BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 2389-2400
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp2389-2400

LOGIC MINING CLASSIFICATION FOR PHONE PRICES DATASET USING DISCRETE HOPFIELD NEURAL NETWORK AND WEIGHTED RANDOM 2 SATISFIABILITY

Nurul Najwa Ahmad Azam [1], Nur Ezlin Zamri [2*], Mohd Shareduwan Mohd Kasihmuddin [3], Nurul Atiqah Romli [4], Mohd. Asyraf Mansor [5]

[1,3]*School of Mathematical Sciences, Universiti Sains Malaysia*
[5]*School of Distance Education, Universiti Sains Malaysia*
11800 USM, Penang, Malaysia

[2]*Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia*
43400 UPM, Serdang, Selangor, Malaysia

[4]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA*
Cawangan, Perlis Branch, Arau, Perlis, 02600, Malaysia

*Corresponding author's e-mail: * **ezlinzamri@upm.edu.my***

Article Info

Article History:

Received: 26th August 2025
Revised: 19th January 2026
Accepted: 12th March 2026
Available online: 8th April 2026

Keywords:

Discrete Hopfield Neural Network;
Logic Mining;
Phone prices classification;
Weighted satisfiability.

ABSTRACT

Smartphones have become essential in today's technology-driven world, with various models offering unique features like camera quality, screen resolution, and storage. Understanding how these features influence smartphone prices can help consumers make informed purchasing decisions. This study introduces a logic mining technique to classify smartphone features that contribute to pricing using Weighted Random k Satisfiability with Modified Reverse Analysis. The model implements a Discrete Hopfield Neural Network, a Modified Niched Genetic Algorithm for training, and the Jaccard Feature Selection Method. The Phone Prices Dataset from Kaggle was used for experimentation, revealing the model's ability to extract optimal patterns in the form of induced logic. The results show that the proposed model outperforms existing methods, achieving an accuracy of 0.8083, precision of 0.8925, specificity of 0.9760, Matthew's correlation coefficient of 0.5334, and an F1-score of 0.5887, demonstrating its effectiveness in analyzing and classifying smartphone pricing factors.

How to cite this article:

N. N. A. Azam, N. E. Zamri, M. S. M. Kasihmuddin, N. A. Romli, and M. A. Mansor, "LOGIC MINING CLASSIFICATION FOR PHONE PRICES DATASET USING DISCRETE HOPFIELD NEURAL NETWORK AND WEIGHTED RANDOM 2 SATISFIABILITY", *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 2389-2400, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

According to Statista, the number of people using smartphones is expected to rise steadily between 2024 and 2029. It is projected that by 2029, there will be 6.2 billion smartphone users worldwide. Approximately 78% of people will use a smartphone on a regular basis. Notably, during the past few years, the number of smartphone users has steadily increased. Nowadays, smartphones have become one of the essential items to survive in this era, where a lot of things can be made digital, such as internet banking. Smartphones have made a huge contribution to the economic impact. The smartphone industry contributes billions of dollars to the global economy through device sales, app purchases, and mobile advertising revenue [1]. The first factor, the Phone Prices dataset, has been chosen for this research because a lot of people will benefit from the prediction of phone prices, such as consumers, retailers, and manufacturers. Secondly, the phone prices dataset contains a variety of attributes. The variety allows for comprehensive analysis and pattern recognition. Lastly, consumers can be better informed when to buy or upgrade their phone based on price trends and predictions. Logic mining is a discrete classification tool that allows computational models to make sense of complex data, identify patterns in datasets, and help in the decision-making process. One of its components is Artificial neural networks (ANNs), where ANNs are computational models that were inspired by the structure of the human brain. The human brain is composed of interconnected nodes (neurons) that process information to recognize patterns and learn from any data. Notably, ANNs can be used as tools within logic mining to identify patterns and relationships in data. A variant of ANN is the Discrete Hopfield Neural Network (DHNN), which was initiated by [2]. Unfortunately, DHNNs are often criticized for their lack of transparency, which leads to difficulty in understanding how inputs translate into outputs.

Many researchers have been using Propositional Satisfiability (SAT) logic to be used in DHNN as a logical rule (also known as the neuro-symbolic approach). Foundational works of neuro-symbolic models were done by [3] and [4]. These works used Horn Satisfiability (HornSAT) as a logical rule in DHNN. Subsequently, 2 Satisfiability (2SAT) and 3 Satisfiability (3SAT) have been introduced by [5] and [6], respectively, as DHNN neuron representation, where both logical rules are classified as systematic SAT rules. The primary limitation of systematic SAT is that the synaptic weight can be easily estimated. To overcome this limitation, [7] introduced the initial non-systematic SAT of Random 3 Satisfiability (RAN3SAT). The property of randomness and the inclusion of higher-order logic proffer better generalizability of SAT as neurons in DHNN. Unfortunately, the structure of RAN3SAT is unable to be generated when there is a need for a specific number of positive (non-negated) or negative (negation) literals. The issue is that RAN3SAT consistently produced N number of clauses that have the exact same set of specified positive and negative literals. Hence, it is highly improbable to analyze the effect of negations and non-negations as symbolic rules in DHNN. Based on the structure of the current RAN3SAT and the findings by [8], most neuro-symbolic practitioners disregard negative literals. This occurs because negative literals tend to be associated with errors. Thus, the investigation of negations as logical rules was poorly discussed.

As mentioned, a logic mining model is a computational approach used to extract logical patterns or rules from data. Initially, the work by [9] was the first to introduce a logic mining model that utilized DHNN and HornSAT alongside a reverse analysis technique, known as Reverse Analysis (RA). The suggested RA prioritized final neuron states according to the overall energy profile of these states before converting the states into induced logic. Some problems have been encountered in research where RA cannot establish continuous or non-categorical data entries. Secondly, the suggested RA generated an infinite number of induced logics without considering the possibility of overlapping formulations among them, and the effectiveness of the retrieved logic cannot be measured. The presence of redundant literals in the structure of HornSAT leads to the interpretability of the induced logic obtained from the DHNN. Over the years, attempts have been made to address these issues. Firstly, [6] suggested the change of logic from HornSAT to 3SAT, which resulted in a logic mining, namely 3 Satisfiability Reverse Analysis (3SATRA). The suggested 3SATRA omitted HornSAT due to redundant neuron connections being trained by DHNN. However, the rigid structure of 3SAT limits a better understanding of the patterns extracted from the dataset. As a result, 3SATRA attained a low accuracy of less than 80%. Then, [10] mentioned that the issues were not from the embedded logical structure but the lack of an optimal attribute selection method. Previous works ignored the need for significant attribute selection and only utilized random selection. Consequently, the full potential of any logic mining model was not fully optimized. In addition, the DHNN trained insignificant attributes, resulting in outputs that lack interpretability. Therefore, [10] proposed a Supervised 2 Satisfiability Reverse Analysis (S2SATRA) with the implementation of the Pearson correlation method as the attribute selection method. As compared to baseline models, S2SATRA performed the best in terms of accuracy for 12 datasets

across different fields. Nonetheless, only a selective number of datasets could be analyzed by S2SATRA because the certainty of all attributes to have a strong association relationship with the class of the datasets is considered almost impossible. Excessive reliance on correlation analysis when selecting significant attributes restricts the exploration of logic mining, limiting its ability to train and recognize patterns across diverse datasets. As a result, effective logic mining is confined to specific characteristics of the dataset. This will degrade the quality of logic mining as a highly interpretable and explainable Artificial Intelligence (AI) approach.

To overcome these problems, the proposed logic mining method should accommodate the following features: logical rules that possess non-redundancy and non-systematic structures to train important and diverse attributes. Subsequently, the need for an unsupervised attribute selection method to identify significant attributes. A new approach to logic mining has been introduced by implementing the Weighted Random 2 Satisfiability Reverse Analysis Method via the Jaccard Feature Selection Method. Therefore, the objective of this paper is as follows: (1) To apply a logical rule, namely Weighted Random 2 Satisfiability in a Discrete Hopfield Neural Network. (2) To construct a logic mining model called Weighted Random 2 Satisfiability Modified Reverse Analysis in extracting the patterns of the Phone Prices dataset. A training algorithm, namely, the Modified Niched Genetic Algorithm, is added to the training phase to optimize the training process. (3) To measure the overall performance of the proposed logic mining by evaluating five different performance metrics, such as accuracy, precision, specificity, Matthew Correlation Coefficient, and F1-Score in comparison to baseline models. (4) To investigate attributes that contribute to the rise and fall of phone prices based on the retrieved induced logic.

The rest of the paper is structured as follows: Section 2 presents the background and methodology of the proposed method. In Section 3, the result will be presented, and the induced logic retrieved by the logic mining will be further analyzed with suitable recommendations. In Section 4, the conclusion of the paper will be presented.

2. RESEARCH METHODS

The background study of each component in the proposed logic mining method will be discussed in this section. There are two subsections: Weighted Random 2 Satisfiability in DHNN and the proposed logic mining, namely, Weighted Random 2 Satisfiability Modified Reverse Analysis, respectively. The first subsection focuses on the foundational background of the components used in the proposed method. It includes the formulation and the embedding process of logical structures within the DHNN. The second subsection details the workflow of the proposed logic mining, which highlights the specific components involved when analyzing the Phone Prices dataset.

2.1 Weighted Random 2 Satisfiability in DHNN (DHNN-*r*2SAT)

The structure of Weighted Random 2 Satisfiability (*r*2SAT) was formulated with two main components. The first component involved the control of b number of literals in each clause, where b can be either 1 or 2 (denoted by $b = \{1,2\}$). The second component is the number of negative literals generated based on the pre-specified r. The general equation in formulating *r*2SAT is presented in Eqs. (1) and (2) as follows:

$$L_r = \wedge_{i=1}^{u} J_i^{(b=2)} \wedge_{i=1}^{v} J_i^{(b=1)}, \tag{1}$$

$$J_i^{(b=1,2)} = \begin{cases} (A_i \vee B_i), & b = 2, \\ C_i \quad , & b = 1, \end{cases} \tag{2}$$

where the total number of first-order $\left(J_i^{(b=1)}\right)$ and second-order clauses $\left(J_i^{(b=2)}\right)$ are indicated by u and v, respectively. The summation of clauses (m) and literals (NN) in the proposed logic can be formulated as in Eqs. (3) and (4) below:

$$m = u + v, \tag{3}$$

$$NN = 2u + v. \tag{4}$$

The allocation of whether the literals in L_r is positive (A_i, B_i, C_i) or negative $(\neg A_i, \neg B_i, \neg C_i)$ is set at random. However, the number of negative literals must be aligned with the pre-defined r. In simpler terms, r is the ratio of negative literals that exist in L_r. The minimization of this process will be done using the Genetic Algorithm (GA). Examples of different L_r for $r = \{0.1,0.5,0.9\}$, for $u = 4$ and $v = 2$ $(NN = 10)$ are presented in Eqs. (5)–(7) as follows:

$$L_{r=0.1} = (A_1 \vee B_1) \wedge (A_2 \vee B_2) \wedge (A_3 \vee \neg B_3) \wedge (A_4 \vee B_4) \wedge C_1 \wedge C_2, \tag{5}$$

$$L_{r=0.5} = (A_1 \vee B_1) \wedge (\neg A_2 \vee \neg B_2) \wedge (A_3 \vee \neg B_3) \wedge (A_4 \vee \neg B_4) \wedge \neg C_1 \wedge C_2, \tag{6}$$

$$L_{r=0.9} = (\neg A_1 \vee \neg B_1) \wedge (\neg A_2 \vee \neg B_2) \wedge (A_3 \vee \neg B_3) \wedge (\neg A_4 \vee \neg B_4) \wedge \neg C_1 \wedge \neg C_2. \tag{7}$$

Once the correct L_r is generated, the implementation of the logic as neurons within DHNN (also referred to as DHNN-*r*2SAT) is being initiated. DHNN is a type of recurrent neural network (RNN) with no hidden layer [2], consisting of a training phase and testing phase [11], and the connections between neurons are represented by the synaptic weights in DHNN [2]. Additionally, the energy profile of the network is based on the Lyapunov energy function, which can be easily identified compared to other networks. During the training phase, the logical rule of L_r is being trained using the WA method to guarantee the generation of optimal synaptic weights. These weights will be stored in the Content Addressable Memory (CAM) of DHNN and retrieved during the testing phase. Eqs. (8) and (9) present the formulation of the cost function in the DHNN-*r*2SAT model.

$$E_{L_r} = \frac{1}{4}\sum_{i=1}^{u}\left(\prod_{j=1}^{2} Q_{ij}\right) + \frac{1}{2}\sum_{i=1}^{v}\left(\prod_{j=1}^{1} Q_{ij}\right), \tag{8}$$

$$Q_{ij} = \begin{cases} (1 - S_{A_i}), & if\ \neg A_i, \\ (1 + S_{A_i}), & if\ A_i. \end{cases} \tag{9}$$

Based on the WA method, the optimal training phase of DHNN can be achieved by successfully minimizing E_{L_r}. When $E_{L_r} = 0$, this indicates that all in the L_r are satisfied (able to locate a satisfied interpretation of the logical rule). Let DHNN $D = \left(W_{ij}^{(b)}, S_j\right)$ composed of P bipolar valued neurons, where $P = \{1,2,3, \dots, P\}$. For each P neuron state, the current neuron state S_i will select 1 if the value of D is equal to or greater than the threshold value (θ). Otherwise, S_i will retain the value of -1. Notably, $W_{ij}^{(b)}$ represents the synaptic weight of the DHNN that corresponds to the defined *b-clausal* order involving neurons i and j.

The retrieval phase of DHNN-*r*2SAT is initiated once all $W_{ij}^{(b)}$ is successfully being stored in the CAM. The local field (h_i) computation to retrieve the final neuron states is presented in Eq. (10). An activation function via the Hyperbolic Tangent Activation Function (HTAF) formulated in Eq. (11) is applied to convert the values of h_i into bipolar values. This step is crucial to ensure that DHNN prevents linearity of the trained data.

$$h_i(t) = \sum_{i=1,j=1}^{NN} W_{ij}^{(2)} S_j + W_i^{(1)}, \tag{10}$$

$$S_i(t) = \begin{cases} 1, & tanh(h_i), \\ -1, & otherwise. \end{cases} \tag{11}$$

The formulation for the Lyapunov energy function $\left(H_{L_r}\right)$ is portrayed in Eq. (12). The absolute minimum energy profile of the DHNN can be evaluated using Eq. (13), which is depicted as $H_{L_r}^{min}$.

$$H_{L_r} = -\frac{1}{2}\sum_{i=1,i\neq j}^{NN}\sum_{j=1,i\neq j}^{NN} W_{ij}^{(2)} S_i S_j - \sum_{i=1}^{NN} W_i^{(1)} S_j, \tag{12}$$

$$H_{L_r}^{min} = -\frac{u + 2v}{4}. \tag{13}$$

Notably, the DHNN-*r*2SAT model must be able to generate final neuron states $\left(S_i^f\right)$ that converge to $H_{L_r} \rightarrow H_{L_r}^{min}$. If the energy difference is less than or equal to the pre-defined tolerance value (tol), the

retrieved S_i^f are classified as a global minimum energy solution. Otherwise, the neurons are depicted as a local minima energy solution. The energy profile of the network can be observed using Eq. (14) below:

$$\left|H_{L_r} - H_{L_r}^{min}\right| \leq tol. \tag{14}$$

In general, the logic of L_r serves as an input and is trained by the DHNN. Once optimal synaptic weights are stored in the CAM, the local field is computed, and the final energy profile of the network is evaluated. The training phase of DHNN-*r*2SAT is optimized using Modified Niched Genetic Algorithm (MNGA).

2.2 Weighted Random 2 Satisfiability Modified Reverse Analysis (*r*2SATMRA)

The proposed logic mining model, namely, *r*2SATMRA, is equipped with three main components: DHNN-*r*2SAT, an attribute selection method, and a modified reverse analysis technique. A detailed explanation of each phase is presented below:

1. Preliminary Phase: This phase focuses on the data pre-processing phase, which includes data preparation, executing the attribute selection approach, data splitting, and 5-fold cross-validation.
   a. All entries of the raw data (e_i) are transformed into bipolar form $\{1, -1\}$ using k-means clustering. The class of the dataset can be defined as 1 (higher than the mean of phone prices) and -1 (lower than or equal to the mean of phone prices).
   b. An attribute selection method using the Jaccard Feature Selection Method (JFSM) is applied to eliminate irrelevant attributes (E_i) based on the evaluation of the Jaccard index value $(Jacc)$ presented in Eqs. (15)–(19). More simply, a high Jaccard value indicates high similarity between the analyzed E_i and the class of the dataset (P). In contrast, a low Jaccard value reflects low similarity between the two items. The aim of JFSM is to identify *N* number of E_i that possessed high Jaccard values.

$$a_j(e_i) = \begin{cases} 1, & if\ (E_i, P) = (1,1), \\ 0, & otherwise, \end{cases} \tag{15}$$

$$b_j(e_i) = \begin{cases} 1, & if\ (E_i, P) = (1,-1), \\ 0, & otherwise, \end{cases} \tag{16}$$

$$c_j(e_i) = \begin{cases} 1, & if\ (E_i, P) = (-1,1), \\ 0, & otherwise, \end{cases} \tag{17}$$

$$d(e_i) = \begin{cases} 1, & if\ (E_i, P) = (-1,-1), \\ 0, & otherwise, \end{cases} \tag{18}$$

$$Jacc = \frac{\sum_{j=1}^{n} a_j}{\sum_{j=1}^{n} a_j + \sum_{j=1}^{n} b_j + \sum_{j=1}^{n} c_j}. \tag{19}$$

   The train to test dataset split is set at a ratio of 60 to 40 ($L_{train}: L_{test} = 60:40$). This ratio has a good agreement with the majority of the existing logic mining models. In addition, 5-fold cross-validation is considered. The average of all folds will be taken as the result for all metrics measured.
2. Logic Phase: This phase focuses on generating L_r for different values of r where $L_r = \{L_{r=0.1}, L_{r=0.2}, L_{r=0.3}, \dots, L_{r=0.9}\}$. The binary GA algorithm is applied to optimize the distribution of negative literals in the logical rule. The algorithm is responsible for producing a unique positioning of the negative literals in L_r to refrain from allowing DHNN to train duplicative logic.
3. Training Phase: Subsequently, this phase evaluates the classification between P (actual) and the outcome of L_r when L_{train} (predicted) is applied to the generated logic. This classification approach is measured using true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP). Any L_r with the highest accumulation of TN and TP will be deemed as the best logic $\left(L_r^{super}\right)$. Each L_r^{super} is then trained using MNGA, and the synaptic weights will be stored in CAM of the DHNN.

4. Testing Phase: The final phase of *r*2SATMRA is responsible for evaluating the local field formulated in Eq. (10). This step is crucial to retrieve the S_i^f that reflect the remembered patterns by DHNN. Once all S_i^f are retrieved, each neuron will be transformed into Boolean logic, where $S_i^f = 1 \rightarrow \{A_i, B_i, C_i\}$ or $S_i^f = -1 \rightarrow \{\neg A_i, \neg B_i, \neg C_i\}$. All logical rules undergo classification between P (actual) and the outcome of the induced logic or $L_r(S_i^f)$ when L_{test} (predicted). Note that, $L_r(S_i^f)$ that acquired the highest accuracy will be selected as the final output of *r*2SATMRA.

The complete implementation of *r*2SATMRA from the preliminary phase until the retrieval phase is illustrated in the flowchart portrayed in Fig. 1. The flowchart illustrates the step-by-step process of the proposed logic mining model, starting from raw data input, feature selection, and training using MNGA, before generating the best logic rules. These rules are then validated through neuron state transformation and testing, which resulted in an optimized logic output. Each colour in the flowchart corresponds to the respective phases.

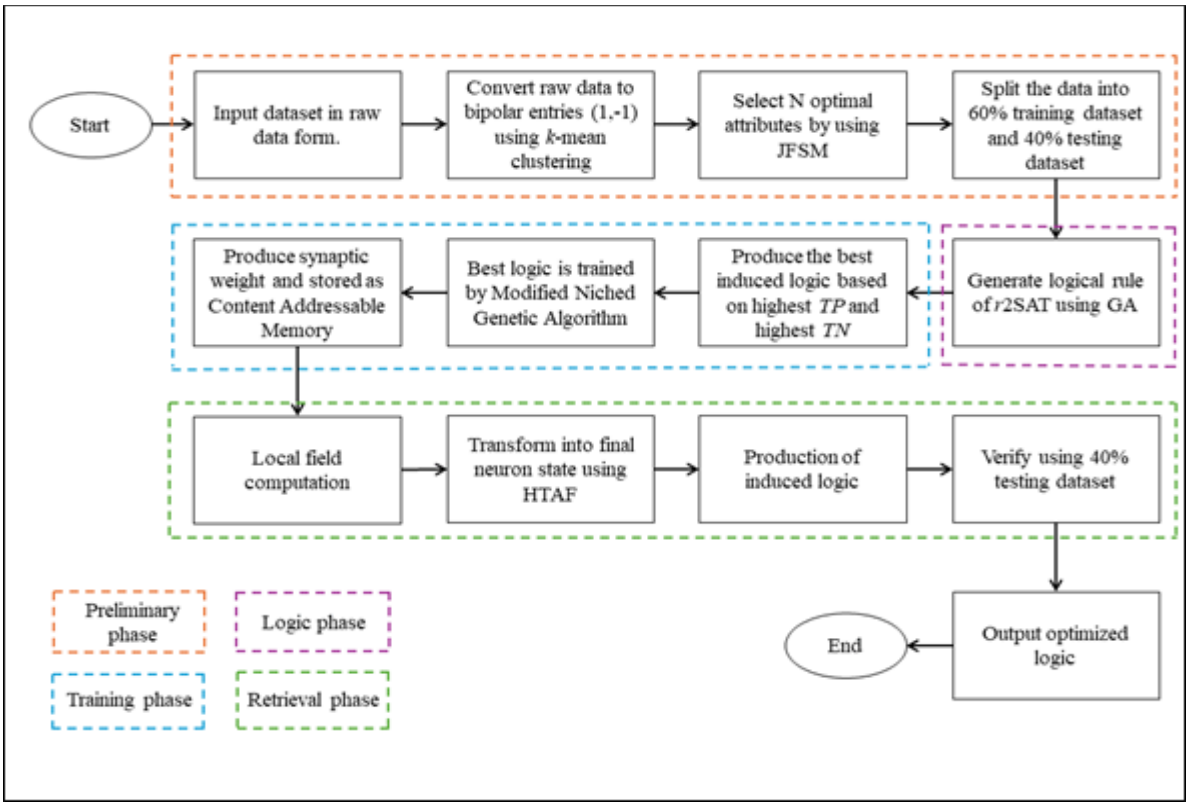

Figure 1. Detailed Flowchart of The Proposed Logic Mining Model, *r*2SATMRA

2.3 Experimental Settings and Dataset Information

All simulations in this experiment were executed using open-source software Dev C++ (Version 5.11). To avoid any bias, only one hardware device was utilized. Table 1 summarizes all parameters involved in the simulation of *r*2SATMRA when analyzing the Phone Prices dataset.

Table 1. List of Parameters in *r*2SATMRA

Parameter/Notation	Value Remark
Preliminary Phase	
Source of the dataset	https://www.kaggle.com/datasets/berkayeserr/phone-prices
Data of access	23rd April 2025
Total number of entries (n)	1512
Simulation threshold time	24 hours [10]
Number of attributes selected using JFSM	10
$\boldsymbol{L_{train}}$: $\boldsymbol{L_{test}}$	**60**: **40** [10]
Logic Phase	
Number of maximum *NN* generated	120
Optimization algorithm	Binary GA

Parameter/Notation	Value Remark
Step size of *NN* ($\Delta \boldsymbol{NN}$)	6
Range of *r*	(0.1, 0.9)
Step size of *r* ($\Delta \boldsymbol{r}$)	0.1
Training Phase	
Number of maximum training iterations	100 [12]
Training algorithm	MNGA
Method to calculate synaptic weight	WA method [3]
Testing Phase	
Number of maximum testing iterations	100 [12]
Value of tol	0.001 [4]
Number of final induced logic $\boldsymbol{L_r}(\boldsymbol{S_i^f})$ selected	1 [12]

There are five classification metrics used to measure the performance of the model *r*2SATMRA, which are accuracy (*acc*) [13], precision (*pre*) [14], specificity (*spe*) [15], Matthew correlation coefficient (*mcc*) [16], and F1-score (*f1s*) [17] that are formulated in Eqs. (20)–(24) below:

$$acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad acc \in (0{,}100), \tag{20}$$

$$pre = \frac{TP}{TP + FP} \times 100\%, \quad pre \in (0{,}100), \tag{21}$$

$$spe = \frac{TN}{TN + FP} \times 100\%, \quad spe \in (0{,}100), \tag{22}$$

$$mcc = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad mcc \in (-1{,}1), \tag{23}$$

$$f1s = \frac{TP}{TP + \frac{1}{2}(FP + FN)}, \qquad f1s \in (0{,}1). \tag{24}$$

The phone prices dataset contains phone features, including the price of popular brands. In this dataset, the discontinued phones are excluded. Out of 22 attributes, we will only consider 20 attributes because 2 attributes contained too many unique values, which are phone name and announcement date. This exclusion was taken to prevent DHNN in training from irrelevant or trivial factors that heavily affect the quality of the final output. The dependent attributes are prices (USD) and independent attributes are brand, OS, inches, resolution, battery, battery type, RAM (GB), weight, storage (GB), video (720p), video (1080p), video (4K), video (8K), video (30 fps), video (60 fps), video (120 fps), video (240 fps), video (480 fps), and video (960 fps). Comparative analysis will be conducted against state-of-the-art systematic SAT-logic classifiers, namely 2SATRA and 3SATRA. Baseline selection is justified by strong reported classification performance within systematic SAT-logic frameworks reported by [5] and [6], respectively. Relevance and effectiveness of the proposed non-systematic *r*2SATMRA will be demonstrated through comparison with these baselines.

3. RESULTS AND DISCUSSION

This section consisted of results for all phases in *r*2SATMRA. Firstly, $Jacc$ values for all 19 attributes can be observed based on Fig. 2 below.

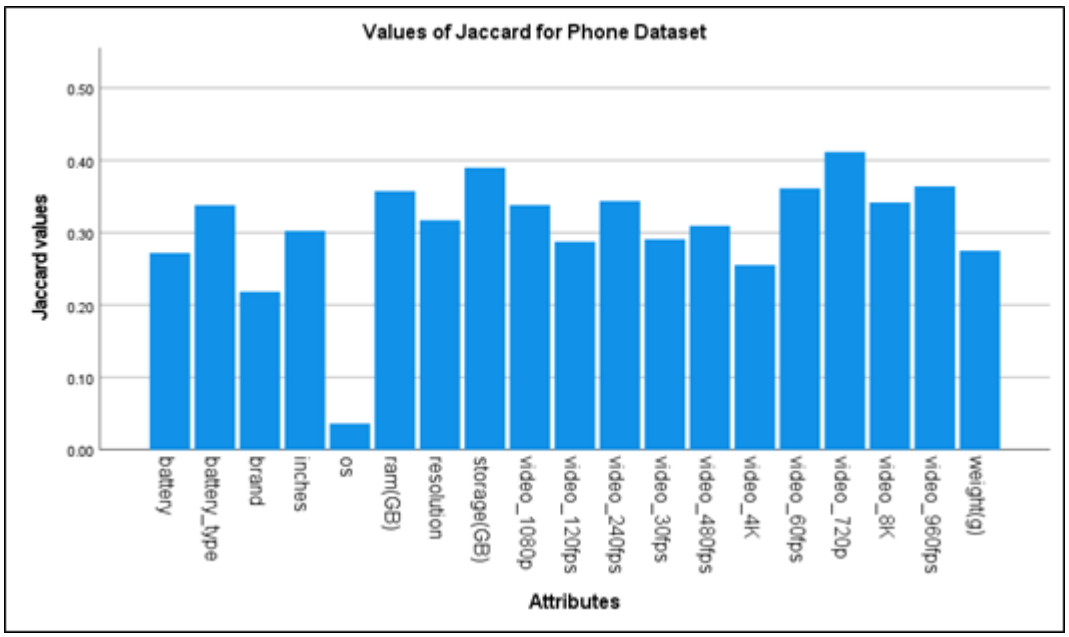

Figure 2. *Jacc* Evaluation for Phone Prices Dataset

The range of the evaluated $Jacc$ presented in Fig. 2 is between 0.0 and 0.42. As mentioned in Table 1, only 10 attributes with the highest $Jacc$ are selected to be trained by *r*2SATMRA. The chosen attributes are resolution, battery type, RAM (GB), storage (GB), video 720p, video 1080p, video 8K, video 60 fps, video 240 fps, and video 960 fps. Table 2 presents $Jacc$ values for all the selected attributes as follows:

Table 2. *Jacc* Values for Selected Attributes

Attribute	$Jacc$
Resolution (A_1)	0.31720
Battery type (B_1)	0.33796
Ram (GB) (A_2)	0.35785
Storage (GB) (B_2)	0.38974
Video 720p (A_3)	0.41170
Video 1080p (B_3)	0.33821
Video 8K (A_4)	0.34198
Video 60 fps (B_4)	0.36146
Video 240 fps (C_1)	0.34391
Video 960 fps (C_2)	0.36403

Secondly, the performance of the proposed *r*2SATMRA is measured when comparing the quality of the retrieved induced logic when compared with baseline models. Table 3 showcases the results of all measured metrics for *r*2SATMRA, 2SATRA [5], and 3SATRA [6], respectively. The bold values in Table 3 represent the highest values for each metric. For all 5 folds of cross-validation, the proposed *r*2SATMRA attained the highest values for all measured metrics as opposed to baseline models. The first reason for this finding is that *r*2SATMRA implemented JFSM to identify relevant attributes based on $Jacc$ values. Attributes with high $Jacc$ reflect the close distance of the measured entry to the class of the dataset. This contributes to the optimal extraction of important patterns of the dataset. Both baseline models only applied random attribute selection, which degrades the performance of DHNN by training insignificant attributes. Consequently, 2SATRA and 3SATRA attained a much lower accuracy with at least 12% differences. In another perspective, *r*2SATMRA achieved the highest *spe* values for all folds due to the structural component of weighted logic. The allocation of negative and positive literals in 2SATRA and 3SATRA was set at random. Hence, the final induced logic retrieved by these models contained an inconsistent representation of the dataset, particularly with respect to the negative class of the dataset (which is phone price less than or equal to 337.85 US dollars). The proposed *r*2SATMRA attained high *mcc* values, which portrayed the model's ability to produce final induced logic that has the ability to classify in both outcomes of the dataset. According to [18], optimal *mcc* values indicate that the model generates solutions that are balanced between the negative and positive classes of the dataset. This balance enables the model to mitigate overfitting and produce solutions that can generalize

well to unseen patterns in the data. Based on this observation, the 2SATRA and 3SATRA models performed poorly in extracting optimal solutions for the Phone Prices dataset.

Table 3. Comparison of *r*2SATMRA, 2SATRA, and 3SATRA Based on Five Performance Metrics for 5-fold Cross Validation

***k*-fold**	**Performance Metrics**	***r*2SATMRA**	**2SATRA** [5]	**3SATRA** [6]
1-fold	*acc*	**0.8830**	0.6876	0.6231
	pre	**0.8925**	0.2650	0.5291
	spe	**0.9760**	0.6908	0.7568
	mcc	**0.5334**	0.0525	0.1848
	f1s	**0.5887**	0.0503	0.4673
2-fold	*acc*	**0.8033**	0.6711	0.5010
	pre	**0.8850**	0.4370	0.6699
	spe	**0.9674**	0.6684	0.7082
	mcc	**0.5506**	0.1372	0.0813
	f1s	**0.6270**	0.0830	0.4775
3-fold	*acc*	**0.8165**	0.6628	0.4165
	pre	**0.9187**	0.5140	0.7944
	spe	**0.9744**	0.6577	0.6508
	mcc	**0.5969**	0.1675	0.0048
	f1s	**0.6706**	0.0974	0.4906
4-fold	*acc*	**0.8579**	0.6843	0.3686
	pre	**0.9453**	0.4500	0.9100
	spe	**0.9827**	0.6795	0.6949
	mcc	**0.6769**	0.1749	0.0178
	f1s	**0.7378**	0.0861	0.4879
5-fold	*acc*	**0.8628**	0.6380	0.3900
	pre	**0.9929**	0.4510	0.9189
	spe	**0.9974**	0.6395	0.6400
	mcc	**0.7160**	0.0781	0.0043
	f1s	**0.7714**	0.0837	0.5251

Once the performance of *r*2SATMRA has been validated by optimal values of *acc*, *pre*, *spe*, *mcc*, and *f1s*, the discussion of the retrieved final induced logic by the model will be discussed below. Eq. (25) formulates the retrieved final induced logic $\left(L_r^{induced}\right)$ by the proposed model, where,

$$L_r^{induced} = (A_1 \vee B_1) \wedge (\neg A_2 \vee B_2) \wedge (\neg A_3 \vee B_3) \wedge (\neg A_4 \vee B_4) \wedge C_1 \wedge C_2 \tag{25}$$

Each variable in Eq. (25) (refer to Table 2) represents an attribute selected from the Phone Prices dataset using JFSM. Based on the applied k-means clustering approach, the outcome of $L_r^{induced} = -1$ corresponds to a low phone price, whereas $L_r^{induced} = 1$ indicates the opposite. One of the possible entries that would make $L_r^{induced} = -1$ is $\{A_1, B_1, A_2, B_2, A_3, B_3, A_4, B_4, C_1, C_2\} = \{-1, -1, 1, -1, 1, -1, 1, -1, -1, -1\}$. One of the possible deductions is that frames per second (fps) refers to a smartphone's capability to capture how many individual frames or images a camera records in one second [19]. Higher fps values allow smoother video playback and are especially useful for slow-motion effects. Phones without 240 fps or 960 fps features are generally cheaper. At 240 fps (C_1), the camera captures 240 frames per second; when played back at a standard frame rate (e.g., 30 fps), the video appears eight times slower. At this rate, phones could capture anatomical details for objective clinical use—demonstrating this feature as a potential medical tool. At 960 fps (C_2), the camera captures 960 frames per second, producing videos that appear 32 times slower than real life. In today's digital era, high-quality cameras are valuable for preserving memories. For instance, parents can record a baby's first steps in slow motion, adding emotional impact. Businesses can also use 960 fps for promotional videos, such as coffee advertisements, to create visually stunning and engaging content that showcases textures, motion, and sensory appeal—benefiting small business owners without the budget for professional videographers. These capabilities require advanced camera sensors and powerful image processing units (ISP) to handle the rapid capture and processing of large amounts of data. Such hardware increases production costs which explain why phones without 240 fps $(C_1 = -1)$, and 960 fps $(C_2 = -1)$ features are generally more affordable $\left(L_r^{induced} = -1\right)$. The recommendation for this extracted knowledge is that budget smartphones without advanced slow-motion

video recording features are ideal for users seeking reliable, cost-effective devices for basic to moderate use. For instance, students typically need phones with essential functions such as communication, internet access, and basic photography. As they generally do not have a stable income, it is wise for them to reduce expenses by avoiding high-priced models and choosing more affordable options. This allows students to allocate their limited funds toward essential needs such as tuition fees, books, rent, food, and transportation, helping them maintain financial stability and avoid overspending on luxury items like high-end smartphones.

Another possible deduction is that smartphones commonly use two types of batteries: lithium-polymer (Li-Po) ($B_1 = 1$) and lithium-ion (Li-Ion) ($B_1 = -1$). Several factors make phones with Li-Ion ($B_1 = -1$) batteries cheaper ($L_r^{induced} = -1$). First, Li-Ion ($B_1 = -1$) batteries have been in development since 1912 and gained popularity after Sony's adoption in 1991. Their long history has led to a standardized, optimized manufacturing process, reducing production costs. Second, Li-Ion ($B_1 = -1$) batteries use a liquid electrolyte, which is simpler and cheaper to produce than the gel or solid polymer electrolyte in Li-Po batteries ($B_1 = 1$). Their cylindrical or rectangular designs are also easier to manufacture. Third, Li-Ion ($B_1 = -1$) batteries have lower production complexity, requiring fewer steps and less specialized handling compared to Li-Po ($B_1 = 1$) batteries. Li-Ion ($B_1 = -1$) batteries also offer high energy density, lightweight design, rapid charging, low self-discharge, long lifespan, and environmental benefits—enhancing user experience and sustainability. Therefore, smartphones using Li-Ion ($B_1 = -1$) batteries are generally more affordable than those with Li-Po ($B_1 = 1$) batteries. As a recommendation, phone companies should use Li-Ion ($B_1 = -1$) batteries to reduce costs through bulk purchasing. Negotiating large-volume deals with battery suppliers can secure lower unit prices, as higher order quantities often come with discounts. Additionally, standardizing Li-Ion ($B_1 = -1$) battery models across multiple smartphone models or product lines can further lower production costs by streamlining manufacturing and inventory management.

The last deduction from the extracted logic formulated in Eq. (25) is that phones are generally cheaper ($L_r^{induced} = -1$) when equipped with 720p resolution ($A_3 = 1$) instead of 1080p ($B_3 = -1$). While both fall under the High Definition (HD) category, they differ in several ways. In terms of image quality, 720p ($A_3 = 1$) is adequate for everyday use, such as browsing, social media, and video streaming, offering clear visuals though less sharp than 1080p ($B_3 = -1$). The 1080p resolution ($B_3 = -1$) delivers superior clarity and detail, making it ideal for gaming and movie viewing, especially on larger screens where the difference is more noticeable. Battery consumption also differs: 720p ($A_3 = 1$) screens use less power due to fewer pixels, while 1080p ($B_3 = -1$) requires more energy to support the higher pixel count. Notably, phones with 1080p ($B_3 = -1$) often include larger batteries to address this, but this increases manufacturing costs and, ultimately, the device price.

In general, phone prices are lower when they lack 240 fps, 960 fps, and 1080p features, and when they use lithium-ion batteries. Consumers should be aware of how these features affect pricing to make informed purchasing decisions. Marketing from smartphone brands can be persuasive and sometimes misleading, so it is important to look beyond promotional claims and rely on logical reasoning and unbiased reviews to assess a phone's true value. By understanding which features matter most, consumers can navigate the market more effectively and choose devices that best match their preferences and priorities.

4. CONCLUSION

This paper presents the pattern extraction of the Phone Prices dataset using a logic mining approach called *r*2SATMRA. The proposed *r*2SATMRA is divided into four phases, which are the preliminary phase, logic phase, training phase, and retrieval phase. Each phase is crucial to ensure optimal knowledge extraction through the embedded logical rule of *r*2SAT. By utilizing the transparency and interpretability properties of SAT, the output produced by DHNN via *r*2SAT can be easily understood by a non-practitioner of AI approaches. This addresses the black box limitation in many AI models. A non-systematic logical rule based on proposed *r*2SAT with $k = 1,2$ is included in *r*2SATMRA to improve interpretability. An attribute selection method of JFSM is included to reduce excessive reliance on correlation-based attribute selection. This supports broader logic mining exploration and strengthens pattern training and recognition across diverse datasets. The reliability of *r*2SATMRA as a classification model was measured using multiple performance metrics such as *acc*, *pre*, *spe*, *mcc*, and *f1s*. As compared to baseline models of 2SATRA [5] and 3SATRA [6], *r*2SATMRA were able to outperform these existing works by attaining high values for all the measured metrics. Currently, *r*2SATMRA is limited to the top 10 attributes selected via JFSM. A larger attribute set

increases DHNN complexity and may hinder efficient optimization during the training phase. The proposed model is currently limited to extracting patterns for binary-class datasets. Under the adopted logical representation, pattern formation is defined in binary form, which prioritizes salient and significant patterns. Extending the framework to continuous-valued targets (e.g., non-binary phone prices) may reduce classification efficiency and is not addressed in the current study. Future extension on the research can consider the inclusion of a topological data perspective as suggested by [20]. In conclusion, the proposed method is a powerful tool for identifying smartphone pricing patterns and sets a new standard for existing classification methods. This research contributes to intelligent pattern analysis and offers a strong model that can be applied to other datasets and challenges.

Author Contributions

Nurul Najwa Ahmad Azam: Writing-Original Draft, Visualization, Data Curation. Nur Ezlin Zamri: Writing-Review and Editing, Conceptualization. Mohd Shareduwan Mohd Kasihmuddin: Resources, Formal Analysis. Nurul Atiqah Romli: Methodology. Mohd. Asyraf Mansor: Validation. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research is financially supported by Universiti Putra Malaysia for the IPM Putra Grant with Project Code: GP-IPM/2024/9806600. All authors gratefully acknowledge the support from Universiti Putra Malaysia.

Acknowledgment

The authors would like to express their sincere appreciation to all individuals and institutions who contributed to this research. Special thanks are extended to colleagues, reviewers, and collaborators for their constructive feedback and valuable assistance throughout the study.

Declarations

The authors declare no competing interests.

Declaration of Generative AI and AI-Assisted Technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript, including writing, editing, data analysis, or the creation of tables and figures.

REFERENCES

[1] K. F. Yang, V. Nalluri, C. C. Liu, and L. S. Chen, “DISCOVERING KEY SUCCESSFUL FACTORS OF MOBILE ADVERTISEMENTS USING FEATURE SELECTION APPROACHES,” Big Data Cogn. Comput., vol. 9, no. 5, p. 119, 2025. doi: https://doi.org/10.3390/bdcc9050119

[2] J. J. Hopfield and D. W. Tank, “NEURAL’ COMPUTATION OF DECISIONS IN OPTIMIZATION PROBLEMS,” Biol. Cybern., vol. 52, no. 3, pp. 141–152, 1985. doi: https://doi.org/10.1007/BF00339943

[3] W. A. T. W. Abdullah, “LOGIC PROGRAMMING ON A NEURAL NETWORK,” Int. J. Intell. Syst., vol. 7, no. 6, pp. 513–519, 1992. [Online]. doi: https://doi.org/10.1002/int.4550070604

[4] S. Sathasivam, “UPGRADING LOGIC PROGRAMMING IN HOPFIELD NETWORK,” Sains Malays., vol. 39, no. 1, pp. 115–118, 2010.

[5] L. C. Kho, M. S. M. Kasihmuddin, M. Mansor, and S. Sathasivam, “LOGIC MINING IN LEAGUE OF LEGENDS,” Pertanika J. Sci. Technol., vol. 28, no. 1, 2020.

[6] N. E. Zamri, M. A. Mansor, M. S. Mohd Kasihmuddin, A. Alway, S. Z. Mohd Jamaludin, and S. A. Alzaeemi, “AMAZON EMPLOYEES RESOURCES ACCESS DATA EXTRACTION VIA CLONAL SELECTION ALGORITHM AND LOGIC MINING APPROACH,” Entropy, vol. 22, no. 6, p. 596, 2020. doi: https://doi.org/10.3390/e22060596

[7] S. A. Karim, N. E. Zamri, A. Alway, M. S. M. Kasihmuddin, A. I. M. Ismail, M. A. Mansor, and N. F. A. Hassan, “RANDOM SATISFIABILITY: A HIGHER-ORDER LOGICAL APPROACH IN DISCRETE HOPFIELD NEURAL NETWORK,” IEEE Access, vol. 9, pp. 50831–50845, 2021. doi: https://doi.org/10.1109/ACCESS.2021.3068998

[8] Z. G. Saribatur and T. Eiter, “OMISSION-BASED ABSTRACTION FOR ANSWER SET PROGRAMS,” Theory Pract. Log. Program., vol. 21, no. 2, pp. 145–195, 2021. [Online]. doi: https://doi.org/10.1017/S1471068420000095

[9] S. Sathasivam and W. A. T. W. Abdullah, "LOGIC MINING IN NEURAL NETWORK: REVERSE ANALYSIS METHOD," Comput., vol. 91, pp. 119–133, 2011. [Online]. doi: https://doi.org/10.1007/s00607-010-0117-9

[10] M. S. M. Kasihmuddin, S. Z. M. Jamaludin, M. A. Mansor, H. A. Wahab, and S. M. S. Ghadzi, "SUPERVISED LEARNING PERSPECTIVE IN LOGIC MINING," Mathematics, vol. 10, no. 6, p. 915, 2022. doi : https://doi.org/10.3390/math10060915

[11] H. Sun and K. Kim, "EVALUATION OF IOT-ENABLED INTERACTIVE UI DESIGN EFFECT BASED ON THE DISCRETE MATHEMATICAL MODEL," Mobile Inf. Syst., 2022. [Online]. doi: https://doi.org/10.1155/2022/2139754

[12] N. Roslan, S. Sathasivam, and F. L. Azizan, "CONDITIONAL RANDOM K SATISFIABILITY MODELING FOR K= 1, 2 (CRAN2SAT) WITH NON-MONOTONIC SMISH ACTIVATION FUNCTION IN DISCRETE HOPFIELD NEURAL NETWORK," AIMS Mathematics, vol. 9, pp. 3911-3956, 2024. doi: https://doi.org/10.3934/math.2024193

[13] S. S. Shafin, "AN EXPLAINABLE FEATURE SELECTION FRAMEWORK FOR WEB PHISHING DETECTION WITH MACHINE LEARNING," Data Sci. Manag., vol. 8, no. 2, pp. 127–136, 2025. doi: https://doi.org/10.1016/j.dsm.2024.08.004

[14] M. R. Salmanpour, *et al.*, "MACHINE LEARNING EVALUATION METRIC DISCREPANCIES ACROSS PROGRAMMING LANGUAGES AND THEIR COMPONENTS IN MEDICAL IMAGING DOMAINS: NEED FOR STANDARDIZATION," IEEE Access, 2025. doi:https://doi.org/10.1109/ACCESS.2025.3549702

[15] Abousaber, H. F. Abdallah, and H. El-Ghaish, "ROBUST PREDICTIVE FRAMEWORK FOR DIABETES CLASSIFICATION USING OPTIMIZED MACHINE LEARNING ON IMBALANCED DATASETS," Front. Artif. Intell., vol. 7, p. 1499530, 2025. doi: https://doi.org/10.3389/frai.2024.1499530

[16] Omazic, G. Grandi, S. Widgren, J. Rocklöv, J. Wallin, J. C. Semenza, and N. Abiri, "AUTOMATED TICK CLASSIFICATION USING DEEP LEARNING AND ITS ASSOCIATED CHALLENGES IN CITIZEN SCIENCE," Sci. Rep., vol. 15, no. 1, p. 24942, 2025. doi: https://doi.org/10.1038/s41598-025-10265-x

[17] Alharthi, M. Alaryani, and S. Kaddoura, "A COMPARATIVE STUDY OF MACHINE LEARNING AND DEEP LEARNING MODELS IN BINARY AND MULTICLASS CLASSIFICATION FOR INTRUSION DETECTION SYSTEMS," Array, p. 100406, 2025. doi: https://doi.org/10.1016/j.array.2025.100406

[18] Chugh, N. Malik, D. Gupta, and B. S. Alkahtani, "A PROBABILISTIC APPROACH DRIVEN CREDIT CARD ANOMALY DETECTION WITH CBLOF AND ISOLATION FOREST MODELS," Alexandria Eng. J., vol. 114, pp. 231–242, 2025. doi: https://doi.org/10.1016/j.aej.2024.11.054

[19] T. Eftimov, I. Kostova, S. Fouzar, D. Brabant, K. Nikolov, and V. Vladev, "SMARTPHONE-READABLE OPTICAL-FIBER QUASI-DISTRIBUTED PHOSPHORESCENT TEMPERATURE SENSOR," Photonics, vol. 11, no. 8, p. 694, Jul. 2024. doi: https://doi.org/10.3390/photonics11080694

[20] K. F. Mojdehi, B. Amiri, and A. Haddadi, "A NOVEL HYBRID MODEL FOR CREDIT RISK ASSESSMENT OF SUPPLY CHAIN FINANCE BASED ON TOPOLOGICAL DATA ANALYSIS AND GRAPH NEURAL NETWORK". IEEE Access, vol. 13, pp. 13101- 13127, 2025. doi: https://doi.org/10.1109/ACCESS.2025.3528373