International Journal Software Engineering and Computer Science (IJSECS)
4 (2), 2024, 576-585
Published Online August 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i2.2864.

RESEARCH ARTICLE Open Access

Estimating Distributor Demand for Fishing Gear Products Using Linear Regression Algorithm

Keswanto *
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra Province, Indonesia.
Corresponding Email: keswanto.zian@gmail.com.

Wahyu Hadikristanto
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra Province, Indonesia.
Email: wahyu.hadikristanto@pelitabangsa.ac.id.

Edora
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra Province, Indonesia.
Email: edora@pelitabangsa.ac.id.

Received: June 28, 2024; Accepted: July 10, 2024; Published: August 1, 2024.

Abstract: Fishing equipment plays a critical role in both recreational and commercial fishing activities across various aquatic environments. The challenge of managing inventory effectively is heightened by the fluctuating demand and the need to avoid overstocking, which can result in increased operational costs. To address this, a linear regression algorithm is utilized to predict demand for fishing products, using relevant independent variables to model the relationship with dependent variables such as monthly sales figures. This predictive model aims to provide actionable insights that can assist businesses in making informed decisions regarding inventory management and distribution strategies. The study employs the RapidMiner Studio application to develop and evaluate the model's performance, with the analysis yielding a Root Mean Square Error (RMSE) of 140.200. This relatively low RMSE value demonstrates the model's accuracy and effectiveness in forecasting demand, suggesting that the algorithm can be a valuable tool for optimizing inventory levels and ensuring product availability while minimizing excess stock.

Keywords: Linear Regression; Inventory Management; Demand Forecasting; Fishing Equipment; RMSE.

1. Introduction

Fishing equipment is essential for catching fish in both freshwater and marine environments. Fishing is not only considered a recreational activity that brings personal satisfaction, but it also serves as a means to connect with nature and learn about aquatic ecosystems. For some individuals, fishing is a primary source of income, making the selection of the right fishing gear crucial for productivity. The correct choice of fishing equipment significantly enhances the chances of a successful catch, leading to either personal fulfillment or economic gain. The fishing gear industry, as part of the growing manufacturing sector, faces increasing competition. The variability and unpredictability of consumer demand require more sophisticated inventory management strategies. Consumer needs are influenced by various factors, including seasonal changes, market trends, and economic conditions, which compel manufacturers to adjust their stock levels promptly. Furthermore, issues in the distribution chain, such as delivery delays and logistical disruptions, can negatively impact product availability in the market, leading to risks of overstock or stockouts.

Ineffective inventory management can lead to higher operational costs and reduced profitability. When consumer demand is not met due to insufficient stock, companies lose potential sales and erode consumer trust. On the other hand, overstocking results in higher storage costs and the risk of product obsolescence. Therefore, optimal inventory management, including determining order quantities, timing, and appropriate safety stock levels, is critical to maintaining a balance between supply and demand. To address these challenges, data-driven predictive methods are essential. One relevant approach is the use of data mining techniques, particularly the Linear Regression Algorithm. This algorithm is used to predict product demand based on historical data and relevant variables, such as previous sales volumes and seasonal trends. Linear Regression models the relationship between dependent and independent variables to provide more accurate estimates.

Previous studies have demonstrated the effectiveness of the Linear Regression Algorithm in forecasting product demand and sales. Iksan, Putra, and Udayanti (2018) explored the use of linear regression to predict the demand for motorcycle spare parts. In their study, linear regression was applied to estimate the stock levels required for items like clutch pads and tires over specific periods. The predictions were evaluated using Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE), showing a high level of accuracy in forecasting spare part demand [1]. Another study by Setiawan, Surojudin, and Hadikristanto (2022) applied the Linear Regression Algorithm to forecast drug sales. By using a dataset of pharmaceutical sales, this study successfully predicted monthly sales in the future. The implemented linear regression model provided a clear projection of future sales conditions and can be effectively used for strategic planning in pharmaceutical sales over the next 12 months [2]. In addition, research by Rusdy, Purnawansyah, and Herman (2022) showed the application of linear regression methods in predicting supply and demand for drugs using a Point of Sales application. The results of this study indicate that the linear regression method can be effectively used to estimate supply and demand, providing significant benefits in more efficient inventory management [3].

The application of data mining techniques using the Linear Regression Algorithm enables companies to identify patterns in sales and demand data, which can be used to improve strategic decision-making. By predicting market needs more accurately, companies can optimize inventory management, reduce the risk of overstocking or stockouts, and ultimately enhance operational efficiency and profitability. Based on this background, this study aims to apply the Linear Regression Algorithm to predict distributor demand for fishing products. Through the analysis of historical sales and distribution data, this research is expected to develop an effective prediction model that will assist companies in better inventory planning. The findings from this study are also expected to provide practical contributions to the fishing gear industry and add to the literature on the application of the Linear Regression Algorithm in supply chain management.

2. Research Method

In this research, a systematic approach is employed to predict pharmaceutical data to ensure the study progresses effectively and meets its objectives. The process begins with data collection from various online sources, including information on fishing equipment and relevant literature to support the theoretical framework. This collection is crucial for maintaining a structured and purposeful research approach. Following data collection, the next step is data selection, which involves sampling data based on attributes and volume to ensure its suitability for modeling. The data is grouped according to distributor requests, and irrelevant attributes are removed to refine the dataset. The data selection process involves sampling data randomly, focusing on attributes with the highest volume to form a dataset. This dataset is then categorized based on distributor requests for fishing equipment. Attributes not required for analysis are removed to ensure that only relevant data is retained.

Table 1. Data Selection

Month	Distribution 1	Distribution 2
2/4/2023	4	4
5/4/2023	4	9
6/4/2023	6	5
7/4/2023	4	5
8/4/2023	6	4
9/4/2023	6	8
10/4/2023	4	4
11/4/2023	4	7
12/4/2023	2	4

Subsequent to data selection, data preprocessing is performed to clean the data by removing missing values, duplicates, and inconsistencies. This manual cleaning ensures high-quality data for further analysis. The following table illustrates the cleaned dataset.

Table 2. Data Preprocessing

Month	Distribution 1	Distribution 2
2/4/2023	4	4
5/4/2023	4	9
6/4/2023	6	5
7/4/2023	4	5
8/4/2023	6	4
11/4/2023	4	7
12/4/2023	2	4

Data transformation follows, where the initial data format is converted into a suitable format for algorithms and tools used in the analysis. The dataset prepared for this phase is displayed in the following table.

Table 3. Data Transformation

Month	Distribution 1	Distribution 2
2/4/2023	4	4
5/4/2023	4	9
6/4/2023	6	5
7/4/2023	4	5
8/4/2023	6	4
9/4/2023	6	8
10/4/2023	4	4
11/4/2023	4	7
12/4/2023	2	4

In the modeling phase, data mining techniques, particularly linear regression, are used for predictions. Linear regression helps determine how a dependent variable can be predicted from independent variables, while multiple regression analysis is employed when two or more independent variables affect the dependent variable. This method provides insights into how changes in independent variables influence the dependent variable. The evaluation phase assesses the effectiveness of the linear regression method and the algorithms used. Tools such as RapidMiner are utilized to check if the results align with expectations, and the Least Squares Method measures the model's accuracy.

3. Result and Discussion

3.1 Results

3.1.1 Data Analysis Results

The initial step in this research involves preparing the data, followed by preprocessing steps such as data cleaning, selection, and transformation of fishing equipment transaction data. Initially, 1,550 rows of data were cleaned, resulting in a final dataset of 1,000 rows. This cleaned data was then aggregated monthly, as shown in the table below.

Table 4. Monthly Sales Aggregation Dataset

Date	Distribution 1	Distribution 2
2/4/2023	23	151
5/4/2023	34	379
6/4/2023	11	242
7/4/2023	4	308
8/4/2023	23	132
9/4/2023	11	217
10/4/2023	32	1,749
11/4/2023	8	1,935
12/4/2023	35	910

3.1.2 Testing and Evaluation of Data Using RapidMiner

The evaluation phase involves assessing the performance of a simple linear regression model for predicting stock requirements for new items. This process in RapidMiner Studio begins with importing the necessary data by selecting the "Import Data" option, choosing the dataset, and defining the appropriate attributes and labels. Next, in the "Design" menu, both the training and testing datasets are added to the process panel. Moving to the "Modelling" menu, the "Linear Regression" operator is selected from the "Predictive" submenu to apply the linear regression algorithm to the data. To apply the model, the "Apply Model" operator is used from the "Scoring" menu and dragged to the process panel. Finally, all operators are connected to form the complete process flow, as depicted in Figure 1.

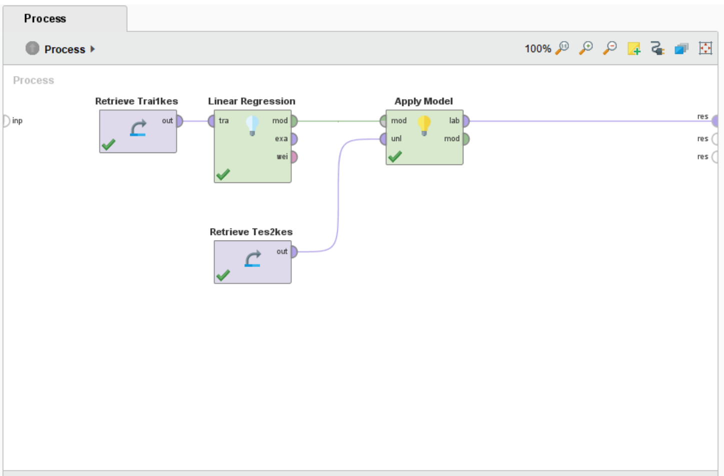

Figure 1. Implementation of Linear Regression Algorithm in RapidMiner Studio

After running the process in RapidMiner, predictions for the testing dataset were generated using the linear regression algorithm, as shown in Figure 2.

1	548	1	16
2	603	2	22
3	539	3	15
4	695	4	32
5	557	5	17
6	997	6	65
7	493	7	10
8	603	8	22
9	594	9	21
10	603	10	22
11	686	11	31
12	539	12	15

Figure 2. Predictions for Testing Dataset Using RapidMiner Studio

3.1.3 Linear Regression Performance Testing
Performance testing evaluates the model's accuracy by comparing the manual calculations with those from RapidMiner. The Root Mean Square Error (RMSE) is computed using RapidMiner’s "Performance Regression" operator.

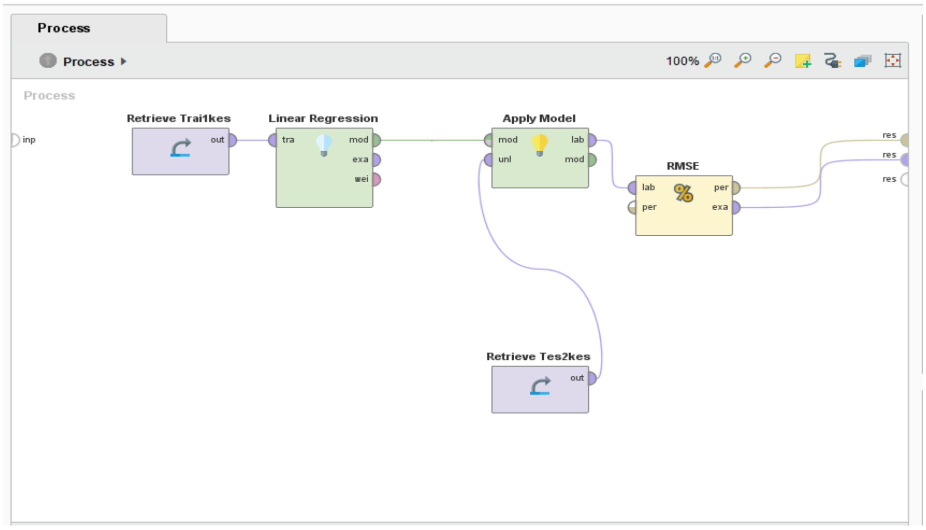

Figure 3. Performance Testing Process in RapidMiner Studio

The RMSE value calculated using RapidMiner is 14.0200 ± 0.000, indicating a small discrepancy between manual calculations and RapidMiner results.

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 140.200 +/- 0.000
```

Figure 4. RMSE Value in RapidMiner Studio

3.1.2 Results Analysis
The linear regression model was applied to a set of 12 testing datasets to predict total sales for the upcoming months, as shown in Table 5.

Table 5 Testing Dataset

No.	Month	Distribution 1	Distribution 2
1	2/4/2023	16	?
2	5/4/2023	22	?
3	6/4/2023	15	?
4	7/4/2023	32	?
5	8/4/2023	17	?
6	9/4/2023	65	?
7	10/4/2023	10	?
8	11/4/2023	22	?
9	12/4/2023	21	?
10	2/4/2023	22	?
11	5/4/2023	31	?
12	6/4/2023	15	?

A comparison between manual calculations and RapidMiner predictions shows no significant differences. The following table compares manual results with those obtained using RapidMiner.

Table 6. Comparison of Manual and RapidMiner Predictions

No.	Month	Distribution 1	Manual Prediction	RapidMiner Prediction
1	2/4/2023	16	548	323.40
2	5/4/2023	22	603	426.15
3	6/4/2023	15	539	393.30
4	7/4/2023	32	695	610.78
5	8/4/2023	17	557	494.50
6	9/4/2023	65	997	1035.32
7	10/4/2023	10	493	501.82
8	11/4/2023	22	603	667.15
9	12/4/2023	21	594	696.89
10	2/4/2023	22	603	747.48
11	5/4/2023	31	686	881.52
12	6/4/2023	15	539	754.80

The predictions for total sales indicate a general increase as the quantity of sales grows over time. This trend is confirmed by the graphical representation shown in Figure 4.

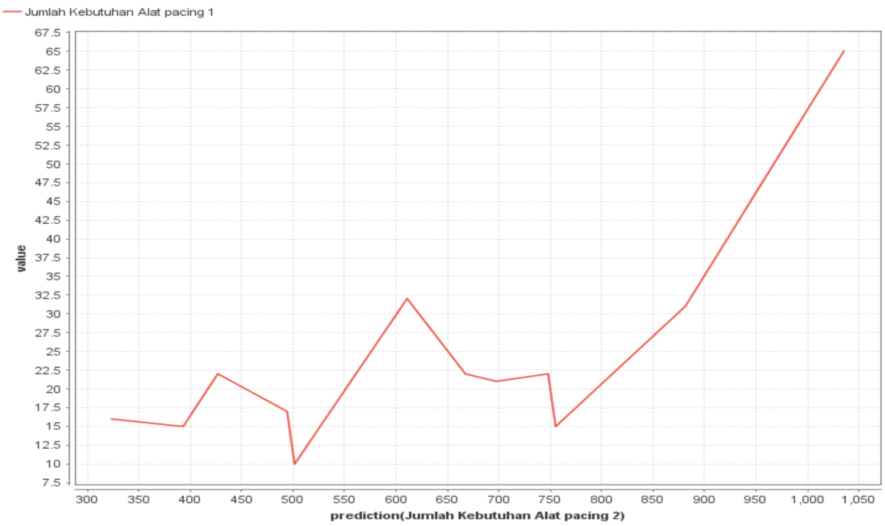

Figure 4. Sales Prediction Graph in RapidMiner Studio

The linear regression model's predictions were evaluated with an RMSE value of 1.40.200 ± 0.000, which is a small discrepancy, indicating that the predictions closely match the manual calculations.

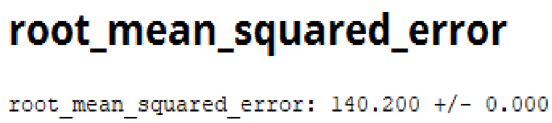

Figure 5. Root Mean Square Error

3.2 Discussion

In the realm of predictive analytics, linear regression remains a foundational method for forecasting future values based on historical data. This is evident from the work of N. Iksan *et al.* (2018), who employed linear regression to forecast the demand for motorcycle spare parts. Their study underscores the effectiveness of linear regression models in predicting future requirements by analyzing past trends and usage patterns [1]. Similarly, D. Setiawan *et al.* (2022) applied linear regression to predict pharmaceutical sales, demonstrating its utility in anticipating future stock needs and guiding inventory management [2].

In the current study, RapidMiner Studio was utilized to implement a linear regression model for predicting stock requirements for new products. The methodological approach involved several key steps. Initially, data was imported into RapidMiner by selecting the "Import Data" option. This process required the identification of relevant attributes and labels necessary for accurate modeling. Following this, both training and testing datasets were incorporated into the process panel via the "Design" menu. This step was crucial for ensuring that the model could be effectively trained and subsequently tested [3][4]. Subsequently, in the "Modelling" menu, the "Linear Regression" operator was selected from the "Predictive" submenu. This operator applied the linear regression algorithm to the dataset, facilitating the prediction of future stock requirements based on historical data. The application of the "Apply Model" operator from the "Scoring" menu allowed the model to be applied to the dataset, generating predictions for new data points. Connecting these operators in the process panel ensured a seamless workflow, as depicted in Figure 1.

The approach employed in RapidMiner Studio aligns with established practices in linear regression modeling, as evidenced by prior research. For instance, A. M. A. Rusdy *et al.* (2022) applied linear regression to predict drug supply and demand, demonstrating its effectiveness in managing inventory levels [3]. Similarly, D. I. Pt *et al.* (2017) used linear regression for inventory control, highlighting its role in optimizing stock levels and reducing excess inventory [6]. Furthermore, E. Rahayu *et al.* (2022) extended the application of linear regression to sales estimation, reinforcing its versatility and applicability across various domains [4]. The use of linear regression in this study not only adheres to established methodologies but also demonstrates the practical application of predictive analytics in stock management. By leveraging historical data to forecast future needs, the study contributes to the broader field of predictive analytics, providing valuable insights into inventory forecasting and stock management practices. The integration of linear regression with tools like RapidMiner Studio exemplifies the practical utility of these models in real-world applications, supported by existing literature and empirical evidence.

4. Related Work

The application of linear regression and other data analysis techniques spans various domains, from inventory management to sales forecasting and beyond. Reviewing the literature reveals both parallels and distinctions between the current study and previous research, each contributing to the broader understanding of predictive analytics. Muhammad & Wahyuni (2022) employed linear regression and single moving average methods to forecast coffee demand. This study is comparable to the current research, which uses linear regression to predict stock requirements for new objects. Both approaches rely on historical data to anticipate future trends, but they apply the methods to different types of products, with coffee versus stock in this case [5]. Similarly, D. I. Pt *et al.* (2017) utilized linear regression in conjunction with the Economic Order Quantity (EOQ) method to optimize inventory levels at PT Indotruck Utama. Their focus on inventory management through predictive modeling aligns with the goals of the current study, which also seeks to enhance stock prediction accuracy. However, their approach integrates EOQ to refine inventory decisions, whereas this study relies solely on linear regression for forecasting [6].

G. N. Ayuni and D. Fitrianah (2019) applied linear regression to predict property sales, demonstrating the method's efficacy in real estate [7]. This is analogous to the current study's application of linear regression for stock prediction, as both use historical sales data to forecast future outcomes. Nonetheless, the specific market dynamics and variables involved in property sales differ from those in stock prediction. In a similar vein, G. Khalda Rifdan *et al*. (2024) utilized linear regression for forecasting residential sales in Jakarta Selatan. This study highlights the adaptability of linear regression across different market sectors, paralleling the current research's focus on stock prediction. The primary difference lies in the product categories and market conditions, which introduce unique variables and challenges [10]. D. Aprilla *et al*. (2013) explored data mining techniques using RapidMiner, a tool also employed in the current study for implementing linear regression models. Their work underscores the tool's capabilities in handling and analyzing large datasets, which facilitates the effective application of linear regression in diverse contexts [12]. This study builds upon this foundation by specifically applying RapidMiner to stock prediction, demonstrating its practical use in a different domain.

In another application, D. Novianty *et al*. (2021) used linear regression to forecast patent applications in Indonesia, showcasing the method's utility in predicting intellectual property trends [13]. While the use of linear regression for patent forecasting shares the predictive aspect with stock prediction, the nature of the data and objectives are distinct, emphasizing different applications of the same technique. A. Hurifiani *et al*. (2024) employed linear regression to predict office supply sales, which is conceptually similar to predicting stock requirements [16]. This study reinforces the versatility of linear regression in commercial forecasting, although the specific products and markets differ from those in the current research. P. Permata (2022) applied linear regression to forecast book orders, illustrating the method's adaptability to educational materials [17]. This bears some resemblance to stock prediction in terms of inventory management, but it focuses on a different product category, thereby highlighting the method's broad applicability. The work of B. S. Kusuma (2015) analyzed the forecasting of bottled water demand using linear regression, comparing least squares and standard error of estimate methods [18]. This study's focus on demand forecasting through linear regression is akin to the current study's aim but targets a different product, illustrating the technique's versatility in handling various forecasting challenges.

Further, studies such as R. Fernando *et al*. (2017), which analyzed the correlation between stroke risk factors and stroke types using the Eclat algorithm, and S. A. J. I. Nugroho's (2020) comparison of fuzzy K-Nearest Neighbor and weighted K-Nearest Neighbor for stroke detection, highlight alternative data analysis approaches that do not directly involve linear regression but demonstrate the broader landscape of predictive analytics [8][11]. These studies underscore the diversity of methodologies available for different types of data analysis and forecasting tasks. Works like R. Planning and D. R. P. Dengan (2011), which focused on optimizing product distribution, and R. Mean *et al*. (n.d.), which emphasized evaluation and validation techniques, provide additional insights into data optimization and validation processes relevant to predictive modeling [20][21]. Although these studies do not use linear regression, they contribute to the understanding of data handling and evaluation in predictive contexts. The current study's use of linear regression for stock prediction builds upon established methodologies found in various fields. While it aligns with previous work in applying linear regression to forecasting, it distinguishes itself by focusing on stock requirements and utilizing RapidMiner Studio for model implementation. This comparison highlights both the shared principles and unique applications of linear regression in predictive analytics.

5. Conclusion

The application of predictive methods to sales transaction data for fishing equipment has proven effective for forecasting future product sales on a monthly basis. This approach enables a systematic analysis of sales data, facilitating accurate predictions and informed decision-making regarding sales strategies and inventory management. The use of a simple linear regression algorithm in this study has demonstrated its capability to model and forecast monthly sales conditions effectively. Evaluations and performance testing using RapidMiner Studio have yielded results that align well with the modeled scenarios. Specifically, the Root Mean Square Error (RMSE) obtained during performance evaluation was 140.200 with a negligible deviation of 0.000. This indicates a high level of accuracy in the predictions made by the model. Overall, this study underscores the value of linear regression methods in sales data analysis and highlights how tools such as RapidMiner Studio can enhance forecasting accuracy and insights into sales trends.

References

[1] Iksan, N., Putra, Y. P., & Udayanti, E. D. (2018). Regresi linier untuk prediksi permintaan sparepart sepeda motor. *Information Technology Engineering Journals, 3*(2), 2548–2157.

[2] Setiawan, D., Surojudin, N., & Hadikristanto, W. (2022). Prediksi penjualan obat dengan algoritma regresi linear. *Prosiding Sains dan Teknologi, 1*(1), 237–246.

[3] Rusdy, A. M. A., Purnawansyah, P., & Herman, H. (2022). Penerapan metode regresi linear pada prediksi penawaran dan permintaan obat: Studi kasus aplikasi point of sales. *Bulletin Sistem Informasi dan Teknologi Islam, 3*(2), 121–126. https://doi.org/10.33096/Busiti.V3i2.1130

[4] Rahayu, E., Parlina, I., & Siregar, Z. A. (2022). Application of multiple linear regression algorithm for motorcycle sales estimation. *JOMLAI: Journal of Machine Learning and Artificial Intelligence, 1*(1), 1–10. https://doi.org/10.55123/Jomlai.V1i1.142

[5] Muhammad, H., & Wahyuni, S. (2022). Generalization of Von-Neumann Regular Rings to Von-Neumann Reg-ular Modules. *Konferensi Nasional Matematika XXI 2022*, *22*, 31.

[6] PT, D. I., Utama, I., & Jakarta, C. (2017). Usulan pengendalian kebutuhan persediaan menggunakan metode Economic Order Quantity di PT. Indotruck Utama Cabang Jakarta. *Spektrum Industri, 15*(1), 1–119.

[7] Ayuni, G. N., & Fitrianah, D. (2019). Penerapan metode regresi linear untuk prediksi penjualan properti pada PT XYZ. *Jurnal Telematika, 14*(2), 79–86.

[8] Algoritma, K., Berbasis, C. P. S. O., & Rohman, R. S. (2020). Komparasi algoritma C4.5 berbasis PSO dan GA untuk diagnosa penyakit stroke. *Vol. 5*(1), 155–161.

[9] Nugraha, D. W., Dodu, A. Y. E., & Chandra, N. (2017). Klasifikasi penyakit stroke menggunakan metode Naive Bayes classifier (Studi kasus pada Rumah Sakit Umum Daerah Undata Palu). *Semantik, 3*(2), 13–22.

[10] Khalda Rifdan, G., Rahaningsih, N., Bahtiar, A., Ali, I., & Dienwati Nuris, N. (2024). Ramalan penjualan rumah menggunakan algoritma linear regresi di Tebet Jakarta Selatan. *Jati: Jurnal Mahasiswa Teknik Informatika, 8*(2), 1847–1851. https://doi.org/10.36040/Jati.V8i2.9022

[11] Nugroho, S. A. J. I. (2020). Naskah publikasi perbandingan metode fuzzy k-nearest neighbor dan neighbor weighted k-nearest neighbor untuk deteksi penyakit stroke.

[12] Aprilla, D., Donny, C. B., Aji, A., Lia, A., & Simri, W. I. (2013). Data mining dengan Rapid Miner. *Produk-Produk Perangkat Lunak Gratis dan Bersifat Open Source*, 1–128.

[13] Novianty, D., Palasara, N. D., & Qomaruddin, M. (2021). Algoritma regresi linear pada prediksi permohonan paten yang terdaftar di Indonesia. *Jurnal Sistem dan Teknologi Informasi, 9*(2), 81. https://doi.org/10.26418/Justin.V9i2.43664

[14] Bayu Anggoro, K., Yuliarty, P., & Anggraini, R. (2020). Analisa kebutuhan produk general lighting di PT X (Distributor lampu LED) dengan metode peramalan. *Industri Inovasi: Jurnal Teknik Industri, 10*(2), 98–104. https://doi.org/10.36040/Industri.V10i2.2652

[15] Fernando, R., Anggraini, L., & Nazir, A. (2017). Analisa keterkaitan risk factor stroke dengan jenis stroke yang diderita menggunakan algoritma Eclat. *Vol. 9789*, 18–19.

[16] Hurifiani, A., Purnamasari, A. I., & Ali, I. (2024). Penerapan algoritma regresi linear untuk prediksi penjualan alat tulis kantor (ATK) di Bumdes. *Jati: Jurnal Mahasiswa Teknik Informatika, 8*(1), 266–273. https://doi.org/10.36040/Jati.V8i1.8305

[17] Permata, P. (2022). Implementasi forecasting pada perencanaan sistem pemesanan buku LKS (Lembar Kerja Siswa) menggunakan algoritma regresi linear (Studi kasus: Toko Buku Darul Ulum, Punggur, Lampung Tengah). *Jurnal Data Mining dan Sistem Informasi, 3*(2), 1. https://doi.org/10.33365/Jdmsi.V3i2.2162

[18] Kusuma, B. S. (2015). Analisa peramalan permintaan air minum dalam kemasan pada PT XYZ dengan metode least square dan standard error of estimate. *Malikussaleh Industrial Engineering Journal, 4*(1), 42–47.

[19] Sederhana, A. R. L. (n.d.). Analisis regresi. *Vol. 0*, 29–52.

[20] Planning, R., & D. R. P. Dengan. (2011). Tugas akhir optimasi distribusi produk melalui pendekatan.

[21] Mean, R., *et al*. (n.d.). Evaluasi dan validasi evaluasi.