Sales Trend Analysis With Machine Learning Linear Regression Algorithm Method

Alwidah Yani Sipahutar[1], Ibnu Rasyid Munthe [2]*, Angga Putra Juledi[3]
[1]Teknologi Informasi, Universitas Labuhanbatu
[2][3]Sistem Informasi, Universitas Labuhanbatu
[1]alwidahyani@gmail.com, [2]ibnurasyid@gmail.com, [3]anggapj19@gmail.com

Abstract: The development of online business in Indonesia is now very rapid, with the process being done by ordering goods through resellers or distributors using one of the social media. Item purchases are made based on product information, prices, discounts and inventory quantities using a decision model. In the sales process, Toko Serbu Aek Batu usually releases several different items to be offered to the market at different prices, but not all items are in high demand. Multiple linear regression is an analysis that describes the relationship between dependent variables and factors that affect more than one independent variable. The purpose of this study is to analyze sales trends using a linear regression method using rapidminer. The results of this study are prediction calculations using manual calculations with rapidminer the same results, predicting the price desired by buyers using a linear regression algorithm with the original price is not much different and rapidminer is very accurate to be used in predicting sales trends at the price desired by customers, so that sellers can pay more attention to things that are very influential in the sales process.

Keywords: Analysis, Trend, Linear Regression, Rapidminer

INTRODUCTION

The development of online business in Indonesia is now very rapid, with the process being done by ordering goods through resellers or distributors using one of the social media (Fabriani & Juanita, 2020). Online shopping is one of the ways to shop through communication tools or social media used in buying and selling transactions, future sales determine sales volume (Ariska et al., 2022). One of the online stores is Serbu Store Aek Batu, the store generates considerable sales every month and year, but the sales results generated are unstable and cannot predict the increase and decrease in sales.

Item purchases are made based on product information, prices, discounts and inventory quantities using a decision model shaped like a tree. In addition, predictions are useful for seeing the state of the products being sold, therefore, the risk of errors due to planning errors must be minimized (Siregar, 2021). Predictions are usually used to find information from a large amount of data that includes satisfied, moderately satisfied and dissatisfied with the sales of Serbu Aek Batu stores. The sales performance of a store, especially the Serbu store of Aek Batu store, is influenced by various factors. In general, these factors are divided into three, namely: location zones, competition, and demographic zones (Mardiatmoko, 2020) and (Supriyatna et al., 2024). In favor of the location and competition zone of Toko Serbu Aek Batu, several studies present contradictory evidence stating that the presence of major competitors or the emergence of many competitors may visibly affect the performance of Toko Serbu Aek Batu. On the other hand, these characteristics may indicate a location with strong economic potential and purchasing power that positively affects the store's performance . Other factors related to the performance of Serbu Aek Batu Shop are demographic zones which include: market potential, demographic characteristics, growth of an area, season, store area, number of goods categories from Serbu Aek Batu Shop, and the number of sales that occur in one period (Hartati et al., 2020).

In this study, the factors that affect sales at the branches of Toko Serbu Aek Batu are analyzed. Transaction activities and services to consumers are increasing every day, so that unknowingly it can result in an increasing pile of data (Purwadi et al., 2019). This big data needs Data Mining processing to be able to produce information that is important for the sustainability of the Aek Batu Serbu Shop, in the sales process the Aek Batu Serbu Shop usually issues several different items to be offered to the market at different prices, but not all items are in high demand (Solihin et al., 2023). The success of its sales determined sustainability for the Aek Batu Serbu Shop itself. In addition to the analysis of these influencing factors, this study also discusses machine learning models that will

be used to predict sales in these stores (Triyanto et al., 2019). Another result that is also expected from this study is a model of the relationship between factors that affect the sales of Toko Serbu Aek Batu.

LITERATURE REVIEW

The prediction system of the number of tonnage of oil palm fruits at PT. Paluta Inti Sawit (PIS) using the web-based PHP program language and MySQL database showed quite good results with the data used in predicting the number of oil palm fruit tonnage, namely the data on the number of oil palm fruit tonnage for 30 days at the PT. Paluta Inti Sawit (PIS). Overall, the predicted results were 99.99%, with the lowest prediction accuracy rate on November 03, 2022 at 88%, while the highest prediction accuracy on November 05, 2022 and November 07, 2022 was 100% (Andrianto & Irawan, 2023).

Predicting the increase in sales turnover at PT Makmur Jaya begins with the calculation process of Multiple Linear Regression then calculating the equation and then producing the desired sales prediction. The system created can be used to predict the increase in sales turnover using the Multiple Linear Regression method with fairly accurate results (Adiguno et al., 2022).

The machine learning system for the prediction model of basic food prices with multiple linear regression methods can be used as a tool to predict daily basic food prices both past, present and future prices (Puteri & Silvanie, 2020).

Predicting vehicle sales using linear regression in the automotive industry in Indonesia using sales data from 2018 to 2023, the Linear Regression model successfully predicts Toyota car sales for the next few months. The prediction results show a number that is quite close to actual sales for January, but in February, March, and April 2024, there is a significant difference between predictions and realizations, which may be caused by seasonal factors or special events that are not captured by the model (Amansyah et al., 2024).

Prediction of vape liquid sales at the Pandaan vapor shop, the linear regression method is very suitable for applying to sales data to predict the sales stock of vape liquid in the coming period, namely in 2021 and in the coming year. This is evidenced by several tests that have been carried out, namely by using the MAPE (Mean Absolute Percentage) accuracy test, the error value obtained is 2 - 3%, so it is not more than 10% (Ababil et al., 2022)

METHOD

The research method used is quantitative with several stages of research which can be seen in figure 1.
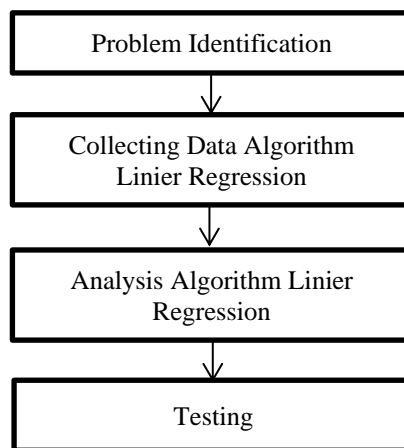

Fig 1. Research Stage

At the Problem Identification stage, data was collected through interviews with Mrs. Indari as the owner of the Serbu shop. After obtaining the data, the next stage is to collect linear regression data. The data used are product prices, sales, quality and also price discounts from 2021 to 2023. The next stage is to analyze the linear regression algorithm using Microsoft Excel and XLSTAT. At the testing stage, the rapidminer application was used to make the measurement results more accurate (Hartono & Widiantoro, 2024).

RESULT

At this stage, data processing and testing techniques are carried out using applications using training data. The following is the training data that will be entered into the application, which can be seen in table 1.

Table 1 Data Training

No	Price (Rp)	Stock (Unit)	Sales (Unit)

* Ibnu Rasyid Munthe

1	50.000	100	30
2	55.000	90	28
3	60.000	85	25
4	45.000	110	35
5	70.000	80	20
6	65.000	75	18
7	50.000	95	32
8	55.000	100	29
9	58.000	88	27
10	48.000	105	33
11	75.000	70	15
12	52.000	92	31
13	62.000	82	22
14	47.000	98	34
15	68.000	78	17
16	56.000	87	30
17	59.000	89	26
18	44.000	115	36
19	63.000	83	21
20	49.000	98	28
21	67.000	76	19
22	54.000	97	32
23	46.000	105	30
24	64.000	84	23
25	51.000	96	27
26	57.000	88	29
27	66.000	79	16
28	53.000	94	31
29	50.000	102	34
30	55.000	91	25

The comparison of data used in training data and testing data is 70 : 30. The Testing data used can be seen in table 2.

Table 2 Data Testing

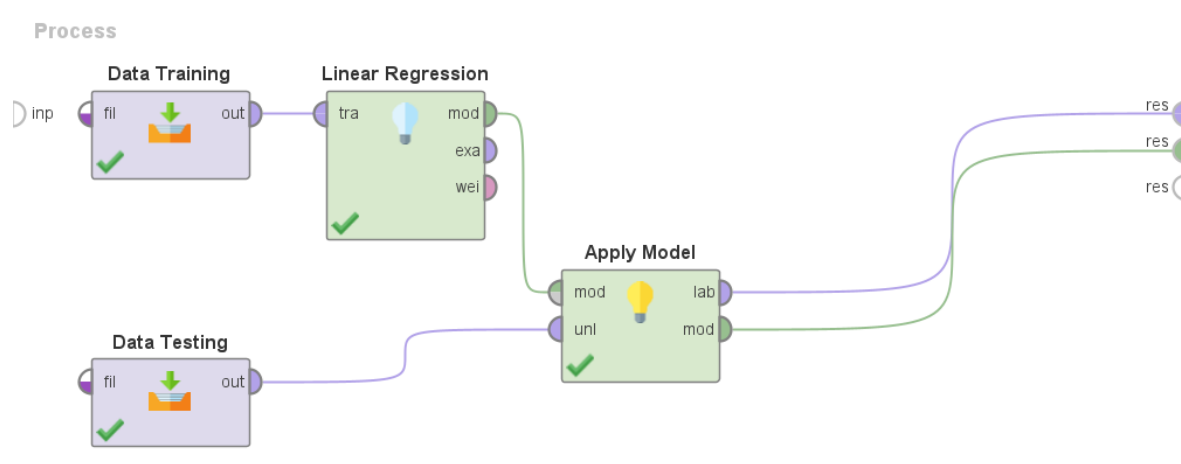| No	Price (Rp)	Stock (Unit)	Sales (Unit)
1	62.000	82	24
2	53.000	97	31
3	67.000	74	17
4	49.000	101	29
5	64.000	80	22
6	52.000	92	27
7	70.000	73	16
8	57.000	89	30
9	60.000	98	26

* Ibnu Rasyid Munthe

Fig 2. RapidMiner Data Testing Process

The operator used in the data processing process to data testing. Some of the operators used are read excel 1 which is used to read excel data (the data used is training data), read excel 2 (the data used is testing data), linear regression is used to process data using the liner regression algorithm, and apply model is used to process the entire data and display the results.

Next, import the training data. In the data input process, determine the attributes that will be used as labels. In this process the average attribute will be used as a label. The data used is the data in table 1. The next step is to enter the testing data that is already in table 2 In the testing data, the average value is left blank to find the price that will be predicted by the system later After all the processes are completed and run, the results of the linear regression that have been obtained will appear. The prediction results obtained from the test data that has been processed by the application are shown in Figure 3.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Stok (Unit)	0.500	0.000	0.659	0.180	∞	0	****
Penjualan (Unit)	0.500	0.000	0.363	0.180	∞	0	****
(Intercept)	-0.000	0.000	?	?	0	1	

Fig 3. Liner Regression Result

DISCUSSIONS

Coefficient results are positive in stock attributes, it can be interpreted that the more stock, the more likely the goods will be sold, otherwise if the results are negative, it can be interpreted that the more stock available, the less likely the goods will be sold and based on testing using the Rapidminer application, there is a star symbol which means that the more stars there are, the more influential the attribute is, and vice versa, if there are fewer stars, the less influential the attribute is on sales.

CONCLUSION

Based on the research that has been carried out on the Aek Batu Serbu Shop, it can be concluded that the prediction calculation using manual calculation with Rapidminer has the same results, the results of the price prediction desired by the buyer using a linear regression algorithm with the original price are not much different and Rapidminer is very accurate to be used in predicting sales trends at the price desired by customers, So that sellers can pay more attention to things that have a great influence on the sales process.

REFERENCES

Ababil, O. J., Wibowo, S. A., & Zahro, H. Z. (2022). Penerapan Metode Regresi Linier Dalam Prediksi Penjualan Liquid Vape Di Toko Vapor Pandaan Berbasis Website. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(1), 186–195.

Adiguno, S., Syahra, Y., & Yetri, M. (2022). Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(4), 275. https://doi.org/10.53513/jursi.v1i4.5331

Amansyah, I., Indra, J., Nurlaelasari, E., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linear : Studi Kasus pada Industri Otomotif di Indonesia. *INNOVATIVE: Journal Of Social Science Research*, *4*(4), 1199–1216.

Andrianto, R., & Irawan, F. (2023). Implementasi Metode Regresi Linear Berganda Pada Sistem Prediksi Jumlah

* Ibnu Rasyid Munthe

Tonase Kelapa Sawit di PT . Paluta Inti Sawit. *Jurnal Pendidikan Tambusai*, *7*(1), 2926–2934.

Ariska, A. M., Irawati, N., & Muhazir, A. (2022). Penerapan Elektronik Customer Relationship Management (E-CRM) Dalam Penjualan Roti Berbasis Web. *Jurnal Media Informatika Budidarma*, *6*(2), 1090. https://doi.org/10.30865/mib.v6i2.4002

Fabriani, S., & Juanita, S. (2020). Implementasi Electronic Relationship Management (E- Crm) Pada Beauty Karlina Salon Untuk Meningkatkan. *Idealis*, *3*(1), 381–385.

Hartati, E., Indriyani, R., & Trianingsih, I. (2020). Analisis Kepuasan Pengguna Website SMK Negeri 2 Palembang Menggunakan Regresi Linear Berganda. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *20*(1), 47–58. https://doi.org/10.30812/matrik.v20i1.736

Hartono, P. C., & Widiantoro, A. D. (2024). Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner. *Journal of Computer and Information Systems Ampera*, *5*(3), 2775–2496. https://journal-computing.org/index.php/journal-cisa/index

Mardiatmoko, G. (2020). Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Berganda (Studi Kasus Penyusunan Persamaan Allometrik Kenari Muda [Canarium Indicum L.]. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *14*(3), 333–342.

Purwadi, P., Ramadhan, P. S., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, *18*(1), 55–61. https://doi.org/10.53513/jis.v18i1.104

Puteri, K., & Silvanie, A. (2020). Machine Learning untuk Model Prediksi Harga Sembako. *Jurnal Nasional Informatika*, *1*(2), 82–94.

Siregar, A. Z. (2021). Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru. *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, *2*(3), 133–137. https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73

Solihin, R. H., Supratman, J., & Muhendra, R. (2023). Pengembangan Model Peramalan Penjualan Menggunakan Metode Regresi Linier dan Polinomial Pada Industri Makanan Ringan (Studi Kasus: CV. Stanley Mandiri Snack). *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, *10*(2), 185–192. https://doi.org/10.37373/tekno.v10i2.456

Supriyatna, A., Irma Purnamasari, A., & Ali, I. (2024). Analisis Penjualan Produk Umkm Di Shopee Pada Toko Agung0Na9 Menggunakan Model Algoritma Regresi Linear. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1911–1915. https://doi.org/10.36040/jati.v8i2.8372

Triyanto, E., Sismoro, H., & Laksito, A. D. (2019). Implementasi Algoritma Regresi Linear Berganda Untuk Memprediksi Produksi Padi Di Kabupaten Bantul. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *4*(2), 66–75. https://doi.org/10.36341/rabit.v4i2.666

* Ibnu Rasyid Munthe