

## Implementation of C4.5 Method and Artificial Neural Networks to Predict Sales

Elsa Trisna Putri

<sup>1,2,3</sup>University of Putra Indonesia YPTK Padang

[elsatrisnaputri01@gmail.com](mailto:elsatrisnaputri01@gmail.com)

### Abstract

Technology and communication is currently growing along with the increasing needs of each individual in various fields such as: business, education, agriculture, health and technology. With the development of technology today, everyone can communicate and obtain and convey various information needed anytime and anywhere quickly, accurately and dynamically. Sales is an activity of buying and selling products or goods carried out between sellers and buyers, can interact within the same scope or online by using legal payment transactions. Building materials are materials used to design buildings such as houses, mosques, schools etc. For buildings, many natural materials are used such as: clay, sand, wood or bamboo, stone and others. In this study, the implementation of the c45 method and an artificial neural network to predict sales for the next year. The sale of building materials at the prayer shop, whose sales are not yet computerized by designing a computerized sales system. Toko Doa Mama is a building shop that sells various building materials such as: sand, paint, wood, or bamboo, saws, cement, roof tiles, gravel, nails, bricks, and others. But sales and marketing are still not computerized which results in frequent errors in calculating sales transactions, data collection of incoming goods and outgoing goods which are still in the form of archives, resulting in accumulation and lack of data security. Therefore, we need a computerized where the computer can help a job to be more effective and efficient.

Keywords: sales, Building Materials, Data Mining C45, Artificial Neural Networks, Bagpropagation.

### 1. Introduction

Technology and communication are currently growing in line with the increasing needs of individuals in various fields such as: business, education, agriculture, health and technology. With the development of technology today, everyone can communicate and obtain and convey various information needed anytime and anywhere quickly, accurately and dynamically.

Along with the development of science and technology which is very fast lately, technology is needed in all fields. One of them is in the field of economics regarding sales, sales are one of the benchmarks for success in a trading business [1]. Sales is an activity of buying and selling products or goods carried out between sellers and buyers, can interact within the same scope or online by using legal payment transactions. Building materials are materials used to design buildings such as houses, mosques, schools etc. To build buildings, many natural materials are used such as: clay, sand, wood or bamboo, stone and others. Toko Doa Mama is a building shop that sells various building materials such as: sand, paint, wood or bamboo, saws, cement, tile, gravel, nails, bricks, and others. But sales and marketing are still not computerized which results in frequent errors in calculating sales transactions, data collection of incoming goods and outgoing goods which are still in

the form of archives, resulting in accumulation and data security.

Therefore, we need a computerized system where computers can help a job become more effective and efficient. Computerization is the use of computers as a tool for completing tasks as a substitute for manual completion of work. The advantage of using a computerized system is that it is more efficient when entering data and calculating total sales without having to use a calculator or stationery. So that stores can improve service to customers [2].

Sales prediction is one way to be able to compete or even increase company profits so that predictions are needed to balance the difference between current and future time of needs [3].

Data mining is the process of extracting and mining knowledge from large amounts of data, databases or other database repositories [4]. Another name for data mining is knowledge discovery in database (KDD), all of which are activities that include collecting, using historical data to find regularities, patterns or relationships with large data sets [5].

Accurate predictions allow the organization to improve market growth with higher level of revenue generation. Data mining techniques are very effective in tuning huge volume of data into useful information for cost prediction and sales forecast, it is the basic of sound budgeting [6]. In order to be competent enough and to generate higher revenue, business organizations are

constantly in search of a better model or technique for data mining and maintenance of critical data [7]. Business industry faces severe challenges to identify an accurate data mining technique and effective predication strategy [8].

One of the data mining techniques known as clustering. According to science, data mining clustering is the grouping of a number of data or objects into clusters (groups) so that each in the cluster will contain data as similar as possible and different from objects in other clusters [9]. The C4.5 algorithm is an algorithm whose use is to form a decision tree (Decision Tree). The decision tree itself is a classification and prediction method that is quite well known and used by many people. The algorithms that are often used in making decision trees are ID3, C4.5 and CART [10].

Therefore, it is necessary to design a technology that can make it easier to predict sales of building materials for the Doa Mama store and provide knowledge in dealing with it using artificial intelligence. Artificial intelligence or also called Artificial Intelligence (AI) is one part of computer science that studies how to make machines (computers) can do work as well as what humans do and can even be better than what humans do [11]. Artificial Intelligence is to find out and model human thought processes and design machines to imitate human behavior. Smart, means having knowledge and experience, reasoning, how to make decisions and take good moral actions [12]. One kind

of field of science that uses artificial intelligence is artificial neural networks, etc. [13].

An artificial neural network (ANN) is an information processing paradigm that is inspired by the biological nervous system, such as information processing in the human brain. The key element of this paradigm is the structure of information processing which consists of a large number of interconnected processing elements (neurons), working together to solve a particular problem. The way ANN works is like how humans work, namely learning by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [14].

## 2. Research Method

Research methodology is a set of steps or procedures used by a researcher or actor of a scientific discipline in solving a problem that occurs so that new knowledge will be obtained.

### 2.1 Research Framework

To clarify the stages that will be carried out in the research, a research framework is needed. The research framework is a sequence of activities to be carried out in a study. So that the steps taken by the author in this design do not deviate from the subject matter, it is easier to understand. So the author forms a research framework as shown in Figure. 1.

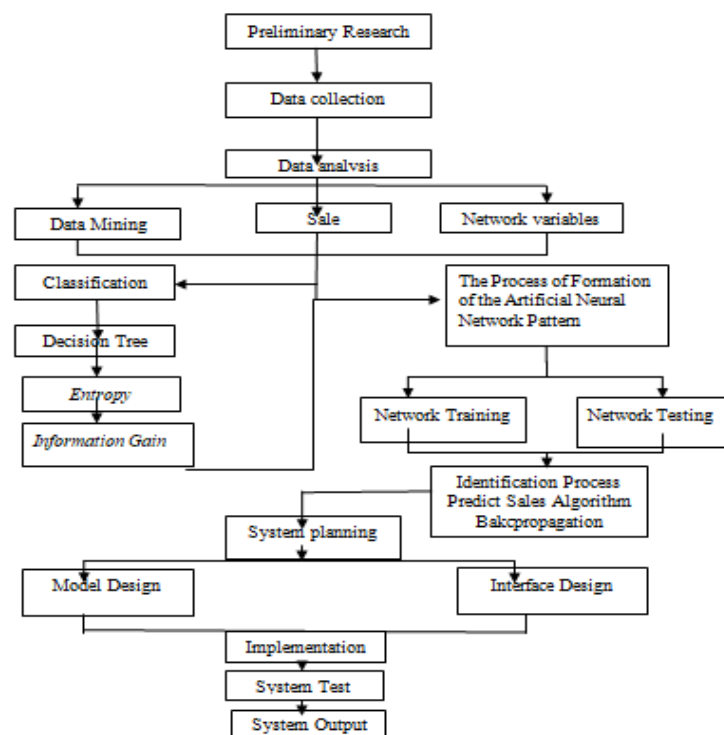


Figure. 1 Research Framework

Figure. 3 describes the stages of the research framework. In collecting data and information in writing this research, there are several ways: Field research with carried out by going directly to the research place. Library Research by studying literature through books, journals, e-books, internet or other materials related to the problems in the process of working on this research in order to obtain theoretical data and information, request reply information packets to the internet network through the proxy feature

## 2.2 Artificial intelligence

Intelligence can be defined as wisdom and ability; what is called artificial intelligence (AI) is a variety of human intelligent behaviors, such as perception, memory, emotion, judgment, reasoning, proof, recognition, understanding, communication, design, thinking, learning, forgetting, creating, and so on, which can realized artificially by machine, system, or network [15].

## 2.3 Backpropagation method

Backpropagation is one of the algorithms that is often used in solving complex problems. Training is carried out repeatedly so that a network is generated that gives the correct response to all of its inputs. The backpropagation network consists of three layers or more processing units, the input layer consists of input variables for nerve cell units, hidden layers and outputs. As shown in Figure 2, the three layers are fully connected [16].

## 2.4 Data Mining

Data mining is the process of discovering interesting knowledge, such as associations, patterns, changes, significant structures and anomalies, from large amounts of data stored in databases or data warehouses or other information repositories [17].

## 2.5 C4.5 Algorithm

C4.5 is a algorithm used to generate a decision trees. It is an extension of the ID3 algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier. The C4.5 algorithm made a number of changes to improve ID3 algorithm[18]. There are several stages to make a decision tree with C4.5 Algorithm [19]:

1. Preparing training data, ordinary training data is retrieved from historical data that have occurred before and have been grouped previously to certain classes.

2. Specifying the root in the tree root to be extracted from the selected attribute, by calculating the gain value of each attribute, the highest gain value which will be the first root, before calculating the gain value of the attribute, compute the entropy value used by the Formula.1&2.

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Description:

S: case set  
A: attribute  
n: the number of partitions S  
pi: the proportion of Si to S

Formula For Finding Gain

Gain (S, A) =

$$Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Description :

S: case collection  
A: attribute  
N: number of cases attribute A  
|Si| : number of cases on partition i  
|S|: number of cases in S

## 3. Result and Discussion

### 3.1. Analysis

System analysis is the initial stage in the design and development of a system to be designed, because this stage will measure and evaluate the performance of the designed system. In conducting a system analysis, you must first know and understand the system, to analyze the system data from the system is needed to be analyzed. Data analysis is the stage to analyze the data needed for the design of the system to be made, in this case the author takes data through the literature related to the research theme, to find information, compile theories related to the discussion so that there is a combination complex one with the other.

### 3.2. Artificial Neural Netrowk

ANN were found to be more efficient and more accurate than other classification techniques [20]. Classification by a neural network is done in two separate phases. First, the network is trained on a dataset. Then the weights of the connections between neurons are fixed so the network is validated to determine the classifications of a new dataset [21] . In this paper, we used about 70% of the total sample data for network training, and 30% for network validation. While many models of ANNs have been proposed, the

feedforward neural networks (FNNs) are the most common and broadly used in many applications. Mathematically, the problem of training an FNN is the minimization of an error function  $E$ ; In another word, to find a minimizer  $w = (w_1, w_2, \dots, w_n)$  such that  $w = \min E(w)$ , where  $E$  is the batch error computed by the sum of square differences over all examples of the training dataset [22].

### 3.3. Network architecture

Previously, the data was processed using an artificial neural network, first, the data was transformed. The results of all the data transformations above can be seen in the following Table.1.

| X1   | X2    | X3          | T     |
|------|-------|-------------|-------|
| 0,25 | 0,74  | 0,318181818 | 0,74  |
| 0,9  | 0,26  | 0,390909091 | 0,228 |
| 0,4  | 0,42  | 0,390909091 | 0,26  |
| 0,2  | 0,836 | 0,5         | 0,74  |
| 0,32 | 0,26  | 0,354545455 | 0,196 |
| 0,35 | 0,9   | 0,463636364 | 0,9   |
| 0,25 | 0,74  | 0,1         | 0,644 |
| 0,3  | 0,74  | 0,1         | 0,74  |
| 0,4  | 0,58  | 0,9         | 0,42  |
| 0,25 | 0,74  | 0,1         | 0,74  |
| 0,65 | 0,164 | 0,572727273 | 0,1   |

|      |       |             |       |
|------|-------|-------------|-------|
| 0,7  | 0,26  | 0,463636364 | 0,196 |
| 0,1  | 0,1   | 0,114545455 | 0,1   |
| 0,65 | 0,164 | 0,5         | 0,1   |

The design and the results of data processing using an artificial neural network with the backpropagation method, it can be seen the relationship that the Artificial Neural Network can be used to predict sales of building materials based on the attributes of Stock, Amount Sold, and Price as input and the next Amount Sold value as its output.

### 3.4. C.45 . Algorithm Calculation Method

C4.5 Algorithm is an algorithm used to form decision tree, decision tree is a methodology and prediction that is very strongest and famous method of decision tree transforming very big fact into decision tree representing rule, rule can be easily understood with natural language, can also be expressed. Rules can be easily understood with natural language, and they can also be expressed in the form of database languages such as Structured Query language to search records in certain categories. The decision tree formed up to this stage is shown in the Figure.2.

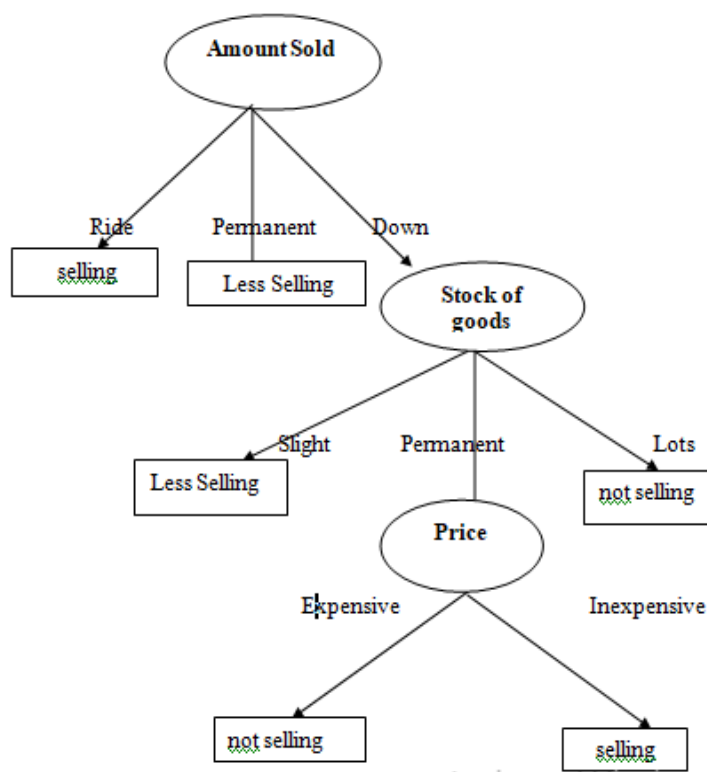


Figure. 2 Node Decision Tree 1.1

Figure. 4. Explaining the results of the decision tree node 1.1 with the amount sold being the highest gain result, then stock and price.

The implementation of this research is to help the Doa Mama building shop to see the selling items that are selling well, not selling well, and not selling well. And can predict the seller against the amount sold in the coming year.

### 3.5 Application Testing

Testing is a stage of verification and evaluation of the results of the implementation. Then, this testing phase also aims to find errors that may occur in the system and to find out whether the application that has been

made is in accordance with the goals that have been set previously. After the application testing is complete, the process that will be carried out next is the interface testing which is the display of the system created in the Figure.3,4,5.

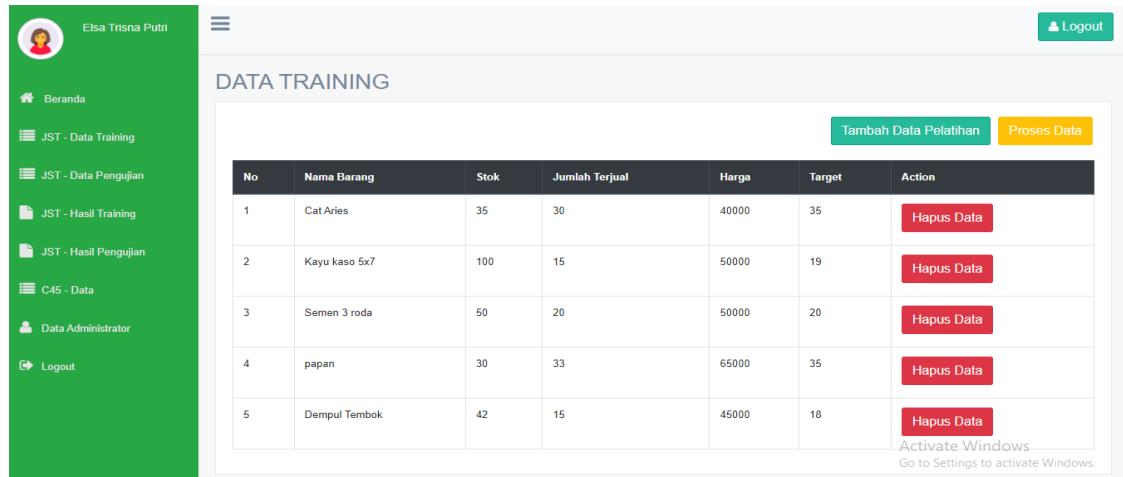


Figure. 3 Training Data Input Page Display

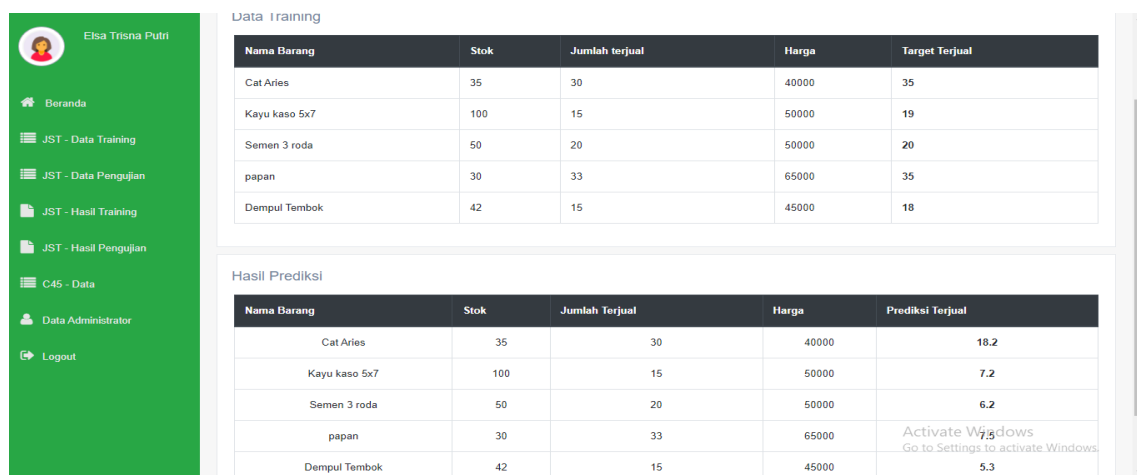


Figure. 4 Display of Training Data Prediction Results Pages

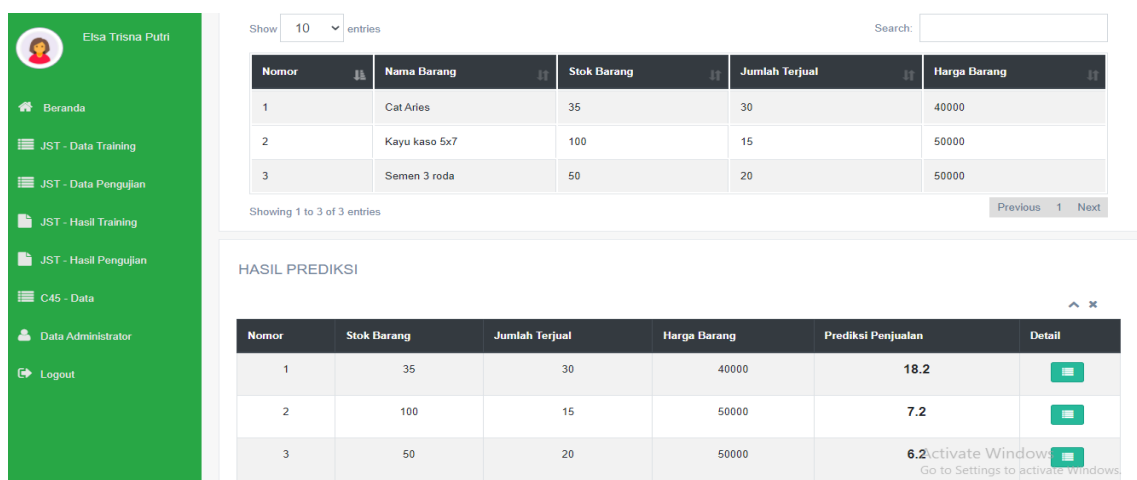


Figure. 5 Prediction results page views

The image above is a page view of the sales prediction at the Doa Mama store, after inputting and processing data. On this page, data and predictions of sales of building materials are displayed.

#### 4. Conclusion

Based on the discussion and description of the problem from the previous chapters, and based on the analysis of existing data, and testing of the system created. So the authors draw the following conclusions: The system that has been designed computerized for the prediction process using the C45 method and the Artificial Neural Network is expected to be able to make a prediction process on the sale of building materials at the Doa Mama Shop. With a system designed to be web-based, it can carry out a predictive process so that it can make it easier to see sales transactions in the next period.

#### References

- [1] A, Hendini. (2016). PEMODELAN UML SISTEM INFORMASI MONITORING PENJUALAN DAN STOK BARANG. *IV*(2), 107–116.
- [2] Dristyan, F., Informasi, S., Buatan, K., Tiruan, S., & Fuzzy, L. (2018). *PREDIKSI JUMLAH PENJUALAN KREDIT SEPEDA MOTOR*. 9986(September)
- [3] Elisa, E., Azwanti, N., Komputer, T., & Batam, U. P. (2019). DOI : <http://dx.doi.org/10.31869/rjt.v2i2.1430>. 2(2).
- [4] Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36. <https://doi.org/10.15575/join.v2i1.71>
- [5] Engineering, I., Hadi, F., Diana, Y., & Komputer, F. I. (2020). *PENGKLUSTERAN PENJUALAN BAHAN BANGUNAN MENGGUNAKAN ALGORITMA K-MEANS*. 4(1), 22–28.
- [6] Fandi, F., & Imaniawan, D. (2019). *Rancang Bangun Aplikasi Penjualan Material Bangunan*. 5(2), 101–112
- [7] Giawa, D. I. (2017). *METODE BACKPROPAGATION ( Studi Kasus : CV . Bryan Ponsel )*. 12, 162–167.
- [8] Ginting, S., & Untuk, C. (2020). Implementasi algoritma c4.5 untuk memprediksi keterlambatan pembayaran sumbangan pembangunan pendidikan sekolah menggunakan python. 10, 36–44.
- [9] Hikma Widayu. (2017). *DATA MINING UNTUK MEMPREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN*. 1(2), 32–37.
- [10] Mardi, Y., Gajah, J., No, M., & Barat, S. (n.d.). *Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatik*
- [11] Penyakit, D., Berdasarkan, M., & Usia, R. (2019). *PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN*. 2(1).
- [12] Pratama, A. S., Darwati, I., & Informatika, M. (2017). *Perancangan Program Berbasis Web Pada Penjualan Bahan Bangunan*. *III*(2), 25–30
- [13] Ramadhan, R., Dyah, S., Wardani, K., Amrozi, Y., & Spiral, M. (2020). *Quo vadis pengembangan rekayasa perangkat lunak*. 3(2).
- [14] Sovia, R., Yanto, M., & Melati, P. (2020). *Prediksi Jumlah Kunjungan Wisata Mancanegara Dengan Algoritma Backpropagation*. 4(April), 355–362. <https://doi.org/10.30865/mib.v4i2.2048>
- [15] Syafitri, Y., Informatika, J. M., Diagram, A., & Diagram, S. (2016). *PEMODELAN PERANGKAT LUNAK BERBASIS UML UNTUK PENGEMBANGAN SISTEM PEMASARAN AKBAR ENTERTAINMENT*. 12(1), 31–39.
- [16] Yanto, M., Praja, E., Mandala, W., & Putri, D. E. (2018). *Peramalan Penjualan Pada Toko Retail Menggunakan Algoritma Backpropagation Neural Network*. 2(3), 110
- [17] Cheriyan, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018, August). Intelligent Sales Prediction Using Machine Learning Techniques. In *2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE)* (pp. 53–58). IEEE.
- [18] Katrina, W., Damanik, H. J., Parhusip, F., Hartama, D., Windarto, A. P., & Wanto, A. (2019, August). C. 45 Classification Rules Model for Determining Students Level of Understanding of the Subject. In *Journal of Physics: Conference Series* (Vol. 1255, No. 1, p. 012005). IOP Publishing.
- [19] Nasser, I. M., Al-Shawwa, M. O., & Abu-Naser, S. S. (2019). A Proposed Artificial Neural Network for Predicting Movies Rates Category.
- [20] Sovia, R., Yanto, M., Budiman, A., Mayola, L., & Saputra, D. (2019). Backpropagation neural network prediction for cryptocurrency bitcoin prices. *Journal of Physics: Conference Series*, 1339(1), 012060. doi:10.1088/1742-6596/1339/1/012060.
- [21] Sugiyarti, E., Jasmi, K. A., Basiron, B., Huda, M., Shankar, K., & Maseleno, A. (2018). Decision support system of scholarship grantee selection using data mining. *International Journal of Pure and Applied Mathematics*, 119(15), 2239–2249