International Journal on Advanced Technology, Engineering, and Information System
https://ojs.transpublika.com/index.php/IJATEIS
Online ISSN 2828-5425

https://doi.org/10.55047/ijateis.v4i1.1587

Original Article

The Use of Single Moving Average and Linear Regression in Spare Part Sales Forecasting at PT. CNC

Eko Prasetiamaolana[1*], Mohammad Syafrullah[2]

[1,2]Master of Computer Science Study Program, Faculty of Information Technology, Budi Luhur University, Indonesia
Email: [1)] 2011600687@student.budiluhur.ac.id, [2)] mohammad.syafrullah@budiluhur.ac.id

Received : 18 December - 2024
Accepted : 24 January - 2025
Published online : 28 January - 2025

Abstract

Given the highly competitive nature of Indonesia's automotive sector, accurate sales forecasting has become a crucial business strategy. This research investigates the application of Single Moving Average and Linear Regression methods for forecasting spare part sales at PT. CNC, an automotive spare parts manufacturer in Indonesia. The study analyzes monthly sales data from January 2019 to December 2022, employing both Single Moving Average and Linear Regression forecasting methods. Model performance was evaluated using multiple accuracy metrics including Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), with data normalized using min-max normalization. The analysis yielded error metrics of MSE = 0.043, RMSE = 0.208, MAE = 0.005, and MAPE = 4.36%, demonstrating the effectiveness of these forecasting methods for spare part sales prediction.

Keywords: Sales, Forecasting, Single Moving Average, Linear Regression, Automotive Sector.

1. Introduction

Over time, the automotive industry in Indonesia has developed significantly, as evidenced by the increasing number of motor vehicles and foreign automotive manufacturers establishing production and business operations in the country. Along with this development, there is also a need for advancements in information technology to assist in problem-solving (Rakasyiwi et al., 2022). The information and communication revolution in this era has given rise to a new civilization that is not bound by time; information can be obtained anytime and anywhere. Likewise, in the business world, technological advancements have led to significant changes, facilitating the execution of various tasks.

PT. CNC is one of the companies in Indonesia engaged in automotive manufacturing, specifically in the production of motorcycle spare parts. It was founded in 2001, located at Jl. Akasia 2 Block AE No 25, Delta Silicon Industrial Area, Lippo Cikarang, Bekasi, West Java. In its operations, PT. CNC produces approximately 480 different spare parts daily to meet customer demand. The spare parts produced daily include items such as Handle Comp, Bracket Engine, Side Stand, Stand Comp Center, Footrest, and many others.

In recent years, PT. CNC's spare parts sales data has experienced significant fluctuations. In 2019, sales reached 14,052,094 units, but decreased in 2020 to 13,319,640 units. In 2021, sales rebounded to 14,359,753 units and then saw a substantial increase in 2022, rising to 25,435,589 units, marking a growth of 77.13%. While this is a positive development for the company, it can become a major problem if the company is not prepared. Sales forecasts have

been made manually, leading to discrepancies between production and actual demand. This mismatch can result in delayed shipments to customers (when production is insufficient) and inventory buildup in warehouses (when production exceeds demand). Therefore, there is a need for a method to predict future spare part sales using data from past sales.

Numerous studies have been conducted to improve sales predictions for products. One such method is K-Nearest Neighbor (KNN), which can predict which products are in high demand and which are less popular (S. P. Dewi et al., 2022). Another commonly used method is the Moving Average (Azis & Kustanto, 2023). For instance, a study predicting the sales of Accu used a moving average with five and seven periods. The results showed an error rate of 7.72% and an accuracy rate of 92.28% for the five-month forecast, and an error rate of 7.28% with an accuracy rate of 92.72% for the seven-month forecast. Another method used for sales predictions is the algorithm model (K. R. Dewi & Mauladi, 2020). Additionally, Linear Regression is widely used (Indarwati et al., 2019). A study predicting smartphone sales using Linear Regression yielded a MAPE value of 0.032 and an MSE value of 5.16.

Based on the above discussion and referring to previous studies on sales prediction, the authors are interested in testing and comparing two forecasting methods: Single Moving Average and Linear Regression, for predicting spare part sales. The purpose of using these two forecasting methods is to compare which method provides a more accurate prediction, closely aligned with actual values.

2. Literature Review

2.1. Data Mining

Data mining is the process of discovering meaningful relationships, patterns, and trends by examining a set of data that is accurate and stored in a database using pattern recognition techniques such as statistical tables from mathematics (Kamal & Ilyas, 2017).

2.2. Cross Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM is a data mining methodology developed by a consortium of European Commission companies in 1996 and has been established as the standard process in data mining (Setiawan, 2016). The CRISP-DM methodology consists of six phases in data mining research and development: Business Understanding Phase, Data Understanding Phase, Data Preparation Phase, Modeling Phase, Evaluation Phase, and Deployment Phase.

2.3. Forecasting

Forecasting is the process of estimating a future event by examining past occurrences (Azis & Kustanto, 2023). Forecasting is performed by taking historical data and projecting it into the future using mathematical models.

2.4. Sales

Sales is an activity in which a seller and a buyer exchange goods. According to Dewi et al. (2022), sales refers to the transaction of goods or services between two parties in exchange for money from one party. On the other hand, Hasmawati and Muchtar (2017) defines sales as an economic activity commonly conducted by companies to earn profits proportional to the efforts undertaken.

2.5. ARIMA (Autoregressive Integrated Moving Average)

The ARIMA model is commonly used in statistical and econometric calculations for time series analysis. ARIMA is an extension of the Autoregressive Moving Average (ARMA) model

combined with the Box-Jenkins method, which is typically used to predict future values based on a range of present and past values (Yuliyanti & Arliani, 2022).

2.6. Linear Regression

$$Y=a+b1X1+b2X2+b3X3+\cdots+bnXn(1)Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n \quad (1)$$

Linear regression is a statistical method used to make predictions by developing a mathematical relationship between the dependent variable (Y) and independent variables (X) (Ayuni & Fitrianah, 2019). The dependent variable is influenced by the independent variables. Predictions for the dependent variable can be made if the independent variables are known. The general formula for calculating linear regression is shown in equation (1) below:

Explanation:
YY = Dependent variable
X1,X2,X3,...,XnX_1, X_2, X_3, \dots, X_n = Independent variables
aa = Constant term
b1,b2,b3,...b_1, b_2, b_3, \dots = Regression coefficients

2.7. Single Moving Average Method

The single moving average method is a forecasting technique that involves taking a group of observations, calculating the average value, and using it as the prediction for the upcoming period. This method is often referred to as a moving average because each time new observation data is available, a new average is calculated and used for prediction (Azis & Kustanto, 2023). A key characteristic of this method is that the prediction for the upcoming period relies on historical data over a certain period. A longer moving average period results in a smoother average. The formula used to calculate the single moving average is equation (2) below:

$$Mt=Yt+1−(Yt+Yt−1+Yt−2+\cdots+Yt−nn)(2)M_t = Y_{t+1} - \left(\frac{Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-n}}{n}\right) \quad (2)$$

Explanation:
MtM_t = Moving average for period tt
Yt+1Y_{t+1} = Predicted value for the next period
YtY_t = Actual value at period tt
nn = Number of periods in the moving average

2.8. Winter's Exponential Smoothing Method

Winter's Exponential Smoothing is a method used to make predictions when time series data exhibits trend and seasonal patterns. The Winter's method is based on three smoothing equations: the overall smoothing equation, the trend smoothing equation, and the seasonal smoothing equation. The advantage of this smoothing method is its ability to provide accurate short-term predictions, with quick adjustments at low costs.

2.9. Prediction Accuracy Calculation

The accuracy of predictions for each model can be evaluated by comparing the predicted values with the actual values. According to Maricar (2019), the smaller the error value, the higher the accuracy of the prediction, and vice versa. The magnitude of the prediction error

can be calculated using several error calculation methods. Generally, there are four types of calculations used to assess the degree of error in predictions:

A. MSE (Mean Squared Error)

Mean Squared Error (MSE) is calculated by summing the squares of all prediction errors for each period and dividing by the total number of prediction periods. A smaller MSE value indicates a smaller prediction error. MSE can be calculated using the formula in equation (3) below:

$$MSE=\sum(At-Ft)2n(3)MSE = \frac{\sum (A_t - F_t)^2}{n} \quad (3)$$

Explanation:
AtA_t = Actual value at period tt
FtF_t = Predicted value at period tt
nn = Number of prediction periods

B. RMSE (Root Mean Squared Error)

Root Mean Squared Error (RMSE) is an error measurement method that assesses the difference between predicted values and observed values. RMSE is the square root of MSE. The accuracy of the RMSE method is indicated by a smaller RMSE value. RMSE can be expressed using the formula in equation (4) below:

$$RMSE=\sum t=1n(At-Ft)2n(4)RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}} \quad (4)$$

Explanation:
AtA_t = Actual value at period tt
FtF_t = Predicted value at period tt
nn = Number of prediction periods

C. MAE (Mean Absolute Error)

Mean Absolute Error (MAE) is an error measurement method that calculates the average of the absolute differences between actual and predicted values. MAE can be calculated using the formula in equation (5) below:

$$MAE=1n\sum t=1n|At-Ft|(5)MAE = \frac{1}{n} \sum_{t=1}^{n} |A_t - F_t| \quad (5)$$

Explanation:
AtA_t = Actual value at period tt
FtF_t = Predicted value at period tt
nn = Number of prediction periods

D. MAPE (Mean Absolute Percentage Error)

Mean Absolute Percentage Error (MAPE) is the mean absolute percentage error. This method is easy to understand and can be used to assess forecasting accuracy. A smaller MAPE percentage error value indicates a more accurate forecast. MAPE can be calculated using the formula in equation (6) below:

$$MAPE=100n\sum|At-FtAt|(6)MAPE = \frac{100}{n} \sum \left|\frac{A_t - F_t}{A_t}\right| \quad (6)$$

Explanation:
AtA_t = Actual value at period tt
FtF_t = Predicted value at period tt
Nn = Number of prediction periods

3. Methods

The method used in this study is experimental, utilizing secondary data and data mining techniques. This research is categorized as quantitative as it involves processing data to extract information for decision-making purposes. The study applies an experimental approach using the Single Moving Average method and Linear Regression method. The final result of this study is the identification of an appropriate method for predicting future sales.

3.1. CRISP-DM Methodology

In the CRISP-DM methodology, there are six phases: Business Understanding, which is the phase where the researcher identifies the problem and solutions to be achieved through user interviews. The next phase is Data Understanding, during which the researcher collects and analyzes quantitative data. Data Preparation follows, which involves processing raw data obtained in file form. Modeling is the phase where predictions are calculated using the Single Moving Average and Linear Regression methods.

The Single Moving Average method is represented by the following equation:

$$M_t = Y_{t+1} - \left(\frac{Yt + Yt\text{-}1 + Yt\text{-}2 + \ldots + Yt\text{-}n}{n}\right)$$

Linear Regression with the following equation

$$Y = a + b1X1 + b2X2 + b3X3 + \ldots + bnXn$$

$$X^2 = (1)*(1) = 1$$

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n\sum X^2 - (\sum X)^2}$$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$F = a + bx$$

The next phase is Evaluation, where the results of applying the Single Moving Average and Linear Regression methods are assessed. The results obtained are expected to align with the research objectives, as outlined in the Business Understanding phase. The system testing or evaluation phase will be carried out by normalizing sales data first using the Min-Max Normalization method, followed by calculating the accuracy of both algorithms. Accuracy testing will be done using MSE, RMSE, MAE, and MAPE metrics with libraries in PHP programming language and testing tools available in MS Excel applications.

Deployment follows, where, based on the evaluation results, the next step is to create a prototype. In this phase, the prototype will be developed using the PHP programming language. The final report of the research will be the output of the calculations performed within the system.

4. Results and Discussion

4.1. Business Understanding

Business Understanding is the first stage in the CRISP-DM methodology. This stage encompasses several steps, including understanding the business objectives of PT. CNC, which focuses on the sales of motorcycle spare parts. The situation assessment is conducted through direct interviews, problem identification, and literature review. A major issue identified in the sales process is the difficulty in predicting sales for the upcoming months. The final stage is to define the data mining objective, which aims to predict future sales based on historical sales data.

4.2. Data Understanding

The data used in this study consists of sales data from the past four years, spanning from January 2019 to December 2022, at PT. CNC. This data will serve as the foundation for predicting future sales of spare parts. The first step involves collecting the sales data to be utilized in the study. This data will be processed using Microsoft Excel and a prototype developed by the author using the PHP programming language, incorporating the Single Moving Average and Linear Regression methods. Accuracy testing will be conducted using MSE, RMSE, MAE, and MAPE.

4.3. Data Preparation

The next step, after obtaining the sales dataset, is to aggregate all the sales data on a monthly basis for each year. The dataset used in this study includes sales data spanning four years, from January 2019 to December 2022. The sales data for motorcycle spare parts from 2019 to 2022 is presented in Table 1.

Table 1. Spare Part Sales Data 2019-2022

Month	Total Sales Per Year			
	2019	2020	2021	2022
January	1.225.009	1.313.826	766.53	2.596.365
February	1.059.752	1.094.576	881.969	2.695.483
March	1.159.489	1.323.849	1.002.420	3.239.672
April	1.243.008	588.456	1.213.749	3.145.561
May	1.270.424	564.809	1.178.646	1.946.590
June	1.018.442	387.754	1.072.591	2.110.323
July	1.238.394	649.895	1.031.920	2.711.323
August	1.198.545	499.278	1.185.419	3.143.299
September	1.225.449	1.918.546	1.463.563	1.348.906
October	1.258.623	2.002.540	1.556.577	972.235
November	1.161.771	952.101	1.509.897	909.13
December	993.188	2.024.010	1.496.472	616.702
Total	14.052.094	13.319.640	14.359.753	25.435.589

4.4. Modelling

In this phase, prediction modeling is conducted on spare part sales data from the past four years using the Single Moving Average and Linear Regression methods. The discussion of each method employed is as follows:

A. Single Moving Average

The first experiment was conducted using the Single Moving Average method. This experiment utilized a 3-period single moving average calculation on spare part sales data from January 2019 to December 2023 to forecast sales from January to July 2023. The process

involved calculating the average sales for the first three months to generate the prediction for the fourth month. Subsequently, the average of months 2 through 4 was calculated to predict the sales for the fifth month, and so on, until predictions for January to July 2023 were obtained. Based on the sales data and following equation (2), the predicted spare part sales for July 2023 were calculated to be 1,011,436 units.

The second experiment was conducted using the Single Moving Average method with a 5-period calculation on spare part sales data from January 2019 to December 2022 to forecast sales from January to July 2023. In this process, the average sales from months 1 to 5 were computed to predict sales for the 6th month, followed by calculating the average of months 2 to 6 to predict sales for the 7th month, and so on, until predictions for January to July 2023 were obtained. Based on equation (2), the predicted spare part sales for July 2023 were 966,699 units.

B. Linear Regression

The next method employed was forecasting using Linear Regression on spare part sales data from January 2019 to December 2022. After calculating the total values of X^2 and XY for months 1 to 48 using equation (1), the results obtained were X^2 = 38,024 and XY = 1,845,359,968. The next step was to determine the values of a (constant) and b (regression coefficient) for Linear Regression calculations. The value of a (constant) was determined using equation (8), yielding a value of 868,019.91. The value of b (regression coefficient) was found using equation (9), resulting in a value of 21,685.47. The third step involved calculating the predicted spare part sales for the first month using equation (10). From this calculation, the predicted spare part sales for January 2019 were 889,705.39 units. Meanwhile, the actual sales for the 49th month, January 2023, were 1,930,608 units.

C. Accuracy Calculation

Before calculating the accuracy of the predictions, the next step involved normalizing the sales data using Min-Max normalization. The normalized data can be seen in Table 2.

Table 2. Normalized Sales Data

Month	Total Sales After normalisation			
	2019	2020	2021	2022
January	0,293576113	0,324719014	0,132814478	0,77443005
February	0,235630197	0,247840927	0,173292149	0,809184906
March	0,2706021	0,328233491	0,215527235	1
April	0,299887304	0,070374394	0,289627893	0,967000804
May	0,309500484	0,062082781	0,277319334	0,546592153
June	0,221145208	0	0,24013208	0,604003692
July	0,298269445	0,091917439	0,225871151	0,814739063
August	0,284296743	0,039104911	0,279694227	0,966207654
September	0,293730395	0,536758771	0,377222978	0,337019508
October	0,305362567	0,566210529	0,40983752	0,20494313
November	0,271402263	0,197883319	0,339346959	0,182815915
December	0,212290115	0,573738796	0,388762229	0,080278605
Total	3,295,692,934	3,038,864,371	340,357,086	7,287,215,481

After normalization, the next stage involves calculating accuracy using the following methods: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These accuracy calculations aim to compare the predicted sales results for spare parts with the actual sales data. The first calculation uses the Mean Square Error (MSE) method, as expressed in Equation (3). The

accuracy results for predictions over three periods using the MSE method yielded an average value of 0.043632723. Meanwhile, the MSE value for five-period predictions using the Single Moving Average method was 0.05223519. Subsequently, the accuracy calculation for the Linear Regression method using MSE resulted in a value of 0.08.

The second calculation employs the Root Mean Square Error (RMSE) method. RMSE is determined by squaring the prediction error, dividing it by the total data count (average), and then taking the square root. The RMSE can be calculated using Equation (4). The accuracy results using the RMSE method for normalized three-period predictions yielded an average value of 0.208884474. For normalized five-period predictions, the RMSE value was 0.228550191. The accuracy result for the Linear Regression method using RMSE was calculated to be 0.28. The third accuracy calculation is conducted using the Mean Absolute Error (MAE) method. MAE calculates the average absolute difference between actual values and predicted or forecasted values. MAE can be derived from Equation (5). The accuracy results for single moving average three-period predictions using MAE yielded an error value of 0.005668385. For five-period predictions, the MAE value was 0.00435331. The accuracy calculation for Linear Regression using the MAE method resulted in an error value of 0.05. The fourth accuracy calculation is conducted using the Mean Absolute Percentage Error (MAPE) method. MAPE is a measure of prediction accuracy in forecasting methods. The MAPE value provides information on the magnitude of forecasting errors relative to the actual series values. The smaller the percentage error (MAPE), the more accurate the forecast. MAPE can be calculated using Equation (6). The accuracy results for single moving average three-period predictions using MAPE yielded a value of 4.36%. For five-period predictions, the MAPE value was 22.86%. The accuracy calculation for the Linear Regression method using MAPE resulted in a value of 5.13%.

4.5. Evaluation

At this stage, an evaluation is carried out from the results of the accuracy calculation using MSE, RMSE, MAE, and MAPE. the results of the application of the research using the Single Moving Average and Linear Regression methods. The results obtained are expected to be in accordance with the research objectives, according to the business understanding stage. Based on the results of the forecasting experiment using the Single Moving Average and Linear Regression methods, the results are as presented in Table 3.

Table 3. MSE, RMSE, MAE and MAPE test results

No	Method	MSE	RMSE	MAE	MAPE
1	Single Moving Average 3 Period	0,043	0,208	0,005	4,36%
2	Single Moving Average 5 Period	0,052	0,228	0,004	22,86%
3	Linear Regression	0,08	0,28	0,005	5,13%

Table 3 shows the results of the analysis that has been done in predicting automotive spare part sales using the Single Moving Average and Linear Regression methods and normalization of the data, the MSE value is 0.043, RMSE is 0.208, MAE is 0.005, and MAPE value is 4.36% in the single moving average method with 3 periods. While the Single Moving Average method with 5 periods gets an MSE value of 0.052, RMSE is 0.228, MAE is 0.004 and MAPE value is 22.86%. In the Linear Regression method, the MSE value is 0.08, RMSE is 0.28, MAE is 0.005 and MAPE is 5.13%. Based on the analysis, the Single Moving Average method proved most suitable for predicting spare part sales, demonstrating high accuracy with a low MAPE value of 4.36%

4.6. Deployment

Based on the results of the evaluation in the previous stage, it was found that the Single Moving Average method with a 3-period configuration demonstrated better accuracy performance compared to the Linear Regression method. This was evident from the experimental results, where the Single Moving Average method with a 3-period configuration achieved significantly lower MSE, RMSE, MAE, and MAPE values than the Linear Regression method.

5. Conclusion

The comparison between the Single Moving Average method and the Linear Regression method revealed that the Single Moving Average method yielded lower error rates. Accuracy calculations, performed after data normalization, resulted in low error values. Sales predictions for spare parts using four years of historical sales data and the Single Moving Average method varied depending on the moving period used. This study only compared two methods; future research is encouraged to incorporate additional methods to gain comprehensive insights into which approach delivers the best results.

6. References

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

Azis, A. K., & Kustanto, K. (2023). Penerapan Moving Average Pada Prediksi Penjualan Accu. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *11*(1), 25–34.

Dewi, K. R., & Mauladi, K. F. (2020). Analisa Algoritma C4. 5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, *4*(3), 109–114.

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 639–648.

Hasmawati, J. N., & Muchtar, M. (2017). *Aplikasi prediksi penjualan barang menggunakan metode k-nearest neighbor (knn)(studi kasus tumaka mart)*. Semantik.

Indarwati, T., Irawati, T., & Rimawati, E. (2019). Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *6*(2).

Kamal, I. M., & Ilyas, R. (2017). Prediksi penjualan buku menggunakan data mining di pt. niaga swadaya. *Semnasteknomedia Online*, *5*(1), 1–2.

Maricar, M. A. (2019). Analisa perbandingan nilai akurasi moving average dan exponential smoothing untuk sistem peramalan pendapatan pada perusahaan xyz. *Jurnal Sistem Dan Informatika (JSI)*, *13*(2), 36–45.

Rakasyiwi, G. R., Witjaksana, B., & Tjendani, H. T. (2022). Project Scheduling Analysis Using The Critical Path Method–Case Study: Subsidized House Construction Project In Hill Mulya Housing, Samarinda City. *International Journal On Advanced Technology, Engineering, And Information System*, *1*(4), 73–88.

Setiawan, R. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, *3*(1), 76–92.

Yuliyanti, R., & Arliani, E. (2022). Peramalan jumlah penduduk menggunakan model arima. *Jurnal Kajian Dan Terapan Matematika*, *8*(2), 114–128.