, and health software safety. In 2021 she was awarded the prestigious Australian Academy of Technology and Engineering, fellowship. Also, Prof Williams was named one of the 2021 Brilliant Women in Digital Health, recognising her contributions to mentoring, technology development, improving health outcomes using digital health, and introducing new digital processes for healthcare delivery.

Mehwish Nasim is a Lecturer in Computer Science at The University of Western Australia. She also hold Adjunct Lecturer positions at School of Mathematical Sciences at University of Adelaide as well as at Flinders University. She is also an Associate Investigator with ARC Centre of Excellence for Mathematical and Statistical Frontiers since 2017 and a member of the Equity and Diversity Committee. Her research interests include: social network analysis, human-centric cyber security, machine learning, medical image processing, health analytics, and human computer interaction. Her current work focuses on predicting population level events in Australia, detecting fake news and misinformation and modelling polarisation on social media.