

# IMPLEMENTATION OF DATA MINING TO PREDICT THE VALUE OF INDONESIAN OIL AND NON-OIL AND GAS IMPORT EXPORTS USING THE LINEAR REGRESSION METHOD

*Elvis Sastra Ompusunggu<sup>\*1</sup>, Wilson Sinaga<sup>2</sup>, Mikael Siahaan<sup>3</sup>, Jepri Banjarnahor<sup>4</sup>, Yonata Laia<sup>5</sup>,  
Oloan Sihombing<sup>6</sup>*

*<sup>1,2,3</sup>Universitas Prima Indonesia*

*Jl.Sampul No.3, Sei Putih Bar., Kec. Medan Petisah, Medan City, North Sumatra 20118*

*Email\* : baruschild2@gmail.com*

**ABSTRACT-** Indonesia's export-import activities in recent years, the value of Indonesia's exports and imports has decreased due to global conditions. The problems that occur are the uncertainty and complexity in estimating the importance of international trade in the oil and gas and non-oil and gas sectors, dependence on just one or a few markets, and the problem of unfair competition, unfair competition between business actors can reduce export-import prices. The value of oil and gas and non-oil and gas exports and imports is influenced by several external factors that are difficult to predict, such as fluctuations in oil and gas prices, changes in trade policies, and global economic factors. The prediction results are obtained monthly from the export value data using the rapid miner application. From the export data, the value of non-oil and gas exports receives a very high value compared to the export data of oil and gas values. Then the results from rapid miner using the linear regression algorithm are obtained. The predicted import value of oil and gas and non-oil and gas value data in June is 209,162,268, and the expected export value of oil and gas and non-oil and gas value data in June is 349,285,781 and non-oil and gas which more are predicted to have the highest value compared to the value of oil and gas in each month.

**KEYWORDS:** Data Mining, Regresi Linear, Rapidminer

## 1. INTRODUCTION

Sustainable economic growth is one of the keys to success in advancing a country's economy. When export demand increases, this will encourage domestic production, affecting all domestic economy sectors to develop and grow. However, there is an increase in imports of goods from abroad. In that case, this can decrease domestic productivity due to reduced demand for similar goods and consequently reduce domestic economic growth. [1] [2]

Indonesia's export-import activities in recent years, the value of Indonesia's exports and imports has decreased due to global conditions. The problems that occur are the uncertainty and complexity in estimating the importance of international trade in the oil and gas and non-oil and gas sectors, dependence on just one or a few markets, and the problem of unfair competition, unfair competition between business actors can reduce export-import prices. The value of oil and gas and non-oil and gas exports and imports is influenced by several external factors that are difficult to predict, such as fluctuations in oil and gas prices, changes in trade policies, and global economic factors.

To maintain economic stability and plan appropriate policies, the government and business players need to predict the value of oil and gas and non-oil and gas exports in the future.

Prediction involves making predictions or estimates based on existing data regarding an event or outcome that may occur in the future. The main objective is to provide estimates of products or circumstances that may arise in the future to support more precise and accurate

decision-making. [3] [4]

In that context, data mining is an approach that can be used to extract patterns and valuable information from existing data. Data mining is a process that uses statistical analysis techniques, mathematics, artificial intelligence, and machine learning to identify patterns and trends in data. The stages involved in this process include data pre-processing, data exploration, modeling, and evaluation. [5] [6]

The linear regression method is a technique in data mining that can be used to model a linear relationship between the dependent and independent variables. The application of the linear regression method is comprehensive and covers various fields, including social sciences, economics, environmental sciences, and natural sciences. For example, in economics, linear regression can predict product sales based on advertising costs incurred.

In environmental sciences, linear regression methods enable the identification of relationships between variables such as temperature, spread of disease, and water density in natural environments. Thus, linear regression provides a valuable and versatile analytical approach to understanding relationships and trends in data from various disciplines. [7] [8] [9]

Previous studies have used data mining methods with linear regression to predict stock prices in shipping companies. Assessment is done by measuring the Root Mean Square Error (RMSE). The evaluation results show that the RMSE value shows a difference of 7.522 from the actual stock closing price data. [10].

Based on the previous research that has been described, linear regression is proven to produce a low level of prediction error, indicating that this method can provide accurate predictions. Therefore, applying linear regression is a potential solution for the government in predicting the direction of future development of oil and gas and non-oil and gas import exports, increasing understanding of international trade trends, and supporting better decision-making in the oil and gas and non-oil and gas sectors.

By utilizing accurate and precise data, the government can develop an economic development strategy based on the country's conditions and situation. In this regard, the authors are interested in conducting a study entitled "Implementation of Data Mining to Predict the Value of Indonesian Oil and Gas and Non-Oil and Gas Exports Using the Linear Regression Method."

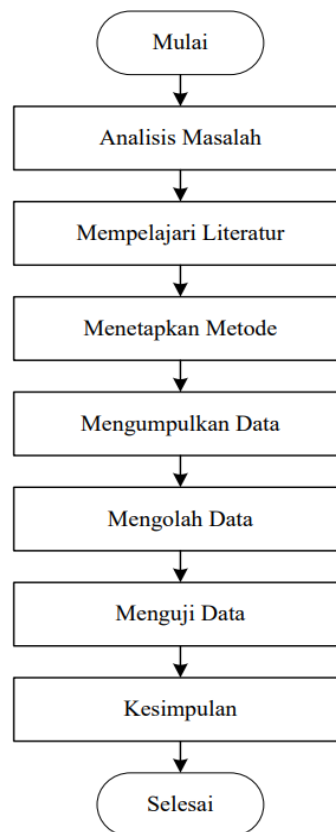
## **2. RESEARCH METHODS**

### **2.1 Data Mining**

Data mining aims to explore valuable information and patterns from large data sets. The stages in data mining include data collection, extraction, analysis, and applying statistics to the data. This process is often called knowledge retrieval, data pattern analysis, exploration, or discovery. It has various other terms depending on the context and field of application.

### **2.2 Research Stages**

This research was planned by observing and collecting data. The collected data will be entered into Microsoft Excel to be processed through calculations and steps of the linear regression method. The results of these calculations will be used in the rapid-miner application to get accurate results.



**Figure 1 Research design**

**Information:**

1. Analysis of the problem analyzing problem-related to data on the value of oil and gas export and import data from 2017-2023
2. In studying research literature, collecting and using appropriate references is very important to obtain accurate and relevant information in research.
3. Establishing a method in research is the process of selecting a method or technique to solve the problem under study.
4. Collect data on the value of oil and gas export and import data from 2017-2023 from the Central Statistics Agency (BPS).
5. Processing data Perform data processing using data mining in the linear regression method.
6. Testing data Data testing is done using rapid-miner tools.
7. The conclusions drawn in grouping the prediction results for the coming year are the value of oil and gas and non-oil and gas imports in Indonesia.

**2.3. Data Mining Processing**

In this study, the data mining process follows the Knowledge Discovery in Database (KDD) steps to produce information by a predetermined sequence. In this study, we apply simple linear regression. Simple linear regression is a statistical analysis method involving the relationship between one independent variable (X) and one dependent variable (Y). The general form of the simple linear regression equation is as follows:

$$y = b_0 + b_1 s$$

Information:

Y = predictor or dependent variable (dependent variable)

X = predictor or independent variable (independent variable)

b0 = constant (intercept)

b1 = independent variable regression coefficient parameter

### 3. RESULTS AND DISCUSSION

#### 3.1 Result

This study uses a dataset of oil and gas export-import value data from 2017-2023, which is proposed to predict the value of Indonesia's oil and gas and non-oil and gas import exports using the linear regression method. The dataset is carried out using the linear regression method.

#### 3.2 Linear Regression Algorithm

After the data values are aligned, the next step is to cluster using the linear regression method. This process is expected to obtain predictions of Indonesia's oil and gas and non-oil and gas export and import values using linear regression. The linear regression process will run as follows :

##### a. Collecting Datasets

The following table uses a dataset of oil and gas export-import value data from 2017-2023 to predict Indonesia's oil and gas and non-oil and gas export import values using the linear regression method. The dataset is carried out using the linear regression method.

**Table 1 Table of Oil and Gas and Non-Oil and Gas Export Values**

No	Bulan	Migas	Non Migas	Jumlah
1	Jan-17	1278.6	12119.1	13397.7
2	Feb-17	1208.6	11407.3	12616.0
3	Mar-17	1516.2	13202.2	14718.5
4	Apr-17	1036.2	12233.5	13269.7
5	May-17	1294.4	13039.5	14333.9
6	Jun-17	1276.3	10385.1	11661.4
7	Jul-17	1165.0	12446.1	13611.1
8	May-22	1498.1	19995.2	21493.3
9	Jun-22	1551.8	24589.3	26141.0
10	Jul-22	1287.6	24185.8	25473.4
....	....	....	....	....
....	....	....	....	....
....	....	....	....	....
71	Nov-22	1101.9	22957.2	24059.1
72	Dec-22	1457.8	22324.9	23782.7

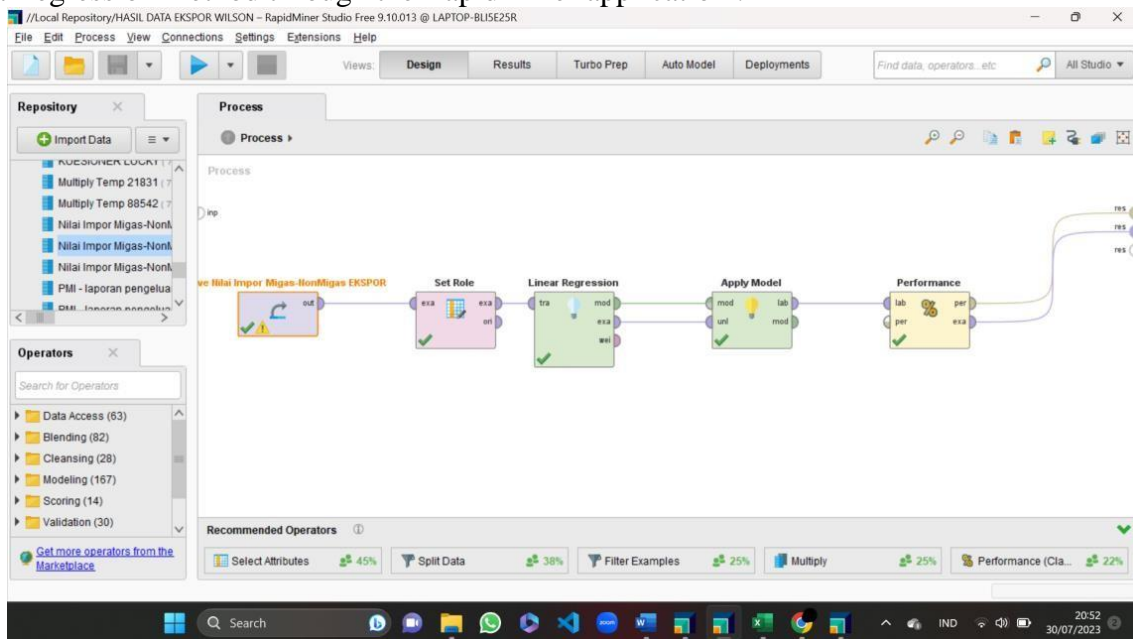
**Table 2 Table of Oil and Gas and Non-Oil Gas Imports**

Bulan	Migas	Non Migas	Jumlah
Jan-17	1828.1	10145.7	11973.8
Feb-17	2473.1	8886.3	11359.4
Mar-17	2277.0	11006.2	13283.2

Apr-17	1646.8	10303.8	11950.6
May-17	1791.6	11980.9	13772.5
Jun-17	1600.6	8391.0	9991.6
Jul-17	1778.7	12111.1	13889.8
.....	.....	.....	.....
.....	.....	.....	.....
Dec-22	3201.0	16662.1	19863.1

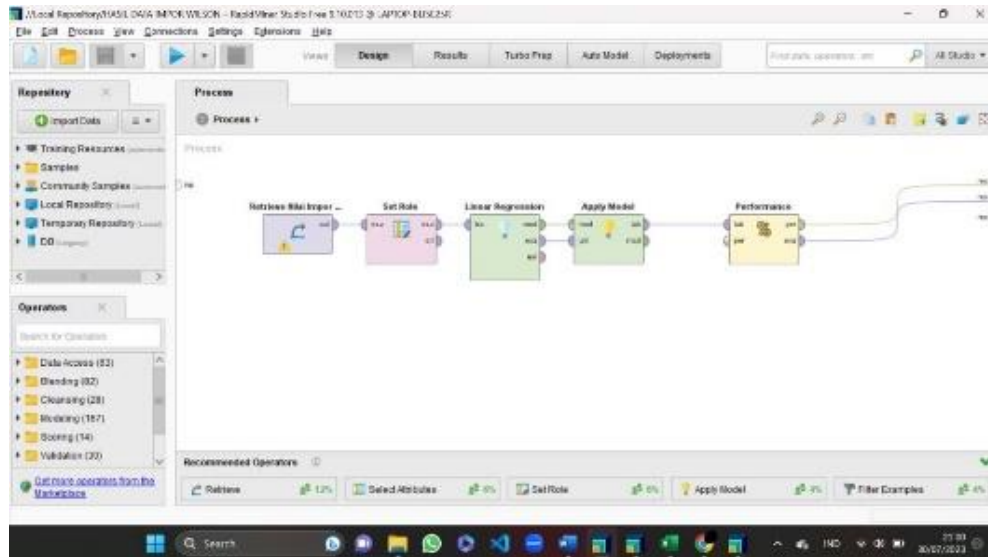
**b. Linear Regression Testing using Rapid Miner**

At this stage, data normalization is carried out to equalize the scale of data attributes within a more specific range. The normalization process aims to make data quickly processed using the linear regression method through the RapidMiner application :



**Figure 2 Display of Oil and Gas Export Value Data Process and Non-Oil and Gas Data From 2017-2023**

Above is a display of data input results from the rapid Miner application in processing Excel data to linear regression in the fast Miner application starting from the set role, which serves as a distinguishing line for naming attribute coordinates and position predictions, which will be entered into the 'label' category then processed into linear regression for testing data on the value of oil and gas exports and imports, the results are then processed by applying the model and directly to performance for the prediction results of the data process for the value of oil and gas exports and non-oil and gas data from 2017-2023 in the image below:



**Figure 3 Data Process Display of Oil and Gas Import Value and Non-Oil and Gas Data From 2017-2023**

Above is a display of input data resulting from the rapid Miner application in processing Excel data to linear regression in the fast Miner application starting from the set role, which serves as a distinguishing line for naming attribute coordinates and position predictions, which will be entered into the 'label' category then processed into linear regression for testing data on the value of oil and gas exports and imports, the results are then applied to the model process and directly to performance for the prediction results of the data process for oil and gas import values and non-oil and gas data from 2017-2023 in the image below :

**c. Export and Import Value Data Results Using the Rapid Miner Application**

**1. Export Value Data Results Using the Rapid Miner Application**

The prediction results are obtained monthly from the export value data using the rapid miner application. From the export data, the value of non-oil and gas exports receives a very high value compared to oil and gas value data.

**Table 3 Table of Export Value Prediction Results Using the Rapid Miner Application**

Bulan	Predection	Migas	Non Migas	Jumlah
Jan-23	22.384.770	1487.0	20835.0	22323.0
Feb-23	213.726.831	1186.0	20134.0	21321.3
Mar-23	235.177.406	1338.0	22077.0	23416.0
Apr-23	192.736.733	1258.0	18025.0	19284.0
May-23	217.622.020	1308.0	20398.0	21706.0
Jun-23	349.285.781	1578.0	2307.0	3835.0
Jul-23	140.129.647	1630.0	12538.0	14169.0
Aug-23	141.109.778	1591.0	12671.0	14263.0
Sep-23	146.092.734	1755.0	13003.0	14759.0
Oct-23	151.844.417	2134.0	13206.0	15340.0
Nov-23	143.325.537	2133.0	12373.0	14506.0
Dec-23	140.968.886	1987.0	12281.0	14268.0

**2. Import Value Data Results Using the Rapid Miner Application**

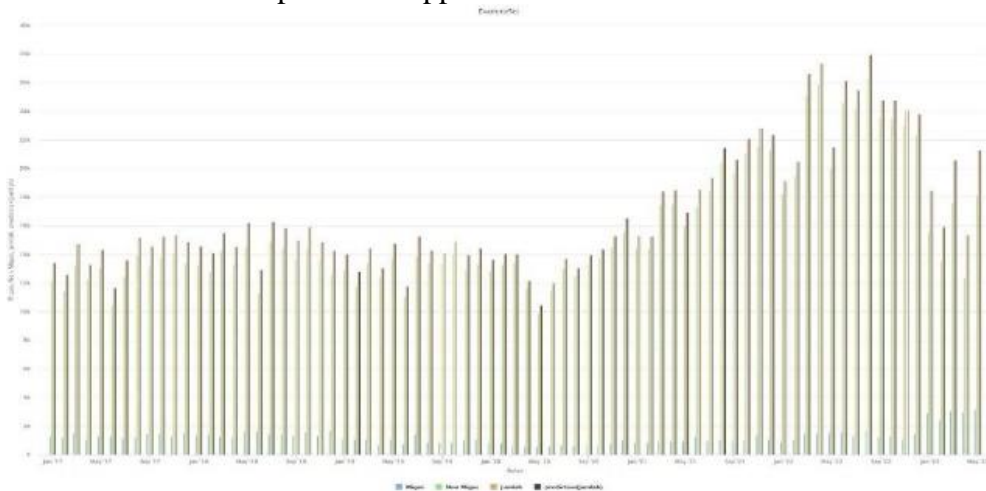
The rapid miner application obtains the prediction results monthly from the import value data. From the import data, the non-oil and gas import value is very high compared to the oil and gas value import data.

**Table 4 Table of Import Value Prediction Results Using the Rapid Miner Application**

Bulan	Predection	Migas	Non Migas	Jumlah
Jan-23	184.430.660	2906.1	15536.8	18442.9
Feb-23	15.919.518	2406.5	13512.7	15919.2
Mar-23	205.882.791	3014.8	17573.3	20588.1
Apr-23	153.476.544	2955.0	12392.6	15347.6
May-23	212.797.448	3135.1	18144.5	21279.6
Jun-23	209.162.268	3353.0	17563.0	20916.0
Jul-23	213.697.507	3567.0	17853.0	21240.0
Aug-23	193.282.943	2568.0	16789.0	19357.0
Sep-23	151.130.639	2689.0	12456.0	15145.0
Oct-23	157.752.303	4379.0	11465.0	15844.0

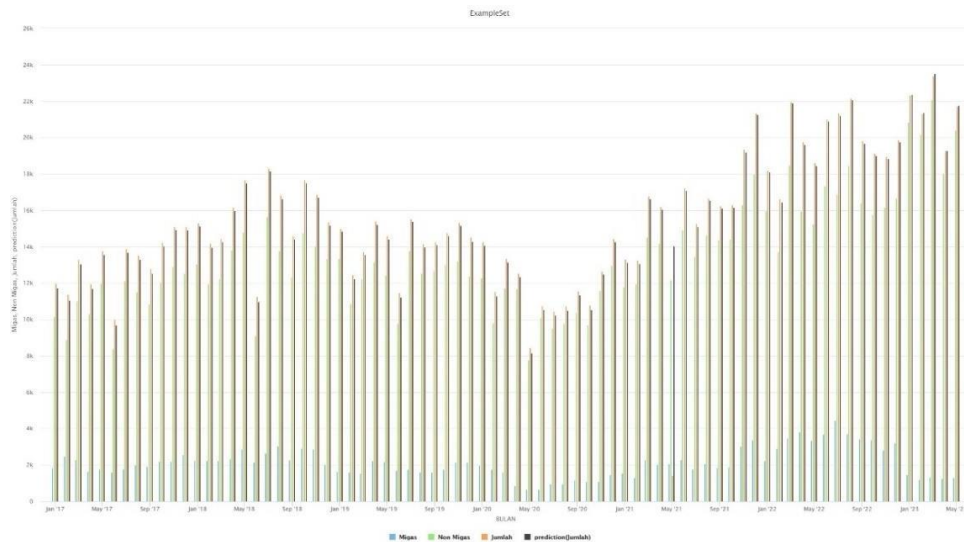
**d. Graph Results in Data Import Export Value Using the Rapidminer Linear Regression Method**

At this stage, the data normalization process is carried out, and the goal is to equate the scale of data attributes in a more specific range so that data can be processed quickly using the linear regression method with the rapid miner application.



**Figure 4 Export Value Data Results with the Rapidminer Linear Regression Method**

From the results obtained, the total value of non-oil and gas exports is a product that always increases every month and is always high compared to the full value of oil and gas, which can be seen in the picture above.



**Figure 5 Import Value Data Results with the Rapidminer Linear Regression Method**

## CONCLUSION

Based on the research results obtained from the algorithm, after receiving the results of the linear regression method.

1. The results from the rapid miner using the linear regression algorithm are obtained. The predicted import value of oil and gas and non-oil and gas value data in June is 209,162,268, and the expected export value of oil and gas and non-oil and gas value data in June is 349,285,781 and non- more predicted oil and gas, which has the highest value compared to the oil and gas value in each month.
2. Analysis results of the linear regression method are a method that can predict the value of imports in the following year.

## BIBLIOGRAPHY

- [1] Astuti, I.P. and Ayuningtyas, F.J., 2018. Pengaruh Ekspor Dan Impor Terhadap Pertumbuhan Ekonomi Di Indonesia. *Jurnal Ekonomi & Studi Pembangunan*, 19(1), pp.1-10.
- [2] Fauziah, Elsa Siti, and Abd Kholik Khoerulloh. "Pengaruh ekspor dan impor terhadap pertumbuhan ekonomi dengan kurs sebagai variabel intervening." *Khazanah Sosial 2.1* (2020): 15-24.
- [3] Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 3(2), 59-66.
- [4] Dewi, Sri Puspita, Nurwati Nurwati, and Elly Rahayu. "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor." *Building of Informatics, Technology and Science (BITS)* 3.4 (2022): 639- 648.
- [5] Mahena, Y., Rusli, M., & Winarso, E. (2015). Prediksi Harga Emas Dunia Sebagai Pendukung Keputusan Investasi Saham Emas Menggunakan Teknik Data Mining. *Kalbiscentia J. Sains dan Teknol*, 2(1), 36-51.
- [6] Ordila, Rian, et al. "Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda)." *Jurnal Ilmu Komputer* 9.2 (2020): 148-153.
- [7] Arafah, S., & Tanjung, Y. (2019). Analisis Faktor Determinan yang Mempengaruhi pemakaian metode JIT. *STUDI KASUS UD. PUSAKA BAKTI*. Bisei: *JurnalBisnisdanEkonomiIslam*, 4(01).



- [8] Suryanto, Andik Adi, and Asfan Muqtadir. "Penerapan metode mean absolute error (MEA) dalam algoritma regresi linear untuk prediksi produksi padi." *Saintekbu 11.1* (2019): 78-83.
- [9] Sarah, Anggi Mei, Bambang Kurniadi, and Endang Warsini. "Implementasi Metode Regresi Linear Dalam Memprediksi Penyakit Anemia Secara Dini" *JURNAL TEKNISI 3.1* (2023): 14-23.
- [10] Akhmad, E. P. A. (2020). Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran. *Jurnal Aplikasi Pelayaran dan Kepelabuhanan*, 10(2), 120-131.
- [11] Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi