Jurnal Sisfotek Global
ISSN (Online): 2721 - 3161, ISSN (Print): 2088 – 1762
DOI: http://dx.doi.org/10.38101/sisfotek.v15i1.15822
Vol. 15, No. 1, March 2025, pp. 9-19

Comparison of the Performance of Multiple Linear Regression and Multi-Layer Perceptron Neural Network Algorithms in Predicting Drug Sales at Pharmacy XYZ

Danang Arifuddin[1], Kusrini[2] & Kusnawi[3]

[1,2,3] Program Studi S2 Teknik Informatika Program Pascasarjana Universitas AMIKOM Yogyakarta
E-mail: [1]danang22551202@students.amikom.ac.id, [2]kusrini@amikom.ac.id, [3]khusnawi@amikom.ac.id

ARTICLE HISTORY

Received : January 24th, 2025
Revised : February 8th, 2025
Accepted : March 25th, 2025

KEYWORDS

Drug Sales Prediction,
Machine Learning Algorithms,
Multiple Linear Regression (MLR),
Multi-Layer Perceptron Neural Network (MLPNN),
Stock Management Optimization

ABSTRACT

The needs of better drugs management tool especially that can predict specific drugs consumption volume are needed by any healthcare facility including retail pharmacies. Thus, finding better prediction algorithm with suitable variable internally and externally becoming this research objectives. The research compares correlation score and histogram of each predictor variable with target variable and further input the selected variable into MLR and MLPNN algorithm to find recommended algorithm with better MSE and MAPE. The findings indicate that MLPNN with backpropagation method slightly outperforms MLR with 'h-7' as single input variable but with unstable predictions with lower MSE of 19588 and MAPE of 22,3%. While MLR's MSE of 22346,129 and MAPE of 25.4% with 'h-7' and 'bm' as input variable perform stable prediction. Finally, the research find 'h-7' is the most significant variable among other variables and both MLR and MLPNN are both need better improvement to perform drugs prediction analysis.

1. Introduction

Drug shortages are a global challenge in healthcare facility, disrupting patient treatment and highlighting the importance of a stable drug supply [1], [2], [3], [4], [5]. Stock management must also consider additional costs such as replacement, shipping, and storage, which can increase drug prices. While the problem is still in there, the lack of in-depth studies on predictive solutions and drug shortage management is a concern, despite the growing need for efficient stock [5], [6].

Research in this field in the future should not only be limited to Explanatory Data Analysis (EDA) but will use a lot of data mining algorithms for forecasting by not only focusing on internal variables but also including external variables that may be triggers for drug consumption levels as part of the openness of artificial intelligence to various factors that affect the supply chain [1]. The COVID-19 pandemic demonstrated that social, economic, and environmental changes triggered surges in drug demand, particularly analgesics [7], [8]. Factors such as commodity prices, inflation, and weather also influence drug consumption [9], [10]. Data mining algorithms that take into account advanced models like MLPNN and hybrid algorithms, have the potential to improve the accuracy of drug stock [11], [12]. Considering all the reason, this research focuses on the use of machine learning algorithms to predict analgesic consumption using both external and internal factors as variable predictor.

Several studies have explored forecasting methods in drug supply management. Neural networks found outperform linear time series models but face challenges in weight selection [1]. With the use of time trend graphical analysis, research highlighted the impact of the Covid-19 pandemic on drug demand, particularly for paracetamol, and emphasized pharmacies as key demand indicators[7]. The used of data mining on French hospital logistics, identifying shipment time as a predictor of future supply conditions [5]. Research analyzed pharmaceutical stock forecasting in South Australia, finding the Holt-Winters Seasonal Additive + damped method more effective than advanced techniques like Genetic Algorithms and BP neural networks [13]. Lastly, research concluded that shallow neural networks are more effective than deep ones for predicting drug demand based on sales data [14].

2. Research Methods

2.1 Type, Nature, and Research Approach

This study employs a quantitative case study approach, with drug sales data from Pharmacy XYZ as the target variable. The relationship between sales and predictor variables is tested, and two algorithms are

compared to determine the one with the lowest error as a recommendation for future sales predictions.

The research is exploratory, with algorithms and predictor variables subject to development in line with environmental and technological changes. The algorithms can be enhanced or replaced in future studies.

2.2 Data Collection Methods

To compare the performance of the two algorithms, this research is outlined as follows:

1. Research Location: XYZ Pharmacy Store.
2. Research Materials: All data is directly obtained from daily medical records from farmacare.id, weather data from the BMKG website, and commodity price data from the East Java Basic Commodity Availability and Price Development Information System website, OJK dan BPS website covering the period from January 1, 2024, to June 30, 2024.

2.3 Data Analysis Methods

The prescription drug sales dataset will be compiled with the assistance of pharmacists, focusing on pain relief medications without distinguishing their types. Figure 1 below is the detailed research step.

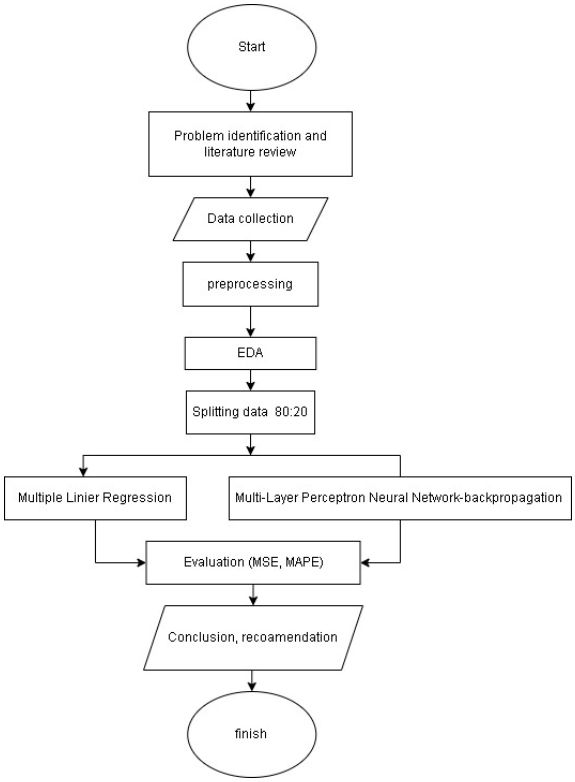


Figure 1. Research Steps

2.3.1. Preprocessing Data

The preprocessing stage is put after data collection in Figure 1 to improve data quality for prediction. The steps include:

1. Data Cleaning: Removing missing values and outliers, as data quality affects model accuracy [15], [16]
2. Data Normalization: Continuous variables, such as temperature and commodity prices, are normalized to the range [0,1] to avoid bias during model training and graphical analysis[16].

2.3.2. Exploratory Data Analysis (EDA)

To further understand the influence of each predictor variable on the target variable, The EDA is put after preprocessing in Figure 1 consisting of a correlation matrix Heatmap and histogram is used [[16].

2.3.3. Model Development

The preprocessing stage improves data quality for prediction and EDA help to understand and select the suitable variable before further process in the prediction model. The two prediction or forecasting models in Figure 1 are below:

1. **Multiple Linear Regression (MLR):**
   MLR is used to identify linear relationships between independent variables (e.g., temperature, and commodity prices) and the dependent variable (drug sales). According to [17], linear regression is an easy-to-implement method suitable for initial analysis.
2. **Multi-Layer Perceptron Neural Network (MLPNN):**
   MLPNN is an artificial neural network algorithm used to capture non-linear patterns in the data. Based on [18], MLPNN has advantages in learning complex relationships between variables.

2.3.4. Model Evaluation

The model's performance is evaluated using the following metrics in Figure 1 before getting a conclusion and recommendations:

1. **Mean Squared Error (MSE):**
   MSE is used to measure the level of prediction error, giving greater weight to larger errors. According to [15], MSE is suitable for models that require optimization for prediction accuracy.
2. **Mean Absolute Percentage Error (MAPE):**
   MAPE provides an overview of the average error in percentage form. [14] indicated that MAPE is highly relevant for comparing model performance across different data scales.

3. Results and Discussion

3.1. Data Collection

Drug data was collected through the Farmacare.id application, which began use at Pharmacy XYZ in December 2023, with classifications confirmed by doctors and pharmacists. This study focuses on

analgesic drugs from January 1 to June 30, 2024. Predictor variables include the prices of basic commodities, weather, inflation, and interest rates within the same period.

Basic commodity price data, such as rice, oil, sugar, and others, was collected from the SISKAPERBAPO JATIM website, using the Jajag market as a reference for the southern Banyuwangi region. Weather data was obtained from the BMKG Online Data website. Inflation data was sourced from the Kab. Banyuwangi website, and interest rate data was retrieved from the OJK website.

Fertilizer-related data was deemed unsuitable for inclusion because fertilizer prices remained constant during the observation period. The collected data will be loaded into a data frame as follows:

```
from google.colab import drive
drive.mount('/content/drive', force_remount=True)
filename = r'/content/drive/MyDrive/rekapoch11b.xlsx'
import csv
open(filename)
```

3.2. Preprocessing Data

3.2.1. Descriptive Data Analysis

This research uses a dataset of internal and external variables that are analyzed descriptively to understand data distribution patterns. Table 1 presents a description of the composite data before it is used. According to [15], descriptive analysis is important for understanding the data before further processing. Data visualization is also performed to detect outliers or anomalies that can disrupt data harmonization.

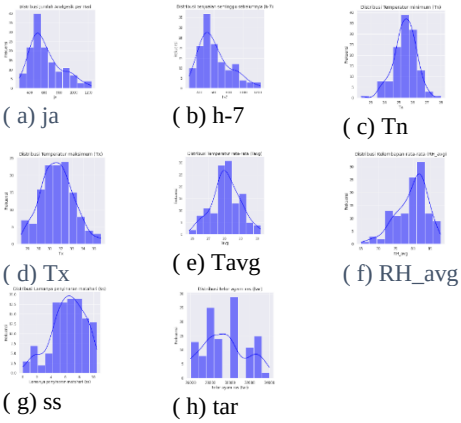
Table 1. Descriptive Statistics for Research Variables

item	tanggal	ja	h-7	r7	Tn	Tx	Tavg	RH_avg
count	140,0	140,0	140,0	140,0	140,0	140,0	140,0	140,0
mean	01/04/2024	619,4	617,9	620,9	25,4	31,6	28,0	79,5
min	08/01/2024	259,0	259,0	505,7	22,3	28,4	25,8	60,0
25%	16/02/2024	472,8	469,3	580,5	25,0	30,6	27,6	76,0
50%	28/03/2024	553,0	553,0	617,6	25,4	31,6	28,0	80,0
75%	15/05/2024	735,0	737,3	666,0	26,0	32,5	28,6	83,0
max	29/06/2024	1242,0	1242,0	745,8	28,0	35,6	30,2	90,0
std		214,2	217,1	58,5	0,9	1,5	0,9	5,1

ss	Ts	brsp	brsm	mgc	mk	mgks	tar	bm
140,0	140,0	140,0	140,0	140,0	140,0	140,0	140,0	140,0
6,3	6,2	14254,3	11422,1	15839,3	15042,9	16657,1	29285,7	34314,3
0,0	2,4	13000,0	10900,0	15000,0	14000,0	16000,0	26000,0	23000,0
4,8	5,1	13400,0	10900,0	15000,0	14000,0	16000,0	28000,0	27000,0
6,5	6,0	14000,0	10900,0	16000,0	15500,0	17000,0	29000,0	32000,0
8,2	7,4	15000,0	12500,0	16500,0	15625,0	17000,0	30000,0	40000,0
10,5	10,2	15600,0	12500,0	17000,0	16000,0	17000,0	34000,0	55000,0
2,4	1,6	883,7	750,3	673,9	814,6	476,4	2238,6	8558,0

bp	gk	dar	cmb	jp	inf	bkm	bkr	bkpr
140,0	140,0	140,0	140,0	140,0	140,0	140,0	140,0	140,0
35278,6	16832,1	32878,6	50535,7	7621,4	0,2	10,9	9,5	9,0
33000,0	15500,0	27000,0	30000,0	7000,0	-0,2	10,8	9,5	9,0
33000,0	16500,0	30000,0	40000,0	7000,0	-0,1	10,9	9,5	9,0
35000,0	17000,0	33000,0	48000,0	7000,0	0,3	10,9	9,6	9,0
37250,0	17000,0	35000,0	60000,0	8000,0	0,5	10,9	9,6	9,1
39000,0	18000,0	40000,0	85000,0	9000,0	0,6	11,0	9,6	9,1
2050,0	604,6	2847,4	13159,5	763,1	0,3	0,1	0,0	0,0

Description:
- ja : Number of analgesics sold per day
- h-7 : Number of analgesics sold the previous week
- r7 : Average analgesics sold during the week
- Tn : Minimum temperature per day
- Tx : Maximum temperature per day
- Tavg : Average temperature per day
- RH_avg : Average humidity per day
- ss : Duration of sunlight per day
- Ts : Maximum minimum temperature difference per day
- brsp : Price of premium rice per day
- brsm : Price of medium rice per day
- brsm : Price of medium rice per day
- mgc : Price of bulk cooking oil per day
- mgk : Price of our oil per day
- mgks : Price of simple packaged cooking oil per day
- tar : Price of chicken eggs per day
- bm : Price of shallots per day
- bp : price of garlic per day
- gk : price of crystal sugar per day
- dar : price of broiler chicken per day
- cmb : price of large red chilies per day
- jp : price of shelled corn per day
- inf : inflation rate per day
- bkm : value of micro credit interest rate per day
- bkpr : value of mortgage interest rate per day



(a) ja (b) h-7 (c) Tn (d) Tx (e) Tavg (f) RH_avg (g) ss (h) tar

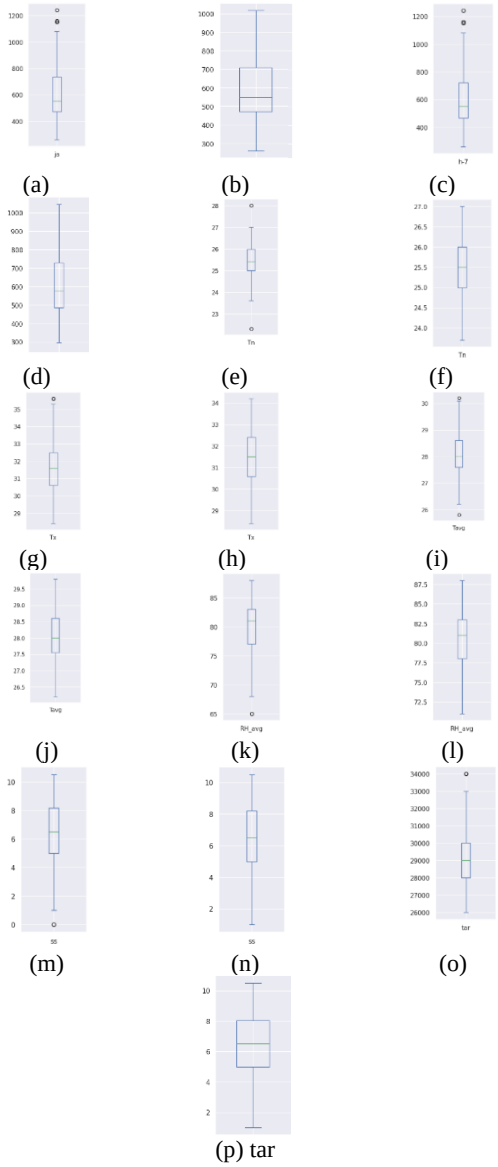
Figure 2. Histogram Distribution image of the variable suffering from outliers

Figure 2 shows several graphs that have skewness either to the left or to the right. This means there is a possibility that the data from those variables have outliers or data anomalies. In Figures 2(a), 2(b), 2(d), and 2(h), variables show that normal distribution skew to the right which means outliers are likely to appear in higher values than 75 % of the data. In Figures 2(f) and 2(g), both variables show that normal distribution skew to the left, which means outliers are likely to appear in lower values than 25% of the data. In Figure 2 (c), 2(e), the normal distribution in both data is not balanced in both sides which means both data need further attention as we just put both of them in outlier detection tools like Boxplot.

3.2.2. Handling Outliers

According to [16], outliers are data that deviate from the general pattern, while [19] suggest their removal to maintain data harmonization. Below is python code with IQR method for outliers removal.

```
k = 1.5
Q1 = df['ja'].quantile(0.25)
Q3 = df['ja'].quantile(0.75)
IQR = Q3 - Q1
lower = Q1 - k*IQR
upper = Q3 + k*IQR
print("batas bawah : ",lower)
print("batas atas : ",upper)
upper_array = df['ja'] >= upper
lower_array = df['ja'] <= lower
df = df[~(upper_array | lower_array)]
```



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o) (p) tar

Figure 3. Boxplot of some variable that suffer outlier (a) 'ja', (c) 'h-7', (e)'Tn', (g) 'Tx', (i)'Tavg', (k)'RH_avg', (m)'ss', (o)'tar' and after outlier clean-up process (b) 'ja', (d) 'h-7', (f)'Tn', (h) 'Tx', (j)'Tavg', (l)'RH_avg', (n)'ss', (p)'tar'

In Figure 3, the state of the boxplot graph before and after the removal of outliers in the attributes that contain outliers in the data can be observed. In Figure 3(a), the outlier of 'ja' is located at values above 1100 and the clean-up process take outlier with value above 1063 in Figure 3(b). In Figure 3(c), the outlier of 'h-7' located at values above 1100 and the clean-up process take outlier with value above1074 in Figure 3(d). In Figure 3(e), the outlier of 'Tn' located at values around 22^0C and the clean-up process take outlier with below 23.5^0C in figure 3(f). In Figure 3(g), the outlier of 'Tx' located at values around 36^0C and the clean-up process take outlier above 35.47^0C in figure 3(h). In Figure 3(i), the outlier of 'Tag' located at values below 26^0C and above 30^0C the clean-up process takes free outlier data within range of 26^0Cand 30^0C in figure 3(j). In Figure 3(k), the outlier of 'RH_avg' located at values around 65^0C and the clean-up process take outlier with value below 68^0C in figure 3(l). In Figure 3(m), the outlier of 'ss' located at values around "0" and the clean-up process take outlier with value below "0.25"^{0}C in figure 3(n). In Figure 3(o), the outlier of 'tar' located at values around 34000 and the clean-up process takes free outlier data below 33000 in figure 3(p).

3.3. Relationship Between Variables (Correlation and histogram comparation)

The next step is to identify the correlation between the independent and dependent variables. According to [20], the correlation coefficient Pearson is capable of measuring the statistical relationship between variables randomly and is the most commonly used linear correlation coefficient. Therefore, correlation analysis is conducted using the Pearson method through the command "df.corr," which will be displayed through a heatmap to measure the linear relationship between variables. Where the value is between the range of -1 to 1. A value of -1 indicates a negative correlation, while a value of zero indicates the absence of a linear relationship signal between two attributes. According to [16] a correlation value that approaches 1 indicates an increasingly perfect positive linear relationship between the two attributes. Below the python program code for obtaining heatmap correlation.

```
#membuat matriks korelasi
cor=df.corr()
plt.figure(figsize=(20,20))
sns.heatmap(cor,annot=True,cmap='coolwarm')
plt.show()
```

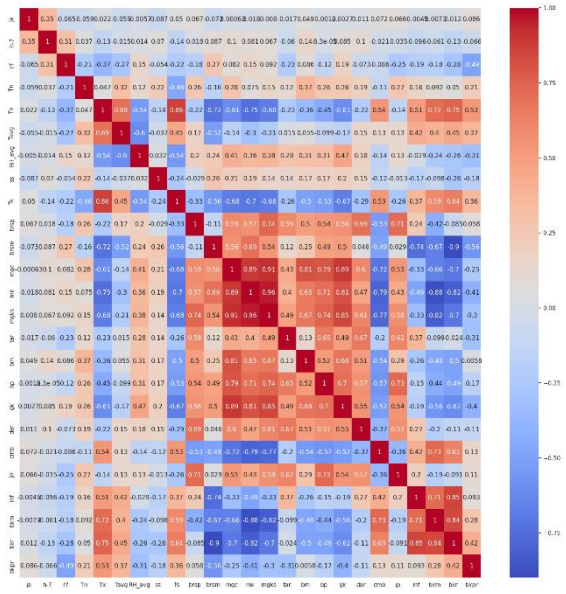
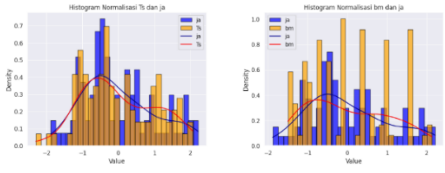


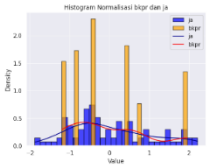
Figure 4. Correlation Matrix between Attributes

Figure 4 shows a correlation value where most attributes have a weak correlation (brighter color) with the "number of analgesics per day (ja)," and thus making linear modeling difficult. Within the heatmap, the highest correlation value is the ‘h-7’ variables. Thus, only this variable has a better linear relationship with ‘ja’ variables and becomes the first candidate for the predictor variable. Other variables are below 0,1 correlation value and that makes them having poor linear relationship with target variables.

As [16] mention, poor linear relationship or poor correlation doesn’t mean there are no relationship. Thus, this study proceeds to examined with two scenarios: (1) variables with a correlation ≥ 0.25 ("h-7") and (2) variables with a graphical unique relation with attribute “number of analgesics per day (ja)” as couple candidate for “h-7” variables. Figure 5 shows the normalization histogram of variable that having visual relationship with ‘ja’ variable.



(a) (b)



(c)

Figure 5. Normalization histogram comparing each predictor variable with target variable (a) ‘Ts’ and ‘ja’, (b)‘bm’ and ‘ja’, (c) ‘bkpr’ and ‘ja’

Within Figure 5, the graphical trend in variables “Ts”, “bm” and “bkpr” show unique relation with variable “ja” histogram. In Figure 5(a), 5(b) and 5(c), the trend line of ‘Ts’, ‘bm’, and ‘bkpr’ flow almost inline. Thus, this research will continue the second scenario with variables “Ts”, “bm” and “bkpr” to be couple of input predictors with variable “h-7”.

3.4. Data Normalization

According to [16], data normalization is performed if the data ranges are significantly different. In this case, data normalization becomes mandatory in the second scenario because, in Table 3, the range values of the attributes 'number of analgesics per day (ja)' and 'mortgage interest rate (bkpr)' are significantly different. Where the attribute 'number of analgesics per day (ja)' has a range from 259 to 1242, while the attribute 'mortgage interest rate (bkpr)' ranges from 8.96% to 9.11%. The normalization that will be used is min-max normalization with the code in Figure 11. Where this normalization works based on the distance between the measured value and the minimum value divided by the range of values. In Python, normalization is generally written using the following command:

```
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(x)
```

According to [16], normalization is necessary if the data ranges differ significantly. In the second scenario, min-max normalization is applied due to the difference in the range. In Python, normalization is performed by calculating the distance of the value from the minimum divided by the range of values. MLPNN requires input in the range of 0–1, and min-max normalization is effective for continuous data because the network tolerates slight boundary violations.

3.5. Prediction Model with MLR and MLPNN

3.5.1. Multiple Linear Regression

Multiple Linear Regression is an algorithm that processes dependent or target data based on independent data (predictor) by utilizing linear mathematical equations [21]. Below the python code for linear regression

```
from sklearn.linear_model import LinearRegression
modelL = LinearRegression()
```

```
modelL.fit(XL_train,yL_train)
```

```
modelL.intercept_
```

```
modelL.coef_
```

```
yL_pred = modelL.predict(XL_test)
```

3.5.1.1 Variable predictor 'h-7'

In the first scenario with the predictor attribute being only the sales amount from 7 days prior (h-7), multiple linear regression obtained from the Python model.intercept = 405,215 and model.coeficient = 0,328 so it result in linear equation as follows

$$y = 405{,}215 + 0{,}328x_1 \dots\dots\dots\dots\dots (I)$$

Explanation:
y = predicted value or target or dependent variable
x1 = independent variable "h-7"

The equation I details as follows:

1. Intercept (b_0) = 405,215 is the initial value of y when the value of x_1 is zero.
2. Coefficient (bi):
   - 0,328 means that every increase of 1 unit in x1 will raise the predicted value of y by 0,328.

The evaluation values can be explained as follows:

1. Mean Squared Error (MSE): 22544,477
   MSE is the average of the squared errors (the difference between actual and predicted values). With an MSE value of 22544,477, this indicates a relatively large prediction error. The size of the MSE also depends on the scale of the target variable, and in this case, an error of approximately 22544,477means the RMSE of 150,148 suggests that the predictions often deviate significantly from the actual values, with an error of around 150 pieces of analgesic per day.
2. Mean Absolute Percentage Error (MAPE): 25.5%
   MAPE measures the average percentage error in predictions relative to the actual values. A MAPE value of 0.255 (or 25.5%) means that the average error of the model is approximately 25.5% of the actual value. In practice, more than 5% error is considered quite high, especially for data with patterns that could be predicted more accurately.

3.5.1.2 Predictor Attributes: 'h-7' and 'Mortgage Interest Rate (bkpr)'

In the second scenario, where the predictor attributes are "h-7" and "bkpr", multiple linear regression in Python produced the following results:
Model intercept = 471,94 and model coefficients = 270,98, 38,34. These can be applied to Equation as follows:

$$y = 471{,}94 + 270{,}98x_1 + 38{,}34x_2 \dots\dots (II)$$

Explanation:
y: Predicted value or target (dependent attribute)
x_1: h-7

x_2: bkpr

b_1: coefficient x_1

b_2: coefficient x_2

The equation II details as follows:

1. Intercept (b_0) = 471,94: This is the initial value of y when both x_1 and x_2 are zero.
2. Coefficients (b_i)
   - 270,98: This indicates that for every 1-unit increase in x_1, the predicted value y will increase by 270,98, assuming other attributes remain constant.
   - 38,34: This indicates that for every 1-unit increase in x_2, the predicted value y will decrease by 38,34, assuming other attributes remain constant.

The evaluation values can be explained as follows:

1. Mean Squared Error (MSE): 22768,427
   MSE is the average of the squared errors (the difference between actual and predicted values). With an MSE value of 22768,427, this indicates a relatively large prediction error. The size of the MSE also depends on the scale of the target variable, and in this case, an error of approximately 22768,427means that the RMSE of 150,89 suggests that the predictions often deviate significantly from the actual values, with an error of around 151 pieces of analgesic per day.
2. Mean Absolute Percentage Error (MAPE): 25,6%
   MAPE measures the average percentage error in predictions relative to the actual values. A MAPE value of 0.256 (or 25.6%) means that the average error of the model is approximately25,6% of the actual value. In practice, more than 5% is considered quite high, especially for data with patterns that could be predicted more accurately.

3.5.1.3 Predictor Attributes: 'h-7' and 'Maximum minimum temperature difference per day (Ts)'

In the second scenario, where the predictor attributes are "h-7" and "Ts", multiple linear regression in Python produced the following results:
Model intercept = 450,157 and model coefficients = 271,883, 76,17. These can be applied to Equation as follows:

$$y = 450{,}157 + 271{,}883x_1 + 76{,}17x_2 \dots\dots (III)$$

Explanation:
y: Predicted value or target (dependent attribute)

x1: h-7

x2: Ts

b_1: coefficient x_1

b_2: coefficient x_2

The equation III details as follows:

1. Intercept **(b_0)** = 450,157: This is the initial value of y when both x_1 and x_2 are zero.
2. Coefficients **(b_i)**
   - 271,883: This indicates that for every 1-unit increase in x_1, the predicted value y will

increase by 271,883, assuming other attributes remain constant.

- 76,17**:** This indicates that for every 1-unit increase in x_2, the predicted value y will decrease by 76,17 assuming other attributes remain constant.

The evaluation values can be explained as follows:

1. Mean Squared Error (MSE): 23.355
   MSE is the average of the squared errors (the difference between actual and predicted values). With an MSE value of 23.355, this indicates a relatively large prediction error. The size of the MSE also depends on the scale of the target variable, and in this case, an error of approximately 23.355 means that the RMSE of 152,8 suggests that the predictions often deviate significantly from the actual values, with an error of around 153 pieces of analgesic per day.
2. Mean Absolute Percentage Error (MAPE): 25,8%
   MAPE measures the average percentage error in predictions relative to the actual values. A MAPE value of 0.258 (or 25.6%) means that the average error of the model is approximately 25,6% of the actual value. In practice, more than 5% is considered quite high, especially for data with patterns that could be predicted more accurately

3.5.1.4 Predictor Attributes: 'h-7' and 'Price of shallots per day (bm)'

In the second scenario, where the predictor attributes are "h-7" and "bm", multiple linear regression in Python produced the following results: Model intercept = 486,224 and model coefficients = 264,136, 12,95193465. These can be applied to the Equation as follows:

$$y = 486{,}224 + 264{,}136x_1 + 12{,}95x_2 \ldots\ldots \text{(IV)}$$

Explanation:
y: Predicted value or target (dependent attribute)

x_1: h-7

x_2: bm

b_1: coefficient x_1

b_2: coefficient x_2

The equation IV details as follows:

3. Intercept (b_0) = 392,5605: This is the initial value of y when both x_1 and x_2 are zero.
4. Coefficients **(**b_i)
   - 0,3256: This indicates that for every 1-unit increase in x_1, the predicted value y will increase by 0,3256, assuming other attributes remain constant.
   - 0,0004**:** This indicates that for every 1-unit increase in x_2, the predicted value y will decrease by 0,0004 assuming other attributes remain constant.

The evaluation values can be explained as follows:

1. Mean Squared Error (MSE): 22346,129
   MSE is the average of the squared errors (the difference between actual and predicted values). With an MSE value of 23.355, this indicates a relatively large prediction error. The size of the MSE also depends on the scale of the target variable, and in this case, an error of approximately 22346,129 means that the RMSE of 149,486 suggests that the predictions often deviate significantly from the actual values, with an error of around 150 pieces of analgesic per day.
2. Mean Absolute Percentage Error (MAPE): 25,4%
   MAPE measures the average percentage error in predictions relative to the actual values. A MAPE value of 0.254 (or 25.4%) means that the average error of the model is approximately 25,4% of the actual value. In practice, more than 5% error is considered quite high, especially for data with patterns that could be predicted more accurately

3.5.1.5 Conclusion from 2 scenarios of Linear Regression model testing

Table 2 show comparison of MSE and MAPE through all scenario. The Linear Regression Model performs poorly with almost the same results by MAPE 25% and MSE around 22300 to 23500.

Table 2. Results of the linear regression evaluation

Item	"h-7", "bm"	"h-7", "Ts"	"h-7", "bkpr"	"h-7"
MSE	22346,129	23355,282	22768,427	22544,477
MAPE	0,254	0,258	0,256	0,255

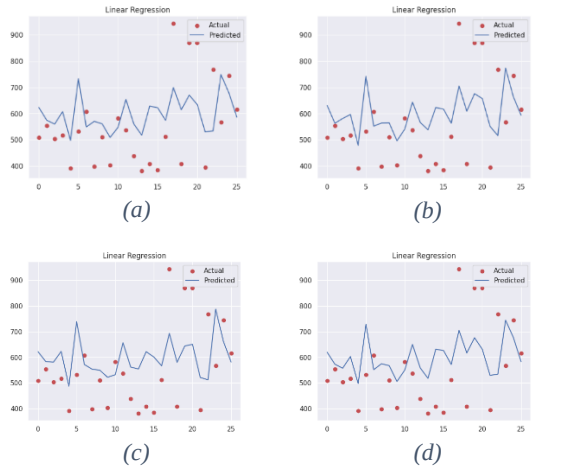


Figure 6. Graph of actual y values and predicted y values from the multiple linear regression model with the predictor being (a) 'h-7', (b) 'h-7' and 'bkpr', (c) 'h-7' and 'Ts' and (d) 'h-7' and 'bm'.

The graph in Figure 6 further compares and describes the MLR result. It's clear that the graphs are almost identical. Thus, the external variables do not

play any role in this MLR due to their lack of correlation value. Here are the complete explanations:

1. Figure 6(a), 6(b), 6(c) and 6(d) show that the model's predictions (blue line) do not fully align with the actual values (red dots). The model does not capture the underlying trend or the variations in the actual data means the models are underfitting. But have lots of sharp turn struggle to get all variation and capture the pattern also means overfitting.
2. The red points scattered randomly around the blue line mostly under blue line in figure 6(a), 6(b), 6(c), and 6(d) indicate the models are overbiased which means the model tends to give a higher value than the actual. The model also underfitting because it seems to struggle to capture most of the data since the data variation is large,
3. A significant difference between actual values and predictions line at several points of figure 6(a), 6(b), 6(c) and 6(d) indicates the presence of error or bias in the model. The sharp turn in the prediction line could make an unstable prediction value. This could be caused by:
   - Data that is insufficient to train the model.
   - variables that are irrelevant or noise in the data.
   - Low correlation value from data
   - Need better model or combination model to capture the dimension of data

3.5.2 Multi-Layer Perceptron Neural Network (MLPNN)

According to [22], MLPNN is an artificial neural network algorithm with a feedforward architecture that can be improve with backward propagation to reduce errors and improve accuracy. This algorithm consists of an input layer, one or more hidden layers, and an output layer, and is effective in learning non-linear data patterns [23]

In this study, MLPNN with backpropagation will be implemented using the TensorFlow library. [24] mention that TensorFlow provides user-friendly tools for implementing neural networks as python code below :

```
import tensorflow as tf
from tensorflow.keras import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.layers import Dense, Flatten
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
# independent feature
x = df.drop(['ja'],axis=1)
# dependent feature, menjadi label atau class
y = df['ja']

from sklearn.preprocessing import PolynomialFeatures
poly = PolynomialFeatures(degree=3)
Xnn_train_poly = poly.fit_transform(Xnn_train)
Xnn_test_poly = poly.transform(Xnn_test)
```

```
n = Xnn_train_poly.shape[1]
print("jumlah input node =", n)
h = (2*n)+1
print("jumlah hidden node =", h)

model5 = Sequential([
    Dense(h, activation='sigmoid', input_shape=(Xnn_train_poly.shape[1],)),
    Dense(1, activation='linear')
])

model5.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.5),
               loss='mse', metrics=['mae'])

TSBS = model5.fit(Xnn_train_poly, ynn_train, epochs=400,
                  validation_split=0.2, batch_size=5, verbose=1)
```

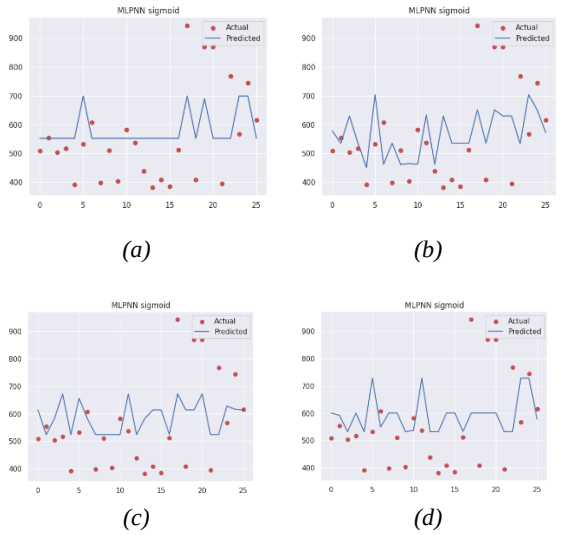


Figure 7. Graph of the MLPNN backpropagation modeling results for (a) 'h-7', (b) 'h-7'and 'bkpr', (c) 'h-7'and 'Ts' and 'h-7'and 'bm'

The results of the Python program with 1 hidden layer consist of hidden node $8n + 1, n\ is\ input\ node$, sigmoid activation and using 4 scenarios are in figure 7. The figure 7 details as follows:

1. Figure 7(a) show that the model's predictions (blue line) do not fully align with the actual values (red dots). The model produces a rigid, simplified pattern that does not capture the underlying trend or the variations in the actual data. In some parts, the model provides predictions that are far from the actual values. This indicates the possibility that the model is unable to capture the patterns well or there is underfitting. Different thing happens to figure 7(b), 7(c) and 7(d) which suffer overfitting from a lot of sharp turn or fluctuation but from perspective of low capability to capture data trend, they are suffer underfitting
2. The red points scattered randomly around the blue line mostly under blue line in figure 7(a), 7(b), 7(c), and 7(d) indicate the models are overbiased which means the model tends to give higher value then the actual. The model also underfitting because it seems struggling to capture most of the data since the data variation is large,

3. A significant difference between actual values and predictions line at several points of figure 7(a), 7(b), 7(c), and 7(d) indicates the presence of error or bias in the model. The sharp turn in the prediction line could make an unstable prediction value. This could be caused by:
   - Data that is insufficient to train the model.
   - variables that are irrelevant or noisy in the data.
   - Need other suitable activation function or hyper-parameter or polynomial order
   - Need a better model or combination model to capture the dimension of data

Table 3. Results of the MLPNN model evaluation

	'h-7'	'h-7' dan 'bkpr'
MSE MLPNN	19588,936	21893,260
MAPE MLPNN	0,223	0,233
	'h-7', 'Ts'	'h-7', 'bm'
MSE MLPNN	24.999	26821,830
MAPE MLPNN	0,287	0,269

The evaluation values in Table 3 can be explained as follows:

1. Mean Squared Error (MSE): 19500–27,000
   MSE measures the average squared error between actual and predicted values. An MSE value of more than 19500 and RMSE of more than 140 indicate that the predictions often deviate by approximately 140 pieces of analgesics per day, potentially exceeding the data's minimum-maximum range.
2. Mean Absolute Percentage Error (MAPE): 22% - 29%
   MAPE value of more than 22% indicates an average prediction error of more than 22% of the actual value. This is considered high, particularly if the data should be more accurately predictable.

3.6 Comparison Between MLR and MLPNN Backpropagation Models

Based on Table 4, the comparison between MLPNN and MLR shows that neither model performs well. MSE values exceeding 19000 indicate prediction errors of more than 130 pieces or more than 20% error. The graphs in Figures 6, 7, 8, 9, 10(a), 10(b), 10(c) and 10(d) also reveal underfitting across all models. But, MLR tends to give a stable output of MSE and MAPE whether work as an individual or work as a couple of other variables

Table 4. Comparison between the MLPNN and MLR models

Item	MSE	MAPE
MLPNN-sigmoid-linear"h-7"	19588,936	0,223
MLPNN-sigmoid-linear"h-7", "bkpr"	21893,260	0,233
MLR "h-7", "bm"	22346,129	0,254
MLR "h-7"	22544,477	0,255
MLR "h-7", "bkpr"	22768,427	0,256
MLR "h-7", "Ts"	23355,282	0,258
MLPNN-sigmoid-linear"h-7", "Ts"	24999,000	0,287
MLPNN-sigmoid-linear"h-7", "bm"	26821,830	0,264

3.7 Influence of Independent Variables on the Model

Based on Table 4, the independent variables especially external variables have not shown a significant influence in the modeling process. The linear regression model failed to detect linear relationships, and MLPNN with backpropagation also did not succeed in capturing non-linear. The only variable with a reasonable correlation and lowest error is "h-7" (sales volume 7 days earlier) with MAPE 22% by MLPNN and 25% by MLR.

3.7.1. Validation of Findings with Literature

These results differ from studies by [23] which indicated that neural networks excel in predicting pharmaceutical stock. The neural network perhaps able to reach better performance, but without suitable conditions the prediction tends to be unstable. The MLR which depends on correlation value tends to give stable predictions.

This study suggests focusing on specific types of medications, exploring other external variables, and collecting data from a broader range and longer periods of observation in healthcare services to improve model accuracy.

3.7.2. Practical Implications

It is recommended that Pharmacy XYZ use an integrated pharmaceutical database connected with other healthcare services within the same region to obtain comprehensive data before implementing predictive models such as linear regression or neural networks.

3.7.3. Research Limitations

The limitations of this study include the following:

1. Pharmacy XYZ has only started recording its pharmaceutical database at the end of December 2023 and that's make low volume of data.
2. The simulation of model work on particularly similar to real conditions where retail pharmacies only have limited computer resource
3. Pharmacy XYZ employees are still adapting to the database recording process and lack an

understanding of the importance of maintaining a pharmaceutical database.
4. The research only works within local pharmacy database.

4. Conclusion and Recommendation

Based on the research conducted on the performance comparison of the Multiple Linear Regression (MLR) and Multi-Layer Perceptron Neural Network (MLPNN) algorithms in predicting analgesic sales at Pharmacy XYZ, several conclusions were drawn. First, the research results indicate that MLPNN slightly outperform Multiple Linear Regression with 'h-7' as single input variable. But the MLR have stable predictions output and simple computation process and resources compare to MLPNN especially when one of the input variables have good correlation value. Second, the MAPE of all model reach more than 20% and MSE reach more than 19000 indicate that all model needs to be further improve. Third, the influence of external variables such as weather and staple goods prices on analgesic sales needs to be further reviewed, particularly in terms of their correlation before being used in modeling. But the looking for other variables are still in need to find more suitable variable such as air pollution, population mobility and others related factor to people health. Fourth, the needs of integrated pharmacy database is urgent to capture data in one region to prevent patient missing data due to patient move to other health facility.

References

[1] G. Emmanuel, Meyliana, A. N. Hidayanto, and H. Prabowo, "A Literature Review: Drug Supply Chain and Neural Networks," in *Proceeding - 6th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Sciences and Artificial Intelligence Technologies for Environmental Sustainability, ICITISEE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 559–564. doi: 10.1109/ICITISEE57756.2022.10057939.

[2] European Association of Hospital Pharmacists, "2019 EAHP Medicines Shortages Report," 2019.

[3] L. Hedman, "Global approaches to addressing shortages of essential medicines in health systems," *WHO Drug Information*, vol. 30, no. 2, 2016.

[4] N. Miljković *et al.*, "Results of EAHP's 2023 shortages survey," *European Journal of Hospital Pharmacy*, 2024, doi: 10.1136/ejhpharm-2024-004090.

[5] A. Nguyen, O. Bougacha, B. Lekens, S. Lamouri, and R. C. Couvreur, "On the use of logistics data to anticipate drugs shortages through," in *Procedia Computer Science*, 2023. doi: 10.1016/j.procs.2023.01.371.

[6] D. Anisya Farmaciawaty, M. Hasan Basri, A. Adiutama, F. Budyanto Widjaja, and I. Nurul Rachmania, "Inventory Level Improvement in Pharmacy Company Using Probabilistic EOQ Model and Two Echelon Inventory: A Case Study," *The Asian Journal of Technology Management*, vol. 13, no. 3, pp. 229–242, 2020, doi: 10.12695/ajtm.2020.13.3.4.

[7] S. Romano, H. Galante, D. Figueira, Z. Mendes, and A. T. Rodrigues, "Time-trend analysis of medicine sales and shortages during COVID-19 outbreak: Data from community pharmacies," *Research in Social and Administrative Pharmacy*, vol. 17, no. 1, pp. 1876–1881, Jan. 2021, doi: 10.1016/J.SAPHARM.2020.05.024.

[8] V. Reinstadler *et al.*, "Monitoring drug consumption in Innsbruck during coronavirus disease 2019 (COVID-19) lockdown by wastewater analysis," *Science of the Total Environment*, vol. 757, 2021, doi: 10.1016/j.scitotenv.2020.144006.

[9] A. Gasparrini *et al.*, "Mortality risk attributable to high and low ambient temperature: A multicountry observational study," *The Lancet*, vol. 386, no. 9991, 2015, doi: 10.1016/S0140-6736(14)62114-0.

[10] S. Cohen, W. J. Doyle, and A. Baum, "Socioeconomic status is associated with stress hormones," *Psychosom Med*, vol. 68, no. 3, 2006, doi: 10.1097/01.psy.0000221236.37158.b9.

[11] N. Tampati, Y. Melita Pranoto, T. Informasi, and I. Sains dan Teknologi Terpadu Surabaya Jl Ngagel Jaya Tengah No, "Prediksi Stok Obat pada Apotik Total Life Clinic Menggunakan Model Kombinasi Artificial Neural Network dan ARIMA," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 1, pp. 49–58, 2022, [Online]. Available: http://jurnal.mdp.ac.id

[12] F. Khairati and H. Putra, "Prediksi Kuantitas Penggunaan Obat pada Layanan Kesehatan Menggunakan Algoritma Backpropagation Neural Network," *Jurnal Sistim Informasi dan Teknologi*, pp. 128–135, Aug. 2022, doi: 10.37034/jsisfotek.v4i3.158.

[13] R. Rushton *et al.*, "Forecasting inventory for the state-wide pharmaceutical service of South Australia," *Procedia Comput Sci*, vol. 219, pp.

1257–1264, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.409.

[14] R. Rathipriya, A. A. Abdul Rahman, S. Dhamodharavadhani, A. Meero, and G. Yoganandan, “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model,” *Neural Comput Appl*, vol. 35, no. 2, pp. 1945–1957, Jan. 2023, doi: 10.1007/s00521-022-07889-9.

[15] D. Kurniawan, *Pengenalan Machine Learning dengan Python*, 3rd ed. Jakarta: PT Elex Media Komputindo, 2022.

[16] Larose, “Discovering Knowledge in Data: An Introduction to Data Mining,” 2005.

[17] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: C.V. ANDI OFFSET, 2009.

[18] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, “Multilayer Perceptron dan Neural Networks Gambar 1: Umpan-maju multilayer jaringan saraf,” *Wseas Transactions On Circuits And Systems*, vol. 8, no. 7, 2009.

[19] Q. Han, X. Xiao, S. Wang, W. Qin, C. Yu, and M. Liang, “Characterization of the effects of outliers on ComBat harmonization for removing inter-site data heterogeneity in multisite neuroimaging studies,” *Front Neurosci*, vol. 17, 2023, doi: 10.3389/fnins.2023.1146175.

[20] Z. Šverko, M. Vrankić, S. Vlahinić, and P. Rogelj, “Complex Pearson Correlation Coefficient for EEG Connectivity Analysis,” *Sensors*, vol. 22, no. 4, Feb. 2022, doi: 10.3390/s22041477.

[21] S. Rath, A. Tripathy, and A. R. Tripathy, “Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model,” *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 14, no. 5, 2020, doi: 10.1016/j.dsx.2020.07.045.

[22] M. Desai and M. Shah, “An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN),” *Clinical eHealth*, vol. 4, pp. 1–11, 2021, doi: 10.1016/j.ceh.2020.11.002.

[23] F. Altamevia and H. Oktafia Lingga Wijaya, “Analisis Pola Penjualan Obat di Apotek Srikandi Menggunakan Algoritma Supervised Learning,” *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 4, no. 1, pp. 170–176, 2023.

[24] T. Filimonova, H. Samoylenko, A. Selivanova, Y. Yurchenko, and A. Parashchak, “Construction Of A Neural Network For Handwritten Digits Recognition Based On Tensorflow Library Applying An Error Backpropagation Algorithm,” *Eastern-European Journal of Enterprise Technologies*, vol. 6, no. 2(126), pp. 25–32, 2023, doi: 10.15587/1729-4061.2023.293682.