

Analysis of Random Forest, Multiple Regression, and Backpropagation Methods in Predicting Apartment Price Index in Indonesia

I NYM Yoga Saputra¹, Siti Saadah², Prasti Eko Yunanto³
^{1,2,3}Informatics Engineering, Telkom University, Jl. Telekomunikasi, Bandung, Indonesia

ARTICLE INFO

Article history:

Received June 28, 2021

Revised July 13, 2021

Accepted July 18, 2021

Keywords:

Backpropagation;
Multiple Regression Forecasting;
Prediction;
Predicted Apartment Prices;
Random Forest Forecasting

ABSTRACT

This study focuses on predicting the apartment price index in Indonesia using property survey data from Bank Indonesia. In the era of the Covid-19 pandemic, accurately predicting the sale and purchase price of apartments is essential to minimize the impact of losses, thus making apartment prices attractive to predict. The machine learning approach used to predict the apartment price index are the Random Forest method, the Multiple Regression method, and the Backpropagation method. This study aims to determine which method is more effective in predicting small amounts of data accuracy. The data used is apartment price index data from 2012 to 2019 in the JABODEBEK area. The research will produce prediction accuracy that will determine the effectiveness of the application of the method. The Random Forest method with parameters $n_estimators=100$ and $max_features="log2"$ produces an R2 accuracy of 0.977. The Multiple Regression method with a correlation between the selling price and rental price variables is 0.746, and the rental inflation variable is 0.042 produces an R2 accuracy of 0.559. The Backpropagation method with a 1000-4000-1 hidden scheme and 20000 iterations produces an R2 accuracy of 0.996. Therefore, the Backpropagation method is more suitable in this study compared to the other two methods. The Backpropagation method is suitable because it gets almost perfect accuracy, so this method will minimize losses in investing in buying and selling apartments in the Covid-19 pandemic era.

This work is licensed under a [Creative Commons Attribution-Share Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



I NYM Yoga Saputra,

Informatics Engineering, Telkom University, Jl. Telekomunikasi, Bandung, Indonesia

Email: nyomanyoga@student.telkomuniversity.ac.id

1. INTRODUCTION

Occupancy is an essential element in life included in primary needs. One of the housings that are often encountered is an apartment. Apartments are attractive to buy and sell because of their minimalist design, luxury, and high selling value. Sales of apartment properties during the Covid-19 pandemic are getting tighter due to the economic downturn. According to a simulation conducted by research [1] that the economic impact that occurred reduced economic growth in 2020 from 5% to between 4.2% and -3.5%. The Covid-19 pandemic has resulted in a decline in income in all economic sectors, and the residential property sector has already felt the impact from 2020 [2]. This impact creates problems for the community to be more careful in deciding between selling or buying apartments. This makes apartment prices attractive to predict during the Covid-19 pandemic. This test is carried out using apartment price data taken from a Bank Indonesia survey. Based on the research buying attitude [3], it is concluded that the price variable is one of the determinants of buying attitude of 68.6%, and is also supported by the variables of apartment facilities, location and access, environmental quality, physical quality, and promotion. Prediction results will significantly affect the decision to sell or rent apartments to maximize profit by looking at price inflation. Price inflation is a condition where there is an imbalance in the value of the flow of goods. A good prediction will produce a minimum accuracy of 80%, so the prediction model will help make decisions during the urgent Covid-19 pandemic. This study

will compare the Random Forest, Multiple Regression, and Backpropagation methods to predict the apartment price index accurately.

Several studies have been conducted on the Random Forest, Multiple Regression, and Backpropagation methods. Based on the analysis carried out by research [4][5][6], apartment rental price predictions in DKI Jakarta with the Random Forest method resulted in an accuracy of 92.12% and on apartment sales in Ljubljana, Slovenia, using this method showed that the algorithm it can identify and predict apartment prices well. Implementation with the time series method shows that the Random Forest method has a high performance. The Random Forest method has an advantage in terms of forecasting. As stated by health research [7], the model has 94% accuracy in predicting Covid-19 patients for case severity and the likelihood of outcome, recovery, or death. In predicting Alzheimer's disease, the Random Forest also has high accuracy with an average mAUC = 0.80 and BCA = 0.74 [8]. The Random Forest method has advantages in complex models with many variables and samples, with a predictive power of 95.06% [9]. The advantage of the Multiple Regression method is that it predicts based on interactive patterns from relevant data in the past, and usually, line patterns will form in the model. Multiple Regression Method is generated from analyzing the pattern of relationships between variables [10]. Research [11] states that the Multiple Regression method has 50% accuracy in predicting demand for bicycle rental, and research [12] has 84.5% accuracy in predicting property prices with several parameters.

The Backpropagation method is interesting to compare because the way it works is similar to the Multiple Regression method, namely by paying attention to the previous data. The Backpropagation method works by studying and calculating the error of a model and repeating it until the iteration is determined to get a small error. The Backpropagation method is one of the Artificial Neural Network algorithms with specifications for identifying input, prediction, pattern recognition. The hallmark of the Backpropagation method is that it has three layers, namely the input layer, the hidden layer where the data will be processed, and the output layer [13]. In research [14], an experiment on human development was carried out. From the results of the research, the Backpropagation method had 100% accuracy. Research [15] states that the Backpropagation method is very well used in the forecasting process. This is supported by the results of research [16] which has an accuracy of 93% with an MSE error value of 0.000956628. Based on this explanation, the Backpropagation method will be used in this study to show the method's accuracy compared to the previous two methods in predicting the apartment price index.

This study aims to determine the best accuracy of the machine learning approach of Random Forest, Multiple Regression, and backpropagation methods in predicting a small amount of data which is expected to have an accuracy of at least 80%. This best method will later be used to minimize the impact of losses in the Covid-19 pandemic in making apartment buying and selling decisions.

2. RESEARCH METHOD

This research methodology examines the steps taken to achieve the desired output. This study allows the Random Forest method, the Multiple Regression method, and the Backpropagation method to predict the apartment selling price index. Therefore, a flowchart will be made to explain the series of processes that are being carried out. Fig. 1 is a flowchart of the system design that has been made.

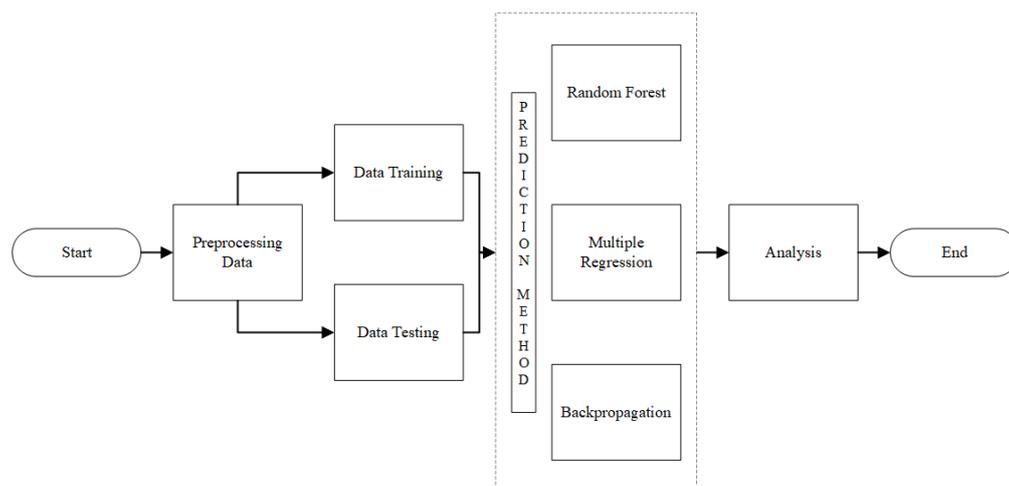


Fig. 1. Prediction System Design

In Fig. 1, the first thing to do is to pre-process the data. Data pre-processing is the process of processing data until the data is ready for use. The processed data is divided into two parts: training data and test data, with a ratio of 60:40, 70:30, and 80:20. The following process is a prediction method that will be carried out using the Random Forest method, the Multiple Regression method, and the Backpropagation method. The final step is the analysis process. The analysis process aims to show the accuracy of the method in predicting the selling price of apartments with several accuracy measurements.

2.1. Pre-process Data

At this stage, apartment price data for 2012 to 2019 are used in the JABODEBEK area, totaling 29 data. The columns used in the test are the column rental price, selling price, rental inflation, and selling inflation. In this process, additional columns, namely inflation columns, will be added to support the prediction process for Multiple Regression and Backpropagation. The sample dataset used for this study will be illustrated in Table 1. This sample represents the first 5 data from a total of 29 data that will be used.

Table 1. Dataset Illustration

Rent	Selling	Rent Inflation	Selling Inflation
95.84	91.86	0	0
100.36	98.56	4.503786	6.79789
101.64	101.37	1.259347	2.772023
102.16	108.21	0.509005	6.321042
102.47	113.15	0.302528	4.365886

2.2. Data Training

The normalized data will be tested three times for each method by dividing the total number of apartment price data. The first test uses a ratio of 60:40, 60% training data, and 40% test data. The second test uses a ratio of 70:30, 70% training data, and 30% test data. While the third test uses a ratio of 80:20, training data is 80%, and test data is 20%. The data distribution is done with a ratio of 60:40, 70:30, and 80:20 because the amount of data used is only 29 data. These three test models aim to see the accuracy of the prediction results with the difference in the amount used as training data.

This is also supported by the study results [17], which concluded that using a ratio of 90:10 indicates that by using the Random Forest method, the percentage deviation of predictions is 5.5%. For comparison, the regression method is 20%. It is essential to choose variables with high data correlation and data training twice in the model learning process to support perfect prediction accuracy.

2.3. Data Testing

After the learning model is obtained from the training data, the next step is testing the data. This process is essential because it determines the model made on the training data is good or overfitting. This process was carried out three times by trying the test data into a model created with training data. The first test uses 40% test data, the second test uses 30% test data, and the third test uses 20% test data. The difference in the test data aims to determine the performance of the trained algorithm model so that the effectiveness of the model in predicting new data is known.

2.4. Prediction Using Random Forest

According to research [18], the Random Forest method has a prediction accuracy rate of 81%, where this accuracy has a reasonably high value. The Random Forest method will be completed in this experiment with the illustration in the flowchart Fig. 2. The first step is to enter data as input data. The data used is data from the rental and sales columns to build a prediction model for apartment selling prices. The second is to build a model, wherein in this process, the input data will be modeled on the Random Forest method by choosing predictions at random. This model is built to predict the apartment price index with a small amount using the features `n_estimators=100` and `max_features="log2"`. This parameter is expected to improve results because it uses 100 random search trees, which will later be selected to calculate forest logs to find the best split feature value. The third step is to calculate the average best value from the results that have been modeled, and later these results will be used as the final prediction results that have high accuracy.

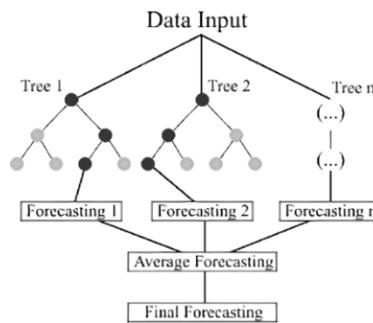


Fig. 2. Random Forest Forecasting Illustration

2.5. Prediction Using Multiple Regression

According to research conducted [19] [20], the use of poor Multiple Regression model predictor variables will reduce the prediction model’s performance, and Multiple Regression has equation calculations that are easy to understand and satisfying in predicting. In this experiment, the first step taken by the Multiple Regression method is the selection of variables that will be used as data input variables. From the selection results, it is decided to use the rental variable and rental inflation to predict the selling price of the apartment. The selling price is the value sought (the dependent variable), the rental price is the independent variable, and the rental inflation is the coefficient of the equation. The selection of parameters is based on the level of data correlation, with the rental variable correlation level of 0.746 and the rental inflation variable of 0.042. The second step is to create a model. At this stage, the training data that has been processed on pre-processing data with a ratio of 60:40, 70:30, and a ratio of 80:20 will be modeled to find the best ratio. Repeated calculations will be carried out to obtain a predictive model of the correct Multiple Regression method on the data. The third step is to evaluate the model results with data testing to get high accuracy. After the Multiple Regression model is built, the next step is to predict the actual data in the model that has been built. The working steps of making the Multiple Regression model will be visualized in the flowchart Fig. 3.

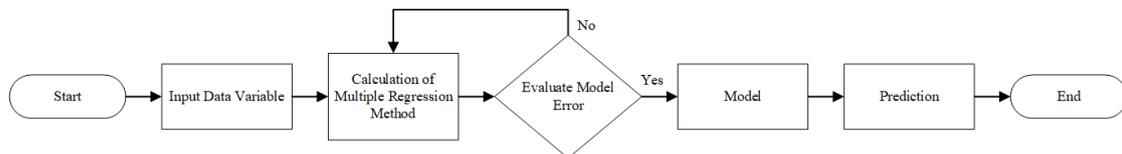


Fig. 3. Multiple Regression Illustration

2.6. Prediction Using Backpropagation

The results of the study [21] showed that the MSE was 0.003901 with 1500 iterations and 21 layers. Based on these results, this experiment carried out with the Backpropagation method is the selection of variables that will be used as data input variables, starting from the decision to decide to use rent and inflation variables to predict the selling price of apartments. The selection of these parameters is based on the variables used in the Random Forest method and the Multiple Regression method, which are expected to produce a comparable model to compare the predicted results. The second step is to create a model. At this stage, the selected variables will be calculated using training data, using the features of 1000 input layers, 4000 hidden layers, and 20000 iterations, the measurement parameter of the Backpropagation method is loss = “mean_squared_errors.” This parameter will calculate the model with training data repeatedly through the main layer and continues to the hidden layer until it gets a low mean squared error accuracy for the specified iteration. Using this feature is expected to get good accuracy results by repeating the error learning and doing it in several different processes. The third step is testing the model with data testing to get a model with reasonable accuracy. After the Backpropagation model is built, the next step is to predict the actual data in the model that has been built. The stages of the work process of making the Backpropagation model will be visualized in the flowchart Fig. 4.

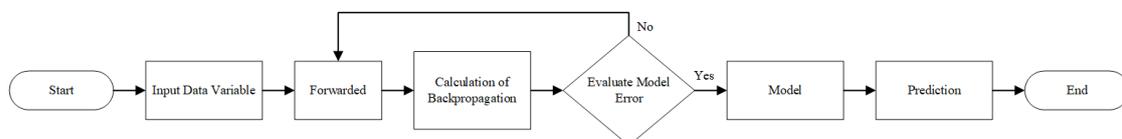


Fig. 4. Backpropagation Illustration

2.7. Analysis

In this study, the results obtained will be analyzed to see the model's accuracy in predicting. The analysis was carried out by comparing the prediction results between the Random Forest, Multiple Regression, and Backpropagation methods to determine the accuracy of apartment selling prices with little data, namely 29 data. The accuracy is represented by using the average accuracy measurement, namely Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Squared Error (RMSE), and R Square (R²).

This accuracy measurement was chosen because MAE is more intuitive in providing a mean error, while MSE is very sensitive to outliers. RMSE has the same measurement scale as the evaluated data. R² is a measure of accuracy where there is a gap in the number of explanatory variables [22].

3. RESULTS AND DISCUSSION

In this study, three scenarios of the test method will be carried out, each of which will use data sharing with a ratio of 60:40, 70:30, and a ratio of 80:20. The difference in the ratio of the data used aims to see the model's performance in producing a significant difference in accuracy from the Random Forest, Multiple Regression, and Backpropagation methods.

The first step is to find the variable with the best correlation. The correlation obtained in testing the data is the rental variable 0.746 and the inflation variable 0.042. This best variable will be used for modeling. The model generated from three methods will be tested on all available data to obtain the best method. The Backpropagation method with a ratio of 80:20 gets the best results, but it takes longer than the other two methods to execute the program. This happens because this method learns more data to produce a more significant error parameter and is repeated in several iterations to produce a model with the lowest error. In comparison, the Random Forest method has results that are not much different and fast execution time. This happens because Random Forest prediction generates random node values and calculates the best gain information to decide the outcome prediction. Meanwhile, The Multiple Regression method is not suitable for predicting small amounts of data because of how this method works by analyzing previously studied data sets to generate patterns in predicting data. It is proven in predicting the apartment price index with a small amount of data that the Backpropagation method and the Random Forest method are far superior to Multiple Regression. The details of the research that has been carried out will be described in detail at the following points.

3.1. First Experiment Data Ratio 60:40

The first test data will be divided into 60:40, where the training data will account for 60%, and the test data will account for 40% of the total data. The training data will use the 2012 data from the first quarter to the first quarter of 2016, while the test data uses data for the second quarter of 2016 to the first quarter of 2019.

The Random Forest algorithm is implemented using the Python programming language. The data that will be used in the prediction is the rental and sale column. This study uses the parameters `n_estimators=100` and `max_features="log2"`. In the first experiment, the Random Forest method had satisfactory results, and these results will be visualized in the graph. Fig. 5 at the top left is a graph of Random Forest's prediction results with the actual selling price visualized using a 60:40 ratio. From the actual data that has been visualized, it is seen that the predicted data follows the pattern in the actual data, even though it has a visible deviation distance. This difference occurs because the test data does not have a lot of training data to build the model. This shows that the Random Forest learning method requires optimal training data for better prediction accuracy in studying gain ratio information.

The Multiple Regression Algorithm is implemented using the Python programming language. This study uses the selling price parameter as the dependent variable, the rental price as the independent variable, and rental inflation as the equation coefficient. The choice of this parameter is based on the correlation value of the variable, the result of the rental correlation is 0.746, and the rental inflation variable is 0.042. Prediction results in the Multiple Regression method produce unsatisfactory predictions where the predictive data are visualized. There is a significant deviation at the beginning of the data, namely from 2012 to 2016. This data deviation shows that the Multiple Regression method in the deviation year failed to study the previous data pattern. Predictive data gradually improved after 2016, which showed success in the training data's learning patterns. These results will be visualized in Fig. 5 at the top right.

The Backpropagation method gets better results than the Multiple Regression method. Experiment with this method using the same column as the Multiple Regression method. This experiment uses 20000 iterations with the first 1000 input layers and 4000 hidden layers. The calculation of the loss used is the mean squared error. Prediction results with the backpropagation method produce an accuracy value above the expected threshold. In comparison, the prediction of this method is quite good from the Multiple Regression method but

is still below the Random Forest method. Fig. 5 at the bottom center shows the prediction results of Backpropagation with the actual selling price, which is visualized in graphical form. The results that have been visualized show that the predictive data still has many undirected data deviations from 2016 to 2019. This data deviation occurs because the number of training data models is only 60% of the total data. Therefore, the data studied is only 60% error, so if the test data is slightly different from the training model data, it will be difficult to predict.



Fig. 5. Approach Model Random Forest, Multiple Regression, and Backpropagation Ratio 60:40

Table 2 is the result of data accuracy from the first test. The results obtained are that the Random Forest method is superior to the other two methods, which has a small accuracy value close to zero and an R2 value of 0.964 or 96.4%. This is supported by how the Random Forest method works by taking trees at random and then evaluating these values to get a good ratio. In the second place, the best is achieved by the Backpropagation method, with an accuracy of 66.6%, but in 2016 there was a very significant deviation. While the Multiple Regression method is in third place, the Multiple Regression method is not recommended as a prediction method because it has a minimum accuracy of 55.9%. This is because the learning data used is only tiny, so this method is challenging to study previous data patterns. From the results obtained in the first experiment, training data was felt to affect subsequent experiments results significantly.

Table 2. Testing Accuracy Ratio 60:40

	MAE	MSE	RMSE	R2
Random Forest	0.021	0.055	0.023	0.964
Multiple Regression	0.049	0.681	0.082	0.559
Backpropagation	0.043	0.515	0.072	0.666

3.2. Second Experiment Data Ratio 70:30

The data will be divided into a ratio of 70:30, where the training data will account for 70%, and the test data will account for 30% of the total data. The training data will use 2012 data from the first quarter to the fourth quarter of 2016, while the test data will use data from the first quarter of 2017 to the first quarter of 2019.

In the second experiment, the Random Forest method gave satisfactory results, and the results will be visualized in a graph Fig. 6. The data used in the prediction is the rental and sale column and uses the parameters $n_estimators=100$ and $max_features="log2"$. Fig. 6 at the top left is a graph of Random Forest’s prediction results with the actual selling price visualized using a 70:30 ratio. From the actual data that has been visualized, it can be seen that the predicted data follows the pattern in the actual data and reduces deviations. The difference in data was seen from 2012 to 2016 and was visible from 2017 to 2018. The deviation occurred right after the start of data testing. This indicates that the Random Forest learning method requires optimal training data from all data for better predictive accuracy in the learning information gain ratio.

The Multiple Regression Algorithm is implemented using the Python programming language using the same variables as the first experiment. The multiple Regression method produces unsatisfactory predictions.

There is a significant deviation at the beginning of the data, namely from 2012 to 2015. This data deviation indicates that the Multiple Regression method in the deviation year failed to study the previous data pattern. Predictive data gradually improved after 2015, which showed success in the training data's learning patterns. These results will be visualized in Fig. 6 at the top right.



Fig. 6. Approach Model Random Forest, Multiple Regression, and Backpropagation Ratio 70:30

The backpropagation method again gets better results than the Multiple Regression method. This experiment still uses layers with a comparison of 1000-4000-1 with 20000 iterations, and the error calculation used is the mean squared error. Prediction results with the backpropagation method produce an accuracy value above the expected threshold. In comparison, the prediction of this method is quite good from the Multiple Regression method but is still below the Random Forest method. The results that have been visualized in Fig. 6 at the bottom center show that the prediction data still has a very significant deviation in the third quarter of 2017. The prediction results generated by the Backpropagation method are excellent. This data deviation occurs because the number of training data models is only 70% of the total data. Therefore, the data learn only 70% error, so if the test data is slightly different from the training model data, it will be difficult to predict.

Table 3 is the result of data accuracy from the second test. The results obtained in the second experiment are that the Random Forest method is superior to the other two methods, which has a small accuracy value close to zero and an R2 value of 0.976 or 97.6%. This is supported by how the Random Forest method works by taking 100 trees at random and then evaluating these values by calculating log2 features up to 70% of the training data to get a good ratio. In the second place, the best was achieved by the Backpropagation method, with reasonably good accuracy of 85.1%, but in 2017 there was a very significant deviation. This happens because the model does not calculate error accuracy from 2017 to 2019. After all, learning is only carried out as much as 70%, namely until 2016. While the Multiple Regression method is in third place, the Multiple Regression method is not recommended as a predictive method because it has an accuracy minimum is 55.9%. This is because only a small amount of learning data is used, so this method is challenging to study the previous data patterns. From the results obtained in the second experiment, training data significantly affected the experimental results.

Table 3. Testing Accuracy Ratio 70:30

	MAE	MSE	RMSE	R2
Random Forest	0.014	0.037	0.0193	0.976
Multiple Regression	0.049	0.681	0.082	0.559
Backpropagation	0.021	0.230	0.048	0.851

3.3. Third Experiment Data Ratio 80:20

The third test data will be divided into 80:20, where the training data will account for 80%, and the test data will account for 20% of the total data, which is 29. The training data will use 2012 data from the first quarter to the fourth quarter of 2017, while the test data from first quarter 2018 to first quarter 2019.

The Random Forest method in the third experiment will use training data with a ratio of 80:20. The addition of data combined with log2 calculations produces a very satisfactory gain ratio. Better prediction results evidence this. Fig. 7 at the top left is the predicted value with the actual data visualized in the form of a graph. From the graph, it can be seen that the addition of 80% dramatically affects the accuracy of the prediction results because the training data for the model is complete so that the optimal model will be obtained from the tested data pattern. This thinking is supported by the deviations that previously occurred in 2017 to 2018 are now very minimal. The difference is still visible at the beginning until 2016. The model from this third experiment is getting better with an accuracy above 80%, which means that the model with the Random Forest method is suitable for actual data in predicting the selling price of apartments so that the decisions taken are more accurate.



Fig. 7. Approach Model Random Forest, Multiple Regression, and Backpropagation Ratio 80:20

The Multiple Regression method still produces unsatisfactory predictions even though additional training data has been added. However, from adding data, there is a slight increase in the accuracy of calculation errors, with MAE being 0.048 and MSE being 0.680. This difference occurs due to 10% of the training data, thus supporting the model's ability to learn the previous data pattern. These different values will be visualized as a graph in Fig. 7 at the top right to see the overall trend of the data.

Fig. 7 at the top right shows the trend of the Multiple Regression prediction with the actual selling price. The actual data that has been visualized shows that the predicted data has a reasonably significant deviation from the first quarter of 2012 to the first quarter of 2016. This deviation occurs again in the first year of the data being tested, as in the errors of the first and second experiments. This result occurs due to the failure of the model to recognize the previous data pattern at the beginning of the year, the lack of data studied at the beginning of the year, and increasing every year makes the model make an improved model pattern. Thus, the study carried out using the Multiple Regression method is very poor with a small amount of data, so it is not suitable for predicting the apartment price index.

The Backpropagation method got better results in the third experiment than the Random Forest and Multiple Regression methods. The prediction results visualized in Fig. 7 at the bottom center show that the prediction results of the Backpropagation method have minimal deviation. This deviation only occurs in the first quarter of 2018 to the first quarter of 2019. This result is fantastic because using 80% training data combined with 1000 input layers, 400 hidden layers, and 20,000 iterations can produce accurate prediction results. The addition of data to this method dramatically affects the results obtained. The Backpropagation method is very feasible from the third experiment in determining decisions in predicting the apartment price index. So that there will be very minimal deviations in results, which helps minimize the impact of losses during the current Covid-19 pandemic.

Table 4 is the result of data accuracy from the third test. The results obtained indicate that the Backpropagation method with a data ratio of 80:20 produces fantastic accuracy, with a minimum accuracy value close to zero and an R2 value of 0.996 or 99.6%. This happens because the error learning performed by Backpropagation is more with the addition of data and is carried out 20,000 times with several hidden layers. Although it produces very accurate results, this method has a reasonably long execution time compared to the other two methods. The Random Forest method also experienced an increase in accuracy of 97.7%, while in

the Multiple Regression Method, there was no significant change. Changes only occur in MAE and MSE, which have increased by 0.001 and have a fixed accuracy of 55.9%. This shows that the Multiple Regression method is unsuitable for predicting small amounts of data because the Multiple Regression method requires extensive training data to predict the data. The selection of strongly correlated parameters dramatically affects the results. In the third experiment, it can be concluded that the use of more training data dramatically affects the quality of the parameters in determining the prediction results so that it will produce perfect accuracy.

Table 4. Testing Accuracy Ratio 80:20

	MAE	MSE	RMSE	R2
Random Forest	0.013	0.035	0.0187	0.977
Multiple Regression	0.048	0.680	0.082	0.559
Backpropagation	0.003	0.006	0.008	0.996

3.4. Discussion

This study proves that the Backpropagation method is superior in forecasting data accuracy with a ratio of 80:20. In contrast, the Random Forest method is superior in terms of time. The Multiple Regression method is not recommended for forecasting small amounts of data.

In this study, the Random Forest method uses a random selection of 100 trees and `max_features = "log2"` allowing the evaluation to produce the best gain ratio. This model produces an accuracy of 0.977 or 97.7%. This value is excellent as the accuracy measurement carried out by previous studies, one of which is stated in research [23] which has an R2 score on data testing of 83.63%.

In this study, the Multiple Regression methods use the column of selling price as the dependent variable, rental price as the independent variable, and rental inflation as the equation coefficient. The choice of this parameter is based on the correlation between variables which shows that the rental variable has a correlation weight of 0.746 and the rental inflation variable is 0.042. This model produces an accuracy of 0.559 or 55.9%, where this result is still below the expected accuracy of 80%. This happens because the dataset used in this study is minimal, so that the Multiple Regression method is inferior in studying previous data patterns to predict existing data. One of the supporting studies is research [24] which states that an essential step in predicting Multiple Regression is to consider the pattern found in the observed data. Research [12] Multiple Regression method has 84.5% accuracy in predicting price properties with several parameters.

The Backpropagation method in this study uses 1000 input layers and 4000 hidden layers by doing as many as 20000 iterations. This model produces an accuracy of 0.996 or 99.6%. This happens because the error learning process will be more for excellent accuracies, such as the accuracy measurements carried out by studies [25] which have an accuracy of 92% on data testing. The comparison of the results obtained in this study with the articles referred to will be visualized in Table 5. The results of this study are expected to be a reference for further research to obtain maximum research results.

Table 5. Comparison of Accuracy Results

	Current Study	Previous Study
Random Forest	97.7%	83.63%
Multiple Regression	55.9%	84.5%
Backpropagation	99.6%	92%

4. CONCLUSION

This research was conducted based on the decline in people's purchasing power during the Covid-19 pandemic. They were still careful in using money in investing, especially in buying and selling apartments. In deciding to predict the apartment price index, a machine learning approach is used using the Random Forest, Multiple Regression, and Backpropagation methods to get the best method to minimize the impact of losses. The data used in the study will be divided into three, namely the ratio of 60:40, 70:30, and the ratio of 80:20. A good ratio is generated by dividing the ratio 80:20 for each method.

The best results of the three machine learning approaches were obtained in the Backpropagation method with input parameters of 1000 layers, 4000 hidden layers. They carried out as many as 20000 iterations with parameter `loss = "mean_squared_errors."` The resulting R2 value is close to one, namely 0.996, and the error accuracy value is 0.003, MSE 0.006, and RMSE 0.008. The Random Forest method with parameters `n_estimators=100` and `max_features="log2"` resulted in superior prediction accuracy in the second experiment

with a ratio of 70:30 compared to the other two methods with an R2 value 0.976 and an MAE accuracy value of 0.014, MSE was 0.037. RMSE was 0.0193, but in the third experiment, the Random Forest method lost to the Backpropagation method even though the accuracy value increased with the R2 value of 0.977 and the MAE accuracy value of 0.013, MSE of 0.035, and RMSE of 0.0187. While the Multiple Regression method, by making selling price as the dependent variable, rental price as the independent variable, and rental inflation as the equation coefficient, the value of R2 is less than 0.80, which is 0.559, and the accuracy value for MAE is 0.048, MSE is 0.680, and RMSE is 0.082. So the conclusion obtained in predicting the apartment selling price index with a small amount of data, namely 29 data, the Backpropagation method, and the Random Forest method, are more suitable than the Multiple Regression method. However, it is superior to using the Backpropagation method.

Suggestions for further research are using the Random Forest method, Multiple Regression method, and Backpropagation method in predicting data. It is better to use a more significant number of datasets and divide the data ratio, which is not much different from the total data. Optimizing the algorithm to increase the correlation value between data parameters can also result in much better accuracy.

REFERENCES

- [1] A. Suryahadi, R. Al Izzati, and D. Suryadarma, "Estimating the Impact of Covid-19 on Poverty in Indonesia*," *Bulletin of Indonesian Economic Studies*, vol. 56, no. 2, pp. 175–192, 2020. <https://doi.org/10.1080/00074918.2020.1779390>
- [2] H. Q. Sari and A. Rahman, "Analisis Pengaruh Pandemi Covid 19 Terhadap Emiten Properti (Studi Kasus Emiten Properti Dalam LQ-45)," *Jurnal Ekonomi, Manajemen, Bisnis, dan Sosial (EMBISS)*, vol. 1, no. 3, pp. 250–254, 2021. <https://www.embiss.com/index.php/embiss/article/view/34>
- [3] A. H. Limbong, "Analisis faktor-faktor yang mempengaruhi buying attitude dan buying intention dalam pembelian apartemen" *Master Thesis, Universitas Pelita Harapan*, 2017. <http://repository.uph.edu/3790/>
- [4] I. L. Mulyahati, "Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python (Studi Kasus: Apartemen di DKI Jakarta Pada Website mamikos. com)," *Undergraduate Thesis, Universitas Islam Indonesia*, 2020. <https://dspace.uui.ac.id/handle/123456789/23970>
- [5] M. Čeh, M. Kilibarda, A. Liseć, and B. Bajat, "Estimating the Performance of Random Forest versus Multiple Regression for Predicting Prices of the Apartments," *ISPRS International Journal of Geo-Information*, vol. 7, no. 5, p. 168, May 2018. <https://doi.org/10.3390/ijgi7050168>
- [6] H. Tyrallis and G. Papacharalampous, "Variable Selection in Time Series Forecasting Using Random Forests," *Algorithms*, vol. 10, no. 4, p. 114, Oct. 2017. <https://doi.org/10.3390/a10040114>
- [7] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Frontiers in Public Health*, vol. 8, Jul. 2020. <https://doi.org/10.3389/fpubh.2020.00357>
- [8] P. J. Moore, T. J. Lyons, and J. Gallacher, "Random forest prediction of Alzheimer's disease using pairwise selection from time series data," *PLoS ONE*, vol. 14, no. 2, Feb. 2019. <https://doi.org/10.1371/journal.pone.0211558>
- [9] Chen, Chuancan, Lulu Hao, and Cong Xu. "Comparative analysis of used car price evaluation models." *AIP Conference Proceedings*. vol. 1839, no. 1, AIP Publishing LLC, 2017. <https://doi.org/10.1063/1.4982530>
- [10] R. W. Abidatul Izah, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier," *KINETIK*, vol. 2, no.3, pp. 141-150, 2017. <https://doi.org/10.22219/kinetik.v2i3.268>
- [11] Y. Feng and S. Wang, "A forecast for bicycle rental demand based on random forests and multiple linear regression," 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017, pp. 101-105. <https://doi.org/10.1109/ICIS.2017.7959977>
- [12] A. Kaushal and A. Shankar, "House Price Prediction Using Multiple Linear Regression," *SSRN Electronic Journal*, 2021. <https://doi.org/10.2139/ssrn.3833734>
- [13] D. Huang and Z. Wu, "Forecasting outpatient visits using empirical mode decomposition coupled with backpropagation artificial neural networks optimized by particle swarm optimization," *PLoS ONE*, vol. 12, no. 2, Feb. 2017. <https://doi.org/10.1371/journal.pone.0172539>
- [14] A. Wanto and A. P. Windarto, "Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 2, no. 2, pp. 37-43, 2017. <https://www.jurnal.polgan.ac.id/index.php/sinkron/article/view/76>
- [15] A. Perdana Windarto Program Studi Sistem Informasi and S. A. Tunas Bangsa Pematangsiantar Jln Jenderal Sudirman Blok No, "Implementasi JST Dalam Menentukan Kelayakan Nasabah Pinjaman Kur Pada Bank Mandiri Mikro Serbelawan Dengan Metode Backpropagation," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 1, no. 1, 2017. <https://doi.org/10.30645/j-sakti.v1i1.25>
- [16] S. P. Siregar, A. Wanto, S. Tunas, and B. Pematangsiantar, "Analysis Accuracy of Artificial Neural Network Using Backpropagation Algorithm in Predicting Process (Forecasting)," *International Journal of Information System & Technology*, vol. 1, no. 1, pp. 34–42, 2017. <https://doi.org/10.30645/ijistech.v1i1.4>
- [17] Hong, Jengei, Heeyoul Choi, and Woo-sung Kim. "A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea." *International Journal of Strategic Property Management* 24.3 (2020): 140-152. <https://doi.org/10.3846/ijspm.2020.11544>

- [18] V. W. Siburian and I. E. Mulyana “Prediksi Harga Ponsel Menggunakan Metode Random Forest,” *Prosiding Annual Research Seminar 2018*, vol. 4, no.1, 2018. <http://seminar.ilkom.unsri.ac.id/index.php/ars/article/view/1992>
- [19] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, “Development of stock market trend prediction system using multiple regression,” *Computational and Mathematical Organization Theory*, vol. 25, pp. 271–301, 2019. <https://doi.org/10.1007/s10588-019-09292-7>
- [20] A. Guardiola Mouhaffel, C. Martínez Domínguez, B. Arcones, F. Morán Redonda, and R. Díaz Martín, “Using Multiple Regression Analysis Lineal to Predict Occupation Market Work in Occupational Hazard Prevention Services,” *International Journal of Applied Engineering Research*, vol. 12, no. 3, pp. 283-288, 2017. http://www.ripublication.com/ijaer17/ijaerv12v3_02.pdf
- [21] K. T. N. Lestari, M. A. Albar, and R. Afwani, “Penerapan Metode Backpropagation Dalam Memprediksi Jumlah Kunjungan Wisatawan Ke Provinsi Nusa Tenggara Barat (NTB),” *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 3, no. 1, 2019.” <https://doi.org/10.29303/jcosine.v3i1.236>
- [22] R. R. Waliyansyah and N. D. Saputro, “Forecasting New Student Candidates Using the Random Forest Method,” *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, 2020. <https://doi.org/10.24843/LKJITI.2020.v11.i01.p05>
- [23] N. Pal, P. Arora, P. Kohli, D. Sundararaman, and S. S. Palakurthy, “How much is my car worth? A methodology for predicting used cars’ prices using random forest,” *Future of Information and Communication Conference*, Springer, Cham, 2018. https://doi.org/10.1007/978-3-030-03402-3_28
- [24] L. Nilawati and Y. E. Achyani, “Optimasi Metode Particle Swarm Optimization (PSO) Pada Prediksi Penilaian Apartemen,” *Paradigma - Jurnal Komputer dan Informatika*, 2019. Vol. 21, no. 2, pp. 227-234, 2019. <https://doi.org/10.31294/p.v21i2.6159>
- [25] A. Mulyani, “Analisis Neural Network Struktur Backpropagation Sebagai Metode Peramalan Pada Perhitungan Tingkat Kemiskinan Di Indonesia,” *Jurnal Techno Nusa Mandiri*, vol. 13, no. 1, pp. 9-14, Mar. 2016. <http://ejournal.nusamandiri.ac.id/index.php/techno/article/view/212>

BIOGRAPHY OF AUTHORS



I NYM Yoga Saputra is currently pursuing an undergraduate program at Telkom University, Bandung. He majored in Informatics. His research interests include machine learning, prediction, and Intelligent Systems. Email: nyomanyoga@student.telkomuniversity.ac.id



Siti Saadah received the Bachelor and Master degree in Informatics Engineering from Telkom Institute of Technology (now Telkom University), Bandung, Indonesia, in 2009 and 2012. Since 2009, she joined Telkom University as a lecturer in the School of Computing. She is Teaching Design and Analysis algorithms, Artificial Intelligence, Theory Automata at Telkom University. Her research interests include machine learning, financial computing, AI healthcare, prediction, and simulation. Scopus ID: 55523371300, Researcher ID: AAD-6187-2021, Publons ID: 4215578. Email: sitisaadah@telkomuniversity.ac.id



Prasti Eko Yunanto received the B.Sc. on Informatics Engineering from Telkom Institute of Technology (now Telkom University), Bandung, Indonesia in 2012, the M.Sc. on Computing from Telkom University, Bandung, in 2015. Since 2019, he has joined Telkom University as a lecturer in the School of Computing. His research interests include Biometrics Security and Intelligent systems. Scopus ID: 57193832286, Researcher ID: AAD-6421-2021, Publons ID: 4216090, Orcid ID: <https://orcid.org/0000-0003-1967-9749>. Email: gppras@telkomuniversity.ac.id