Int. J. Advance Soft Compu.Appl, Vol. 11, No. 1,March 2019
ISSN 2074-8523

Copula-Based Feedforward Neural Network Genetic Algorithm Cargo Forecasting

Toni Toharudin[1†],Rezzy Eko Caraka[1,2‡]
Achmad Bachrudin[1], Sakhinah Abu Bakar[2], Dewi Ambarsari[1]

[1]Department of Statistics, Padjadjaran University, Bandung, Indonesia
[2]School of Mathematical Sciences, Faculty of Science and Technology, The National University of Malaysia
[†]e-mail:toni.toharudin@unpad.ac.id
[‡]email: rezzyekocaraka@gmail.com

Abstract

The presence of revolution industry 4.0 made business activities increasing one of them is transaction and online shopping. Moreover, it makes the amount of export and import Cargo significantly increased. Bandung is one of the largest city in Indonesia who often visited foreign tourists and became fashion city of Indonesia and impact to the increases of cargo amount. In this paper we proposed copula and soft computing neural network with genetic algorithm also to analysis as well as forecasting of cargo amount that is affected by planes variable by reaching MAPE FFNN-AG 1.09% and Neuro-Copula 1.17% respectively.

Keywords: *Copula; Neural Network; Cargo; Genetic Algorithm*

1 Introduction

HuseinSastranegara airport is an airport which located in Bandung, West Java, Indonesia.Number of cargo shipmentsat HuseinSastranegara Airport is strongly influenced by the number of aircraft available at the airport, so the forecasting process takes into account the variable of the aircraft as the corresponding variable and influences the amount of cargo at HuseinSastranegara Airport. There is also a marginal distribution in the cargo data and the plane also does not follow the normal distribution. The copula is a method that can combine the function of marginal distribution into distributed functions that are not normally distributed and can provide characteristics of the relationships between variables[1]. In the heart of statistics, there are non-parametric techniques can gives robust accuracy in case time series as well as neural network [2], Support vector regression [3], Localized Multi kernel support vector regression [4][5]. In general, FFNN is trained using the Back propagation algorithm to get the weights. Training using Back propagation includes three stages: feed forward of the input pattern, error calculation, and weight adjustment. Back propagation works well on simple training issues but its performance will decline and be trapped in local minimums when applied to data with great complexity. The solution to the problem is to train FFNN using Genetic Algorithm (AG)[6].Combination of copula with neural network will yield information that is not obtained when compared with using single method or traditional method.

2 Methods

2.1 COPULA

Copula is a method first introduced by Sklar[7]. Copula is a tool that can combine marginal functions into shared functions and can describe variable dependency structures[8]. The copula theory serves to model a combined distribution without the need for normality assumptions[9]. Elliptical Copula is a type of copula with eliptical-shaped distribution. If there are dimensionless copula dimensions, then the parameters are at least d (d-1)/2.Archimedean copula[10] is also a group of Laplace transform[11][12]. A copula generator, assuming that the generator has only one parameter, i.e. θ. Copula Archimedean is most widely used in bivariate cases. Archimedean copula itself has several density functions in it, namely Frank, Clayton, and Gumbel.

Definition.1: Let φ: $[0,1] \rightarrow [0,\infty]$ is continuous, decreasing and convective functions such that $\varphi(1) = 0$ and $\varphi(0) = \infty$. The function of φ has inverse φ^{-1}: $[0, \infty] \rightarrow [0,1]$ which is same like φ, except $\varphi^{-1}(0) = 1$ and $\varphi^{-1}(\infty) = 0$.

Definition 2: Function of C: $[0,1]^2 \longrightarrow [0.1]$ can be defined:

$$C(u_1, u_2) = \phi^{-1}\big(\phi(u_1) + \phi^{-1}(\phi(u_2)\big) \quad (1)$$

$\phi$is derived from a random number with the following assumptions:

1. $\phi(1) = 0$;
2. ϕ continous;
3. For all$t \in (0,1), \phi^{-1}(t) < 0, \phi\ decrease\ function$;
4. For all$t \in (0,1), \phi^{-1}(t) \geq 0, \phi\ conveks\ function$;

Archimedean copula has several advantages, which can measure non-linear correlations and tail dependency and can build new distributions but only limited by one or two parameters to be able to analyze their dependency structure[10]

$$\phi_C(u) = \frac{1}{\theta_C}\big(u^{-\theta_C} - 1\big), \theta_C > 0 \quad (2)$$

Archimedean copular flexibility is given by generator functions, eg from Clayton, frank and Gumbel copula[13]

$$\phi_F(u) = \log\left(\frac{e^{\theta_F u}-1}{e^{\theta_F-1}}\right), \theta_F \neq 0 \quad (Frank) \tag{3}$$

$$\phi_G(u) = (-log^u)^{\theta_G}, \theta_G \geq 1 \tag{4}$$

For Archimedean copula in the case of bivariate equations (4) can be written as follows:

$$C(u_1, u_2) = \phi^{-1}\big(\phi(u_1) + \phi(u_2)\big) \tag{5}$$

Table 1. Archimedean Copula

Type	Bivariat Copula $C_\theta(u, v)$	Parameter
Al-Mikhail Haq	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta\epsilon[-1,1)$
Clayton	$[\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]$	$\theta\epsilon[-1,\infty)\{0\}$
Frank	$-\frac{1}{\theta}log\left[\frac{1+(\exp(-\theta u)-1)(erxp(-\theta v)-1)}{\exp(-\theta)-1}\right]$	$\theta\epsilon\mathbb{R}\{0\}$
Gumbel	$\exp[-((-\log(u)^\theta + (-\log(v))^\theta)^{-1/\theta}]$	$\theta\epsilon[1,\infty)$

Tabel 2. Copula generator

Type	Generator $\phi_\theta(t)$	GeneratorInverense$\phi_\theta(t)^{-1}$
Al-Mikhail Haq	$\log\left[\frac{1-\theta(1-t)}{t}\right]$	$\frac{1-\theta}{\exp(x)-\theta}$
Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$(1-(\theta_t)^{-1/\theta}$
Frank	$-log\left[\frac{(\exp(-\theta t)-1)}{\exp(-\theta)-1}\right]$	$\frac{1}{\theta}log(1+\exp(-t))$
Gumbel	$(-\log(t))^\theta$	$\exp(-t^{1/\theta})$

2.2 Neural Network Genetic Algorithm

In FFNN modeling for time series data[14], the input model is past data $(X_{t-1}, X_{t-2}, \ldots, X_{t-p})$ and the target is the current X_t. data. The general form of the FFNN model for time series data is written in the following equation:

$$X_t = \psi_o\{w_{bo} + \sum_{j=1}^{H} w_{jo}\,\psi_j\,(w_{bj} + \sum_{i=1}^{p} w_{ij}\,X_{t-i})\} \tag{6}$$

With:

ψ_o : the activation function used in the output layer

ψ_j : activation function used in the hidden layer

w_{ij} : the weight of the i-th neuron at the input layer leading to the j-th neuron in the hidden layer
w_{bj} : the bias weights in the input layer into the j-th neuron in the hidden layer
w_{jo} : the j-th junction weights in the hidden layer to the output layer
w_{bo} : the bias weight on the hidden layer to the output layer

Network training method is a network training process or procedure that is a sequence of integrated steps (algorithm) to modify the values of weight and bias for a network to get the appropriate weight and bias values to produce the desired network output[2]. If the error on the network output is very small, then it can be said has been obtained the amounts of weight and bias accordingly and the network has achieved excellent performance.Prior to the training of artificial neural networks, it is often necessary scaling on inputs and targets so that data entered in a certain range. It is intended that the data processed following the activation function used. This process is called Pre-Processing. Then after the training process is done, the data is returned to its original form (Post-Processing)[15]. In this paper the activation function used in the hidden layer to the output layer is the binary sigmoid (sigmoid logistics), then the data must be transformed first into the interval [0, 1]. However, it would be better if the data is transformed into smaller intervals, e.g. at intervals [0.1, 0.9]. This is given that the sigmoid function is an asymptotic function whose value never reaches 0 or 1. Pre-Processing for transforming data into intervals [0.1, 0.9] as follows:

$$x' = \frac{0.8\,(x-a)}{b-a} + 0.1 \tag{7}$$

While Post-Processing to restore data to its original form as follows:

$$x = \frac{(x' - 0.1)(b-a)}{0.8} + a \tag{8}$$

Where$a$minimum of data and b is maximum data

3Methods

In this study, we used several time series methods such as ARIMA, FFNN with genetic algorithm and copula. In Copula, we are using two types of variables namely number of cargo shipments and number of the airplane at HuseinSastranegara Airport. The analysis phase can be seen Figure 1. First, Modelling and get ARIMA against cargo data, Second, modeling cargo data using Copula (Clayton, Frank, and Gumbel) and thirdly is conducting FFNN-AG. In this model we can justify input variable by the significant lag in PACF (Partial Autocorrelation Function).

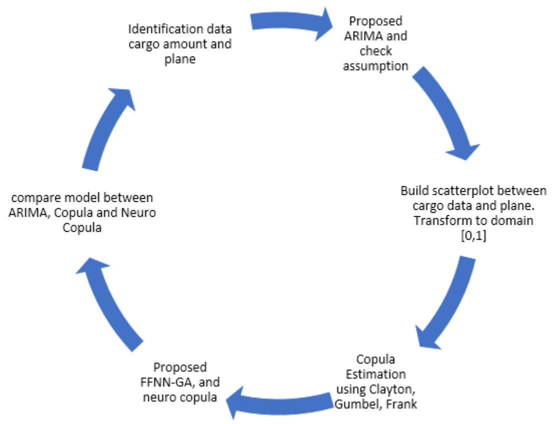

Fig 1. Flowchart

At the same time, Selection of the best method based on the prediction error rate. To justify the accuracy, we using Mean Absolute Percent Error (MAPE) which is smaller MAPE amount indicates that the model used for forecasting is gives more accurate with the following equation as bellow:

$$MAPE = \frac{\sum_{t=1}^{n} |\left(\frac{X_t - F_t}{Xt}\right) x100\%|}{n} \quad (9)$$

With,

X_t : Actual data at period-t
F_t : Prediction data at period-t
n : Number of observation
t : Period to 1,2,3,...,t

Tabel 3. MAPE Interpretation

MAPE	Interpretation
<4.9%	Highly Accurate Forecasting
5%-9.9%	Accurate Forecasting
10%-14.9%	Good Forecasting
15%-19.9%	Reasonable Forecasting
>20%	Inaccurate Forecasting

4 Analysis

4.1 Fitting Copula

The first step is to determine the order *p* and *q* in the ARIMA model which can be done by looking at significant ACF and PACF plots[16]. So, at this stage will be the model identification in view plot ACF and PACF table 3 show the ACF cut off on 1st lag and PACF also cut off on lag to 1, 2, and 3.

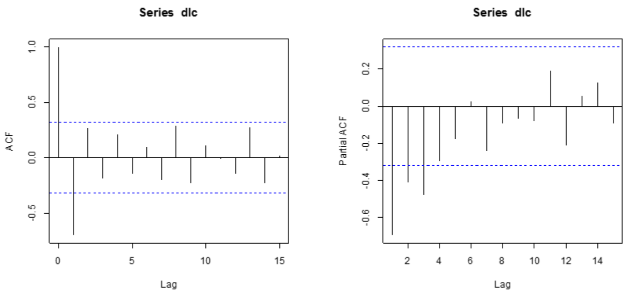

Fig 2. Plot ACF and PACF Cargo Data

ARIMA model with differencing= 2. Here are some models of ARIMA conjecture on cargo data model that can be seen in Table 4

Tabel 4. ARIMA

No	Model ARIMA		*p-value*	Decision	AIC
1	ARIMA (0,2,1)	MA(1)	0.000*	Significant	-326.5136
2	ARIMA (1,2,1)	AR(1)	0.004*	Significant	**-332.311**
		MA(1)	0.000*	Significant	
3	ARIMA (1,2,0)	AR(1)	0.000*	Significant	-315.5508
4	ARIMA (2,2,0)	AR(1)	0.000*	Significant	-319.9355
		AR(2)	0.001*	Significant	
		AR(1)	0.000*	Significant	
5	ARIMA (3,2,0)	AR(2)	0.0001*	Significant	-327.2448
		AR(3)	0.0019*	Significant	

The criteria in choosing the best model are to see the smallest AIC, from the analysis results obtained ARIMA model (1, 2, 1) is the best model for cargo and airplane variables by looking at the amount of AIC obtained because it has the smallest value compared with other models. Furthermore, the copula is a method that can measure dependencies. The advantages possessed copula can see how the characteristics in the relationship between variables[1]. The process of the copula method itself was analyzed from the standardized residual ARIMA model and transformed into Uniform [0,1]. However, before knowing what copulate functions are suitable for data and knowing parameter estimates, standardized residual plots can be seen to compare them with the contour plots of selected copula functions

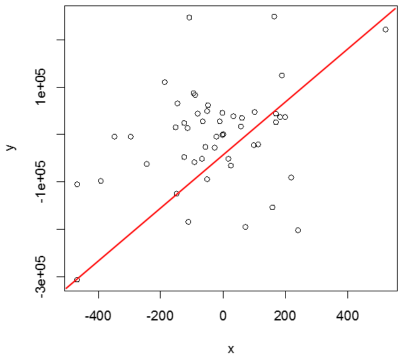

Fig 3. Residual Standardized Cargo VS Numbers of Airplane

In Figure 3 it can be seen that residual cargo and plane that has been standardized gathered at one point, this cannot be explained by ordinary correlation to see what symptoms are happening. Then using copula relationships and characters between cargo and plane can be seen, either in symmetric, asymmetric, lower tail or upper tail relationships. In the copula analysis, the residuals obtained from the best ARIMA model were the values applied in the analysis. Residuals obtained after standardized are then transformed. The first step is to look at the Archimedean family copula fit test against the data and also the estimated copula parameters estimated with MLE, and the results can be seen as follows.

Table 5: Copula Parameter

Copula	**Parameter**	**Cramer Von-Mises**	***p-value***
Clayton	0.43376	0.019175	0.7475
Frank	1.3778	0.023724	0.6816
Gumbel	1.2047	0.022461	0.6018

Table 5 shows that by using α = 5%, the null hypothesis is in the empirical model of the cargo and plane variables the same as the estimation model using the copula parameter (suitable model). The third copula family of Archimedean does not reject the null hypothesis (p-value> α) means that the three copula families have a match with the data. In determining the copula with the best family is to see the value of p-value is the largest. Among the three copulas, Clayton copula has the most significant p-value, which is 0.7475. Thus, Clayton copula is the preferred family because it has a pattern that is more in line with empirical data with $\hat{\theta}$ =0.43376. Where Clayton bivariate copula function, itself has a function$\left[\max\left\{u^{-\theta}+v^{-\theta}-1;0\right\}\right]$.The function of Clayton copula Archimedean has characteristics as copula that have the lower tail dependence. That is that the relationship between cargo and airplane variables means strong when it is at the bottom or low point. Clayton also has an asymmetric property, when associated with data is the variable of the plane has a relationship and influence variable cargo that is one way. Here will be shown how 3D visualization and also contour of Clayton function by generating data as much as N = 5000.

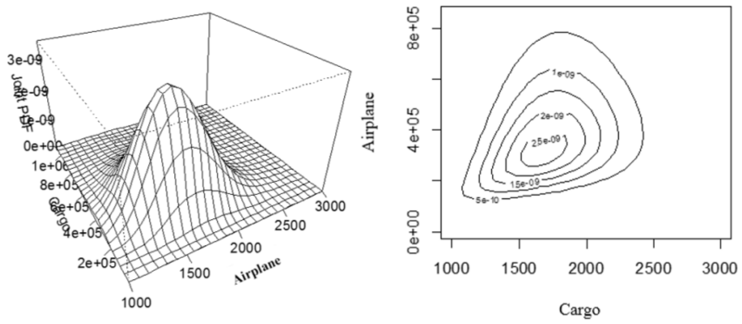

Fig 4. Contour Clayton (left) Clayton Dependency (right)

Through the contour plot and 3D visualization, the characteristics of the Clayton function are visible in the character of lower tail dependency, can be interpreted with at least a lot of cargo is closely related when many aircraft at Hussein Sastranegara Airport is also at a low point. The θ value obtained on the Clayton function is 0.43376. The estimated value of the Clayton function can be defined in the dependencies itself through the equations shown in Table 5. According to this Clayton, function copula method can be determined the amount of lower tail dependency on the Clayton function is 0.202301 which has a positive and significant relation between variables by using Tau Kendall approach, got the value of relationship of 0.174.

4.2 FFNN-GA

In this paper, the neural network genetic algorithm technique using a certain interval, if the lower boundary of r_b and the upper limit of r_a, the code of genetic algorithm compose $g_1, g_2, \ldots, g_n$ gene, the decoding can be done in the following way:

Real-number encoding

$$x = r_b + (r_a - r_b)g \tag{10}$$

Discrete decimal encoding

$$x = r_b + (r_a - r_b)(g_1 \times 10^{-1} + g_2 \times 10^{-2} + \cdots + g_n \times 10^{-n}) \tag{11}$$

Binary encoding

$$x = r_b + (r_a - r_b)(g_1 \times 2^{-1} + g_2 \times 2^{-2} + \cdots + g_n \times 2^{-n}) \tag{12}$$

If the coding scheme for discrete decimal encoding and binary encoding solution is apparently beyond the upper limit of the given interval, then the decoded formula needs to be changed to:

Discrete decimal encoding:

$$x = r_b + \left(\frac{(r_a - r_b)}{\sum_{i=1}^{n} 10^{-i}}\right)(g_1 \times 10^{-1} + g_2 \times 10^{-2} + \cdots + g_n \times 10^{-n}) \tag{13}$$

Binary encoding

$$x = r_b + \left(\frac{(r_a - r_b)}{\sum_{i=1}^{n} 2^{-i}}\right)(g_1 \times 2^{-1} + g_2 \times 2^{-2} + \cdots + g_n \times 2^{-n}) \tag{14}$$

In some cases, straightforward fitness values can be found easily, but in some other cases a very complex and challenging fitness value is required.

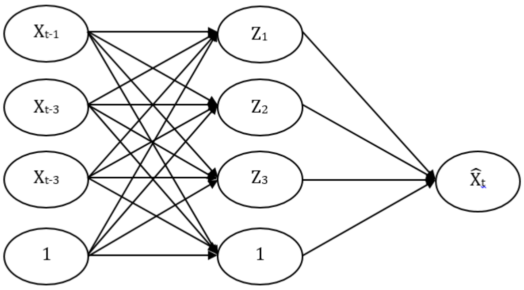

Fig 5. Architecture Neuro-AG Cargo Data

The network architecture to be used in the FFNN model is a multilayer network consisting of an input layer, a hidden layer, and an output layer. In the determination of the number of units in the input layer, there is no standard provision. Likewise, with the number of units in the hidden layer and output layer. At the same time, the problem is limited to the number of hidden layer units equal to the number of units in the input layer. However, the network architecture that is formed consists of 4 units of variable inputs that are considered influential, 1 hidden layer consisting of 4 neurons, 1 neuron at the output layer and bias. While the activation function used in the hidden layer to the output layer is the binary sigmoid (sigmoid logistics) and the activation function used for the output signal is the function of density (purlin). Based on the FFNN network architecture that has been formed then the number of weights or parameters to be estimated with using AG. As many as 16 units consisting of 9 weights of the neuron to give input signal to hidden layer (w_{ij}), 3 bias weight for hidden layer (w_{bj}), 3 weights of neuron to produce output layer (w_{jo}) and 1 bias weight for output layer (w_{bo}). The following shows the performance of AG training with a combination of crossover probability (p_c) = 0.8, population number 50, number of generation 500 and size of tournament (k) = 4:

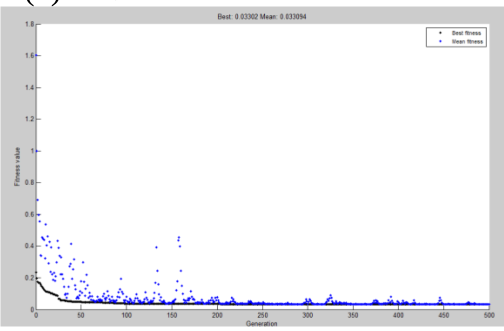

Fig 6. Genetic Algorithm 500 Generation

Based on Figure 6 it appears that the AG process was terminated after reaching the 500th generation. In addition, the resulting fitness value has converged and reached the global optimum that is with the best fitness value of 0.0302 and the average fitness value of 0.033904. After getting the ARIMA model, Copula Clayton, Copula Gumbel, Copula Frank, FFNN, and FFNN AG. We did the forecasting for the next 5 months and found that ARIMA and Copula models have low accuracy compared to FFNN, and FFNN AG. In addition, a combination of Copula model with neural network has been done and the result that Neuro Copula and FFNN-AG can follow the pattern of actual data.

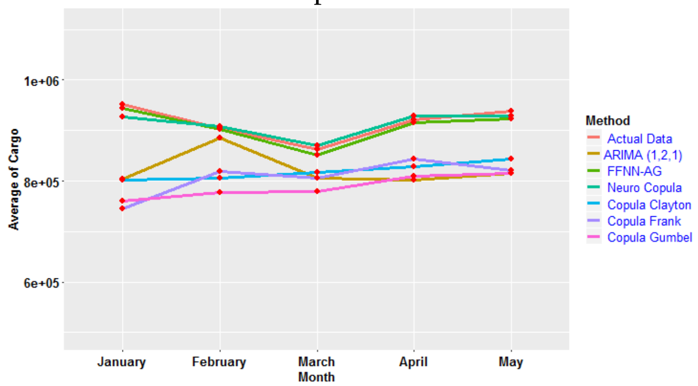

Fig 7. Comparing Prediction

Table 5: Accuracy Models

Model	MAPE	Category
ARIMA (1,2,1)	11.28%	Good Forecasting
Copula Clayton	10.29%	Good Forecasting
Copula Frank	12.36%	Good Forecasting
Copula Gumbel	13.55%	Good Forecasting
FFNN-AG	1.09%	Highly Acurate Forecasting
Neuro-Copula	1.17%	Highly Acurate Forecasting

5Conclusion

Based on the analysis it is found that the best model that can be used is FFNN-AG and Neuro-copula because it produces MAPE 1.09% and 1.17%. Noteworthy in training and testing of FFNN-AG is determining the number of generations to be achieved. Researchers may use trial and error to obtain high accuracy. In addition, traditional time series models such as ARIMA do not provide good accuracy in forecasting. However, since the principle of FFNN is like a regression function, the input lag used is derived from PACF. If the researcher wrongly specifies the lag value on PACF, then FFNN-AG will not be valid as it will over fitting or under fitting. Although the best copula model is Clayton, comparisons have been made to Copula Frank and Copula Gumbel and found that the accuracy can still be quite good. Subsequent research can perform a combination of heteroscedasticity (ARCH/GARCH/FIGARCH/APARCH) models with neural network.

Acknowledgments
This is research supported by GUP 2017-118 School of Mathematical Sciences, The National University of Malaysiaand Department of Statistics Padjadjaran University, West Java, Indonesia.

References

[1] R. E. Caraka, Supari, and M. Tahmid, “Copula-Based Model for Rainfall and El-Niño in Banyuwangi Indonesia,” *J. Phys. Conf. Ser.*, vol. 1008, no. 1, 2018.

[2] M. T. Hagan, H. B. Demuth, and M. H. Beale, “Neural Network Design,” *Bost. Massachusetts PWS*, vol. 2, p. 734, 1995.

[3] A. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Stat. Comput.*, vol. 14, pp. 199–222, 2004.

[4] R. E. Caraka and S. A. Bakar, “Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations,” *J. Eng. Appl. Sci.*, vol. 13, no. 17, 2018.

[5] R. E. Caraka, S. A. Bakar, B. Pardamean, and A. Budiarto, “Hybrid support vector regression in electric load during national holiday season,” in *Proceedings - 2017 International Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017*, 2018, vol. 2018–Janua.

[6] R. E. Caraka, “Prediction of Euro 50 Using Back Propagation Neural Network (BPNN) and Genetic Algorithm (GA),” *Int. J. Eng. Bus. Manag.*, vol. 1, no. 1, pp. 62–67, 2017.

[7] A. Sklar, “Fonctions de répartition à n dimensions et leurs marges,” *Publ. Inst. Stat. Univ. Paris*, vol. 8, pp. 229–231, 1959.

[8] B. Rayens and R. B. Nelsen, “An Introduction to Copulas,” *Technometrics*, vol. 42, no. 3, p. 317, 2000.

[9] R. E. Caraka, H. Yasin, W. Sugiyarto, Sugiarto, and K. M. Ismail, “Time Series Analysis Using Copula Gauss and AR(1)-N.GARCH(1,1),” *MEDIA Stat.*, vol. 9, pp. 01–13, 2016.

[10] U. Nacional and M. R. Melchiori, “Which Archimedean Copula is the Right One?,” *SSRN Electron. J.*, no. October, 2003.

[11] R. E. Caraka *et al.*, “Ecological Show Cave and Wild Cave : Negative Binomial Gllvm ’ s Arthropod Community Modelling,” *Procedia Comput. Sci.*, vol. 135, pp. 377–384, 2018.

[12] S. Ahmad and A. Ambrosetti, “Laplace transform,” in *UNITEXT - La Matematica per il 3 piu 2*, vol. 88, 2015, pp. 225–249.

[13] C. Meyer, “The bivariate normal copula,” *Commun. Stat. - Theory Methods*, vol. 42, no. 13, pp. 2402–2422, 2013.

[14] B. Warsito, R. Santoso, Suparti, and H. Yasin, “Cascade Forward Neural Network for Time Series Prediction,” in *Journal of Physics: Conference Series*, 2018, vol. 1025, no. 1.

[15] Suhartono, B. S. S. Ulama, and A. J. Endharta, “Seasonal time series data forecasting by using neural networks multiscale autoregressive model,” *Am. J. Appl. Sci.*, vol. 7, no. 10, pp. 1372–1378, 2010.

[16] R. E. Caraka, H. Yasin, and T. Toharudin, “The Step Construction of Copula Gaussian Multivariate and AR(1)-N.GARCH(1,1) Models,” in *Joint Workshop KO2PI and The 1st International Conference on Advance & Scientific Innovation*, 2018.