Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

Artificial Neural Network (ANN) Classification: Titanic Passenger Safety

Juanda[1*], Khoirin Nisa[2]

[1,2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia
[*]corresponding author: juanda007@gmail.com

Received dd Month yy; Received in revised form dd Month yy; Accepted dd Month yy (9pt)

Abstrak. *Inovasi sains dan teknologi selalu menjadi pendorong utama pertumbuhan ekonomi dan kemajuan social. Pesatnya perkembangan teknologi dan kemajuan internet telah memungkinkan menyerbarkan informasi, maupun interaksi semakin mudah. Dengan pesatnya perkembangan teknologi, banyak informasi yang dibagikan setiap detik sehingga menghasilkan big data baik segi variabel yang berbeda, kompleks. JST merupakan hasil karya dalam bidang computer yang diinspirasi dari kemampuan otak manusia yang terdiri dari jaringan saraf biologis, Dalam beberapa tahun terakhir, penggunaan jaringan saraf tiruan (JST) semakin meningkat. Penelitian yang dilakukan bertujuan untuk menganalisa kemampuan bertahan hidup penumpang titanic yang mengalami kecelakaan pada saat berlayar dan tenggelam. Penelitian ini menggunakan data awal sebanyak 1309 observasi dengan 14 variabel. Dari hasil penelitian variabel tersembunyi berjumlah 2 merupakan yang paling akurat dengan akurasi 80,5 %, dibandingkan dengan jumlah variabel tersembunyi sebesar 3 (79 %) dan 4 (79 %). Sehingga dapat disimpulan jumlah variabel tersembunyi dengan jumlah layar tersembunyi yang sama tidak memiliki perbedaan akurasi yang signifikan.*

Kata kunci*: ANN; Big Data; Machine Learning; Titanic*

Abstract. Scientific and technological innovation has always been the main driver of economic growth and social progress. The rapid development of technology and advances in the internet have made it possible to disseminate information and interact more easily. With the rapid development of technology, a lot of information is shared every second, resulting in big data in terms of different, complex variables. ANN is the result of work in the computer field that is inspired by the capabilities of the human brain which consists of biological neural networks. In recent years, the use of artificial neural networks (ANN) has increased. The research carried out aims to analyze the survival capabilities of Titanic passengers who experienced an accident while sailing and sank. This research uses initial data of 1309 observations with 14 variables. From the research results, 2 hidden variables are the most accurate with an accuracy of 80.5%, compared to the number of hidden variables of 3 (79%) and 4 (79%). So it can be concluded that the number of hidden variables with the same number of hidden screens does not have a significant difference in accuracy.
Keywords: ANN; Big Data; Machine Learning; Titanic

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

INTRODUCTION

Scientific and technological innovation has always been a key driver of economic growth and social progress. With the advent of the digital era, continuous breakthroughs in information technologies such as artificial intelligence (AI), blockchain, and big data have brought many new opportunities for patent system improvement (Custers, 2022). Meanwhile, digital technology and related innovations have many characteristics that differ from those of the past, posing further challenges to the innovation system. How to effectively utilize these new technologies to solve or mitigate the problems of legacy systems, and avoid the associated difficulties, has become an important issue for scholars and policymakers in the digital era (Ouyang et al., 2023). In 2011, McKinsey & Company believed that big data could independently mine and analyze information with the help of computers and produce effective information (Cui et al., 2022). Big data has 5 characteristics, including: volume, speed, variety, truth and value (Geerts & O'Leary, 2022). Big data analytics (BDA) has become increasingly available in recent years. Companies can collect and analyze big data with low-cost service packages (Article et al., 2020). One Examines the impact of the use of big data analytics on corporate social performance in corporate social responsibility initiatives, Identifying the importance of a company's ability to provide data-based insights through big data analytics capabilities (Choi & Park, 2022).

ANN is the result of work in the computer field which is inspired by the capabilities of the human brain which consists of biological neural networks (Management & Homes, 2019). Through a processing unit better known as a neuron (Nikoo et al., 2023), where ANN imitates and is trained like the way the human brain works in finding the function value of the goal (Chaki & Biswas, 2023), thus getting better prediction results compared to with conventional models (Betiku et al., 2015). In recent years, the use of artificial neural networks (ANN) has increased (Esfe et al., 2022), this is because ANN modeling is considered a reliable model for non-linear data, where complex relationships between dependent and independent can be found (Teke et al., 2023). Novelty in this study, the use of ANN to analyze Titanic accident data is a new approach in understanding the factors that influence passenger survival in emergency situations. The emphasis on hidden variables in the ANN model and the finding that a certain number of hidden variables did not make a significant difference in accuracy, provide new insights regarding the optimal configuration of the ANN model for ship accident cases. The relationship between research and the use of big data in analyzing and understanding social phenomena such as ship accidents, which increases understanding of the potential of big data and its analysis in the context of public security and safety.

The application of ANN in this study as a model that is considered able to overcome the complexity of non-linear data, shows the relevance and progress in using machine learning techniques to understand and predict events involving many variables and factors.Research purposes is analyzing the survival capabilities of Titanic passengers who experienced an accident and sank using an Artificial Neural Networks (ANN) approach. Assessing the accuracy of the ANN model in predicting passenger survival capabilities based on initial data including 14 variables.

METHOD

The research carried out aims to analyze the survival capabilities of Titanic passengers who experienced an accident while sailing and sank. This research uses initial data of 1309 observations with 14 variables. Where the safety status of passengers after the drowning

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

tragedy is the dependent variable and the other thirteen variables are independent variables. The variables used in this research can be seen as in Table 1 below:

Table 1. Initial data variables

No	Variable	Description
1	Happy	explain safe (1) and not safe (0)
2	pclass	passenger class level
3	Name	the name of each passenger taking part in the cruise
4	sex	passenger gender
5	age	age of the passenger
6	siblings	number of passengers' siblings
7	dry	body condition at the time of rescue
8	ticket	ticket number used
9	tariff	the fare of the ticket used
10	cabin	cabin where passengers rest
11	leave	place of departure
12	boat	number of rescue ship used
13	body	body temperature at the time of rescue
14	home.destination	home or destination address of the passenger

This research specifically uses the R Studio 4.1.2 application, with the Artificial Neural Network (ANN) algorithm method with the nnet library. The nnet algorithm uses 1 hidden screen stage with k variables, where this hidden screen is the link between 13 independent variables and 1 dependent variable.
The classification steps using the ANN method are as follows:
1. Carry out selection on the data, by taking variables that are deemed to have sufficient influence and eliminating observations whose data are incomplete and dividing the selection data into training data and test data.
2. Build an ANN model using training data with a number of hidden variables of 2,3,4.
3. Evaluate the ANN model using test data
4. Compare the accuracy levels of the five models with hidden variables of 2,3 and 4.
5. Draw conclusions

RESULT AND DISCUSSION

In this research, researchers will create an ANN model to classify whether Titanic passengers were safe or not according to the variables used and compare the accuracy level of the ANN model with different numbers of hidden screens.

The following are the results of the safety classification of Titanic passengers based on the ANN algorithm:

1. Data selection

   The initial data taken initially had 14 variables and 1309 observations. However, there were several observations whose data were missing (blank) so data selection was carried out both from the number of variables and the number of observations so that

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

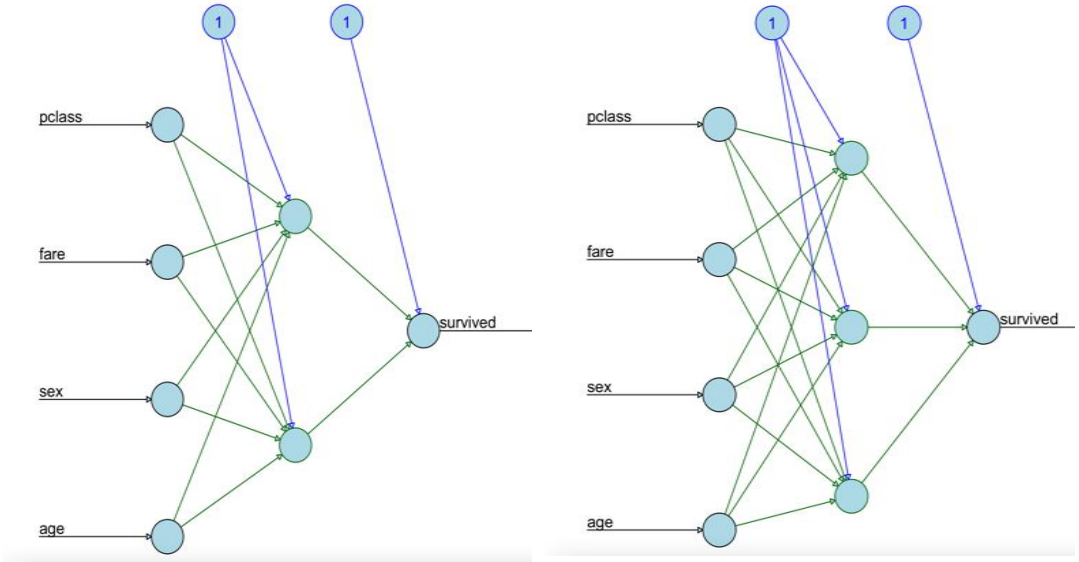
complete data was obtained (nothing was missing). The selection results obtained 5 variables with a final number of observations of 1045 observations

Table 2. Data variables after selection.

No	Variable	Explanation
1	Happy	explain safe (1) and not safe (0)
2	pclass	passenger ticket class level
3	sex	passenger gender
4	age	age of the passenger
5	dry	body condition at the time of rescue

In Table 2 there are 5 variables used, including: safe, pclass, sex, age, and dry. Where the safe variable is the dependent variable while the pclass, sex, age and dry variables are independent variables. In building this model the researcher will use training data of 836 observations (80%) and test data of 209 observations (20%).

2. Building a model using the ANN method (single screen)
The ANN method consists of multiple hidden screens and single hidden layers. In this research a single hidden screen will be used with many variables in the hidden screen of 2,3,4 variables. The purpose of the hidden screen is to connect the dependent variable with several independent variables. The process of a single hidden screen ANN with 2, 3, and 4 hidden variables can be seen in Figure 1.



Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

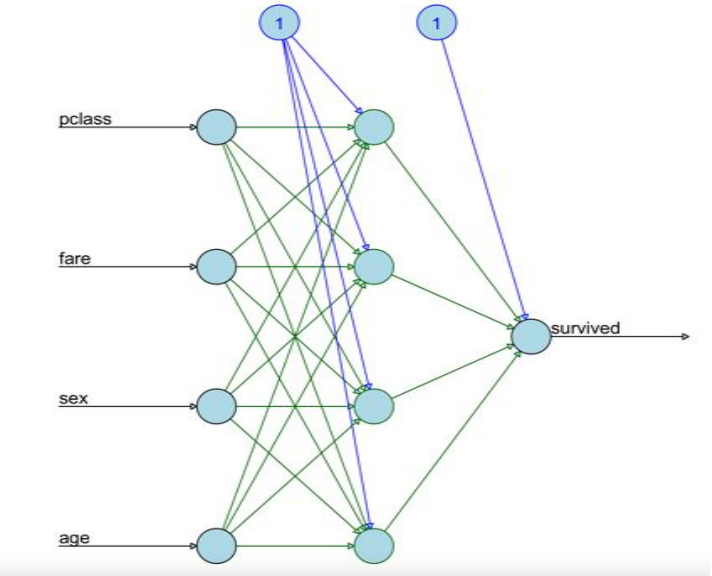


Figure 1. Classification of survivors using the ANN method k=2,3,4

In Figure 1 we can see that each independent variable is connected to all the hidden screen variables 2,3,4 then these hidden variables are connected back to the dependent variable which is the final goal value of the ANN process.

3. Evaluate the ANN model with test data
   After modeling with ANN, the number of variables in the hidden screen has been set to 2 variables, 3 variables and 4 variables. The final score obtained from the classification is as shown in table 3.

Table 3. Comparison of the final ANN prediction results with test data values

No	Data Test	P.K2	P.K3	P.K4
1	1	1	1	1
2	0	0	0	0
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
……	……	……	……	……
	0	0	0	0
205	0	0	0	0
206	0	0	0	0
207	0	0	0	0
208	0	0	0	0
209	1	1	1	0

4. Analyze the accuracy of the ANN model

Sciencestatistics
Journal of Statistics, Probability, and Its Application
Available at https://scholar.ummetro.ac.id/index.php/sciencestatistics/index

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

After carrying out ANN modeling, we will compare the accuracy of each method used to find the ANN model with the maximum accuracy. The accuracy values of each method can be seen in Table 4.

Table 4. Accuracy value of the ANN method with variables 2,3 and 4.

		Data Test		Acuration
		0	1	
P.K2	0	120	33	0.80382775
	1	8	48	
P.K3	0	110	24	0.79904306
	1	18	57	
P.K4	0	113	27	0.79904306
	1	15	54	

CONCLUSION

The ANN method is used to classify with adequate accuracy in determining data classification. This research examines a comparative analysis of the accuracy of the ANN algorithm for classifying the safety of Titanic passengers with a number of variables on a single hidden screen. From the research results, 2 hidden variables are the most accurate with an accuracy of 80.5%, compared to the number of hidden variables of 3 (79%) and 4 (79%). So it can be concluded that the number of hidden variables with the same number of hidden screens does not have a significant difference in accuracy.

REFERENCE

Article, F. L., Morimura, F., Morimura, F., Morimura, F., Sakagawa, Y., & Lecturer, S. (2020). Journal of Retailing and Consumer Services The intermediating role of big data analytics capability between responsive and proactive market orientations and firm performance in the retail industry. *Journal of Retailing and Consumer Services*, *29*(October 2022), 260–287. https://doi.org/10.1016/j.jretconser.2022.103193

Baygin, M. (2021). An accurate automated schizophrenia detection using TQWT and statistical moment based feature extraction. *Biomedical Signal Processing and Control*, *68*(April), 102777. https://doi.org/10.1016/j.bspc.2021.102777

Betiku, E., Okunsolawo, S. S., Ajala, S. O., & Odedele, O. S. (2015). Performance evaluation of artificial neural network coupled with generic algorithm and response surface methodology in modeling and optimization of biodiesel production process parameters from shea tree (Vitellaria paradoxa) nut butter. *Renewable Energy*, *76*, 408–417. https://doi.org/10.1016/j.renene.2014.11.049

Chaki, S., & Biswas, T. K. (2023). An ANN-entropy-FA model for prediction and optimization of biodiesel-based engine performance. *Applied Soft Computing*, *133*, 109929. https://doi.org/10.1016/j.asoc.2022.109929

Choi, H. Y., & Park, J. (2022). Do data-driven CSR initiatives improve CSR performance? The importance of big data analytics capability. *Technological Forecasting and Social Change*, *182*(July 2021), 121802. https://doi.org/10.1016/j.techfore.2022.121802

Cui, Y., Ma, Z., Wang, L., Yang, A., Liu, Q., Kong, S., & Wang, H. (2022). A Survey on Big Data-Enabled Innovative Online Education Syst ems During the COVID-19 Pandemic. *Journal of Innovation & Knowledge*, 100295. https://doi.org/10.1016/j.jik.2022.100295

Custers, B. (2022). New digital rights: Imagining additional fundamental rights for the digital era. *Computer Law and Security Review*, *44*, 105636. https://doi.org/10.1016/j.clsr.2021.105636

Vol. 1, No. 2, 2023, 84-90
ISSN 2963-9875 (Online)
ISSN 2964-2884 (Print)

Esfe, M. H., Toghraie, D., & Amoozad, F. (2022). Optimization and design of artificial neural network with Levenberg-Marquardt algorithm to increase accuracy in predicting the dynamic viscosity of SAE40 oil-based hybrid nano-lubricant. *Powder Technology*, *415*(October 2022), 118097. https://doi.org/10.1016/j.powtec.2022.118097

Geerts, G. L., & O'Leary, D. E. (2022). V-Matrix: A wave theory of value creation for big data. *International Journal of Accounting Information Systems*, *47*(January), 100575. https://doi.org/10.1016/j.accinf.2022.100575

Management, D., & Homes, S. (2019). Analytics-Assisted Smart Power Meters Considering. *Sensors*, *19*(9), 1–26.

Nikoo, M., Abbasi Malekabadi, R., & Hafeez, G. (2023). Estimating the mechanical properties of Heat-Treated woods using Optimization Algorithms-Based ANN. *Measurement: Journal of the International Measurement Confederation*, *207*(March 2022), 112354. https://doi.org/10.1016/j.measurement.2022.112354

Ouyang, X., Sun, Z., & Xu, X. (2023). Patent system in the digital era - Opportunities and new challenges. *Journal of Digital Economy*. https://doi.org/10.1016/j.jdec.2022.12.003

Pei, Y., Tang, X., Zhang, Y., Huang, Y., & Dong, L. (2022). Brief introduction of machine learning on cSDH patients. *Clinical Neurology and Neurosurgery*, *213*(98), 106982. https://doi.org/10.1016/j.clineuro.2021.106982

Teke, C., Akkurt, I., Arslankaya, S., Ekmekci, I., & Gunoglu, K. (2023). Prediction of gamma ray spectrum for 22Na source by feed forward back propagation ANN model. *Radiation Physics and Chemistry*, *202*(October 2022), 110558. https://doi.org/10.1016/j.radphyschem.2022.110558

Yang, Z., Tang, R., Zeng, W., Lu, J., & Zhang, Z. (2021). Short-term prediction of airway congestion index using machine learning methods. *Transportation Research Part C: Emerging Technologies*, *125*(September 2020), 103040. https://doi.org/10.1016/j.trc.2021.103040