

# Dividend Payout Forecast : Multiple Linear Regression vs Genetic Algorithm-Neural Network

**Lidya Chitra Laoh**

Jurusan Sistem Informasi , Fakultas Ilmu Komputer, Universitas Klabat,  
Jl. A. Mononutu, Airmadidi - Manado  
e-mail: [lidya.laoh@unklab.ac.id](mailto:lidya.laoh@unklab.ac.id)

## *Abstract*

*This research aims to compare two methods of forecasting, i.e Multiple Linear Regression (MLR) and Genetic Algorithm-Neural Network (GA-NN), in forecasting dividend payout of Indonesian manufacturing company listed on Indonesia Stock Exchange from 2010-2014. Having collected 1384 firm-year observations, the result shows that these two methods could be used to predict dividend payout by considering earnings, free cash flow, growth opportunity, leverage, liquidity and size. This resesarch finds that even though both methods are powerful in prediction, yet in this case, MLR outperforms GA-NN.*

**Keywords** – *Forecasting, Genetic Algorithm-Neural Network (GA-NN), Multiple LinearRegression (MLR), Dividend Payout Policy , Indonesian Manufacturing Companies*

## 1. INTRODUCTION

A dividend is a form of return over investment distributed by a company towards its stockholder. A dividend can be seen as a return generated by investor over the investment made on the stock of a particular company . [1] argue that a dividend conveys an important information used by investors to differentiate the profitable firm from the maximizing-wealth firm. The amount distributed to stockholders carries information about the expectation of the manager about the future profitability and investment direction. Amount to be distributed in form of dividend is determined by company's dividend policy. This term is used in relation to the profit of a company. As a dividend maximizes the wealth of the investor, then forecast information on dividend is considered important for investor in decision making. Eventhough dividend forecasting is essential in investments, but according to [2] it is difficult to find good theoretical models in forecasting dividend.

In forecasting a dividend, the financial accounting data is used. Multiple Linear Regression (MLR) is one of the commonly used statistical methods to process accounting data. MLR is used to see the effect of the multiple independent variables over the dependent variable. The effect of each independent variable is used to forecast the direction as well as the magnitude of the dividend payout.

Another useful method that can be used to forecast a dividend payout is computational intelligence method. One of the powerful computational intelligence methods that have the ability to do forecasting is the Artificial Neural Network (ANN). This method has been successfully used in business application [2][3]. ANN is a network that consists of interconnected nodes inspired by the brain that generates output through the training process. In the training process, ANN will learn the pattern of data given to the network by comparing the actual output produced by the network with the target or desired output. The result of the comparison is known as the error.

Weights of the network will be iteratively modified until it reaches the minimum error. When the network reaches the minimum error, training will stop. After the training and testing, the network is ready to be used as a machine that produces forecasting value.

Several studies show that ANN works well when used together with Genetics Algorithm (GA) ([4],[5],[6]). These studies shows that both of these methods can be combined to solve forecasting problems in many applications and is proven to improve the forecasting accuracy compared with only using GA or ANN alone. Another study shows that the combination of GA and ANN performs better than Support Vector Machine (SVM) and Vector Autoregression (VAR) [7].

GA can be used in different ways. Several recent studies use GA as an optimization algorithm to produce optimal weights for the network. This weights is then fed to the network for training using backpropagation. The results show that this method helps increase the accuracy of the forecast. ([8], [9]) .

Both MLR and Genetics Algorithm-Neural Network (GA-NN) can be used to forecast dividend. Several studies had been conducted in regard to comparing multiple MLR with ANN in financial forecast such as stock or bankruptcy predictions, but not on dividend payout. Therefore, the comparison of the predictive power of MLR and GA-NN is an empirical issue. This study will compare MLR with ANN in forecasting dividend payout and GA will be used to train ANN to produce optimal weights that will then be fed to the network for training using backpropagation. MLR and GA-NN will be compared in terms of their Mean Square Error (MSE).

## 2. METHODOLOGY

### 2.1 Data Collection

In doing the comparison, the purposive sampling method is used. Criteria to be fulfilled for the chosen sample are:

- Listed on Indonesia Stock Exchange from 2000-2013
- Paying dividend at least once during the observation period
- Providing all data needed as variables of research

Manufacturing companies are the object of this research for several reasons. Firstly, it is the largest sector in Indonesia Stock Exchange compared to other sectors. Secondly, it provides enough data to run multiple linear regressions. Choosing other sectors that have fewer companies might be at risk as some companies in that sectors might not provide all the data required for this research. Thirdly, the firms in the similar industry share many things in common. This make the results more generalizable and applicable for the investors.

Data of all variables are taken from *Datastream*. Data used for this research are dividend per share as the proxy for dividend policy, earnings per share (the proxy for earnings), free cash flow per share (free cash flow), market capitalization (firm size), change in sales (growth opportunity), quick ratio (liquidity) and debt to asset ratio (leverage).

### 2.2 Forecasting with MLR

The developed model used to predict dividend payout has to go through several test i.e multicollinearity test, heteroscedasticity test, and autocorrelation test. Figure 1 describes the steps to perform these tests. Having detected the existence of multicollinearity, heteroscedasticity or autocorrelation, the model has to be treated until the problem is omitted. By addressing all the existence of the aforementioned problems, the model is expected to generate a best, linear, and unbiased estimator.

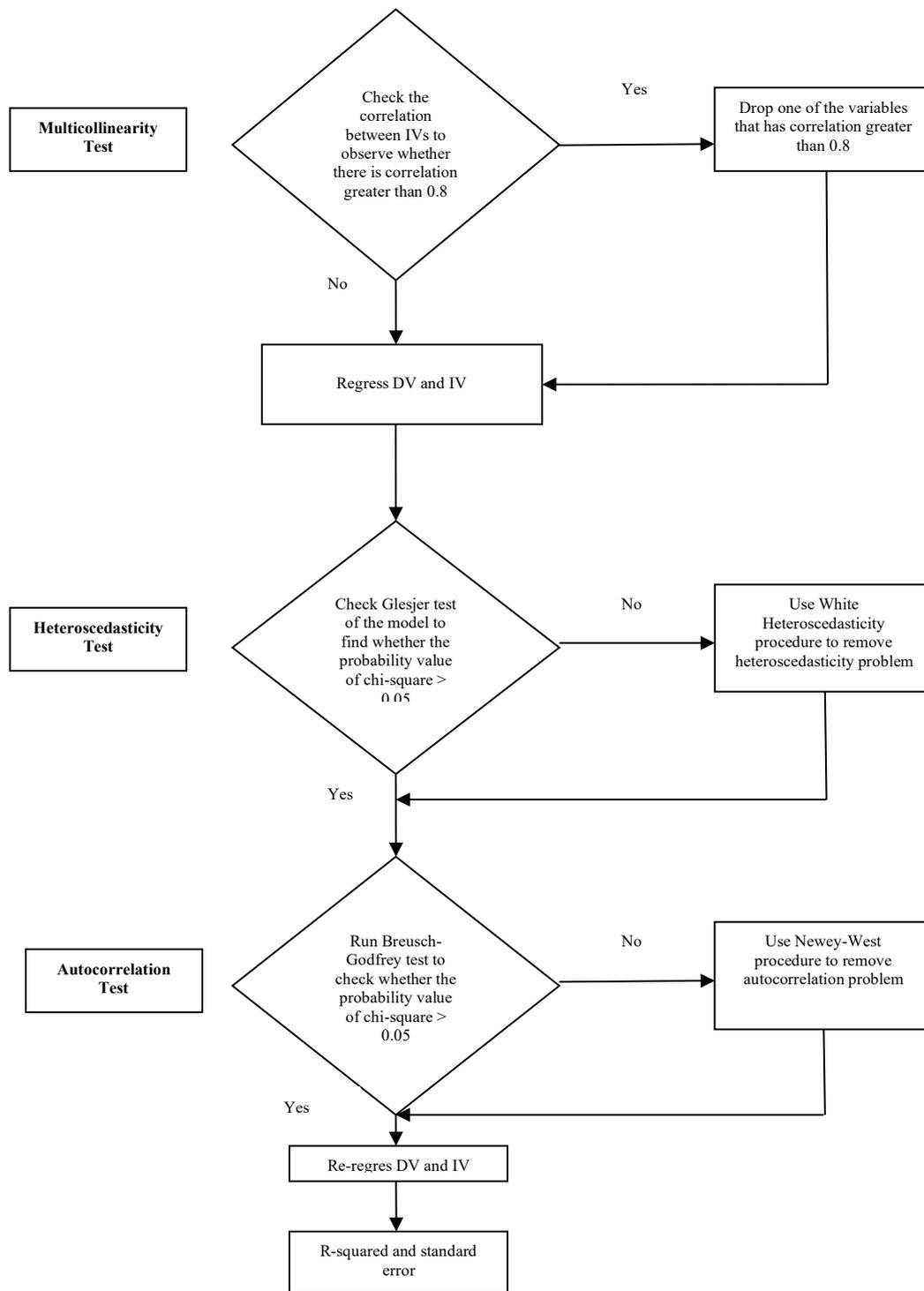


Figure 1 The Multicollinearity, Heteroscedasticity, and Autocorrelation Test

### 2.2.1 *Multicollinearity Test*

Multicollinearity test is conducted to identify whether there is an exact collinearity among independent variables. The existence of multicollinearity can be observed through correlation between independent variables as well as coefficient significance of each independent variables. If correlation between variables is greater than 0.8 then multicollinearity holds. High R-squared yet statistically insignificant coefficient is another indicator of this problem. Therefore, dropping one of the independent variables or transforming the variables could be done as treatment. [10] identifies some the consequences of multicollinearity. Those are precise estimation of estimator becomes difficult, readily acceptance of “zero null hypothesis”, tendency of coefficient to be statistically insignificant yet the goodness of fit could be very high. Therefore, conducting multicollinearity test is indispensable.

### 2.2.2 *Heteroscedasticity Test*

Heteroscedasticity test is conducted to identify whether the population has the same variance or homoscedastic. Heteroscedasticity refers to the variability of variances. This problem is possible when regressing cross-sectional data since each company has its own uniqueness or characteristics. To test the existence of heteroscedasticity, Glejser test will be used. Having run the model, Glejser test will be run to find out if the probability chi-square is greater or less than the level of confidence (0.05). If probability chi-square is less than 0.05, then heteroscedasticity exist. If the problem exists, then the t-test of F-test might give us an inaccurate result. What is statistically insignificant might be significant. Therefore, should this problem exists, then White Heteroscedasticity procedure will be employed to eliminate this problem.

### 2.2.3 *Autocorrelation Test*

Autocorrelation test is used to identify serial correlation in the observation. The existence of autocorrelation will cause the estimators to be inefficient. Breusch-Godfrey test is to be exercised to check the existence of autocorrelation problem in the developed model. If the probability chi-square of Breusch-Godfrey test is less than the level of confidence (0.05), autocorrelation holds. To eliminate its existence, Newey-West will be used.

Having conducted all the aforementioned tests, the model will be run to test the significance of the model as well as the independent variables. If the F-test is less than 0.05, then the model has predictive ability. Finally, the adjusted R-squared and standard error of this model will be compared to the hybrid method.

Dependent variable of this model is dividend payout (represented by DivPay in the model), while earnings (EARN), free cash flow (FCF), firm size (SIZE), growth opportunity (GrOp), liquidity (LIQ) and leverage (LEV) are independent variables. Each variable would be proxied by data taken from *Datastream* as explained in figure 1. The model to statistically test the significance of independent variables is as follow:

$$DivPay_t = \alpha_1 + \alpha_2 EARN_t + \alpha_3 FCF_t + \alpha_4 SIZE_t + \alpha_4 GrOp_t + \alpha_4 LIQ_t + \alpha_5 LEV_t + e_t \quad (1)$$

To statistically test the model, EVIEWS 6 software is used. Afterward, the automatically-generated adjusted R-squared and standard error will be compared to results generated by hybrid models.

## 2.3 *Forecasting with GA-NN*

This research uses ANN multilayer neural network combined with GA. GA will be used to train the ANN in order to achieve minimal Mean Square Error (MSE). Figure 2 describes the steps to be performed by GA-NN.

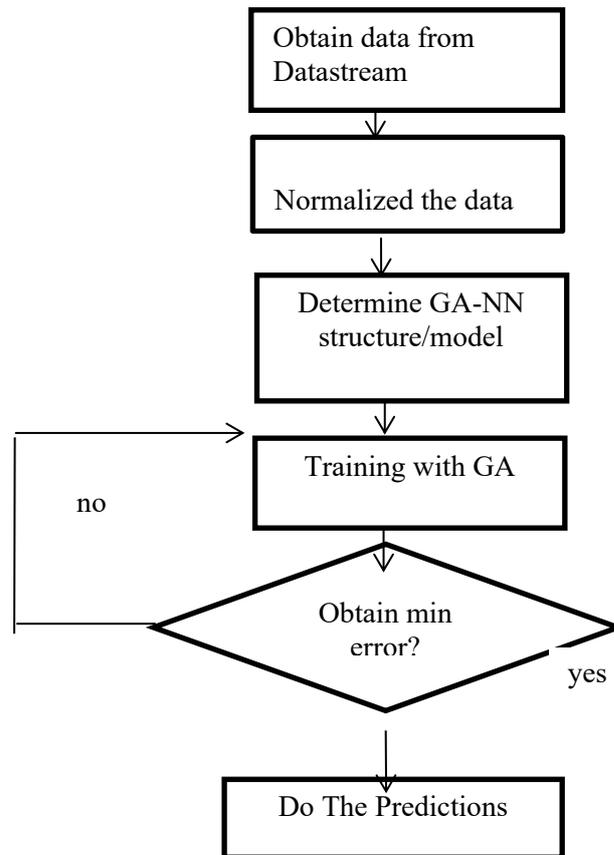


Figure 2 GA-NN steps to forecast

### 2.3.1 Data Normalization

The input and output data should be normalized before it is used by ANN. The purpose of normalization is to transform the data so it will be in accordance with the Sigmoid function format. The Sigmoid function uses values between -1 and 1. The Sigmoid function is used as the activation function and is used to determine the output of each neuron.

The max-min method with the following formula will be used to normalized the data :

$$I = I_{min} + [(I_{max} - I_{min})x (D - D_{min})] / (D_{max} - D_{min}) \quad (2)$$

where :

$I_{min}$  = the smallest number in the list

$I_{max}$  = the largest number in the list

$D$  = value to be normalized

$D_{min}$  = lower range

$D_{max}$  = upper range

### 2.3.2 The GA-NN Structure/Model

The ANN structure describes the framework that shows the connections and arrangements of neurons. These arrangements will be the parameters for the ANN. Moreover, the number of layers and neurons on each layer are depicted in the ANN structure. The ANN structure of this research is shown in Figure 3, with the input neurons are earnings, free cash flow, firm size, growth opportunity, liquidity and leverage.

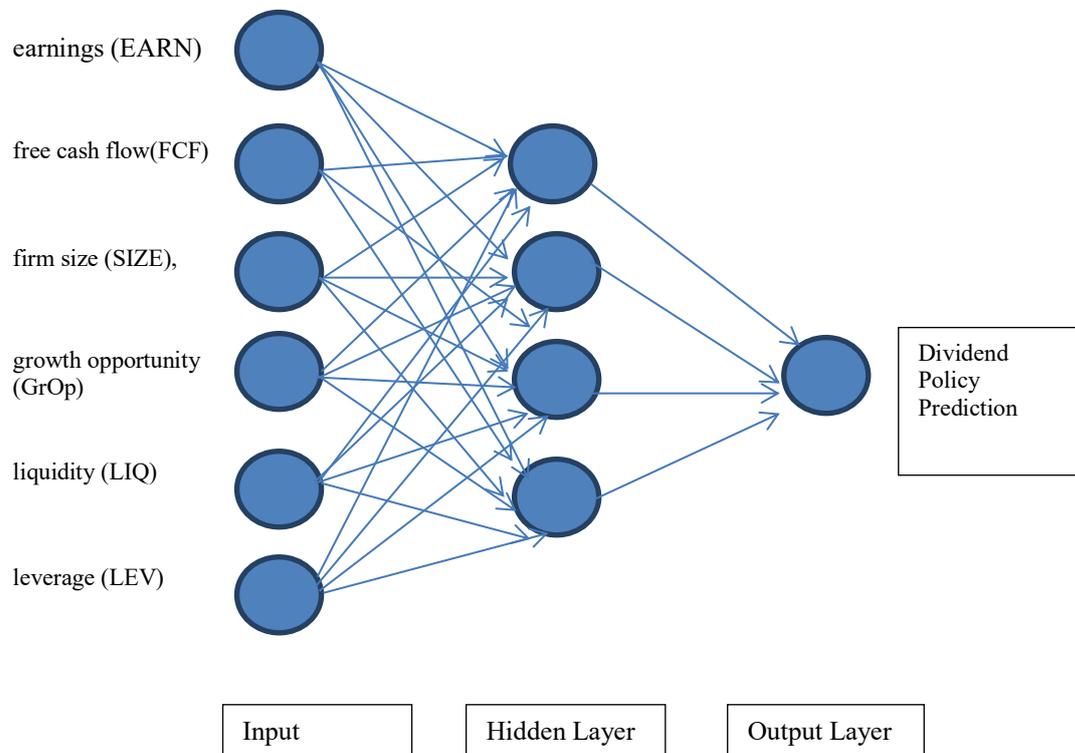


Figure 3 The Artificial Neural Network Structure

### 2.3.3 Training with GA

The training process is essential in ANN. In this process, ANN will learn from the input-output pattern given to the network. Weights in each layer will be updated until the given input produced the desired output. On this research, GA is used as the training algorithm. Training ANN with GA was first done by [11]. The result of this study shows that GA outperformed Backpropagation algorithm in training ANN. Several studies had been done in order to deal with some issues in using GA as a training method for ANN. These issues are related to the representation schemes of the connection weights and chromosomes representation of the connectivity. Studies show that real numbers might be used to represent connection weights and a chromosome might represent the connectivity, weights, and biases in the form of a matrix ([12],[13], [14]). Therefore, this study will use the same approach in representing chromosomes to be used by GA. GA parameters include selection options, mutation options, and crossover options. The fitness function for GA is the MSE, where the goal is to find the optimal weights that generate the least MSE. Figure 4 describes the process of training using GA.

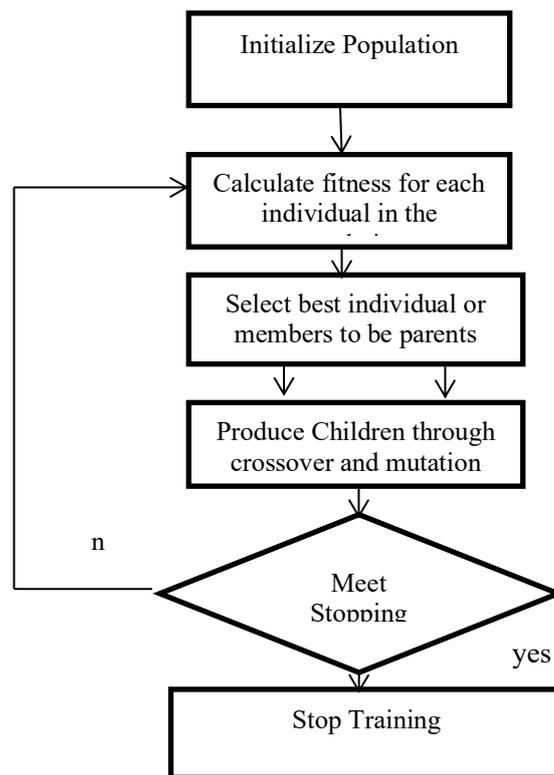


Figure 4 Training with GA

The following description describes the training process as shown in Figure 4 :

- 1) *Initialize Population* : The training with GA starts with initializing the population randomly. The members of the population in this study are the weights and biases for the network and are encoded using weight matrices.
- 2) *Calculate Fitness* : After defining the population, each member that represents the weights and bias will be scored according to their fitness, based on the value returned by the fitness function. The fitness function is the mean squared error (MSE), in which the lower the value of the MSE the larger the fitness value
- 3) *Select Best Members* : GA will select weights and biases that have better fitness to be the parents that will produce members for the next population. The selection algorithm used in this study is the stochastic uniform that uses a line where each parent are directly related to a part of the line proportional with its value. The algorithm will move along the line and will select parents based on the section it lands on.
- 4) *Produce Children* : The selected parents will be used to create children in which the process of forming children are done either through crossover or mutation. Crossover is the process of combining 2 parents from the current population to produce children, while mutation is the process of randomly changing the genes of the members. The crossover function used in this study is a single point crossover. This function will choose a random number between 1 and the number of variables that represents the points in which the crossover will occur. The mutation operator used in this study is mutation Gaussian, where this algorithm works by adding a random number to each parent. The amount of mutation will usually decrease at new generation. After forming the children, GA will replace the members of the current population with the children that it produces. GA will stop iterating the process to modify the population of the solution when the stopping criterion is met.

### 3. RESULTS & DISCUSSION

#### 3.1 MLR Results

EViews software is used to statistically test the model. Table 1- 4 shows the MLR results for the correlation test, heterokedasticity test , autocorrelation test, and the final result.

Table 1 Correlation Result

Correlation t-Statistic	EARN	FCF	SIZE	GrOp	LIQ	LEV
EARN	1.000000					
<i>p-value</i>	-----					
FCF	0.319649	1.000000				
<i>p-value</i>	12.55457	-----				
SIZE	-0.007506	-0.012548	1.000000			
<i>p-value</i>	-0.279356	-0.467009	-----			
GrOp	-0.004836	-0.008754	0.005562	1.000000		
<i>p-value</i>	-0.179974	-0.325784	0.206989	-----		
LIQ	0.052446	0.035427	-0.109790	-0.008739	1.000000	
<i>p-value</i>	1.954487	1.319266	-4.110755	-0.325245	-----	
LEV	-0.010695	-0.029186	0.022045	-0.003194	-0.017203	1.000000
<i>p-value</i>	-0.398062	-1.086643	0.820634	-0.118876	-0.640324	-----

The result in Table 1 shows that there is no variable has greater (lesser) than 0.8 (-0.8) correlation with other variable. Moreover, the correlation between each independent variable is insignificant. Since there is no highly correlated variable exists, then multicollinearity problem does not hold. Therefore, no variables should be dropped from the proposed model.

Table 2 Heterokedasticity

Glesjer Test

F-statistic	309.4264
Obs*R-squared	
Prob. F(6,1377)	0.0000
Prob. Chi-Square(6)	0.0000

Test Equation:

Dependent Variable: ARESID

Method: Least Squares

Sample: 1 1384

Included observations: 1384

Variable	Coefficien t	Prob.
C	0.010603	0.0800
EARN	1.67E-05	0.0000
FCF	-7.37E-06	0.0000
GROP	-4.80E-05	0.6718
LEV	8.09E-08	0.7819

LIQ	-1.98E-05	0.8300
SIZE	-0.000288	0.3214
<hr/>		
R-squared	0.574153	
Adjusted R-squared	0.572298	
S.E. of regression	0.016596	
Sum squared resid	0.379264	
F-statistic	309.4264	
Prob(F-statistic)	0.000000	

Table 2 shows that there are two independent variables (EARN and FCF) whose p-value is less than 0.05. It signifies that as the independent variable size increase, the variance of the dependent variable will also increase significantly. Therefore, Table 2 shows that the p-value of Obs\*R-squared of Glesjer test is 0.0000 or less than 0.05. In other words, the data used in this research contains heteroskedasticity problem. If the existence of heteroskedasticity is not treated, then it would affect the significance of statistical test. Therefore, White Heteroskedasticity is used to address this issue. The result of removing heteroskedasticity is presented in Table 4.

Table 3. Autocorrelation test  
Breusch-Godfrey Serial Correlation LM Test

F-statistic	85.97755	
Obs*R-squared	153.8415	
Prob. F(2,1375)	0.0000	
Prob. Chi-Square(2)	0.0000	
<hr/>		
Test Equation:		
Dependent Variable: RESID		
Method: Least Squares		
Sample: 1 1384		
Included observations: 1384		
<hr/>		
Variable	Coefficient	Prob.
EARN	-1.35E-06	0.0266
FCF	1.51E-06	0.1160
GROP	-1.37E-05	0.9376
LEV	3.22E-08	0.9430
LIQ	-1.38E-06	0.9922
SIZE	0.000103	0.8178
C	-0.001750	0.8513
RESID(-1)	0.331237	0.0000
RESID(-2)	0.022292	0.4112
<hr/>		
R-squared	0.111157	
Adjusted R-squared	0.105986	
S.E. of regression	0.025580	
Sum squared resid	0.899687	
F-statistic	21.49439	
Prob(F-statistic)	0.000000	

The result of Breusch-Godfrey test in Table 3 shows that autocorrelation exists as one of the independent variables (EARN) has significant effect to the residual (p-value is 0.0266 or less than 0.05). Moreover, the Obs\*R-squared Prob.Chi-Square is 0.0000 or less than 0.05. The

Newey-White test would be used to address this issue. The result of treating the autocorrelation issue is presented in Table 4.

Having seen the existence of all problem, the treatment have been conducted. The following result shown in Table 4 is the final result after the treatment of the heteroskedasticity and the autocorrelation issues by including White Heteroskedasticity and Newey-West to run the multiple linear regression.

The result of Breusch-Godfrey test in Table 3 shows that autocorrelation exists as one of the independent variables (EARN) has significant effect to the residual (p-value is 0.0266 or less than 0.05). Moreover, the Obs\*R-squared Prob.Chi-Square is 0.0000 or less than 0.05. The Newey-White test would be used to address this issue. The result of treating the autocorrelation issue is presented in Table 4.

Having seen the existence of all problem, the treatment have been conducted. The following result shown in Tabel 4 is the final result after the treatment of the heteroskedasticity and the autocorrelation issues by including White Heteroskedasticity and Newey-West to run the multiple linear regression.

Table 4. Multiple Linear Regression

Dependent Variable: Div\_Pay

Included observations: 1384

Newey West

Variable	Coefficient	Prob.
EARN	6.29E-05	0.0000
FCF	-1.19E-05	0.0153
GROP	-7.32E-05	0.0069
LEV	-2.93E-07	0.2138
LIQ	0.000116	0.0466
SIZE	-0.000376	0.5312
C	0.003232	0.7946
R-squared	0.881753	
Adjusted R-squared	0.881238	
S.E. of regression	0.027112	
Sum squared resid	1.012201	
F-statistic	1711.358	
Prob(F-statistic)	0.000000	

### 3.2 Forecasting with GA-NN

Table 5 shows the GA parameter. Global optimization toolbox provided by Matlab is used to run GA. This toolbox could be used to search global solutions for problems that contains multiple minima. GA solvers in this toolbox can be customized by modifying its initial population or by delineating its selection, crossover and mutation function. The best fitness value and the variables values for that fitness will be returned by the GA solver.

Table 5 GA Parameter

GA parameters	Value
Population Type	Double vector
Population Size	60

Selection Function	Stochastic Uniform
Population Initial Range	[0..1]
Crossover Function	Scattered
Mutation Function	Gaussian

After configuring the GA-NN structure, GA is used to train the network. Data used in training is 80% of the sample data and 20% is for testing. After running GA, the minimum MSE is achieved on generation 87 with the value of 0.0001458. The optimization terminates after the average change in the fitness value is less than the error goal. The weights produced by GA is then be fed to ANN for training using backpropagation.

### 3.3 Comparison Result

MLR and GA-NN are compared in terms of how well their model do a forecast. Forecasting results of each model for the dividend payout value will be compared with actual values of dividend payout. In this study, 2 methods will be used to analyze and compare the results of both models. The 2 methods are: 1) Comparing Graph that compares the actual dividend payout with the result of the forecast of both models using sample data, 2) Comparing Mean Squared Error of both methods.

#### 1). Comparing Graph of Actual vs Forecast values

Figure 5-7 shows the comparison of the forecasting results with actual values in graph. Figure 5 shows the comparison of GA-NN forecast results with the actual values, while figure 6 shows the comparison for MLR result. In figure 7, the result of the forecast by both GA-NN and MLR are compared to the actual values. Based on this graph, GA-NN fitness to the actual values are poor compared to MLR. The quantification of these fitness is measured by MSE.

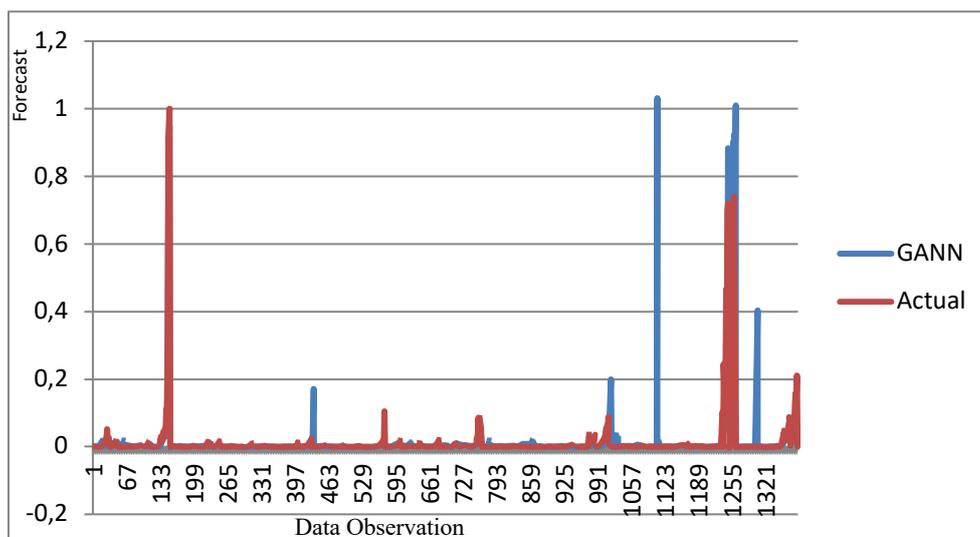


Figure 5. Comparison of GA-NN Forecast with Actual Values

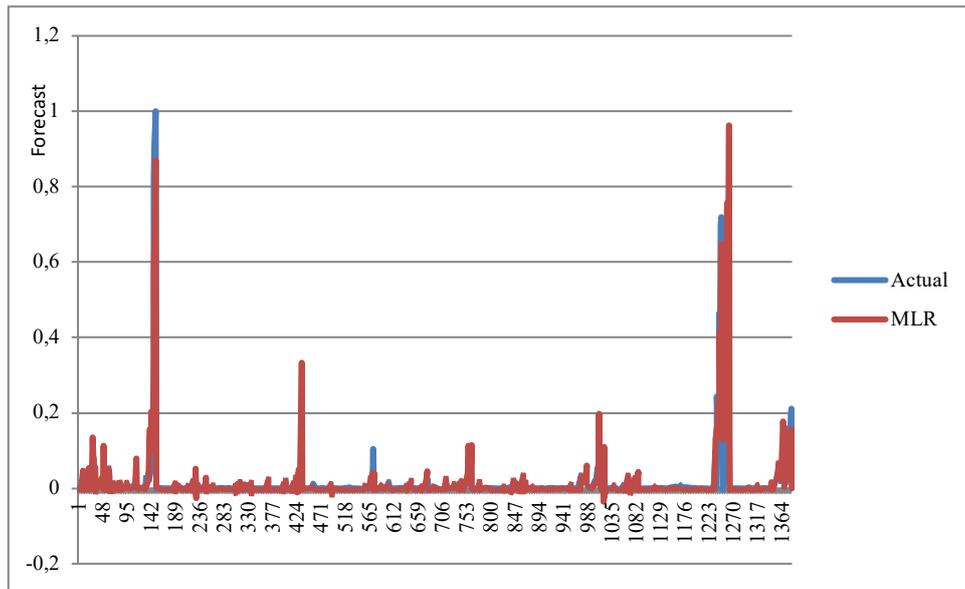


Figure 6. Comparison of MLR Forecast with Actual Values

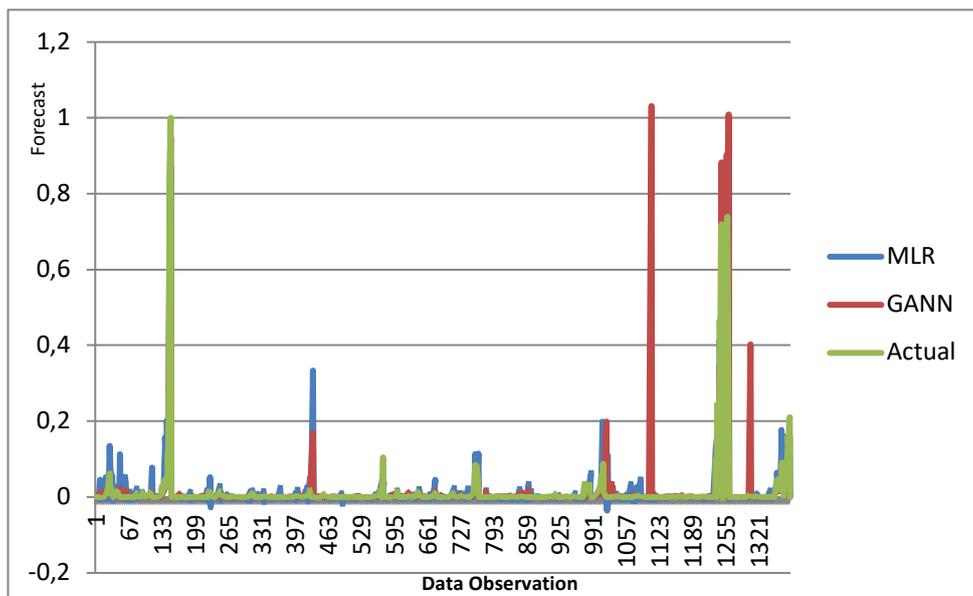


Figure 7. Comparison of GA-NN and MLR Forecast with Actual Values

2). Comparing Errors and Fit of Data

Table 7 shows the comparison in terms of MSE and  $R^2$  from both methods. This table confirms the graph shown above. It suggests that MLR model outperforms GA-NN, where MSE of MLR model is lower than GA-NN. This table suggests that MLR model performs better in terms of its errors and the fits of its data.

Table 7. Comparison of GA-NN and MLR

Methods	MSE	$R^2$
GA-NN	0.0016	0.88
MLR	0.00073	0.88

#### 4. CONCLUSION

MLR results show that EARN, FCF, GrOp, and LIQ significantly affect company dividend payout decision. The model used to assess those factors significance toward company dividend payout decision is significant (p-value of F-test is 0.0000 and model reliability is 0.88). Therefore, MLR could be used by investor to assess the possibility of generating return over investment made on particular company.

GA-NN results show that the target data has a significant relationship with the output data (this is shown through the  $R^2$  value of 0.88). Furthermore, GA can be used as a training algorithm for ANN. This is confirmed by the low MSE value achieved by GA. Using the sample data collected from *Datastream*, GA-NN model can be used to forecast the dividend payout of Indonesian manufacturing companies.

By comparing the result of MSE from both methods, we can conclude that MLR produces better model compares to GA-NN because MLR produces lower MSE value than GA-NN. Therefore, MLR forecast values are closer to the actual values. This result indicates that MLR generates better fit and forecast than GA-NN model.

#### 5. RESEARCH LIMITATION AND RECOMMENDATION

Several limitations of this research are firstly, GA optimization has some drawbacks such as it is slow and in need of algorithm parameter tuning. Secondly, the MLR has to fulfill certain assumptions, such as autocorrelation, multicollinearity, and heteroskedasticity. Thirdly, the comparison of performance of both models are only based on mse and regression.

Having considered the limitations, some considerations could be considered by future research. This considerations include using other training algorithms that runs faster. If the data have a small ratio of input to target, then according to [15], it is probably best to use resilient backpropagation algorithm. Generalized Least Square (GLS) might be used as the statistical method to do forecasting, since GLS is not restricted by assumptions used by MLR. In comparing the statistical method and computational intelligence performance in forecasting dividend payout, other evaluations might be used.

#### REFERENCES

- [1] Gitman, L. J. , Zutter, C. J., 2018, *Principles of managerial finance*, 14<sup>th</sup> edition Pearson Education South Asia Pte Ltd, Singapore .
- [2] Kim, J., Won, C., and Bae, J.K., 2010, A knowledge integration model for the prediction of corporate dividends, *Expert Systems with Applications*, vol. 37, pp. 1344-1350.

- [3] Ghoddusi, H., Creamer, G.G., and Rafizadeh, N., 2019, Machine learning in energy economics and finance : A review, *Energy Economics*, vol.81, pp.709-727
- [4] Kant, G., and Sangwan, K.S., 2015, Predictive modelling and optimization of machining parameters to minimize surface roughness using artificial neural network coupled with genetic algorithm," *Procedia Cirp*, vol. 31, pp. 453-458.
- [5] Ping, F.F., and Fei, F.X., 2013, Multivariant forecasting mode of Guangdong province port throughput with genetic algorithms and Back Propagation neural network, *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 1165-1174
- [6] Yang, F., and Yue, Z., 2014, Kernel density estimation of three-parameter Weibull distribution with neural network and genetic algorithm, *Applied Mathematics and Computation*, vol. 247, pp. 803-814.
- [7] Chiroma, H., Gital, A. Y. u., Abubakar, A., Usman, M.J., and Waziri, U., 2014, Optimization of neural network through genetic algorithm searches for the prediction of international crude oil price based on energy products prices, *Proceedings of the 11th ACM Conference on Computing Frontiers*, Cagliari, May 20.
- [8] Torregoza, M. L. R., and Dadios, E. P., 2014, Comparison of neural network and hybrid genetic algorithm-neural network in forecasting of Philippine Peso-US Dollar exchange rate, *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, Palawan, November 12.
- [9] Zhang, Y., Gao, X., and Katayama, S., 2015, Weld appearance prediction with BP neural network improved by genetic algorithm during disk laser welding, *Journal of Manufacturing Systems*, vol. 34, pp. 53-59.
- [10] Gujarati, D.N., 2009, *Basic econometrics*: Tata McGraw-Hill Education.
- [11] Whitley, L.D., and Hanson, T., 1989, Optimizing Neural Networks Using FasterMore Accurate Genetic Search, *Proceedings of the 3rd international conference on genetic algorithms*, Virginia, June.
- [12] Montana, D. J., and Davis, L., 1989, Training Feedforward Neural Networks Using Genetic Algorithms, *Ijcai*, pp. 762-767.
- [13] Harp, S. A., Samad, T., and Guha, A., 1989, The genetic synthesis of neural networks, *Proceedings of the International Conference on Genetic Algorithms*, Virginia, June.
- [14] Dasgupta, D., and McGregor, D.R., 1992, Designing application-specific neural networks using the structured genetic algorithm, [*Proceedings*] *COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks*, Maryland, June 6.
- [15] Blomström, 2012, Benchmarking an artificial neural network tuned by a genetic algorithm,<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.4785&rep=rep1&type=pdf>., accessed on December 9, 2019.