Jurnal Mantik, 5 (4) (2022) 2694-2701

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Application of Data Mining to Predict Work Accident Rates using Rapid Miner

Sapna Indah Br. Ginting[1], Muhammad Iqbal[2]

[1,2]Computer system, Faculty of Sains and Technology, Universitas Pembangunan Panca Budi , Jl. Gatot Subroto KM. 4,5, Medan 20122, Indonesia

Email: sapnagtgjawak@gmail.com, muhammadiqbal@dosen.pancabudi.ac.id

ARTICLEINFO

Article history:

Received: Jan 15, 2022
Revised: Jan 30, 2022
Accepted: Feb 19, 2022

Keywords:

Work accidents,
Data mining,
Manufacturing,
Prediction,
Rapid Miner

ABSTRACT

Working safety is strongly linked to machines, aircraft, and work tools in the workplace runway and its surroundings, as well as how to work. The corporation must provide this protection since it is a human right. According to ILO (International Labor Organization) estimates, 2 million people die each year as a result of work-related issues around the world. A total of 354,000 persons died as a result of these accidents. The rate of fatal accidents in underdeveloped countries is four times that in developed countries. The agriculture, construction, mining, forestry, and fisheries industries all have hazardous jobs that account for the majority of accidents. This research is descriptive in nature, examining work accidents that occur based on secondary data and making predictions to estimate the amount of work accidents using a Data Mining approach utilizing Rapid Miner to determine the level of work accidents. Rapid miner is a data mining processing software that includes tools for creating decision trees and a data mining engine that may be used in its own products. The data utilized was collected from the Industrial Safety and Health Analytics Database as secondary data. The database's content consists primarily of accident records from 12 distinct factories in three different nations, with each row representing a 439-data-row accident incident. According to the findings, 11 of the 12 factories have an accident rate of level I; the third factory (level 03) has an accident rate of level IV on the Risco Critico power lock; and factory 11 (local 11) does not have a crash lift.

1. Introduction

Occupational Health and Safety (K3) is a concept and endeavor aimed at ensuring the integrity and perfection of both the physical and spiritual workforce, as well as humans in general, as well as the outcomes of work and culture in order to achieve a just and prosperous society[1]. Workplace safety refers to the attempt to avoid workplace accidents. Workplace safety entails identifying and exposing operational flaws that allow accidents to occur. According to the International Labor Organization (ILO), 2 million people die each year as a result of work-related difficulties around the world. A total of 354,000 persons died as a result of these accidents. In addition, 270 million workers are involved in work-related accidents each year, and 160 million are impacted by work-related diseases. The costs of dealing with this work-related danger are significant. The International Labour Organization (ILO) estimates that yearly losses due to accidents and occupational diseases total more than US$1.25 trillion, or 4% of GDP (GDP).[2]

The rate of fatal accidents in underdeveloped countries is four times that in developed countries. Agriculture, construction, mining, forestry, and fisheries are some of the most dangerous jobs in developing countries, accounting for the bulk of occupational accidents and diseases. These five industries are frequently found in those with the highest risk of occupational hazards. [3]

The mining industry, for example, oil and gas mining, carries a significant level of risk. Because of the significant number of accidents that occur in the oil and gas industry, such as fires, explosions, pollution, and

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

others, the oil and gas industry has a high risk of workplace accidents [4]. This research uses the Data Mining technique with Rapid Miner to investigate work accidents that occur based on secondary data and make predictions to determine the accident rate. [5]

2. Method

Occupational health and safety, from a scientific standpoint, is science and its application in an endeavor to prevent workplace accidents and occupational diseases [6]. Occupational diseases such as loss of life or limb, injuries induced by repetitive activities, back pain, carpal tunnel syndrome, cardiovascular diseases, various types of cancer such as lung cancer and leukemia, emphysema, and arthritis are examples of physiological-financial problems. White lung illness, brown lung disease, black lung disease, infertility, central nervous system damage, and chronic bronchitis are all known to be caused by an uncomfortable work environment. Physiological issues resulting from workplace stress and a poor quality of life at work. Dissatisfaction, apathy, disengagement, self-assertion, narrow view, forgetfulness, uncertainty about responsibilities and obligations, suspicion of others, indecisiveness in making decisions, inattention, impatience, continually procrastinating on work, and a tendency to break up easily are some of these symptoms [7] . I'm hoping for little more than a blip on the radar. The following are some of the goals of workplace safety and health:

a. Preventing workplace mishaps
b. Prevent sickness from spreading as a result of your job.
c. Avoid or decrease death
d. Avoid/reduce long-term impairment
e. Securing supplies, building construction, use, and maintenance, work tools, machines, and installations, among other things.
f. Improve work productivity without putting a strain on the personnel and ensuring their long-term viability.
g. Avoid squandering labor, capital, tools, and other manufacturing assets.
h. Providing a healthy, clean, comfortable, and safe work environment that inspires excitement and passion.
i. Streamlining, expanding, and safeguarding industrial expansion and output

The data utilized was collected from the Industrial Safety and Health Analytics Database as secondary data. The goal of this database is to give a database of real-world manufacturing plants from one of Brazil's and the world's largest industries [8]. The main reason IHM Stefanini is disclosing this industrial work accident data is that the company needs to know why employees are still suffering from multiple injuries/accidents at the factory. Some people even died. This database's content consists primarily of accident reports from 12 distinct factories in three different nations, with each row representing an accident event. There are 439 data and column entries in this file. Table 1 shows how to fill out the form.

TABLE 1
DATA SAFETY

Data (1)	*Countries (2)*	*Local (3)*	*Industry Sector (4)*	*Accident Level (5)*	*Potential Accident Level (6)*	*Genre (7)*	*Employee ou Terceiro (8)*	*Risco Critico (9)*
01/01/2016 00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party	Pressed
02/01/2016 00:00	Country_02	Local_02	Mining	I	IV	Male	Employee	Pressurized Systems

The following are the contents of the column description:

a. Information on the time and date of the accident, which spans the years 2016 and 2017.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

b. Countries, the location of the accident; the information is anonymized. It is made up of three countries.
c. The data is anonymised at the local level, in the city where the factory is located. There are 12 factories in this complex.
d. Industry Sector: What industry does the factory fall under?
e. The number of accidents (Accident level) Record the severity of the accident on a scale of one to six (I means not severe but VI means very severe). The severity level is shown in Table 2.
f. Accidental Risk Level, The database also captures the severity of the accident based on the accident rate (due to other factors involved in the accident).
g. Gender, whether it's a man or a woman.If the injured person is an employee or a third party,
h. Employee ou Terceiro (Employees or Third Parties) is used. Some descriptions of the hazards involved in an accident.
i. Risco Critico (Critical Risk)

TABLE 1
DATA SAFETY

Severity	Description
Level I	An error occurs, but it is harmless
Level II	Needs increased oversight, no changes needed
Level III	Improve control, temporary changes in vital parts, but not harmful
Level IV	Improve supervision, complete changes in vital parts, maintenance required
Level V	Improve supervision and care, Long-term change, pose a danger of dying
Level VI	Cause death

Data mining is a method of automatically analyzing and extracting knowledge using one or more computer learning algorithms (machine learning). The use of scientific methods to data mining is known as Knowledge Discovery in Databases (KDD) [9].

Estimation (Linear Regression), Forecasting (Neural Network), Classification (C4.5 Algorithm), Clustering (K-Means), and Association (FP-Growth) are the different types of data mining methodologies, according to Larose [10]. Figure 1 depicts data mining's primary function. The steps of data mining are broken down into six categories:

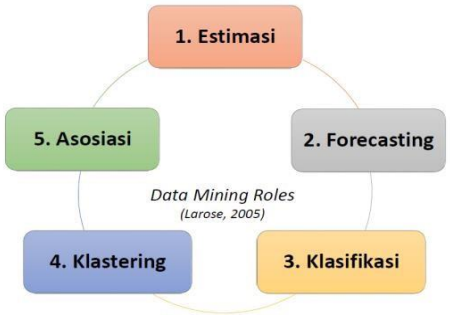

Figure 1. The Main Role of Data Mining

a. Cleansing the data Cleaning the data that is the subject of KDD is required before the data mining process can begin. The cleaning process entails deleting duplicate data, checking for inconsistencies, and repairing data problems such as typographical errors, among other things. Enrichment, or the process of "enriching" data or other important and necessary information for KDD, such as external data or information, is also carried out.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

b. Integration of data (data integration) The merging of data from several databases into a single new database is known as data integration. Frequently, the data required for data mining originates from multiple databases or text files, rather than from a single database. Data integration is done on attributes that identify distinct entities, such as name attributes, product kinds, and customer numbers. Data integration must be done carefully since errors in data integration can lead to skewed results and potentially lead to future action being taken in the wrong direction. If data integration based on product type results in the combination of products from various categories, for example, a correlation between products will be discovered that does not exist
c. Data selection (Data selection) Because not all of the data in the database is used, only the data that is acceptable for analysis will be extracted from it. In the case of market basket analysis, for example, a case that analyses people's buying tendencies does not require the client's name, only the customer id.
d. Data transformation (Data Transformation) In data mining, data is translated or merged into a format that can be processed. Some data mining techniques necessitate the use of specific data formats before they can be used. Some standard methods, such as association analysis and clustering, for example, can only handle categorical data. As a result, data in the form of continuous numeric numbers must be partitioned into many intervals. Data transformation is a term used to describe this process. 5. The mining process is particularly important when the procedure is used to uncover hidden and useful information in data.
e. Pattern analysis To find intriguing patterns in the knowledge based data that have been discovered. The findings of the data mining technique are examined in this stage in the form of typical patterns and predictive models to see if the present hypothesis is correct.
f. Knowledge presentation (knowledge presentation) is the presentation and representation of the knowledge gained from the analysis. There are instances when you'll need to work with people who aren't familiar with data mining. As a result, presenting data mining results in the form of knowledge that everyone can understand is an important step in the data mining process. In this presentation, visualization can also be used to help communicate data mining results [11]. nformation on the ways for acquiring user-acquired knowledge How to formulate decisions or actions based on the analytical results gained is the final stage of the data mining process. There are instances when you'll need to work with people who aren't familiar with data mining. As a result, presenting data mining results in the form of knowledge that everyone can understand is an important step in the data mining process. In this presentation, visualization can also be used to help communicate data mining results [12].

A decision tree , often known as a decision tree, is a method that humans have evolved to locate and make solutions for specific situations by taking into account many elements connected to the scope of the problem.[13]

A decision tree is a tree-like flowchart structure in which each internal node represents an attribute test, each branch indicates the test result, and the leaf node represents a class or class distribution. From the base of the decision tree to the leaf node that contains the forecast [14], the path is traced.

Rapid miner is a data mining processing package that includes decision tree capabilities. When compared to manually creating a decision tree, this will make it much easier for us to use Rapid Miner to create one. Open-source software (open source). Rapid miner is a data mining, text mining, and predictive analysis service. Rapid miners deliver insights to users through a range of descriptive and predictive methodologies, allowing them to make the best decisions possible. Rapid miner includes around 500 data mining operators, which include input, output, data preparation, and visualization operators. Rapid miner is a data mining engine that may be integrated into its own products as well as stand-alone tools for data analysis. Rapid miner is a data mining processing software that includes tools for creating decision trees and acting as a machine to evaluate data. Data mining that can be included into the company's own goods The data utilized was collected from the Industrial Safety and Health Analytics Database as secondary data [15].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

3. Result and Discussion

The database built by IHM Stefanini consists primarily of accident reports from 12 different factories in three different countries, with each row representing a 439-data-row accident incident. As indicated in Figure 2, the data mining process begins with the data set, data mining methods, expertise, and evaluation.

.

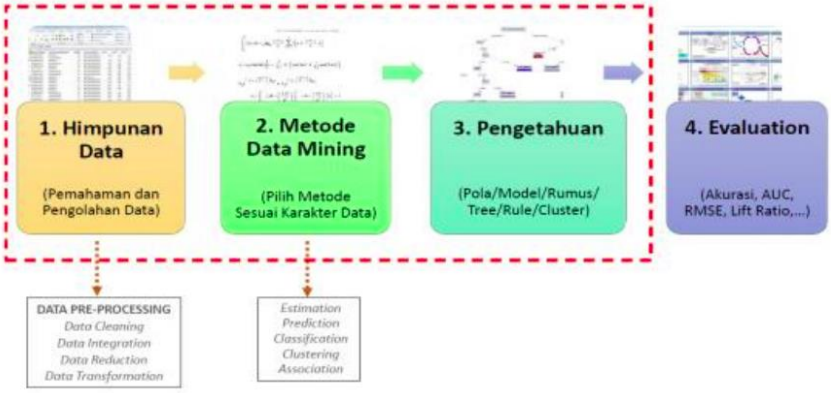

Figure 1. Data Mining Process

On the official website, www.rapid-i.com, you may get the rapid miner program installation. The Rapid Miner main screen looks like this in Figure 3. Figure 4 depicts the results of the database retrieval. The secondary data for this study was taken from the Industrial Safety and Health Analytics Database, which contains 439 records.

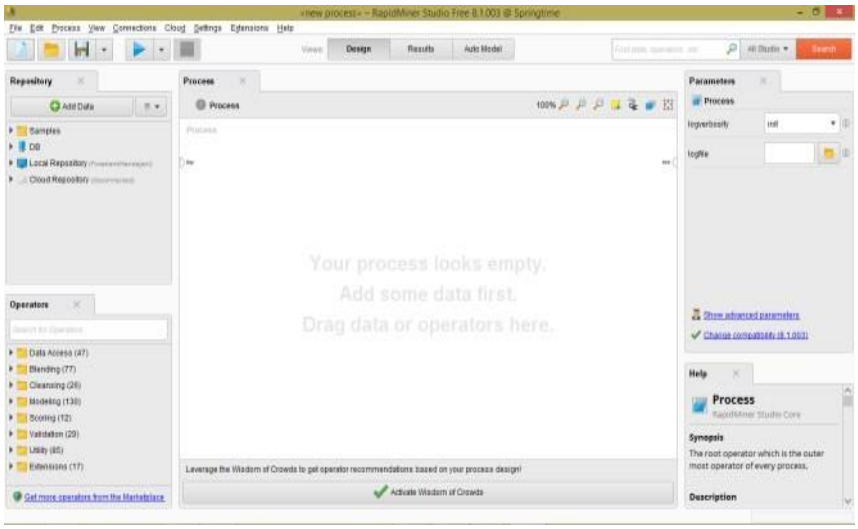

Figure 3 Rapidminer Home Screen

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

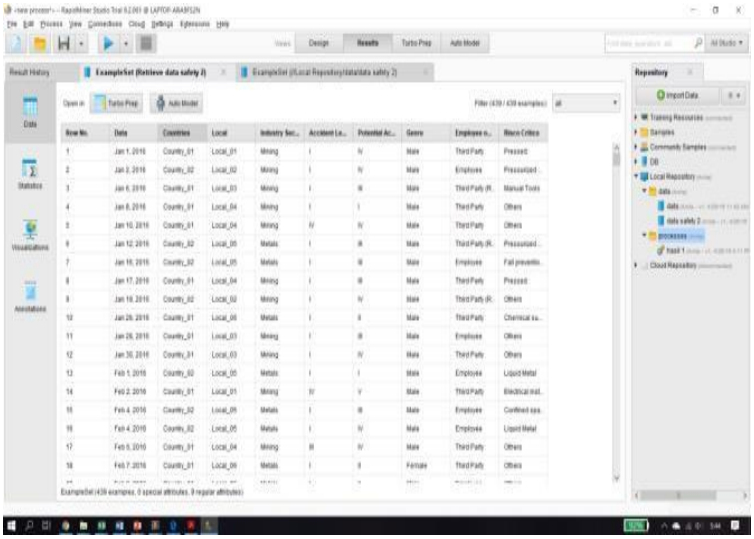

Figure 2. Results from a Database

Figure 5 shows the design for predicting the accident rate. Starting with data retrieval, selecting attributes, assigning responsibilities, collecting samples, and finally modeling with a Decision Tree.

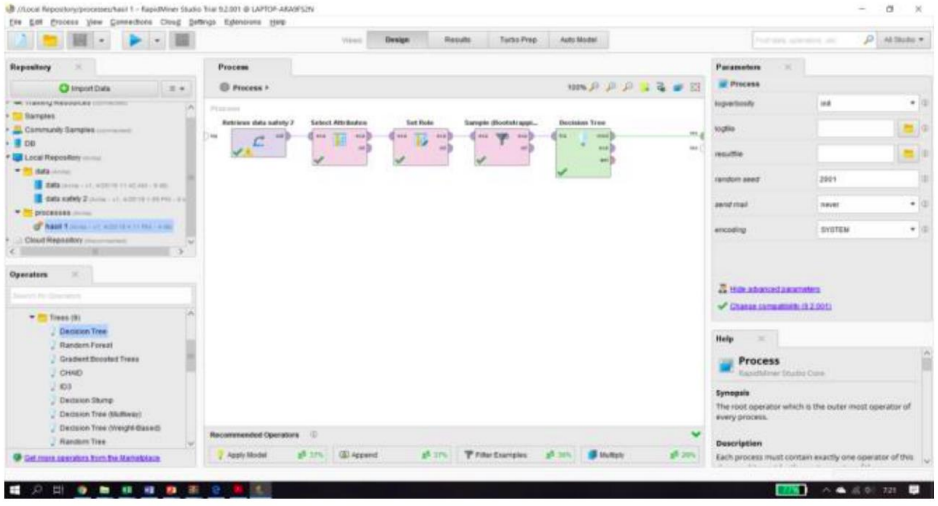

Figure 5. Design for Accident Rate Prediction

The decision tree's outcomes are shown in Figure 6. From level I to level VI, the size of the accident rate can be seen from each local (I means not severe but VI means very severe). The following are the outcomes of employing a data mining strategy using quick miners to forecast the rate of work accidents based on secondary data:

a. In total, 11 of the 12 manufacturers had an average accident rate of level 1, indicating that an error occurred but was not harmful.
b. The third factory (local 03) has a level IV accident rate, indicating that more supervision is required, as well as complete changes to critical parts and specific maintenance on the Risco Critio power lock.
c. There is no accident rate for the 11th factory (local 11).

The outcome, in addition to being in the form of a tree, can also take the form of a description, as seen in Figure 7.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

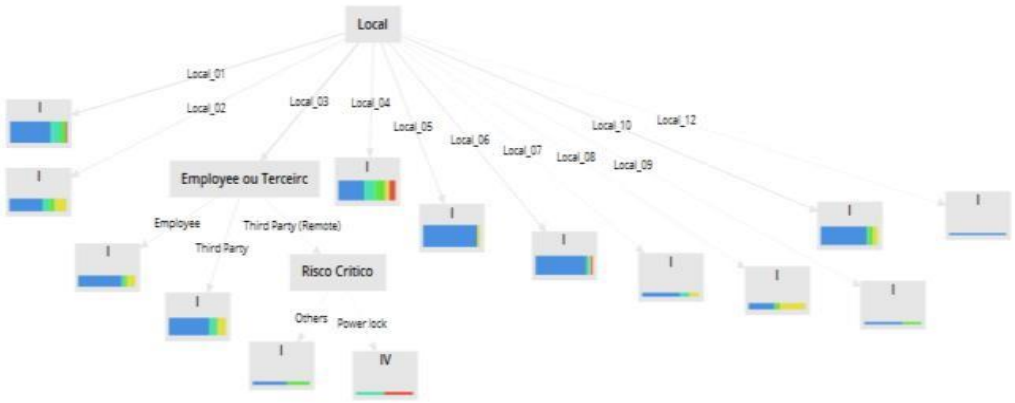

Figure 6. Decision Tree Results

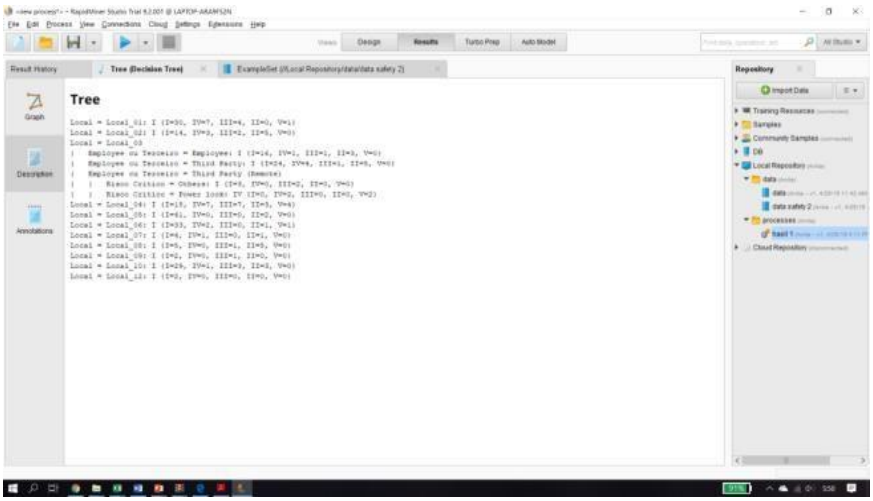

Figure 7. Description Tree

4. Conclusion

The results of forecasts made utilizing a data mining approach and quick miners to determine the rate of work accidents, including Overall, 11 of the 12 manufacturers had an average accident rate of level 1, indicating that an error occurred but was not harmful. The third factory (local 03) has a level IV accident rate, indicating that more supervision is required, as well as complete changes to critical parts and specific maintenance on the Risco Critio power lock. There is no accident rate for the 11th factory (local 11).

References

[1] D. Puspasari and S. M. Rindi, "ANALISIS KESEHATAN DAN KESELAMATAN KERJA BIDANG PENANGGULANGAN KEBAKARAN PADA DINAS PEMADAM KEBAKARAN KABUPATEN SUMEDANG," *SINTESA STIE Sebel. April SUMEDANG*, vol. 10, no. 2, pp. 145–155, 2020.

[2] J. MADELEY, "BIG BUSINESS POOR PEOPLES Bisnis Besar, Masyarakat Miskin: Bagaimana Perusahaan-Perusahan Transnasional Merusakkan Kehidupan Orang Miskin." PT Elex Media Komputindo, 2013.

[3] M. Stergiou-Kita *et al.*, "Danger zone: Men, masculinity and occupational health and safety in high risk occupations," *Saf. Sci.*, vol. 80, pp. 213–220, 2015.

[4] A. N. Ananda and M. Ibrahim, "Pengaruh Perputaran Modal Kerja Dan Perputaran Piutang Terhadap Profitabilitas Pada Perusahaan Pertambangan Sub Sektor Minyak Dan Gas Bumi Listing Di Bei." Riau University, 2017.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

[5] G. Kaur and H. Kaur, "Prediction of the cause of accident and accident prone location on roads using data mining techniques," in *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017, pp. 1–7.

[6] R. Rafika, D. Lestantyo, and S. Suroto, "Pengetahuan tentang Keselamatan Kerja Ditinjau dari Inspeksi K3, Media Poster, Sikap dan Kondisi Supervisi: Studi Literatur," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 11, no. 2, pp. 301–310, 2021.

[7] R. R. Moningka, "ANALISIS SENTIMEN TERKAIT PEMENUHAN KEBUTUHAN DASAR DI TEMPAT EVAKUASI SEMENTARA PASCA ERUPSI MERAPI." UAJY, 2018.

[8] A. Lombardi Netto, V. A. P. Salomon, M. A. Ortiz-Barrios, A. K. Florek-Paszkowska, A. Petrillo, and O. J. De Oliveira, "Multiple criteria assessment of sustainability programs in the textile industry," *Int. Trans. Oper. Res.*, vol. 28, no. 3, pp. 1550–1572, 2021.

[9] Y. Syahra, M. Syahril, and Y. Yusnidah, "Implementasi Data Mining Dengan Menggunakan Algoritma Fuzzy Subtractive Clustering Dalam Pengelompokan Nilai Untuk Menentukan Minat Belajar Siswa Smp Primbana Medan," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 17, no. 1, pp. 54–63, 2019.

[10] P. Subarkah, A. N. Ikhsan, and A. Setyanto, "The effect of the number of attributes on the selection of study program using classification and regression trees algorithms," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, 2018, pp. 1–5.

[11] Y. Mahena, M. Rusli, and E. Winarso, "Prediksi Harga Emas Dunia Sebagai Pendukung Keputusan Investasi Saham Emas Menggunakan Teknik Data Mining," *Kalbiscentia J. Sains dan Teknol*, vol. 2, no. 1, pp. 36–51, 2015.

[12] Y. Asriningtias and R. Mardhiyah, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," *J. Inform.*, vol. 8, no. 1, pp. 837–848, 2014.

[13] H. Pratiwi, *Buku Ajar: Sistem Pakar*. STMIK Widya Cipta Dharma, 2019.

[14] H. Sulistiani, "Pemilihan Fitur Untuk Klasifikasi Loyalitas Pelanggan Terhadap Merek Produkfast Moving Consumer Goods (Studi Kasus: Mie Instan)." Institut Teknologi Sepuluh Nopember Surabaya, 2016.

[15] S. Jeble, S. Kumari, and Y. Patil, "Role of big data in decision making," *Oper. Supply Chain Manag. An Int. J.*, vol. 11, no. 1, pp. 36–44, 2017.