JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Mohd Kasihmuddin, M. S., Abdul Halim, N. S., Mohd Jamaludin, S. Z., Mansor, M.. A., Alway, A., Zamri, N. E., Azhar, S. A., & Marsani, M. F. (2023). Logic mining approach: Shoppers' purchasing data extraction via evolutionary algorithm. *Journal of Information and Communication Technology, 22*(3), 309-335. https://doi.org/10.32890/jict2023.22.3.1

# Logic Mining Approach: Shoppers' Purchasing Data Extraction via Evolutionary Algorithm

[1]Mohd Shareduwan Mohd Kasihmuddin,
[2]Nur Shahira Abdul Halim, [3]Siti Zulaikha Mohd Jamaludin,
[4]Mohd. Asyraf Mansor, [5]Alyaa Alway, [6]Nur Ezlin Zamri,
[7]Siti Aishah Azhar & *[8]Muhammad Fadhil Marsani
[1,2,3,7&8]School of Mathematical Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia
[4,5&6]School of Distance Education, Universiti Sains Malaysia, Pulau Pinang, Malaysia

[1]shareduwan@usm.my
[2]nurshahirahalim@student.usm.my
[3]szulaikha.szmj@usm.my
[4]asyrafman@usm.my
[5]alyaalway@student.usm.my
[6]ezlinzamri@student.usm.my
[7]saishahazhar96@student.usm.my
*[8]fadhilmarsani@usm.my
*Corresponding author

Received: 18/7/2022 Revised: 4/1/2023 Accepted: 22/2/2023 Published: 24/7/2023

## ABSTRACT

Online shopping is a multi-billion-dollar industry worldwide. However, several challenges related to purchase intention can impact the sales of e-commerce. For example, e-commerce platforms are unable to identify which factors contribute to the high sales of a product. Besides, online sellers have difficulty finding products that align with customers' preferences. Therefore, this work will utilize an artificial neural network to provide knowledge extraction for the online shopping industry or e-commerce platforms that might improve their sales and services. There are limited attempts to propose knowledge extraction with neural network models in the online shopping field, especially research revolving around online shoppers' purchasing intentions. In this study, 2-satisfiability logic was used to represent the shopping attribute and a special recurrent artificial neural network named Hopfield neural network was employed. In reducing the learning complexity, a genetic algorithm was implemented to optimize the logical rule throughout the learning phase in performing a 2-satisfiability-based reverse analysis method, implemented during the learning phase as this method was compared. The performance of the genetic algorithm with 2-satisfiability-based reverse analysis was measured according to the selected performance evaluation metrics. The simulation suggested that the proposed model outperformed the existing model in doing logic mining for the online shoppers dataset.

**Keywords:** 2-satisfiability, genetic algorithm, Hopfield neural network, logic mining, online shopping.

## INTRODUCTION

Artificial intelligence (AI) is a computer program designed to obtain information in a more specific and analogous manner like the human brain. AI allows researchers to identify the pattern of the raw data precisely, learn from those data and adapt to display. According to Davenport et al. (2019), AI has the tendency to influence future strategies on marketing, customer behaviors, and customer services. One of the most notable techniques of AI is the artificial neural network (ANN). ANN is a representation in the form of machine learning emerging from the concept of simulating the human brain (Zhou et al., 2015). The Hopfield neural network (HNN) serves as a good illustration and model of ANN. HNN has a feature known as content-addressable memory, abbreviated as CAM, which acts as a central memory to store satisfied interpretation and synaptic weight.

The structure of HNN is a set of neurons without hidden neurons. On top of that, HNN is a variant of ANN that represents the human brain (Hopfield & Tank, 1985). According to Brette et al. (2007), the dynamic of HNN can be seen when the output of each neuron becomes feedback to each other. HNN has a symmetrical weight with no self-connection. Effective synaptic weight management will determine the capability of HNN to be combined with the learning algorithm and as the main network of logic mining or knowledge extraction paradigms.

In the 21st century, online shopping has become incredibly popular due to the emergence of smartphones and the Internet. Based on Coppola (2020), in 2020, about two billion individuals across the world bought products online. The development of online shopping is incredible and does not show any sign of slowing down any time soon. Online shopping is a mechanism for purchasing products and services from vendors who sell online. Shoppers can access web stores and shop just at their fingertips or in front of the computer. Furthermore, online shopping is designed to enable individuals to sell and publicize their goods or to purchase any items they desire through a website.

According to osCommerce Ltd (2002), browsing for information about products or buying goods online has become a common activity. The rationale for selecting online shopping might differ from comfortable to competitive prices. Online businesses give their utmost to ensure that the experience of online shopping matches the consumers' desires. Besides, they will provide consumers with product information and pictures, comprising a significant advance, such as a 360-degree item viewpoint. In addition, for online shopping, the details of sizes will be provided as this helps shoppers to buy. Therefore, information on online shopping behavior will be guided to specific parties like marketers or online stores in formulating and implementing an efficient marketing strategy to promote their goods and services. Up to this point, the development of the data extraction model in HNN for the online shopper analysis has received less research effort, despite the promising capability of HNN in extracting the shopper purchasing trend.

Propositional satisfiability (SAT) logic is comprehensively utilized by many practitioners in the field of prediction (Hutter et al., 2014), machine learning (Hireche et al., 2020), and ANN (Abdullah, 1992).

In the domain of ANN, SAT logic is a promising formulation to govern the structure of activated neurons. Abdullah (1992) pioneered the use of logic in ANN by minimizing the cost function of the logical representation in HNN. The proposed work embedded Horn SAT into HNN through the cost function comparison with the Lyapunov energy function. As a result, the proposed HNN had the capability to produce a final neuron state with high compatibility with the "learned" logical representation. Recently, Kasihmuddin et al. (2017) engineered the 2-satisfiability (2SAT) logical rule into HNN by means of capitalizing the two neurons per clause. The proposed network achieved promising results in terms of global minima ratio for a lower number of neurons. This motivates practitioners to utilize this network to perform knowledge extraction through logic mining. Alway et al. (2020) proposed a 2-satisfiability-based reverse analysis method (2SATRA) in analyzing the influence of cocoa beans, crude petroleum, and gold on the trend of palm oil prices. Based on the experimental value, the proposed 2SATRA was also able to obtain an acceptable quality of the induced logic. Even though the work achieved more than 70 percent accuracy, the use of an exhaustive search (ES) algorithm during the learning phase reduces the capability of the proposed 2SATRA to obtain optimal induced logic. Therefore, an optimal metaheuristics algorithm that has a simple and effective structure is required to optimize the learning phase of HNN.

Recently, Zamri et al. (2020) proposed a 3-satisfiability-based reverse analysis method (3SATRA) that was optimized by the clonal selection algorithm (CSA). The proposed 3SATRA managed to identify the significant factors that contributed to the employees' resources applications. The finding showed that the proposed model achieved higher precision and fewer errors compared to ES. By choosing the right learning algorithm, 2SATRA can obtain optimal synaptic weight, leading to optimal induced logic. By utilizing the attributes chosen from the online shopping dataset, the data were represented based on the bipolar neurons in the 2SATRA model. Therefore, the output from 2SATRA is an optimal logical rule that is being used to analyze the behavior of customers with a given set of logical rules. The contribution of this research embarks on the objective to represent a variant of the systematic logical representation, 2-satisfiability logic in HNN. In addition, this research aims to incorporate the 2-satisfiability logical representation with a hybrid genetic algorithm as the learning algorithm in HNN. Therefore, another impetus of this work is to propose the 2-satisfiability-based reverse analysis method in HNN in extracting the optimal association of online shoppers' purchasing intentions.

## MATERIALS AND METHODS

This section outlines the methodology used to provide knowledge extraction to the online shopping industry. First, the method of extracting the formulation of the 2-satisfiability logical rule to represent the information of the dataset is described. Then, the study explains the process of 2-satisfiability logic into HNN. Next, an enhanced reverse analysis method, namely the 2-satisfiability-based Reverse Analysis method (2SATRA), and how the genetic algorithm can be implemented to optimize the logical rule during the learning phase in 2SATRA are clarified. The extracted logic in the online shopping datasets is also discussed.

### 2-satisfiability Logic (2 SAT)

2-satisfiability logical representation or shortened as 2SAT denotes the logical representation (rule) of strictly two literals per clause (Kasihmuddin et al., 2017). Then, 2SAT representation can be expressed as 2CNF (2-conjunctive normal form). According to Kowalski (1979), 2SAT is a case of systematic Boolean satisfiability, which incorporates a restriction on two variables. 2SAT is coined to be a non-deterministic polynomial (NP) problem. There are three components of the 2SAT logic:

1. Comprises a set of $m$ variables, $x_1, x_2, ..., x_m$.
2. A set of literals where a variable can be either negated or non-negated of the variable.
3. Given a set of $n$ definite logical clauses: $C_1, C_2, ..., C_n$. Each consists of two literals and is combined by just logical OR ( $\vee$ ).

Each $m$ variable can hold a neuron representation of 1 or -1 that corresponds to TRUE and FALSE. Next, the main task of 2SAT $\left(P_{2SAT}\right)$ is to represent the information in terms of variables, where each variable will be assigned a truth value that leads to $P_{2SAT} = 1$.

In other words, $P_{2SAT} \neq 1$ when at least one clause in the formula is not satisfiable. For example, the following 2SAT logic $\left(P_{2SAT}\right)$ consists of three clauses $\left(C_i\right)$ and two literals for each clause will be used, which can be referred to in Equation 1. For instance, Equations 1–3 are formulated as follows:

$$P_{2SAT} = (A \vee B) \wedge (C \vee \neg D) \wedge (\neg E \vee F) \tag{1}$$

Then, the general structure of $P_{2SAT}$ is formulated as in Equations 2 and 3:

$$P_{2SAT} = \wedge_{i=1}^{n} C_i \tag{2}$$

$$C_i = \vee_{j=1}^{k} \left(x_{ij}, y_{ij}\right), \; k = 2 \tag{3}$$

where $C_i$ and variable $x_{ij}, y_{ij}$ in $P_{2SAT}$ are connected with AND $(\wedge)$ and OR $(\vee)$, respectively. Note that $x_i \in \left\{k_i, \neg k_i\right\}$ and $y_i \in \left\{r_i, \neg r_i\right\}$, $\neg k_i$, where $\neg r_i$ are negations of literals. Table 1 demonstrates the outcome for two different cases of the P2SAT model. From the table, P2SAT is satisfiable when $(A,B,C,D,E,F) = (1, -1, 1, -1, -1, -1)$, which gives $P_{2SAT} = 1$, resulting in P2SAT to be true. On the contrary, P2SAT is unsatisfiable when $(A,B,C,D,E,F) = (1, -1, -1, 1, -1, 1)$, which gives $P_{2SAT} = -1$.

**Table 1**

*Example of Cases of the 2SAT Logical Rule*

| Case | Instances of $P_{2SAT}$ | Outcome |
|---|---|---|
| 1 | $(A,B,C,D,E,F) = (1,-1,1,-1,-1,-1)$ | *Satisfiable* $\left(P_{2SAT} = 1\right)$ |
| 2 | $(A,B,C,D,E,F) = (1,-1,-1,1,-1,1)$ | *Unsatisfiable* $\left(P_{2SAT} \neq 1\right)$ |

## Hopfield Neural Network

Hopfield neural network (HNN) is a type of recurrent ANN introduced by Hopfield and Tank (1985). Based on Hopfield and Tank (1985), the bipolar form of the neuron state in HNN is firing at the optimum rate of $S_i = 1$, whereas the neuron is not firing at $S_i = -1$. The possibility of two neurons firing at the same time is zero. This dynamic enables the neuron to update consequently. According to Kasihmuddin et al. (2017), the standard HNN neuron update is as given in Equation 4:

$$S_i = \begin{cases} 1, & \text{if } \sum_j W_{ij} S_j \geq \psi_i \\ -1, & \text{Otherwise} \end{cases} \tag{4}$$

where $W_{ij}$ is the notation of the synaptic weight from unit $j$ to unit $i$. $S_j$ is the state of the unit $j$ and $\psi_i$ is the pre-defined threshold value of the unit $i$. The absence of connection of each neuron in HNN, where $W_{ii} = W_{jj} = 0$, leads the network to be symmetrical. Besides, HNN incorporates logical representation represented by the neuron state in notifying the behavior of the network by utilizing synaptic weight (neuron connection). In fact, synaptic weight is a connection between variables and clauses in the logical formula. The integration of the 2SAT logic representation in HNN is represented by HNN-2SAT. The main objective of the network is to minimize the possible logical inconsistency by reducing the cost function of HNN. The general formulation of the corresponding 2SAT cost function, $E_{P_{2SAT}}$, can be represented as Equations 5 and 6:

$$E_{P_{2SAT}} = \sum_{i=1}^{NC} \prod_{j=1}^{NV} L_{ij} \tag{5}$$

whereby $NC$ is defined as the number of clauses and $NV$ refers to the number of variables in the network. The inconsistencies of the 2SAT logical clause, $L_{ij}$, is shown:

$$L_{ij} = \begin{cases} \frac{1}{2}\left(1 - S_x\right), & \text{if} \neg x \\ \frac{1}{2}\left(1 + S_x\right), & \text{Otherwise} \end{cases} \tag{6}$$

where $S_x$ refers to the state of the neuron in HNN. In obtaining the optimal synaptic weight values of HNN, the WA method (Abdullah, 1992) is used directly by comparing the cost function based on the logical representation and Lyapunov energy function. Note that the HNN model is isomorphic to the Ising magnetism model. As a consequence, neurons in HNN will spin until they reach an equilibrium state. The formulation of the local field of 2SAT logical representation in HNN is enclosed as Equations 7–8:

$$h_i(t) = \sum_{j=1, i \neq j \neq k}^{N} W_{ij}^{(2)} S_j + W_i^{(1)} \tag{7}$$

$$S(t) = \operatorname{sgn}\left[h_i(t)\right] \tag{8}$$

where the updating rule of $h_i(t)$ is given by in Equation 7 and $W_{ij}$ refers to the synaptic weight representation from unit $j$ to unit $i$ and "sgn" denotes the signum function. Note that this paper utilizes the hyperbolic activation function (HTAF) to provide an output squashing mechanism to HNN. The formulation of the Lyapunov energy function for HNN-2SAT is given as Equation 9:

$$H_{P_{2SAT}} = -\frac{1}{2} \sum_{i=1, i \neq j}^{N} \sum_{j=1, i \neq j}^{N} W_{ij}^{(2)} S_i S_j - \sum_{i=1}^{N} W_i^{(1)} S_j \tag{9}$$

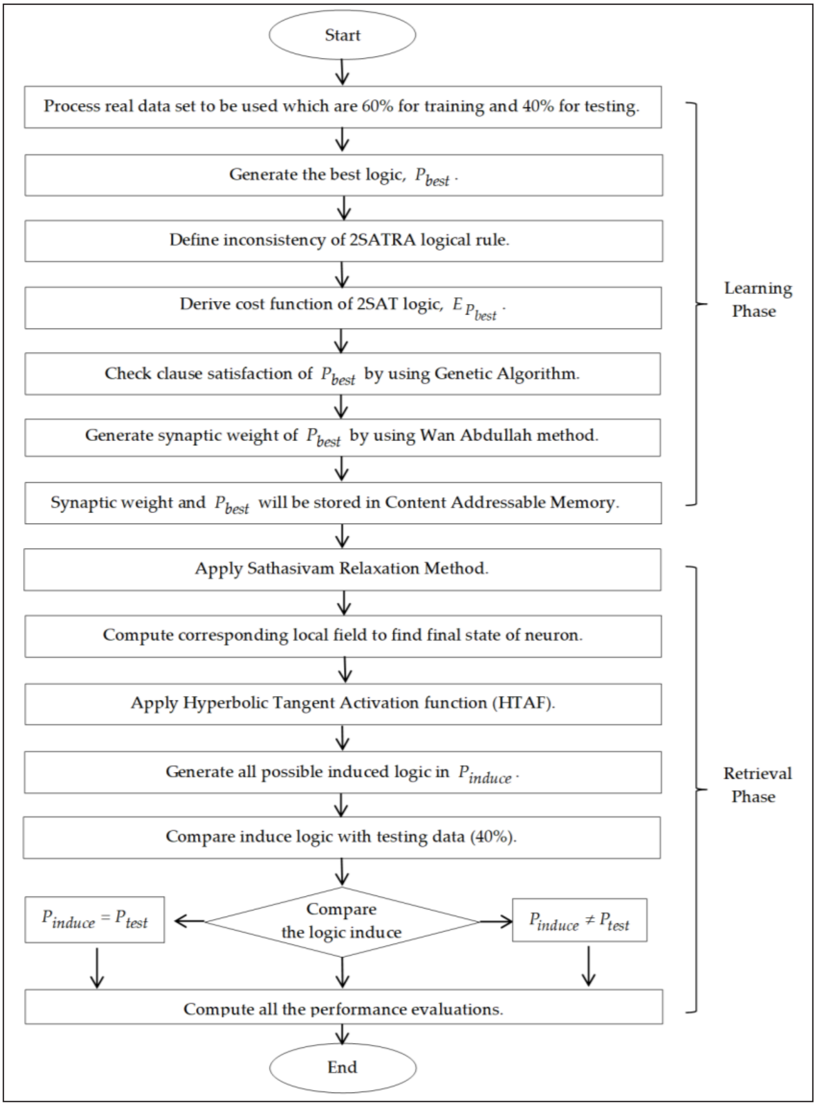
The global minimum energy of HNN often relies on the efficiency of the learning and retrieval phases. Ineffective learning and retrieval phases will possibly lead HNN to be trapped in local minimum energy. By utilizing HNN-2SAT, the attribute chosen from the online shopping data will be characterized in terms of the bipolar neurons in 2SATRA.

**2-satisfiability-based Reverse Analysis Method**

The primary goal of satisfiability logic mining is to retrieve useful information or pattern from an actual set of data entries. Sathasivam and Abdullah (2011) effectively employed the reverse analysis approach to extract the relationship between each individual dataset. Furthermore, the existing reverse analysis method encourages this work to take into account a more directed logical rule and teaching strategy that can precisely generalize the relationship of the given data. The brief process of the enhanced reverse analysis method involving the systematic logical representation, namely 2-satisfiability-based reverse analysis method (2SATRA) will be elaborated in this section. Specifically, 2SATRA is a variant of logic mining paradigm incorporated with HNN-2SAT to retrieve valuable information in the form of satisfiability logical rules that generalize the dataset. Then, the respective attributes of the dataset can behave accordingly to bipolar representations, which are -1 and 1, and be characterized in terms of variables in the systematic 2SAT clause. 2SATRA has the capability to unearth the connection level between two neurons in the data by optimally acquiring corresponding synaptic weights. The WA method will be utilized by 2SATRA during the learning phase to identify the correct (optimal) synaptic weight between two attributes of the logical representations. The full implementation of the 2SATRA model involves phases, such as the learning phase and retrieval phase, as can be seen in Figure 1. First, in the learning phase, all real datasets are converted into the bipolar form by using the k-mean clustering method. The datasets will be segregated accordingly 60 percent on the learning dataset and 40 percent on the retrieval dataset. Next, 2SATRA converts all the entries of the data into 2SAT logical representation due to the nature of the attributes. The goal of each dataset is decided by applying two different goals $\left(P_{PV} \,\&\, P_R\right)$in this work. Accordingly, 2SATRA will evaluate each 2SAT logic and choose the 2SAT logic with a higher frequency in the dataset. This 2SAT logic is called $P_{best}$. Accordingly, the generated $P_{best}$ from 2SATRA can be embedded into CAM and the learning phase of the proposed HNN-2SAT model. Therefore, the suggested genetic algorithm will be embedded in the HNN-2SAT model as HNN-2SATGA to learn from $P_{best}$. The process is continued by computing the synaptic weight to be stored in CAM by using the proposed HNN-2SATGA model. All the information stored in CAM will proceed to the retrieval phase. In the retrieval phase, the computation of the local field is conducted, followed by updating the final neuron states by using the hyperbolic tangent activation function (HTAF). This action is to ensure the final neuron states will produce a list of induced logic. HNN-2SAT will retrieve induced logic $P_{induce}$ by retrieving the state of the neuron. The logical representation with the highest value of success will be chosen as the best $P_{induce}$ and will be represented as the optimal logical rule for the online shopping dataset. The optimal induced logic will explain the relationship between the output and attributes of the datasets.

**Figure 1**

*Brief Implementation and Process of 2SATRA in HNN*




## Genetic Algorithm

Genetic algorithm (GA) is a conventional learning algorithm inspired by the concept of natural selection and genetics. The chromosome in GA is the solution to the problem, and GA employs the likelihood search algorithm by using the chromosome to find an approximate solution to numerous optimization problems. Hoque et al. (2007) stated that since chromosome (solution) in GA is always continuing to improve in every generation, many researchers solve real-life problems using GA. This algorithm comprises three primary operators, namely selection, crossover, and mutation operators. Kasihmuddin et al. (2017) utilized a hybrid GA in HNN to solve the *k*-SAT problem. In this work, during the learning phase, GA will be integrated with HNN-2SAT and represented as HNN-2SATGA to generate consistent interpretation. In addition, the objective function of GA in $P_{2SAT}$ can be formulated as Equation 10:

$$\max\left|f_{2SATGA}\right| \tag{10}$$

where $f_{2SATGA}$ is defined as the number of satisfied clauses (fitness) of chromosomes. In other words, GA is utilized to maximize and capitalize the number of the satisfied clause and minimize the $E_{P_{2SAT}}$ in Equation 5 . The following Equations 11 and 12 are used to evaluate the fitness of any $P_{2SAT}$ logic (Kasihmuddin et al., 2017):

$$f_{2SATGA} = \sum_{i=1}^{NC} C_i \tag{11}$$

$$C_i = \begin{cases} 1, & \textit{Satisfied} \\ 0, & \textit{Otherwise} \end{cases} \tag{12}$$

$C_i = 1$ when the states of variable in the $P_{2SAT}$ clause have consistent interpretations. The stages involved in the proposed GA are as follows:

<u>Stage 1: Initialization</u>
The initial population of *n* was obtained by utilizing bit string representation as a chromosome unit. Each candidate chromosome comprises a possible consistent interpretation of the $P_{2SAT}$ clause. The state of the neuron in each of the states $S_i$, is represented by 1 (TRUE) and -1 (FALSE).

<u>Stage 2: Fitness Evaluation</u>
The fitness values of the corresponding candidate chromosome in GA, which corresponds to $P_{2SAT}$, can be assessed using Equation 11.

Note that the learning phase will be complete when $f_{2SATGA}$ is equal to the number of clauses (maximum fitness).

Stage 3: Selection

In this stage, GA will choose the best $n$ chromosomes with the highest fitness $f_{2SATGA}$ that corresponds to $P_{2SAT}$. Note that selection is vital to ensure that chromosomes with high fitness will improve further.

Stage 4: Crossover

During this stage, two chromosomes that were selected in Stage 3 will be randomly chosen. The offspring chromosome is developed by randomly swapping two genes of the two parent chromosomes. In this context, information of $S_i$ from different locations in the chromosome will be transferred to the other chromosomes. The offspring chromosome is expected to have higher $f_{2SATGA}$ compared to the parent chromosomes.

Stage 5: Mutation

The mutation operator can be implemented in the chromosomes by randomly flipping the state of $S_i$ into the opposite state. The main reason for mutation is to create genetic diversity and avoid possible local minima of $\left|f_{2SATGA} - f_i\right|$. In addition, mutation has the potential to reduce the value of $f_{2SATGA}$ when the wrong $S_i$ is flipped in $P_{2SAT}$. Therefore, the mutation rate remains low to avoid such a problem.

By considering all the stages involved, the pseudocode of the proposed GA in doing $P_{2SAT}$ can be summarized in Algorithm 1.

**Algorithm 1: Pseudocode for the Proposed HNN-2SATGA**

Initialize a population of randomly generated $Q$ chromosomes, $S_i$ for $i=\{1,2,3,...,Q\}$,
where $S_i \in \{1,-1\}$;
*while* $f_{2SATGA} = NC$ or iteration $\geq NH$
{Selection}
*for* $i \in \{1,2,3,..,Q\}$, *do*
Calculate $f_i$ of each $S_i$ by using Equation (11);
*end*
{Crossover}
*for* $S_i \in \{1,2,3,..,Q\}$, *do*
Exchange the states of the selected two $S_i$ at a random point;
*end*
{Mutation}
*for* $S_i \in \{1,2,3,..,Q\}$, *do*
Flipping states from 1 to -1 or vice versa of $S_i$ at a random location;
Evaluate $f_i$ of $S_i$ according to Equation (11);
If $f_{2SATGA} = NC$ or meets the termination criteria, then stop the process.
*end*
*end while*
Return the output of final neuron states.

## Online Shopping Datasets

E-commerce is a multi-billion-dollar industry worldwide. E-commerce refers to the transaction of purchasing and selling various goods and services through the online medium (Internet). E-commerce has been very convenient to use due to its around-the-clock accessibility. Furthermore, stores offer a wider range of products online, increasing the choices of products. On top of this, online shopping is where shoppers can purchase goods online. In this work, the online shoppers' purchasing intention data were used by the open-source repository access of the UCI machine learning repository. The main purpose of using the online shopping dataset is to retrieve the optimal logical rules that reflect the specific behavior of shoppers. Besides, the extracted logic in the online shopping dataset will be used to determine whether shoppers make a purchase. In this work, the 2SATRA with HNN-2SAT model will be utilized to retrieve suitable logical rules from the dataset. Next, the logical rule gives classification, prediction, and knowledge extraction of the whole online shopping dataset. The output from 2SATRA is an optimal logical rule that can be used to explain the behavior of the dataset and the shoppers' behavior in purchasing goods. This research will assist the e-commerce players in developing a more productive sales process by upgrading the websites and product details. Equally significant, the induced logic obtained can be used to increase the number of potential customers purchasing goods by improving their services and strategies.

## IMPLEMENTATION

This section discusses the implementation of the experimental setup and the related performance metric used to evaluate the performance of the proposed model in performing logic mining for the online shoppers dataset.

### Performance Evaluation Metrics

The evaluation of HNN was from the perspective of error analysis. In this context, root mean square error (RMSE) and sum of squared error (SSE) were used to evaluate the competency of the proposed hybrid model. RMSE measured the deviation of the error between the desired fitness, $f_{NC}$, and the existing fitness, $f_i$. Equation 13 of the RMSE used in this experiment is as follows (Alway et al., 2020):

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{1}{n}\left(f_{NC} - f_i\right)^2} \tag{13}$$

On the other hand, SSE calculated the ability of the proposed model to reduce the difference between $f_{NC}$ and $f_i$. This metric was crucial to ensure the proposed model was sensitive to the error. The value of SSE can be obtained by the following Equation 14 (Zamri et al., 2020):

$$SSE = \sum_{i=1}^{n} \left( f_i - f_{NC} \right)^2 \tag{14}$$

Based on the metric interpretation, the capability and performance of the proposed model were considered worst if a higher SSE value was observed.

## Experimental Setup

Two simulations were performed by utilizing $P_{2SAT}$ logic rules. The first simulation, $P_{PV}$, represented the mean value of the page a visitor visited before completing an e-commerce purchase, while the second simulation, $P_R$, denoted the revenue signifying whether a visit had been finalized with the transaction. The real dataset was represented in $P_{2SAT}$ and embedded in 2SATRA for logical extraction. This dataset consisted of 12,330 instances that belonged to different users in one year and had been divided into 60 percent training data and 40 percent retrieval data. In this execution, 7,398 data points were integrated into 2SATRA as training data and 4,932 as retrieval data. Furthermore, two different training algorithms were used in the training phase of 2SATRA, namely HNN-2SATGA and the benchmark algorithm. The benchmark algorithm, i.e., exhaustive search algorithm (ES), was implemented in the HNN-2SAT model, known as HNN-2SATES, which operated based on the trial and error searching technique. HNN-2SATES contained no operator widely used by many researchers when dealing with a small number of neurons (Alway et al., 2020, Kasihmuddin et al., 2017). The hybrid models were run on Windows 7 with an Intel Core i3 and 2GB of RAM using Dev C++ Version 5.11. Tables 2 and 3 reveal the list of parameters in HNN for ES and GA, respectively.

**Table 2**

*List of Parameters in HNN-2SATES*

| Parameter | Parameter Value |
|---|---|
| Neuron Combination $(COMBMAX)$ | 100 (Alway et al., 2020) |
| Tolerance $(Tol)$ | 0.001 (Alway et al., 2020) |
| Number of Trials $(Trial)$ | 100 (Alway et al., 2020) |

(cotinued)

| Parameter | Parameter Value |
|---|---|
| Number of Learning $(NH)$ | 100 (Zamri et al., 2020) |
| Ratio of Learning and Testing | 60:40 (Alway et al., 2020) |
| Learning Method | Exhaustive Search (Kasihmuddin et al., 2017) |
| Activation Function | (HTAF) (Zamri et al., 2020) |

**Table 3**

*List of Parameters in HNN-2SATGA*

| Parameter | Parameter Value |
|---|---|
| Neuron Combination $(COMBMAX)$ | 100 (Kasihmuddin et al., 2017) |
| Tolerance $(Tol)$ | 0.001 (Kasihmuddin et al., 2017) |
| Number of Trials $(Trial)$ | 100 (Kasihmuddin et al., 2017) |
| Number of Learning $(NH)$ | 100 (Zamri et al., 2020) |
| Ratio of Learning and Testing | 60:40 (Alway et al., 2020) |
| Learning Method | Genetic Algorithm (Kasihmuddin et al., 2017) |
| Activation Function | (HTAF) (Zamri et al., 2020) |
| Selection Rate | 0.1 (Zamri et al., 2020) |
| Mutation Rate | 0.001 |
| Number of Generation | 100 |

The simulation dataset was provided to predict the probability of shoppers' purchasing intention in e-commerce. Table 4 shows the simulation data on the average value of the page visited by the visitor before completing an e-commerce transaction ( $P_{PV}$ ). The simulation dataset consisted of six attributes that can be categorized into: i) informational shopping websites (information); ii) the total quantity of time spent by a visitor (informational duration); iii) informational shopping websites on related products (product-related); iv) the total quantity of time spent by a visitor on related products (product-related duration); v) the percentage of visitors that leave a webpage without taking action (bounce rate); and vi) special day. Knowing how these attributes are related to forecasting the number of page values is helpful. The number of page values can be used to determine whether visitors complete their shopping.

**Table 4**

*The Attributes of $P_{PV}$ and the Range for the Attributes*

| Attribute Name | Attribute | Bipolar Representation | Range |
|---|---|---|---|
| Informational | $I$ | 1 | $\geq 1.6735$ |
| | | -1 | $< 1.6735$ |
| Informational Duration | $ID$ | 1 | $\geq 414.146$ |
| | | -1 | $< 414.146$ |
| Product-Related | $PR$ | 1 | $\geq 86.031$ |
| | | -1 | $< 86.031$ |
| Product-Related Duration | $PD$ | 1 | $\geq 3446.5565$ |
| | | -1 | $< 3446.5565$ |
| Bounce Rate | $BR$ | 1 | $\geq 0.095$ |
| | | -1 | $< 0.095$ |
| Special Day | $SD$ | 1 | $\geq 0.338$ |
| | | -1 | $< 0.338$ |

Table 5 depicts the simulation data on the revenue, signifying whether a visit had been finalized with the transaction ($P_R$). The simulation dataset comprised six attributes that can be categorized into: i) visitor type (new or existing visitor); ii) shopper visiting online shopping during the weekend (weekend); iii) informational shopping websites (information); iv) the total quantity of time spent by a visitor (informational duration); v) informational shopping websites on related products (product-related); and vi) the total quantity of time spent by a visitor on related products (product-related duration). Note that the range value in Tables 4 and 5 was obtained from the mean value in each attribute. For the bipolar value, 1 referred to a value less than the mean value, while -1 signified a value more than and equal to the mean value.

**Table 5**

*The Attributes of $P_R$ and the Range for the Attributes*

| Attribute Name | Attribute | Bipolar Representation | Range |
|---|---|---|---|
| Visitor Type | $VT$ | 1 | Returning |
| | | -1 | New Visitor |
| Weekend | $W$ | 1 | True |
| | | -1 | False |
| Informational | $I$ | 1 | $\geq 1.6735$ |
| | | -1 | $< 1.6735$ |
| Informational Duration | $ID$ | 1 | $\geq 414.146$ |
| | | -1 | $< 414.146$ |
| Product-Related | $PR$ | 1 | $\geq 86.031$ |
| | | -1 | $< 86.031$ |
| Product-Related Duration | $PD$ | 1 | $\geq 3446.5565$ |
| | | -1 | $< 3446.5565$ |

## RESULTS AND DISCUSSION

In this study, the hybrid network was trained using the best logical rule that was extracted using 2SATRA. The proposed hybrid models, HNN-2SATES and HNN-2SATGA, were trained using a visitor's intention to perform online shopping. In this work, six types of attributes were used in the $P_{2SAT}$ logical rule in both simulations. In achieving the highest precision, different clause combinations were obtained using permutation. Figures 2 and 3 illustrate the performance error achieved by both HNN-2SATES and HNN-2SATGA models for $NC = 1$ until $NC = 10$ of $P_{PV}$.

**Figure 2**

*RMSE Evaluation for Both HNN-2SAT Models for* $P_{PV}$

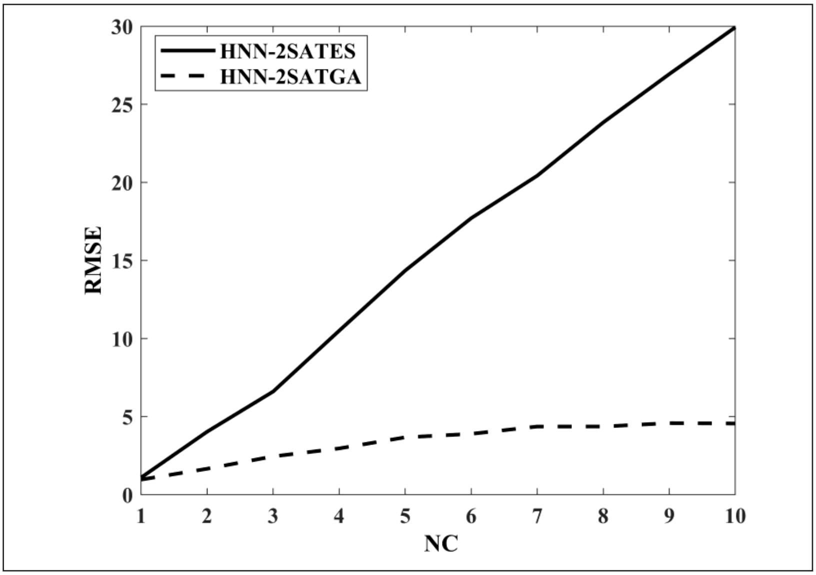
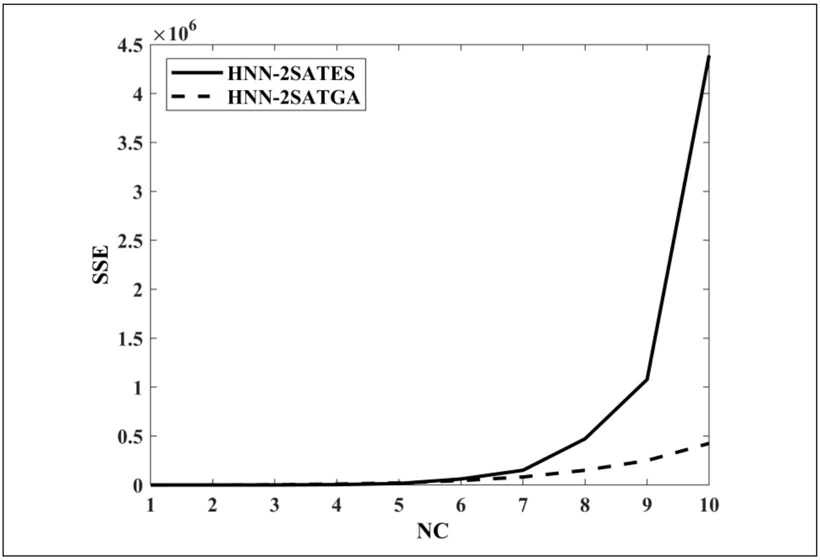



**Figure 3**

*SSE Evaluation for Both HNN-2SAT Models for* $P_{PV}$



From Figure 2, HNN-2SATGA managed to achieve an RMSE value less than 5 for all $NC$. On the other hand, in Figure 3, after $NC = 6$, the increase in error for HNN-2SATES started to leave a huge gap between both models. These results were due to the two optimization operators in GA, which were the crossover and mutation operators. During the crossover stage, the fitness of the offspring chromosomes was expected to be higher than the parent chromosome after the exchange of the gene between two parents' chromosome. Then, the mutation operator improved the fitness of the chromosomes by gene flipping. Therefore, the effectiveness of the optimization operator in GA lowered the error produced in each iteration for HNN-2SATGA during the learning phase. The experiments ended at NC = 10 due to the effectiveness of the proposed model in terms of the sensitivity analysis, HNN-2SATGA, with the benchmark model, HNN-2SATES, which can only be seen when the logic mining simulations of $P_{PV}$ was conducted in fewer neurons. The significant differences could be seen from NC = 6 until NC = 10, indicating the effect of GA as the learning algorithm in optimizing the logic mining model. On the other hand, Figures 4 and 5 illustrate the performance errors achieved by both HNN-2SATES and HNN-2SATGA models for $NC = 1$ until $NC = 10$ of $P_R$.

**Figure 4**

*RMSE Evaluation for Both HNN-2SAT Models for* $P_R$

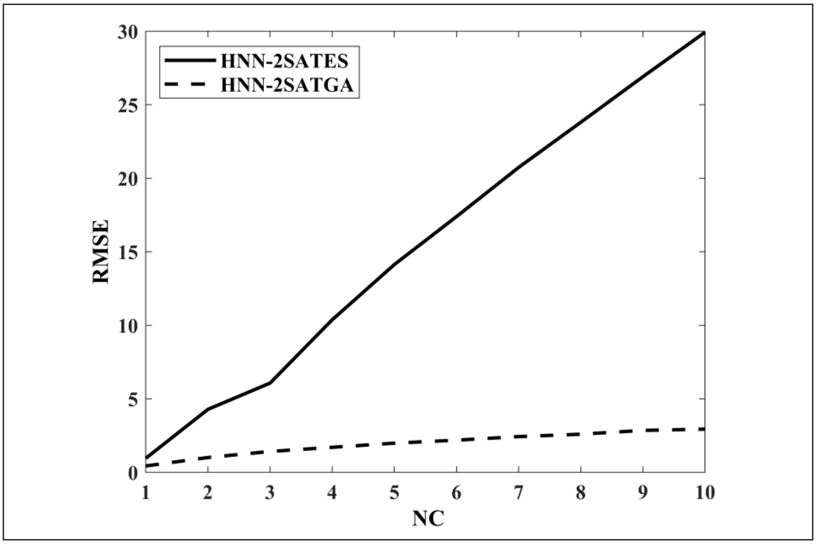
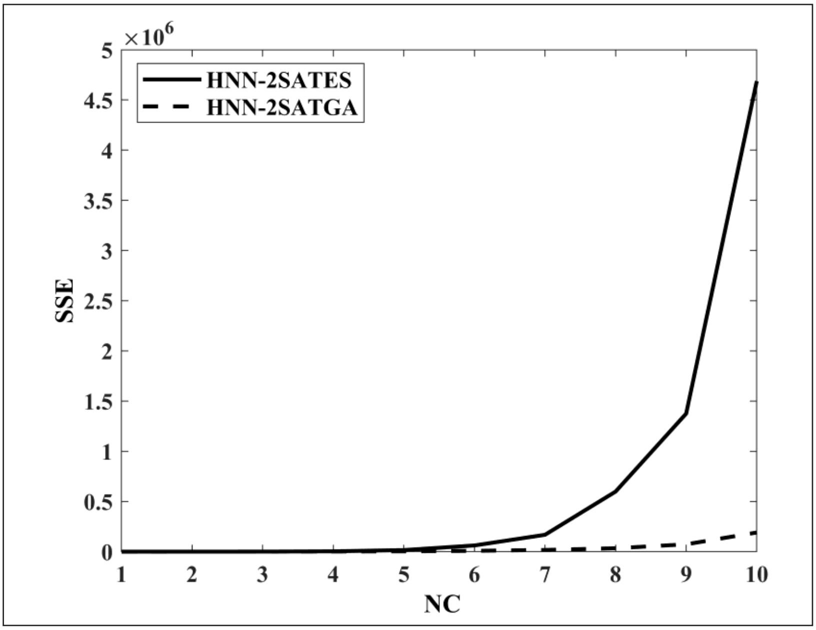


**Figure 5**

*SSE Evaluation for Both HNN-2SAT Models for* $P_R$




Figures 4 and 5 demonstrated that HNN-2SATGA outperformed HNN-2SATES by achieving lower error as the number of clauses increased. It was observed that HNN-2SATES had a higher RMSE value compared to HNN-2SATGA, which were 29.9260131 and 2.94066, respectively for $NC=10$. Note that ES was run by trial and error. Therefore, as the number of $NC$ increased, the complexity of finding the correct solutions increased, thus raising the value of error achieved by HNN-2SATES. However, there was a slight gap of SSE value from Figure 5 for both models from $NC=1$ until $NC=5$. This finding showed that, for a lower $NC$ value, HNN-2SATES managed to achieve a lower SSE value.

In the first simulation ($P_{PV}$), the induced logic obtained by HNN-2SATGA is in Equation 15:

$$P_{PV}=\left(\neg I\vee ID\right)\wedge\left(PR\vee\neg PRD\right)\wedge\left(BR\vee SD\right) \quad (15)$$

The relationship between these attributes is useful for predicting the number of page values. The number of page values can be used to identify whether visitors end up making a purchase. Based on Equation 15, when informational duration, product-related, bounce rate, and special day increased and informational and product-related decreased, the page value would increase. Additionally, it has become an important task to design an efficient platform that attracts as well as retains web users, as stated by Vijayasarathy (2004). $P_{PV}$ specified that the number of pages visited by a visitor about shopping sites, communication, and address information on searching websites decreased subsequently, indicating this website is a perfect website full of information and has fewer pages, which is more effective. The willingness of the customer to go back and extend the duration of each visit to a website is employed to effectively determine the stickiness (Pappas et al., 2017). As mentioned previously, informational duration, or in other words total quantity of time spent by a visitor on informational duration, is very vital in considering page value. According to Chen (2016), the most significant factors affecting online shopping purchasing intent could be a product (quality, price, and characteristics). From the best induced logic, $P_{PV}$ predicted that product-related was very essential. When product-related increased, the page value also increased. The value of the bounce rate feature for a webpage referred to the percentage of the visitor who from that page entered the site and then exited ("bounced") during that session without causing any other request to the analytics server. Hussein (2016) pinpointed that the higher awareness of the individual toward a specific event, the higher their desire to revisit it. From the best induced logic, $P_{PV}$, the special day was one of the important attributes of online purchase intention. On the other hand, the induced logic obtained by HNN-2SATGA in the second simulation for $P_R$ is in Equation 16:

$$P_R = (\neg VT \vee W) \wedge (I \vee \neg ID) \wedge (PR \vee PD) \tag{16}$$

Based on Equation 16, online shopping purchasing intention was influenced by weekend, information, product-related, and product duration. Visitor type was also another influence as they tended to make purchases during online shopping, particularly new visitors. Another factor influencing intention was the awareness of individuals regarding the existence of the visitors' revisit event, as stated by Hussein (2016). From the best induced logic, the time visitors visit online shopping websites was during the weekend. The excess time from online shopping turns into a weekend activity for relaxing, entertainment, and socializing activities (Türk, 2019). According to Chen (2016), numerous information and descriptions of high-quality products must be provided to customers to help them make a well-informed decision. Information is seen as a vital attribute in online purchasing intention. Based on $P_R$, the total quantity of time (in seconds) spent by visitors on informational pages had decreased. This observation was due to visitors' focus on products compared to informational shopping websites. Moreover, promotions and product quality aspects are the major factors in purchasing decisions in many developed countries, such as Saudi Arabia (Al Hamli & Sobaih, 2023). The notable reason for e-commerce success is the ability of the website to retain online customers and extend their duration of stay (Pappas et al., 2017). If the number of pages visited by the visitor on product-related pages and the total quantity of time (in seconds) spent by the customer on the respective product-related pages increase, the possibility of the visitor finalizing the transaction is higher.

In conclusion, this work implemented 2SATRA to extract the online shoppers' purchasing intention dataset. The extracted information or knowledge was used to identify online consumers' behavior. The purpose of this dataset was to decide the end behavior of visitors by capitalizing on several attributes. By using 2SATRA, the optimal induced logical rule obtained by the completion of this simulation successfully described the behavior of online shoppers based on the data. This finding can be utilized to explain to non-practitioners, such as industrial players.

## CONCLUSION

In conclusion, the three primary objectives stated were accomplished throughout this work. The first objective was to propose $P_{2SAT}$ in HNN. 2SAT logic was integrated with HNN and represented in terms of HNN-2SAT. The second objective was to incorporate $P_{2SAT}$ with hybrid GA in HNN. During the learning phase, HNN-2SATGA was a superior metaheuristic that helped to improve the solution of the proposed model. Next, the efficiency of the proposed model in doing $P_{2SAT}$ was assessed by using performance evaluation metrics, such as RMSE and SSE. The effective performance of 2SATRA by using

HNN-2SATGA successfully outperformed the benchmark model HNN-2SATES in doing logic mining. Based on the findings, online stores were able to improve sales based on the best induced logic obtained. It is recommended for online stores to update product information with interesting images to capture shoppers' attention. Other than that, online stores can create special promotions, such as happy hour sales for customers. In terms of the model architecture, this work can be further extended by using different variants of ANN, ranging from the radial basis function neural network, ensemble feed-forward neural network, and support vector machine (Jimoh et al., 2022) and convolutional neural network (Ong et al., 2022). Other structures of SAT can also be implemented in the selected ANN as a symbolic logical rule, such as Boolean satisfiability with majority logic (Amarú et al., 2015). Additionally, this research can apply a new metaheuristic to enhance the learning phase capability of HNN by using the sine cosine algorithm (Yang et al., 2023) or particle swarm optimization algorithm (Minh et al., 2023).

## ACKNOWLEDGMENT

This work was funded by the Short Term Grant, Universiti Sains Malaysia with the grant number 304/PMATHS/6315390.

## REFERENCES

Abdullah, W. A. T. W. (1992). Logic programming on a neural network. *International Journal of Intelligent Systems*, *7*, 513–519. https://doi.org/10.1002/int.4550070604

Al Hamli, S. S., & Sobaih, A. E. E. (2023). Factors influencing consumer behavior towards online shopping in Saudi Arabia amid COVID-19: Implications for e-businesses post pandemic. *Journal of Risk and Financial Management*, *16*(1), 36. https://doi.org/10.3390/jrfm16010036

Alway, A., Zamri, N. E., Kasihmuddin, M. S. M., Mansor, M. A., & Sathasivam, S. (2020). Palm oil trend analysis via logic mining with discrete Hopfield neural network. *Pertanika Journal of Science & Technology*, *28*(3), 967–981.

Amarú, L., Gaillardon, P. E., Chattopadhyay, A., & De Micheli, G. (2015). A sound and complete axiomatization of majority-$ n $ logic. *IEEE Transactions on Computers*, *65*(9), 2889–2895. https://doi.org/10.1109/TC.2015.2506566

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., Diesmann, M., Morrison, A., Goodman, P. H., Harris, F. C., Jr, Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., … Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience*, *23*(3), 349–398. https://doi.org/10.1007/s10827-007-0038-6

Chen, R. Y. (2016). Online marketing and innovative business services: Cloud business and cases of the internet of things. Gotop Information.

Coppola, D. (2020, September 17). *E-Commerce worldwide - Statistics & facts*. Statista. https://www.statista.com/topics/871/online-shopping/#topicHeader wrapper.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Hireche, C., Drias, H., & Moulai, H. (2020). Grid based clustering for satisfiability solving. *Applied Soft Computing*, *88*, 106069. https://doi.org/10.1016/j.asoc.2020.106069

Hopfield, J. J., & Tank, D. (1985). "Neural" computation of decisions in optimization problems. *Biological Cybernetics, 52*(3), 141–152. https://doi.org/10.1007/BF00339943

Hoque, M. T., Chetty, M., & Dooley, L. S. (2007). Generalized schemata theorem incorporating twin removal for protein structure prediction. In *IAPR International Workshop on Pattern Recognition in Bioinformatics* (pp. 84–97). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-75286-8_9

Hussein, A. S. (2016). How event awareness, event quality and event image creates visitor revisit intention?: A lesson from car free day event. *Procedia Economics and Finance*, *35*, 396–400. https://doi.org/10.1016/S2212-5671(16)00049-6

Hutter, F., Xu, L., Hoos, H. H., & Leyton-Brown, K. (2014). Algorithm runtime prediction: Methods & evaluation. *Artificial Intelligence*, *206*, 79–111. https://doi.org/10.1016/j.artint.2013.10.003

Jimoh, R. G., Abisoye, O. A., & Uthman, M. M. B. (2022). Ensemble feed-forward neural network and support vector machine for prediction of multiclass malaria infection. *Journal of*

*Information and Communication Technology*, *21*(1), 117–148. https://doi.org/10.32890/jict2022.21.1.6

Kasihmuddin, M. S. M., Mansor, M. A., & Sathasivam, S. (2017). Hybrid genetic algorithm in the hopfield network for logic satisfiability problem. *Pertanika Journal of Science & Technology*, *25*(1), 139–152. https://doi.org/10.1063/1.4995911

Kowalski, R. (1979). Algorithm=logic+control. *Communications of the ACM*, *22*(7), 424–436. https://doi.org/10.1145/359131.359136

Minh, H. L., Khatir, S., Rao, R. V., Abdel Wahab, M., & Cuong-Le, T. (2023). A variable velocity strategy particle swarm optimization algorithm (VVS-PSO) for damage assessment in structures. *Engineering with Computers*, *39*(2), 1055–1084. https://doi.org/10.1007/s00366-021-01451-2

Ong, J. H., Ong, P., & Lee, W. K. (2022). Image-based oil palm leaves disease detection using convolutional neural network. *Journal of Information and Communication Technology*, *21*(3), 383–410. https://doi.org/10.32890/jict2022.21.3.4

osCommerce Ltd. (2002, June 29). *The TNS interactive - Global eCommerce report 2002*. osCommerce. https://www.oscommerce.com/blog/the-tns-interactive-global-ecommerce-report-2002

Pappas, I. O., Kourouthanassis, P. E., Giannakos, M. N., & Lekakos, G. (2017). The interplay of online shopping motivations and experiential factors on personalized e-commerce: A complexity theory approach. *Telematics and Informatics*, *34*(5), 730–742. https://doi.org/10.1016/j.tele.2016.08.021

Sathasivam, S., & Abdullah, W. A. T. W. (2011). Logic mining in neural network: Reverse analysis method. *Computing*, *91*(2), 119–133. https://doi.org/10.1007/s00607-010-0117-9

Türk, E. (2019). Factors affecting online shopping decision: Customers in Turkey. *Journal of International Trade, Logistics and Law*, *5*(1), 35–43.

Vijayasarathy, L. R. (2004). Predicting consumer intentions to use on-line shopping: The case for an augmented technology acceptance model. *Information & Management*, *41*(6), 747–762. https://doi.org/10.1016/j.im.2003.08.011

Yang, X., Wang, R., Zhao, D., Yu, F., Huang, C., Heidari, A. A., ... & Chen, H. (2023). An adaptive quadratic interpolation and rounding mechanism sine cosine algorithm with application to constrained engineering optimization problems. *Expert Systems with Applications*, *213*, 119041. https://doi.org/10.1016/j.eswa.2022.119041

Zamri, N. E., Mansor, M. A., Kasihmuddin, M. S. M., Alway, A., Jamaludin, S. Z. M., & Alzaeemi, S. A. (2020). Amazon employees resources access data extraction via clonal selection algorithm and logic mining approach. *Entropy*, *22*(22), 596. https://doi.org/10.3390/e22060596

Zhou, C., Zeng, X., Jiang, H., & Han, L. (2015). A generalized bipolar auto associative memory model based on discrete recurrent neural networks. *Neurocomputing*, *162*, 201–208. https://doi.org/10.1016/j.neucom.2015.03.052