ISSN: 2338-8412 e-ISSN : 2716-4411

Ekombis Review – Jurnal Ilmiah Ekonomi dan Bisnis
Available online at : https://jurnal.unived.ac.id/index.php/er/index
DOI: https://doi.org/10.37676/ekombis.v14i1

Sales Profit Forecasting In Indonesian State Owned Enterprise: A Comparative Study Of Machine Learning Algorithms

Afifah Dyah Puspa [1], Sunu Widianto [2], Samidi [3]
[1] Master of Science in Management, Universitas Padjadjaran
[2] Master of Science in Management, Universitas Padjadjaran
[3] Master of Science in Management, Universitas Padjadjaran
E-mail : [1] afifah21007@mail.unpad.ac.id ,[2] sunu.widianto@unpad.ac.id ,[3] samidiunpad@gmail.com

How to Cite :
Puspa, A, D., Widianto, S., Samidi. (2026). Sales Profit Forecasting In Indonesian State Owned Enterprise: A Comparative Study Of Machine Learning Algorithms . EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi Dan Bisnis, 14(1). doi: https://doi.org/10.37676/ekombis.v14i1

ARTICLE HISTORY
Received [24 June 2025]
Revised [04 January 2026]
Accepted [20 January 2026]

KEYWORDS
Profit Forecast, Machine Learning, Linear Regression, Neural Network, Gradient Boost Regression .

Abstract

Forecasting is a crucial step in management planning to predict future condition of the business. Company needs to find forecasting method with best performance to create company's strategy and avoid inaccuracy. Recent studies have attempted to find the best prediction method using machine learning to predict sales demand and sales forecast in various industries. By using machine learning algorithm, accuracy rate can be measured to evaluate prediction method. This study aims to find best sales profit forecast method by comparing three machine learning model: Linear Regression, Neural Network, and Gradient Boost Regression. Data source for this study was from 32 branch offices of Indonesian State Owned Enterprise PPI (Perusahaan Perdagangan Indonesia / Indonesian Trade Center), with data range of year 2017 to 2022. Study finds that Neural Network has the highest performance with the smallest error rate, compared to Linear Regression and Gradient Boost Regression, with 96.97% accuracy rate.

INTRODUCTION

Planning is an important action to achieve company goals by providing ideas of future directions (Nafarin, 2008). This management function includes managerial activites related to preparation for future conditions, such as projecting trends, defining goals, crafting strategies, and establishing company policies (David & David, 2017). Successful strategic implementation and evaluation heavily relies on forecast accuracy, as management activities depend on how well the company can formulate short-term and long-term plan. Forecasting is an activity to estimate or create assumptions about future conditions (Griffin, 2020) Knowledge about future condition will influence the company's actions and decisions. Accurate predictions can identify opportunities and threats, while providing a competitive advantage. Contrarily, Incorrect assumptions may lead to significant prediction errors, resulting inaccurate decision-making process.

Projecting future condition is a challenging field especially in the State-Owned Enterprises (SOEs). Every aspect of establishment, management, supervision, and budgeting of Indonesian SOEs is controlled under goverment policies. The planning process regulated through the Decree of the Minister of State-Owned Enterprises (Keputusan Menteri Badan Usaha Milik Negara Nomor KEP-101/MBU/2002 Tentang Penyusunan Rencana Kerja Dan Anggaran Perusahaan Badan Usaha Milik Negara, 2002). Based on the article three of the ministrial decree, SOEs are required to prepare company's work and budget plan RKAP (Rencana Kerja dan Anggaran Perusahaan), as well as financial forecast by comparing future projections with current year's prognosis, with approval from the shareholders. This decision process have disadvantage by having large gap from real data. Based on the 2022 annual report from PT Perusahaan Perdagangan Indonesia (Indonesian Trade Center), SOE that focusing in trading industry, the revenue and net profit was recorded to achieve 79.27% and 72.15% from the target (PT PPI, 2023). The condition leads to find profit forecast methods with higher accuracy.

Similar to the other companies in general, one of the primary objectives of establishing State-Owned Enterprises is to pursue profit. Generally, total profit (TP) is an equation of revenue (TR) substracted by Total Cost (TC), which can be written by formula $TP = \sum TR - \sum TC$ (Noor, 2017). Meanwhile, net profit is a final result from sales activities, defined by difference between gross sales profit and indirect costs or operating expenses (Evmenchik et al., 2021). In order to find significant predictor for sales net profit, it can be concluded that the formula consists of revenue, cost, and operating expenses. This claim is supported by study that found sales revenue has a positive impact to net profit, while operating expense has negative impact to net profit (Suzan & Siagia, 2023). The higher sales made by the company, the higher company will get profits, and vice versa. However, other study with case of pharmaceutical manufacturing company found that sales revenue does not have a significant impact to net profit. Other factors such as production cost influenced to net profit (Suzan & Nabilah R., 2020). This findings from literature study needs to be studied further.

Cost of Goods Sold (COGS) and material cost is one of the main cost that has large portions of operational cost (Mun & Jang, 2018). Study found that cost of goods sold has singnificant impact to sales and profitability (Dewi et al., 2021; Goestjahjanti & Widayati, 2020; Mun & Jang, 2018). Research in restaurant industry indicates that food cost can significantly impact full-service restaurant and cause lower profitability (Mun & Jang, 2018). Meanwhile in manufacturing industry, cost of goods sold and inventory has significant effects on sales where good cost will be followd by high levels of profitability (Dewi et al., 2021). Other study in gas industry found that distribution cost has positive relationship with significant impact to sales turnover, where every increase in distribution costs will be followed by an increase in sales (Dumadi et al., 2023). Meanwhile, study from tyre industry found that distribution cost has no significant impact to profit, where every increase of variables will not be followed by sales performance (Dharmayuni et al., 2021).

Promotion is a marketing tools to make products can be known to wider audience (Listiawati et al., 2021). Promotion consists of various marketing activities such as advertising, exhibitions, and other sales promotions. Large amounts of promotion costs are expected with sales return of increased sales demand. In garment industry, research shows that there is an influence of promotion cost on increasing sales (Listiawati et al., 2021). In the other hand, research conducted by Mun & Jang in restaurant industry states that advertising cost does not have significant impact on the company profits, as high influence came from food cost and employee cost (Mun & Jang, 2018). The finding is supported by a research in aviation industries, where cost per employee had positive influence to the economic performance that leads to profitability (Ginieis et al., 2020).

Profits can be increased through various type of cost efficiency. Study found that operating expense had significant negative impacts to profitability (Mun & Jang, 2018). Thus, efficiency in operating expenses will be followed with increasing profit. In farming industry, storage cost has an impact to profitability (Jones et al., 2014). Efficiency of warehouse related and storage management is a challenge in operational decisions. Utilization of space and cost effectiveness of material handling can have impact to increasing profits (Ajol et al., 2018). In other research, general & administrative expense (Mun & Jang, 2018; Suzan & Siagia, 2023) and distribution cost (Prahada et al., 2022) has influence to sales and profit. Another study found that depreciation and amortization can also have influence to profitability (Elise & Daryanto, 2018; Lukić et al., 2016). Choosing the right depreciation calculation method and effieciency in shrink management can increase profits in retail companies (Elise & Daryanto, 2018). Recent studies has identified several variables that influence business profitability. However, it remains unclear which variable have significant impact on net profit and how these predictor variables can fit as suitable parameters to create predictive model with high performance. This study aims to address this research gap by identifying variable with significant relation to net profit, thereby enabling the development of predictive model with high accuracy.

In recent decades, traditional prediction methods have been widely used in various business industry. However, this method usually has limitation on data collecting and processing. The use of limited factors in traditional method can affect variations and patterns outside these factors cannot be explained in the predictive models (Kharfan et al., 2021). In the other hand, machine learning technique have advantages over traditional prediction methods. Machine learning can utilize various unlimited data sources to identify the most significant variables and have the ability to identify hidden patterns (Kharfan et al., 2021). Machine learning is a branch of data science, which is an interdisciplinary field that extracts value from data with data mining (Kotu & Deshpande, 2019). Machine learning works by utilizing historical data to find patterns. The development of machine learning combines statistical science techniques, machine learning technology, pattern recognition, database systems, algorithms, visualization, and other fields of study (Han et al., 2012). Machine learning studies how computers or machines can learn and improve performance based on data (Han et al., 2012). Supervised learning is a technique for predicting the value of output variables based on input variables (Kotu & Deshpande, 2015). The variable that will be predicted is called label or target variable. Supervised learning will search for data pattern and functions based on examples in the training data set and uncovers hidden patterns from the data (Han et al., 2012).

The use of machine learning to make predictions has been widely used in predictive analysis study and contributed to finding prediction methods with the best performance. Such as sales and product demand forecast in fashion and apparel industry (I. F. Chen & Lu, 2021; Kharfan et al., 2021; Yasir et al., 2022), prediction in catch fishery (Zhang et al., 2022), stock market prediction (Panwar et al., 2021), predictions of MSME bankruptcy (Altman et al., 2020), predictive analysis for telecommunications industry (Wisesa et al., 2020), and forecast of advertising effectiveness for renewable energy (Sharifi et al., 2019). Study found that predictions using machine learning methods have higher accuracy compared to manual prediction methods (Wisesa et al., 2020)

Previous research has been studied, several studies have compared algorithm performance to find the best method. Study about demand forecast in the textile industry found that linear regression has a high accuracy (Yasir et al., 2022). Linear regression was also the algorithm with the best performance in studies studying the stock market (Panwar et al., 2021). However, other study about fish catches forecast found that Neural Network is the algorithm with the best performance and accuracy (Zhang et al., 2022). Study that predicting MSME financial classification (Altman et al., 2020), analyzing advertising effectiveness (Sharifi et al., 2019), and consumer behavior analysis (Hew et al., 2019) also shows that Neural Networks have a high performance. Meanwhile, research about sales predictive analysis in the telecommunications industry found that Gradient Boost Regression had higher accuracy, compared to manual predictions (Wisesa et al., 2020).

Based on the explanation above, further testing is needed to find the best algorithm method with high performance for the case studies of sales profit forecast. Machine learning predictive models will be assessed by comparing accuracy rates. Another challenge of finding model predictions with best performance is determining the significant predictor variables. The efficiency and accuracy of predictive models are highly related to the selection of significant predictor variables in relation to the target variable (P. Chen et al., 2021; Fitni & Ramli, 2020). This paper will discuss sales profit forecast using machine learning method with case study of State-Owned Enterprise in Indonesia. The purpose of this paper is to find predictive model with highest accuracy by comparing machine learning algorithms. This paper structured as follows: section (1) introduction, included reviews the latest literature related to the machine learning and sales predictions, section (2) presents the data source and methodology, section (3) presents results and evaluation, and section (4) provides conclusion and research limitation.

METHODS

The data source of this study is based on sales report from Perusahaan Perdagangan Indonesia (PPI), a State-Owned Enterprise under ID-FOOD SOE holding company that focused their business in trading. The data covered report of PPI branch office in 32 cities from year of 2017 to 2022. Raw data originally obtained in Excel worksheet for each branch report. Due to the manual formatting method, raw data need to be cleaned and transformed into CSV (Comma Separated Value) format. After processing data preparation steps by cleaning missing value and data outliner, dataset consists of 4,660 sales commodity data and 191 branch sales data. Description of columns used in this study are mentioned as follows with Y as target variable.

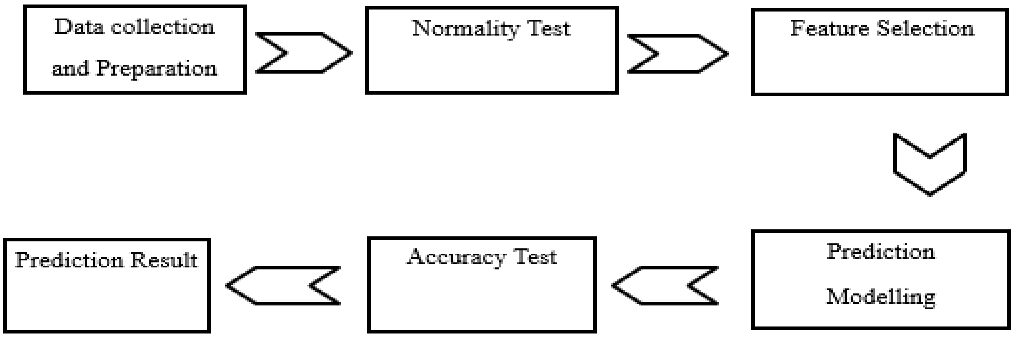
Table 1. Variable Description

Variables	Description
X1	Sales revenue
X2	Cost of goods sold
X3	Direct selling cost
X4	Warehouse and inventory cost
X5	Promotion cost
X6	Employee cost
X7	Vehicle and transportation cost
X8	Office and utility cost
X9	General and administration cost
X10	Depreciation and amortization cost
Y	Net profit from sales

This research is using Python version 3.10.5, accessed through Jupyter Notebook, as the tool for data analysis and modelling. Python is a programming language created by Guido Von Rossum in 1991 (Nelli, 2018) and has become the lingua franca in data science (Muller & Guido, 2017). Python was chosen because of its variety of open access libraries for data processing, visualization, and statistics. The Python libraries used in this research including Pandas, Numpy, Matplotlib, Seaborn, and Sci-Kit Learn.

To provide a clearer overview, The following diagram outlines the steps involved in the research methodology.

Figure 1. Research Design



Source: modified from various sources

Feature selection is an important step in machine learning modelling to improve performance. Data set may contain feature that has weak correlation or no significance to target the variable. Feature selection method allows only selecting features that have significant correlation to target variable. Therefore, accuracy and effiency can be improved (P. Chen et al., 2021; Fitni & Ramli, 2020). Study from Fitni & Ramli proposed ensemble learning with filter method by defining threshold from Spearman correlation coefficient result. The study only selected 23 from 80 features and achieved high performance with 98.8% accuracy (Fitni & Ramli, 2020). Filter method allows selection from statistical evaluations. The threshold was specified to decide which variables would be included in modelling process.

Before processing the filtering method, data normality test is conducted using Kolmogorov-Smirnov test to decide which correlation metrics can be used. If the data distribution is normal, the appropriate method is parametric test such as Pearson correlation test. Meanwhile, if data has non-normal distribution, the appropriate testing is using non-parametric tools such as Spearman correlation test (Mooi & Sarstedt, 2010; Schober et al., 2018). Spearman correlation test is calculated by ranking the values of each variable and converting nonlinear relationship to linear (Schober et al., 2018). With this ranking method, Spearman correlation coefficient test is relatively robust against outliers. Spearman coefficient value is generally referred to as ρ (rho) or rs. Spearman's correlation formula is shown in the following formula:

$$\rho = 1 - \frac{6 \sum d_i^{\ 2}}{n\ (n^2 - 1)}$$

Spearman correlation coeffiecient valued between -1 to +1. ρ = 0 indicates there is no significant relationship between variables. The closer ρ to -1 or +1, the more significant relationship between variables. Feature selection method for this research is based on classification of correlation coefficient. If variable coeffiecient has less than 0.3, the relationship between variables has very low significance and almost no correlation (H. Chen & Chang, 2021). Therefore, features with < 0.3 coefficient will be eliminated.

Table 2. Correlation Coefficient Classification

Coefficient	Correlation classification
> 0.8	High
0.5 – 0.8	Moderate
0.3 – 0.5	Low
< 0.3	Very low

Modelling Methods

This study compares performance of three different modelling methods, linear regression, Neural Network, and Gradient Boost Regression. Linear regression is a classic method commonly used in predictive analysis techniques for datasets with linear relationships. Linear regression equation formula with multiple variables is shown in the following mathematical notation (Yasir et al., 2022):

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 \dots \alpha_k x_k + e$$

Artificial Neural Network, also known as NN, is an algorithm inspired by the Neural Network of the human brain. The human brain consists of neurons, small interconnected units that have the ability to transmit signals one anothers. In 1943, McCulloch and Pitts applied concept of neurons to artificial Neural Network (ANN) (Swamynathan, 2017). NN consists of nodes and connectors and uses mathematical functions to create predictive model. NN used mathematical function similar to linear regression as follows:

$$y = w_0 + w_1 x_1 + w_2 x_2 \dots w_n x_n$$

In the mathematical equation above, x is the independent variable, and w is the weight value. Once the input variables are processed, learning will begin by determining the connection weights between neurons and evaluating the model error. Then, the weights will be adjusted to reduce the error value. This process is repeated until the predictive model finds the error rate approaching value near 0 (Yasir et al., 2022)

In addition, Gradient Boost Regression is an ensemble method, combining many simple models called weak learners (Muller & Guido, 2017). This algorithm works with a random forest approach to build many trees, where each trees corrects errors of the previous tree to create a stronger predictive model (Swamynathan, 2017). The more trees, the greater the algorithm performance. Another parameter of gradient boost is learning rate, which controls how many times the tree will try to correct errors in the previous tree. The greater learning rate, the more correction processes will be, making the model become more complex (Muller & Guido, 2017). This boosting algorithm can identify weak learner by gradients and work with regression function. The steps follows to fit classifier to the training data to initial model with a constant value $y_0(x) = argmin\delta \sum_{i=1}^{n} \quad L(y_m, \delta)$. Then each iteration will calculate error loss, use it to

fit a new base learner, and update the estimate the output (Swamynathan, 2017).

Evaluation Metrics

There are 3 metrics that are commonly used in testing the performance of predictive models, R-Square, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) (Swamynathan, 2017). A predictive model with the smallest RMSE and MAE values is consideres as a good model because it has the lowest error rate. Similarly, predictive model with R-Square value closer to +1 or -1 is regarded as a strong predictive model.

MAE is the absolute average of the difference between predicted values and actual values (Kotu & Deshpande, 2019). If the prediction result has MAE value of 1,000,000, it means that the predicted value of the average net sales profit has a difference of ±1,000,000 to the actual value. The smaller MAE value, the better the predictive model performance. MAE is shown by the following mathematical formula, where y is the value of the target variable and yˆis the predicted value.

$$MAE = \frac{1}{n} \Sigma |y - \hat{y}|$$

RMSE is the square root of the squared average error. The difference of predicted and actual data can have negative or positive values, and there is a possibility that the value could be

0. Therefore, it is necessary to calculate by root mean square error to eliminate negative values (Swamynathan, 2017). The smaller the RMSE value, the better the model performance. RMSE is written in mathematical notation as follows, where y is the value of the target variable and y ̂ is the predicted value.

$$RMSE = \sqrt{\frac{1}{n}\Sigma(y - \hat{y})^2}$$

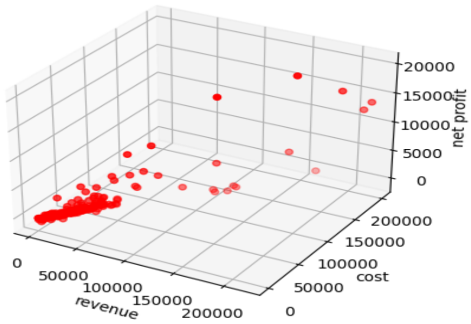
R-square or coefficient of determination is a metric for evaluating goodness of fit. The R-square value shows the total proportion of variance in the dependent variable explained by the independent variable, which is shown at a value between 0 and 1. A value close to 1 indicates a good R-square value. The R-square calculation is the division of the Total Sum of Square Residual (Σ SSR) by the Total Sum of Square (Σ SSR), or written using the following mathematical formula:

$$R^2 = \frac{\Sigma SSR}{\Sigma SST} = \frac{\Sigma(\hat{y}_i - \underline{y})^2}{\Sigma(y_i - \underline{y})^2}$$

RESULTS AND DISCUSSION

The graph shows data visualization from branch data with 3D scatterplot across 3 variables: revenue, cost, and net profit. profit and gross profit. Based on the graph, it is seen that the data pattern can give understanding of branch office size. The data point mostly has similar grouping towards small branch office. Several datapoints have high revenue, cost, and net profit came from big city of branch office. This spatial relationship can give understanding of data distribution from dataset.

Figure 2. Scatterplot of Dataset by Revenue, Cost, and Net Profit



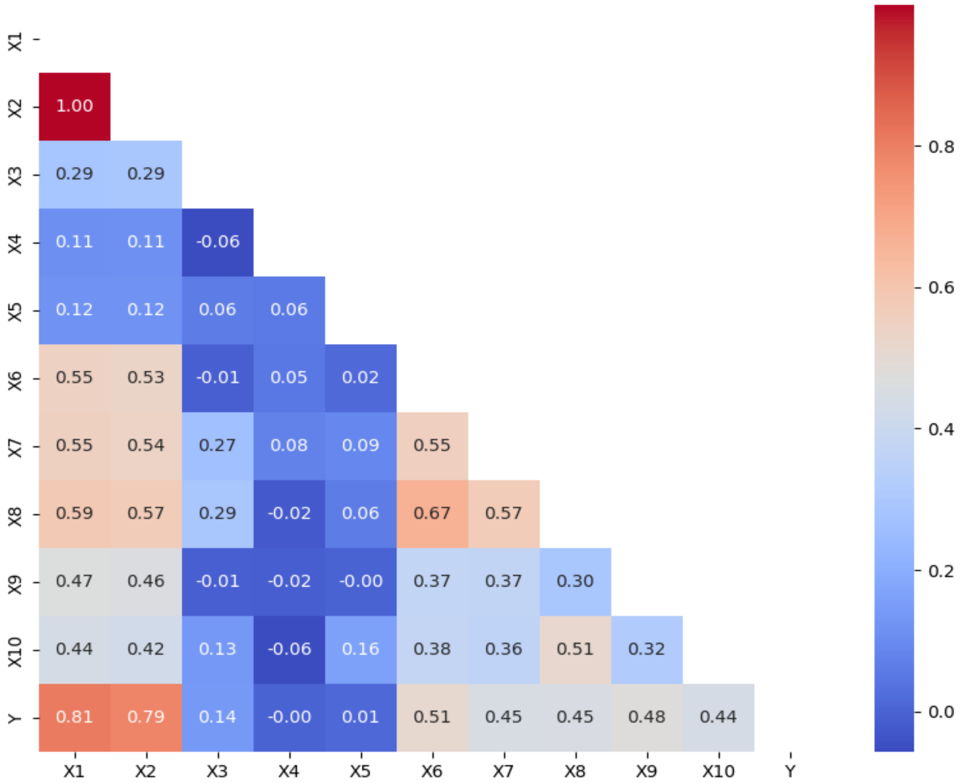
Kolmogorov-Smirnov test is conducted to assess data normality. The table follows shows result of the test.

Table 3. Data Normality Test Result

Column Name	K-S Statistics	P-value
X1	0.240	0.000
X2	0.231	0.000
X3	0.238	0.000
X4	0.329	0.000
X5	0.396	0.000
X6	0.122	0.006
X7	0.209	0.000
X8	0.213	0.000
X9	0.178	0.000
X10	0.301	0.000
Y	0.294	0.000

Based on the table, it is seen that each variable has p-value < 0.05. It means that the data is non-normally distributed. The graph also shows an asymmetrical distribution with long tail on the right side and concentrated on the left side. This result is due to only a few branch are located in the large cities, such as Jakarta and Surabaya, which have larger office sizes. Dealing with non-normally distributed data requires the use of non-parametric test, which do not assume normality and robust against outlier (Schober et al., 2018). Therefore, Spearman correlation test is used for feature selection.

Figure 3. Correlation Matrix with Spearman Correlation Coefficient



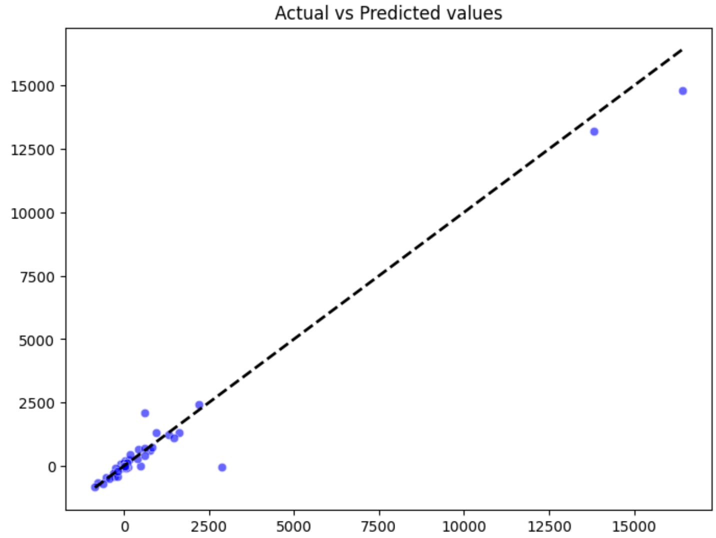
According to the result of Spearman correlation test, revenue (X1) has a highest correlation coefficient. This means that X1 has significant impact to net profit (Y) with strong correlation. Cost of Goods sold (X2) and employee cost (X6) has significant influence to net profit with moderate correlation. Transport cost (X7), office & utility cost (X8), general & administration cost (X9), and depreciation (X10) also has significant influence with low correlation. Meanwhile, direct selling cost (X3), warehouse cost (X4), and promotion cost (X5) has very low correlation coefficient and almost irrelevant to net profit (Y). Therefore, in this paper, X3, X4, X5 is eliminated. Only 7 variables are used as final parameter for modelling.

Modelling

Modelling process is conducted by comparing performance result of 3 algorithms to find the best predictive model with the highest accuracy. Each dataset has its own characteristics, it is necessary to find predictive model that is suitable with the dataset. The algorithm chosen for this study is based on the previous research methods: linear regression (Panwar et al., 2021; Yasir et al., 2022), Neural Network (Altman et al., 2020; Hew et al., 2019; Sharifi et al., 2019; Zhang et al., 2022), dan Gradient Boost Regression (Wisesa et al., 2020). Dataset is divided into data training to train the machine, and data test for evaluation. Usually about 70% to 90% data in a dataset is used for training, while the rest is used for data testing (Kotu & Deshpande, 2019). This research used a ratio of 80% training data and 20% testing data.

Linear Regression

Figure 4. Actual vs Predicted Values of Linear Regression Model



Linear regression is an algorithm that uses training data to create a linear model with the simple equation y=ax+b (Yasir et al., 2022). In this equation, a is the coefficient value and b is the constant value (intercept). Linear regression is applied with sklearn library from Python. The mathematical formula for linear regression result is:

$$y = 1.03x_1 - 1.04x_2 - 0.81x_3 - 1.96x_4 - 0.86x_5 - 0.06x_6 - 1.00\ x_7 - 153.14$$

y = net profit

x_1 = revenue

x_2 = cost of goods sold

x_3 = employee cost

x_4 = transportation cost

x_5 = office & utility cost

x_6 = general & administration cost

x_7 = depreciation & amortization cost

Linear regression result shows that intercept value is -153.14, represents the value of net profit when all variables are zero. The coefficient value explains relationship between variables to net profit. Sales revenue is the only variable with positive coefficient value to net profit. An increase of revenue correlates with predicted Y increase of 1.03 units. Meanwhile, negative coefficient value indicates inverse relationship to net profit. An increase of cost of goods sold, employee cost, transportation cost, office & utility cost, general & administration cost, and depreciation & amortization cost sold linked to predicted Y decrease from coefficient value units, and vice versa. Evaluation test is conducted by calculationg MAE, RMSE, and r-square. The MAE shows that linear regression predictive model has average difference of ±294.34 to the actual value. The prediction error rate of RMSE metric value is 615.49. Meanwhile the goodness of fit value of linear regression predictive model using r-square metric is 0.9666.

Neural Network

Figure 5. Actual vs Predicted Values of Neural Network Model

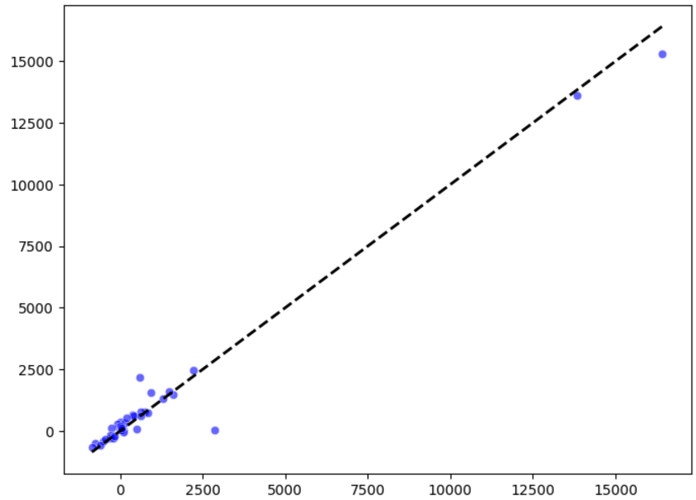
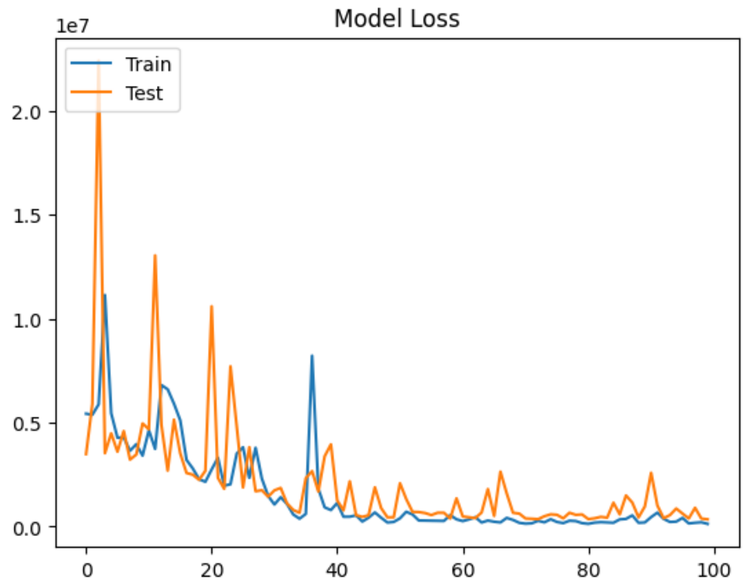


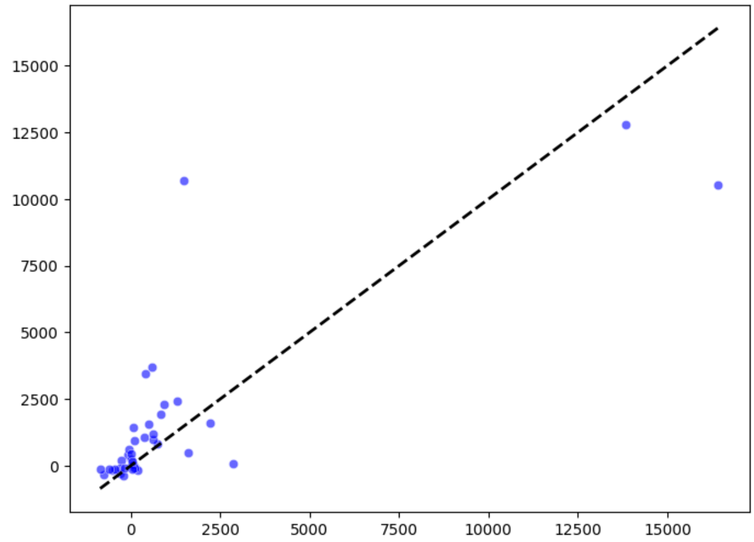
Figure 6. Model Loss From Neural Network



Predictive model with Neural Network are using Keras library with ReLU (rectified linear unit) activation function and linear function as final layer. The layers are defined to have 4 layers with learning rate 0.01 in every sequential cycle. The algorithm will try to minimize loss value by evaluate and modify the weight value repeatedly on each cycle until the algorithm finds the best result. The visualization of learning cycle of Neural Network model is shown in figure 7. X-axis shows the number of epoch or learning model cycle, and y-axis is the loss value of each cycle. The algorithm will use lowest loss point as a reference to make prediction. Evaluation test is conducted by calculationg R-square, MAE, and RMSE metrics. The evaluation results show that Neural Network predictive model has average difference of ±298.34 to the actual value. The prediction error rate of RMSE metric value is 586.44. Meanwhile the goodness of fit value of Neural Network predictive model using r-square metric is 0.9697.

Gradient Boost Regression

Figure 7. Actual vs Predicted Values of Gradient Boost Regression Model



The process of creating Gradient Boost Regression predictive model begins by importing the “GradientBoostingRegressor” class from sklearn.ensemble library. Then, determine the parameter values to create a predictive model. Evaluation test is conducted by calculationg R-square, MAE, and RMSE metrics. The evaluation results show that prediction from gradient boost regression model has average difference of ±1069.31 to the actual value. The prediction error rate of RMSE metric value is 2031.68. Meanwhile the goodness of fit value of Gradient Boost Regression predictive model using r-square metric is 0.636.

CONCLUSION

This study proposed machine learning-based technique for sales profit forecasting. Feature selection was conducted with filter method by using Spearman correlation coefficient. The feature selection process leaves 7 variables out of a total of 10 variables, by eliminating variables with no significance to net profit. The correlation test result shows that the revenue has the highest significant influence to net profit with positive relationship. If company wants to have a high net profit, the company should focus on creating strategy to increase sales revenue. The higher the sales made by the company, it will achieve optimal profits (Suzan & Nabilah R., 2020). However, direct selling cost, warehouse cost, and promotion cost has very low correlation coefficient and almost irrelevant to net profit based on the data source.

Overall, the performance of the predictive model with the Neural Network algorithm has the best performance, followed by Linear Regression in second place with slight difference of accuracy rate, and Gradient Boost regression in the last place. Linear Regression has the lowest MAE. Meanwhile Neural Network has the smallest RMSE. The metrics explaned the difference between predicted result and actual value, the smallest result has the better performance compared to other algorithms. Neural Network also has highest value of r-square with 96.97% accuracy rate. Which means that the model has a high degree of accuracy to explain the variability in the data. Therefore, it can be concluded that predictive model algorithm with the best performance in this case study is Neural Network with predictors of cost of goods sold, sales revenue, employee costs, vehicle and transportation costs, office and utility cost, general and administrative costs, and depreciation and amortization costs. This findings can help business company to find alternative of profit forecast method with a high performance and accuracy.

Due to the limitation to access data source, this research only conducted with existing internal report data of PPI in range of year 2017 to 2022. As a result, the research only discuss features from dataset with limited data. However to get better performance, larger dataset is needed. The larger dataset used, the more machine learning can learn data patterns. Future study can employ forecast with larger data and features. This study also only compared performance from three algorithms. It is possible that other algorithms can have good accuracy but not discussed on this paper. Future research can explore additional machine learning techniques and test various parameter variable to gain more insights in profit forecasting.

REFERENCES

Ajol, T. A., Gran, S. S., & Ali, A. N. A. (2018). Minimizing Warehouse Operation Cost. In R. Saian & M. A. Abbas (Eds.), *Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017* (pp. 625–634). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-10-8471-3_62

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2020). A Race for Long Horizon Bankruptcy Prediction. *Applied Economics*, *52*(37), 4092–4111. https://doi.org/10.1080/00036846.2020.1730762

Chen, H., & Chang, X. (2021). Photovoltaic power prediction of LSTM model based on Pearson feature selection. *Energy Reports*, *7*, 1047–1054. https://doi.org/10.1016/j.egyr.2021.09.167

Chen, I. F., & Lu, C.-J. (2021). Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms. *Processes*, *9*(9). https://doi.org/10.3390/pr9091578

Chen, P., Li, F., & Wu, C. (2021). Research on Intrusion Detection Method Based on Pearson Correlation Coefficient Feature Selection Algorithm. *Journal of Physics: Conference Series*, *1757*(1). https://doi.org/10.1088/1742-6596/1757/1/012054

David, F. R., & David, F. R. (2017). *Strategic Management: A Competitive Advantage Approach, Concepts and cases* (S. Wall (ed.); 16th ed.). Pearson Education.

Dewi, P. E. D. M., Devi, S., & Masdiantini, P. R. (2021). *Analysis of Cost of Sold and Production Costs on Company Profit BT - Proceedings of the 6th International Conference on Tourism, Economics, Accounting, Management, and Social Science (TEAMS 2021)*. 388–391. https://doi.org/10.2991/aebmr.k.211124.056

Dharmayuni, L., Sunarsi, D., Suranta Karina Sembiring, E., Satata, S., Bahrudin, U., Erlangga, H., Anwar, A., Fuad Salam, A., Sapari Kahpi, H., & Purwanto, A. (2021). Effect of Distribution Cost and Promotion Cost on Tyre Industries Sales Performance. *Annals of R.S.C.B.*, *25*(4), 12672–12684. http://annalsofrscb.ro

Dumadi, Sugiani, Kharisma, A. S., Afrida, N., & Wulandari, H. K. (2023). Effect Of Production Costs, And Distribution Costs On Turnover Sales (Case Study at PT Sandana Istana Multigas). *Jurnal Ekonomi*, *12*(01), 806–811.

Elise, E., & Daryanto, W. M. (2018). The Effect of Depreciation Methods on the Profitability and Net Present Value (NPV): A Case Study of Nam Con Son 2 Phase Pipeline Project, Vietnam for the Period …. *South East Asia Journal of …*, *17*(3), 8–12. https://repository.ipmi.ac.id/553/

Evmenchik, O. S., Niyazbekova, S. U., Seidakhmetova, F. S., & Mezentceva, T. M. (2021). *The Role of Gross Profit and Margin Contribution in Decision Making BT - Socio-economic Systems:*

Paradigms for the Future (E. G. Popkova, V. N. Ostrovskaya, & A. V Bogoviz (eds.); pp. 1393–1404). Springer International Publishing. https://doi.org/10.1007/978-3-030-56433-9_145

Fitni, Q. R. S., & Ramli, K. (2020). Implementation of Ensemble Learning and Feature Selection for Performance Improvements in Anomaly-Based Intrusion Detection Systems. *2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 118–124. https://doi.org/10.1109/IAICT50021.2020.9172014

Ginieis, M., Hernández-Lara, A. B., & Sánchez-Rebull, M. V. (2020). Influence of airlines' size and labour costs on profitability. *Aviation*, *24*(4), 157–168. https://doi.org/10.3846/aviation.2020.12539

Goestjahjanti, F. S., & Widayati, C. C. (2020). *Significance Effect Cost of Goods Sold and Inventory on Sales PT. Nippon Indosari Corpindo Tbk BT - Proceedings of the 4th International Conference on Management, Economics and Business (ICMEB 2019)*. 200–205. https://doi.org/10.2991/aebmr.k.200205.037

Griffin, R. (2020). *Manajemen, Jilid 2* (G. Gania & W. C. Kristiaji (eds.); 7th ed.). Erlangga.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concept and Techniques* (3rd ed.). Morgan Kaufmann.

Hew, J. J., Leong, L. Y., Tan, G. W. H., Ooi, K. B., & Lee, V. H. (2019). The age of mobile social commerce: An Artificial Neural Network analysis on its resistances. *Technological Forecasting and Social Change*, *144*, 311–324. https://doi.org/10.1016/j.techfore.2017.10.007

Jones, M., Alexander, C., & Lowenberg-DeBoer, J. (2014). A simple methodology for measuring profitability of on-farm storage pest management in developing countries. *Journal of Stored Products Research*, *58*, 67–76. https://doi.org/https://doi.org/10.1016/j.jspr.2013.12.006

Keputusan Menteri Badan Usaha Milik Negara Nomor KEP-101/MBU/2002 tentang Penyusunan Rencana Kerja dan Anggaran Perusahaan Badan Usaha Milik Negara, 1 (2002) (testimony of Kementerian BUMN).

Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, *303*(1–2), 159–174. https://doi.org/10.1007/s10479-020-03666-w

Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts and Practice* (2nd ed., Vol. 1, Issue 69). Elsevier Inc.

Listiawati, Karta Negara Salam, Retno Wulansari, & Pramudi Harsono. (2021). Promotion Costs Analysis To Increased Volume Sales In The Convection Companies. *International Journal of Science, Technology & Management*, *2*(5), 1542–1551. https://doi.org/10.46729/ijstm.v2i5.300

Lukić, R., Hanic, A., & Hanic, A. (2016). The impact of depreciation expense on performance of trade in Serbia. *International Review*, *2016*, 123–137. https://doi.org/10.5937/intrev1604123L

Mooi, E., & Sarstedt, M. (2010). Hypothesis Testing & ANOVA. In *A Concise Guide to Market Research*. https://doi.org/10.1007/978-3-642-12541-6_6

Muller, A. C., & Guido, S. (2017). Introduction to Machine Learning with Python: A Guide for Data Scientist. In *O'Reilly Media, Inc.* (1st ed.).

Mun, S. G., & Jang, S. (Shawn). (2018). Restaurant operating expenses and their effects on profitability enhancement. *International Journal of Hospitality Management*, *71*, 68–76. https://doi.org/https://doi.org/10.1016/j.ijhm.2017.12.002

Nafarin, M. (2008). *Penganggaran Perusahaan* (3rd ed.). Salemba Empat.

Nelli, F. (2018). Python data analytics: With Pandas, NumPy, and Matplotlib: Second edition. In *Python Data Analytics: With Pandas, NumPy, and Matplotlib: Second Edition* (2nd ed.). Apress. https://doi.org/10.1007/978-1-4842-3913-1

Noor, H. F. (2017). *Ekonomi Manajerial, Edisi Revisi*. Rajawali Pers.

Panwar, B., Dhuriya, G., Johri, P., Yadav, S. S., & Gaur, N. (2021). Stock Market Prediction Using Linear Regression and SVM. *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 629–631. https://doi.org/10.1109/ICACITE51222.2021.9404733

Prahada, M. H., Siregar, M. Y., & Sugito, S. (2022). The Effect of Distribution Costs and Distribution Channels on the Sales of PT. Pasha Jaya Medan. *Jurnal Ilmiah Manajemen Dan Bisnis (JIMBI)*, *3*(1), 42–50. https://doi.org/10.31289/jimbi.v3i1.992

PT PPI. (2023). *Laporan Tahunan 2022 PT Perusahaan Perdagangan Indonesia*. https://www.ptppi.co.id/tata-kelola/good-corporate-governance/

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5). https://journals.lww.com/anesthesia-analgesia/fulltext/2018/05000/correlation_coefficients__appropriate_use_and.50.aspx

Sharifi, M., Khazaei Pool, J., Jalilvand, M. R., Tabaeeian, R. A., & Ghanbarpour Jooybari, M. (2019). Forecasting of advertising effectiveness for renewable energy technologies: A neural network analysis. *Technological Forecasting and Social Change*, *143*, 154–161. https://doi.org/https://doi.org/10.1016/j.techfore.2019.04.009

Suzan, L., & Nabilah R., S. (2020). Effect of production Costs and Sales on the Company's Net Profit. *Jurnal Akuntansi*, *24*(2), 169. https://doi.org/10.24912/ja.v24i2.689

Suzan, L., & Siagia, H. A. F. (2023). Analysis The Effect Of Production Costs, Operational Costs And Sales Volume On Net Profit (In Pharmaceutical Subsector Companies Listed on the Indonesia Stock Exchange in 2014-2021). *JHSS (Journal of Humanities and Social Studies)*, *07*(02), 367–373. https://doi.org/10.33751/jhss.v7i2.7880

Swamynathan, M. (2017). Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python. In *Scandinavian Journal of Information Systems*. Apress. http://aisel.aisnet.org/sjis%0Ahttp://aisel.aisnet.org/sjis/vol19/iss2/4

Wisesa, O., Adriansyah, A., & Khalaf, O. I. (2020). Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm. *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP)*, 101–106. https://doi.org/10.1109/BCWSP50066.2020.9249397

Yasir, M., Ansari, Y., Latif, K., Maqsood, H., Habib, A., Moon, J., & Rho, S. (2022). Machine learning–assisted efficient demand forecasting using endogenous and exogenous indicators for the textile industry. *International Journal of Logistics Research and Applications*, 1–20. https://doi.org/10.1080/13675567.2022.2100334

Zhang, Y., Yamamoto, M., Suzuki, G., & Shioya, H. (2022). Collaborative Forecasting and Analysis of Fish Catch in Hokkaido From Multiple Scales by Using Neural Network and ARIMA Model. *IEEE Access*, *10*, 7823–7833. https://doi.org/10.1109/ACCESS.2022.3141767