Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

Linear Regression Algorithm Analysis for Predicting Electrical Panel Painting Quality

Arif Susilo[1*], Edy Widodo[2], Elkin Rilvani[3], Syahro Suryana[4]
[1,2,3,4]Informatics Engineering Study Program, Faculty of Engineering, Pelita Bangsa University, Indonesia
[1]arif.susilo@pelitabangsa.ac.id, [2]ewidodo@pelitabangsa.ac.id, [3]elkin.rilvani@pelitabangsa.ac.id,
[4]syahrosur@gmail.com

open access

***Corresponding Author**

Article History:
Submitted: 13-06-2024
Accepted: 14-06-2024
Published: 05-06-2024

Keywords:
Data Mining; Electrical Panel; Linear Regression.

ABSTRACT

Rapid industrial development often results in increasingly fierce competition, which tends to compete to maintain and increase its market share. They implement aggressive marketing strategies, develop new products or services, or improve operational efficiency to reduce production costs. This research uses one of the methods in data mining, namely the SVM algorithm to determine predicted values and Linear Regression to predict data. Data mining is used to predict data categories based on a set of data attributes that have been collected. Analysis of linear regression algorithms in the context of predicting the quality of electrical panel painting can be carried out abstractly using several basic steps. After carrying out the analysis, significant variables can be identified in influencing the quality of electrical panel painting. The model that has been developed can describe the linear relationship between these variables and painting quality. Based on evaluation using metrics, it can be evaluated how well the linear regression model is able to predict paint quality. The results of this evaluation provide an idea of how accurate and reliable the model that has been developed is. The results of research analyzing linear regression algorithms to predict the quality of electrical panel painting will provide a deeper understanding of the influencing factors and how the use of linear regression can be applied to improve the painting process and results. Testing using rapidminer produces performance that is relevant to the modeled scenario. After comparing the results of manual calculations and the Rapid Miner application, the simple Linear Regression equation model generally shows the same data. The RMSE value was also obtained when evaluating the performance of the applied model, with an RMSE value of 0.273 with a standard deviation of +- 0.0.

INTRODUCTION

Rapid developments in the industry often result in increasingly fierce competition between the companies operating in it. Companies tend to compete to maintain and increase their market share (Muhammad Rafly Taufiqurahman et al., 2023). They may pursue aggressive marketing strategies, develop new products or services, or improve their operational efficiency to reduce production costs. As a result, companies will compete to create the best service to customers (Yasin, 2017). The painting process is a form of coating an object where the coating material used usually has a certain color. In general, painting is often used to finish products made of metal, iron, wood, plastic. Painting is the process of painting a product, but several other things that must be considered are improving the quality of the painting product (Bimamurti & Sukawi, 2017).

An aspect that supports consumer satisfaction is the quality of the electrical panel. Reliable and durable electrical panels are the main thing that consumers are looking for (Bintoro & Safwandi, 2018). Material quality, reliability in performance, and resistance to environmental conditions (such as humidity, heat, or dust) are key factors. Linear regression algorithm analysis for predicting the quality of painting on electrical panels involves the use of statistical techniques to understand the relationship between independent variables (for example: type of paint, painting technique, environmental conditions during painting) and dependent variables (quality or durability of electrical panel paint). Identify relevant variables that may influence the quality of electrical panel painting, type of paint, painting technique, environmental conditions, and surface preparation. Linear regression analysis (Alifi et al., 2022) is a powerful tool for understanding and predicting the quality of painting on electrical panels based on relevant variables. By understanding this relationship, companies can improve their painting processes to produce better quality, more durable electrical panels (Adyatama & Handayani, 2018).

The analysis begins with the selection of relevant variables that can influence the quality of the painting (Casban & Zulfikar, 2022). Relevant data is collected and processed to ensure cleanliness and consistency, as well as transformations where necessary such as data normalization. Linear regression models are used to model linear relationships between independent variables and dependent variables. Model evaluation was carried out to check

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

suitability and accuracy, including coefficient significance tests and regression assumption tests (PRASMONO & Atina Ahdika, 2023).

By using linear regression analysis, companies can optimize their production processes, increase customer satisfaction, and improve the reputation of their products in the electrical panel industry. The prediction results will later be used as a definite reference regarding painting quality research activities, improving product quality and the data resulting from the painting process can be used to determine whether to improve the quality of the product and the results of data processing become information, knowledge, it is hoped that potential or knowledge can be explored. which is more accurate, precise and fast than this data so that it can analyze product quality and find new opportunities and find strategic plans to increase profits and maintain quality.

LITERATURE REVIEW

The linear regression method is a statistical technique used to understand the relationship between two or more variables. Basically, linear regression tries to find a linear relationship between a dependent variable (also called dependent variable) and one or more independent variables (also called independent variables or predictors) (Lestari, 2023). In linear regression, a mathematical model is formed using straight lines to describe the relationship between these variables (Ayuni & Fitrianah, 2020). This model can be used to predict or estimate the value of the dependent variable based on the known value of the independent variable (Fajri Harits Muzaki & Wawan Joko Pranoto, 2024).

S. Rahayu and A. S. RMS in the research "Application of the Naive Bayes Method in Selection of Quality Types of Garden Grass CV. Our Grass Landscape" explains that arranging an attractive, cool and beautiful garden requires a high budget (Rahayu & RMS, 2018). The choice of garden grass type is an important decision in landscape design, affecting the appearance, maintenance and sustainability of the garden (Rahayu & RMS, 2018). The Naive Bayes method can be used to assist in selecting the appropriate type of grass based on soil characteristics, climate and use requirements (Bravo et al., 2023).

In this problem the Naïve Bayes method is used as a Decision Making System (DSS). In this problem the Naïve Bayes method is used as a Decision Making System (DSS). Naïve Bayes as a Decision Making System can be used for various applications in decision making, such as data classification, prediction, or in the broader context of analysis and data-based decision making.

Implementation of Predictive Modeling in Predicting the Amount of Household Electricity Use (Case Study: PLN Lubuk Pakam Area) (Ishak Iskandar, 2016). Electricity is the foundation of modern life, almost all aspects of daily life, from lighting, use of electronic devices, to transportation, are highly dependent on electricity. Industry, business and the public sector need stable and reliable access to electricity to run their operations. Electricity is the main foundation in driving innovation and the modern economy (Sapthu, 2023). With the availability of reliable and efficient electrical energy, various sectors can develop advanced technology, improve production processes and create new jobs. The technology and communications industries rely on electricity to operate their infrastructure. From large data centers that store and manage global information to telecommunications networks that facilitate global communications, they all require a steady supply of electricity to operate efficiently. Furthermore, the manufacturing sector uses electricity to support automation and mass production processes. Modern factory machines and equipment cannot operate without a reliable electrical power source. This not only increases productivity but also reduces production costs, thereby creating incentives for companies to develop and retain jobs.

METHOD

Research Object

The stages that will be used in making predictions on electrical panel data and determining attributes to facilitate research so that research can run well and systematically, and meet the desired objectives. The descriptive research category is a category used to describe or analyze research results but is not used to make broader conclusions. A quantitative approach requires a lot of use of numbers, starting from collecting data, interpreting that data, and displaying the results, by describing something based on the data collected in the form of numbers regarding existing facts.

Research Stage

In carrying out analysis and looking for consumption patterns in the use of electrical energy to facilitate research and be able to run systematically and meet the desired goals, the following steps are made in the research stages:

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

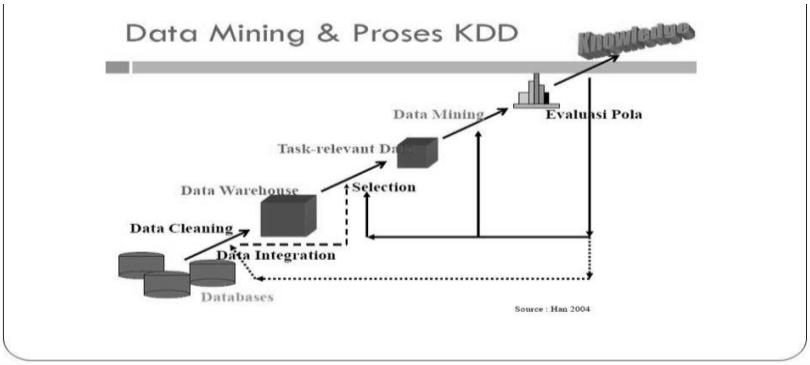

Figure 1. Research Stages

a. Data Cleaning

The data cleaning stage is the process of cleaning unused data. Data deletion by removing missing values, duplicate data, checking and correcting errors in the data. The data cleaning process was carried out manually with the help of Microsoft Excel.

Table 1. Data Cleaning

Date of Production	Temperature	Electrostatis	Minute	Thickness Results
2/11/2019	180º C	90 µA	10	56
2/11/2019	180º C	90 µA	10	76
2/11/2019	180º C	90 µA	10	67
2/12/2019		90 µA	10	88
2/12/2019	180º C	90 µA	10	89
2/12/2019	180º C		10	90
2/13/2019	180º C	90 µA	10	67
2/13/2019	180º C		10	45
2/14/2019	180º C	90 µA	10	45
2/14/2019	180º C	90 µA	10	67
2/14/2019	180º C	90 µA	10	69

b. Data Warehouse

Data warehouse is the process of selecting data from a collection of existing operational data before entering the data and information mining stage. The selected data will be used for the data mining process and stored in a separate file from the operational database.

Table 2. Data selection

Date of Production	Temperature	Electrostatis	Minute	Thickness Results
2/11/2019	180º C	90 µA	10	56
2/11/2019	180º C	90 µA	10	76
2/11/2019	180º C	90 µA	10	67
2/12/2019		90 µA	10	88
2/12/2019	180º C	90 µA	10	89
2/12/2019	180º C		10	90
2/13/2019	180º C	90 µA	10	67
2/13/2019	180º C		10	45
2/14/2019	180º C	90 µA	10	45
2/14/2019	180º C	90 µA	10	67
2/14/2019	180º C	90 µA	10	69

c. Task Relevant Data

Relevant data tasks are the process of changing data from one format to another according to needs which aims to simplify the process of reading data in the program or tool used.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

Table 3. Task Relevant Data

Date of Production	Minute	Thickness results
2/11/2019	10	56
2/11/2019	10	76
2/11/2019	10	67
2/12/2019	10	88
2/12/2019	10	89
2/12/2019	10	90
2/13/2019	10	67
2/13/2019	10	45
2/14/2019	10	45
2/14/2019	10	67
2/14/2019	10	69

d. Data Mining

The linear regression algorithm, which is a classifier, assumes that the value of an attribute in a given class does not depend on the values of other attributes. The following is an overview of data modeling. The impact of using regression analysis can be used to decide whether increasing and decreasing the dependent variable can be done by increasing and decreasing the state of the independent variable, or whether increasing the state of the dependent variable can be done by increasing the independent variable/and vice versa.

Least squares method: the most popular method for establishing simple linear regression equations. General Form of Simple Linear Regression:

Determining the Simple Linear Regression Equation for the values of a and b

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$a = \bar{y} - b\bar{x} \quad \text{so that} \quad a = \frac{\sum_{i=1}^{n} y_i}{n} - b\frac{\sum_{i=1}^{n} x_i}{n}$$

n : the number of data pairs

yi : value of the ith dependent variable Y

xi : value of the ith independent variable X

Calculate the linear regression equation

Y = a + bX

Y : dependent variable X : free variable

a : constant b : slope

e. Evaluation of Test Results

Method testing is carried out with the aim of knowing the results to be analyzed and measuring whether the methods and algorithms used are functioning well or not. The testing process uses the rapidmener tool and sees whether the data matches the results obtained through the tool, by measuring the accuracy, precision and recall results. Evaluation can be done by observing and analyzing the results of the Naïve Bayes algorithm used to ensure that the test results are truly in accordance with the discussion. Analyze each attribute value and model that has been built and ensure that the test results produce accuracy, precision, recall values. Accuracy is defined as the level of closeness between the predicted value and the actual value of the model that has been built.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

RESULT

Simple Linear Regression Algorithm Modeling

a. Data analysis

This research uses a linear regression algorithm to identify retail product stock needs and will obtain root mean square error (RMSE) results and predictions that can be used in decision making regarding the quality of painting results. The initial stage carried out in this research was preparing the data and then carrying out data pre-processing stages such as data cleaning, data selection, and data transformation on sales transaction data. A total of 1200 rows of data were cleaned so that for the next process we only need to use clean data for a total of 1000 rows of data. Then data is aggregated per month to produce data that will be modeled as in the following table.

Table 4. Aggregation of sales dataset per month

Month Name	Minute	Thickness Results
January	10	56
January	12	76
January	12	67
February	15	88
February	15	89
February	15	90

Henceforth, the results of the data aggregation will be used as training data at the modeling stage using the Linear Regression algorithm

b. Data validation

Validation technique that divides data into two parts randomly, some as training data and the other part as testing data. The data used is stock data and data carried out in the selection process, the next process determines the dataset that will be used as training data and testing data, the data used as prediction data is 15 data selected randomly from 1,200 data that carry out the selection process stages. The resulting data will be modeled as in the following table.

Table 5. Product Stock Dataset

Month Name	Minute	Thickness Results
January	10	56
February	12	76
February	12	67
March	15	88
March	15	89
March	15	90
April	12	67
May	10	45
June	10	45
June	12	67

Linear Regression Calculations

The results of calculating the XY and XX values from the dataset records which are used as training data. The data that will become the X variable is Minute, while the Y variable is the Thickness Result. Calculations are carried out on the training data rows, including by getting the value (XY) which is the product of the two variables X and Y, In applying the formula below, the values obtained for Coefficient A and Coefficient B from the training dataset used in this research are as follows:

Coefficient A : $\frac{(\sum y)(\sum x^2)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$

$a = \frac{(1635)(71848)-(1744)(59188)}{100(71848)-(1744)^2}$

$a = \frac{14247608}{4143264}$

$a = 3.44$

So it can be seen that the value of a is 3.44

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

Coefficient B : $\frac{n(\sum xy)-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2}$

$b = \frac{100(59188)-(1744)(1635)}{100(71848)-(1744)^2}$

$b = \frac{3067360}{4143264}$

b = 0,74

So it can be seen that the value of b is 0,74

Implementation of Linear Regression Equations

From the calculation results of searching for coefficient A and coefficient B values, the resulting linear regression equation model is as follows:

Y = a + bx

Y = 3.44 + 0,74 (x)

Where the information is as follows:

Y = prediction outcome variable

3.44 = intercept regression coefficient (constant)

x = independent variable

0,74 = slope regression coefficient (slope)

Evaluation of Test Data on Rapid Miner

Evaluation stages of test data used to assess the simple Linear Regression algorithm equation model applied in establishing Painting Product Quality of a new object with proper accuracy. In this research, the author utilized the Rapidminer Studio application, the results of testing carried out using the Rapidminer Studio application were by applying the following steps:

a. Import the data needed for the process in the Rapidminer tool. In the Rapid Miner application, select and click Import Data, then select the data that will be used and then determine the attributes and labels that will be used.
b. Click the Design menu, in the process view, add the training dataset and test dataset in the folder to the process display screen.
c. Next, on the Modeling menu, in the Predictive submenu, select the Linear Regression function operator, to apply a simple Linear Regression algorithm equation to the object prediction process that will be carried out.
d. Next, in the Scoring menu, select the Apply Model operator and drag it to the process display screen. Through this function, you can set the application of the model from the dataset used in this process to the predicted labels that will be applied.
e. Connect all the function operators from the prepared design to form a process flow as in Figure 4.1 below:

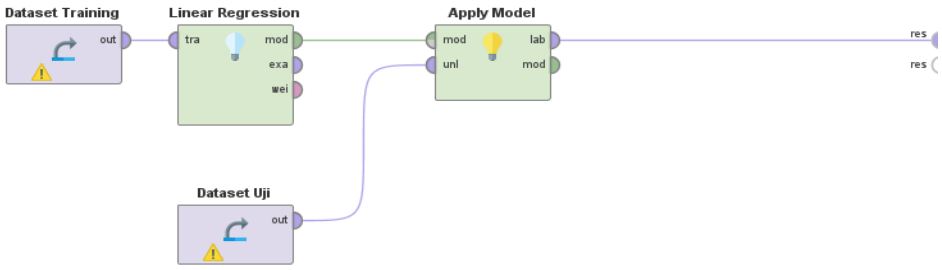

Figure 2. Implementation of the Linear Regression Algorithm in the RapidMiner Studio application

Linear Regression Performance Testing

Performance testing of models and algorithms is carried out with the aim of knowing the results of the calculations being analyzed and measuring whether the methods and algorithms used are functioning well or not. By applying this model to the results of calculating the RMSE (Root Mean Square Error) value via the Rapidminer Studio application, it produces an RMSE figure of 0.273 +/- 0.000, a value that is small enough so that only small differences occur from the manual calculation simulation process with the Rapidminer Studio application. prepared test data.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

ROOT_MEAN_SQUARED_ERROR
root_mean_squared_error: 0.273 +/- 0.000
PERFORMANCEVECTOR
PerformanceVector:
root_mean_squared_error: 0.273 +/- 0.000

DISCUSSION

The linear regression equation will then be implemented in making predictions on testing data. In general, the application of the linear regression equation is applied to predict 20 dataset records as predetermined test data. From calculations carried out manually and also compared with the process in the rapid miner application, the results shown do not have a significant difference, in other words, both manual calculations and those processed in the application show similar results. Below is a comparison table between the results of manual calculations and those in the rapid miner application for variable Y. Meanwhile, the comparison of the actual variable Y value from testing (observation) data with the predicted variable Y value in the rapidminer application can be seen in the following graph.

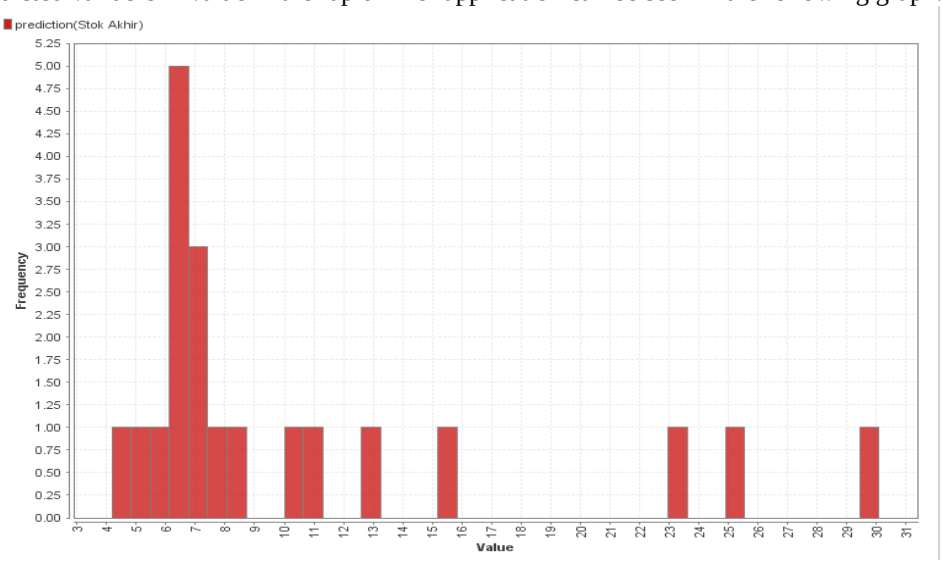

Figure 3. Graph of Comparison of Observed (Y) Values with Predicted (Y) Values

Based on the results of the tests that have been carried out, the variables or attributes used in this research (minute and thinkness results) have a significant effect on this research. It is proven that using the linear regression algorithm is able to provide good results with a Root Mean Squared Error value of 0.273 +/- 0.000. This is because there is a correlation or functional relationship (cause - effect) between one variable (dependent or criterion) and another variable (independent or predictor). This testing process is carried out to identify stock needs using a linear regression algorithm.

CONCLUSION

This research utilizes several painting production data through a predictive method approach and can be applied to analyze data to predict future quality needs based on production quality. A simple Linear Regression algorithm prediction process can be implemented where the results also show new insights into the need for predictions on the quality of painting production. Testing using rapidminer produces performance that is relevant to the modeled scenario. The simple Linear Regression equation model after comparing the results of manual calculations and also with the Rapid Miner application generally shows the same data. The RMSE value was also obtained when evaluating the performance of the applied model, with an RMSE value of 0.273 with a standard deviation of +- 0.0.

REFERENCES

Adyatama, A., & Handayani, N. U. (2018). PERBAIKAN KUALITAS MENGGUNAKAN PRINSIP KAIZEN DAN 5 WHY ANALYSIS: STUDI KASUS PADA PAINTING SHOP KARAWANG PLANT 1, PT TOYOTA MOTOR MANUFACTURING INDONESIA. *J@ti Undip : Jurnal Teknik Industri*, *13*(3), 169. https://doi.org/10.14710/jati.13.3.169-176

Alifi, M. R., Hayati, H., & Fauzi, C. (2022). Penerapan Algoritma Regresi Linier pada Prediksi Tarif Influencer Media Sosial. *Journal of Information System Research (JOSH)*, *4*(1), 210–218. https://doi.org/10.47065/josh.v4i1.2361

Ayuni, G. N., & Fitrianah, D. (2020). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4096

XYZ. *Jurnal Telematika*, *14*(2), 79–86. https://doi.org/10.61769/telematika.v14i2.321

Bimamurti, H., & Sukawi, S. (2017). ENERAPAN MATERIAL FINISHING INTERIOR KAFÉ DI TEMBALANG, SEMARANG. *MODUL*, *16*(2), 94. https://doi.org/10.14710/mdl.16.2.2016.94-100

Bintoro, A., & Safwandi, S. (2018). Implementasi Data Mining Penentuan Daya Pelanggan Baru untuk Klasifikasi Subsidi dan Non Subsidi di Wilayah PLN Kota Lhokseumawe. *Sisfo: Jurnal Ilmiah Sistem Informasi*, *2*(2). https://doi.org/10.29103/sisfo.v2i2.1011

Bravo, A., Tursina, T., & Sastypratiwi, H. (2023). Penerapan Metode Naive Bayes Untuk Penentuan Bibit Kelapa Sawit Berdasarkan Kondisi Daerah Tanam dan Perawatan Tanaman. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *11*(1), 101. https://doi.org/10.26418/justin.v11i1.52277

Casban, C., & Zulfikar, S. R. (2022). Analisis Cost of Poor Quality Proses Painting Produk Pan Oil TD. *Jurnal INTECH Teknik Industri Universitas Serang Raya*, *8*(1), 9–16. https://doi.org/10.30656/intech.v8i1.4458

Fajri Harits Muzaki, & Wawan Joko Pranoto. (2024). ANALISIS REGRESI LINEAR DALAM DATA MINING UNTUK PREDIKSI SIJIL OFF DI KSOP KELAS I SAMARINDA. *Jurnal Ilmu Teknik*, *1*(2), 261–266.

Ishak Iskandar. (2016). Implementasi Predictive Modelling Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga (Studi Kasus: PLN Area Lubuk Pakam). *JURIKOM (Jurnal Riset Komputer)*, *5*(6), 621–628.

Lestari, S. (2023). Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan Album KPOP. *INSOLOGI: Jurnal Sains Dan Teknologi*, *2*(1), 199–209. https://doi.org/10.55123/insologi.v2i1.1692

Muhammad Rafly Taufiqurahman, Averusani Putra W, Muhamad Syahajifany SL, & Gavanico Alnata Verbasov. (2023). Pengaruh Teknologi Informasi Dalam Perkembangan Bisnis. *Jurnal Pendidikan Dan Konseling*, *5*(1), 5302–5307.

PRASMONO, A. S. P., & Atina Ahdika. (2023). Analisis Regresi Berganda pada Faktor-Faktor yang Mempengaruhi Kinerja Fisik Preservasi Jalan dan Jembatan Di Provinsi Sumatera Selatan. *Emerging Statistics and Data Science Journal*, *1*(1), 47–56. https://doi.org/10.20885/esds.vol1.iss.1.art6

Rahayu, S., & RMS, A. S. (2018). Penerapan Metode Naive Bayes Dalam Pemilihan Kualitas Jenis Rumput Taman CV. Rumput Kita Landscape. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, *9*(2), 162–171. https://doi.org/10.31849/digitalzone.v9i2.1942

Sapthu, A. (2023). LISTRIK DAN PENGARUHNYA TERHADAP PERTUMBUHAN EKONOMI DI PROVINSI MALUKU. *Jurnal Cita Ekonomika*, *17*(2), 199–207. https://doi.org/10.51125/citaekonomika.v17i2.11315

Yasin, A. (2017). STRATEGI PENINGKATAN KUALITAS PELAYANAN PADA PT SAFINA ASSALAM TOUR GAMBUT KALIMANTAN SELATAN. *Jurnal Ilmiah Ekonomi Bisnis*, *3*(2). https://doi.org/10.35972/jieb.v3i2.98