

A Study of Prediction Model for Capture Fisheries Production in Indonesian Sea Waters Using Machine Learning

Ganjar Adi Pradana

Program Studi Perangkat Lunak Sistem Informasi, Jurusan Manajemen Sistem Informasi,
Fakultas Pascasarjana, Universitas Gunadarma
Email: georgius_ganjar@yahoo.co.id

Abstract – The potential for capture fisheries in Indonesia is a priceless wealth. This wealth has not been explored optimally. Fisheries resources are included in the category of renewable resources whose sustainability needs to be considered. This is important in maintaining food security which will increase over time, due to population growth. Capture Fisheries Production Prediction Model is needed to find out what determining variables affect capture fisheries production. There are many methods for predicting, the method that is widely used today is using machine learning since its ability to handle complex jobs with large input data. This research is a literature study, which aims to: (1) identify and analyze machine learning methods that are suitable for predicting capture fisheries production, and (2) identify variables that can affect capture fisheries production. The results of the study show that the Neural Network method is most widely used as a predictive model. In addition, the Random Forest and Linear Logistics methods provide better accuracy results. The results of the study also succeeded in finding 12 determining variables for the capture fisheries production prediction model.

Keywords – *Capture Fisheries Production, Machine Learning, and Prediction Models.*

I. INTRODUCTION

Indonesia is the largest archipelagic country in the world. The great marine and fisheries potential is a priceless wealth [1]. However, this wealth has not been explored optimally, especially capture fisheries resources. Fisheries resources are included in the category of renewable resources, then the question often arises of how to maximize the potential of these fishery resources without causing negative impacts in the future. This is important in maintaining food security which will increase over time, due to population growth.

Sustainability is the key in fisheries development which is expected through wise management to improve the condition of resources and the welfare of the fishing community itself [1, 2]. The main problem is structure of capture fisheries in Indonesia that is still dominated by small-scale fishermen, this affects the amount of production of the main commodities. In addition, Illegal, Unreported and Unregulated Fishing (IUUF) is the biggest threat to the sustainability [3]. In fishing activities at sea, there are factors that can affect the number of catches [2, 4, 5, 7]. Thus it is necessary to know the factors that influence fisheries production. After knowing these factors, analysis and prediction of capture fisheries production can be carried out. This allows stakeholders to receive input to make plans and policies to increase production in this sector.

Prediction models are one of the strategies that are commonly used in most companies or organizations in the world to plan their work before it actually happens. The essence of forecasting is predicting future events based on past patterns and applying judgment to the projections. To carry out the assessment process with lots of data, we need a system that is able to predict in order to increase

effectiveness. There are many methods for predicting, the method that is widely used today is using machine learning. Machine Learning can be used as a tool to analyze big data, find patterns in the past, to make predictions for the future. The Machine Learning method was chosen in this study because it can handle very complex jobs with large amounts of input data. This can offer a solution to predicting capture fisheries production.

The description of the problem and possible solutions above, made the writer interested in making it into a research. This initial research is a literature study, which aims to: (1) identify and analyze machine learning methods that are suitable for predicting capture fisheries production, and (2) identify variables that can affect capture fisheries production. This is expected to be a guideline for further technical research, in the use of appropriate model variables and also appropriate machine learning prediction methods. The results of this study in general are also expected to contribute to being a scientific reference for stakeholders in the field of capture fisheries to determine appropriate policies in order to improve the quality and quantity of sustainable capture fisheries production.

II. RESEARCH METHODOLOGY

This research is a literature study that summarizes some of the relevant literature on the prediction and analysis of capture fisheries production using machine learning. The stages in the research are as follows: literature identification, literature selection, method analysis and model determining variable analysis. The flowchart of the research stages can be seen in Figure 1.

The first step is to identify the literature through a search on the Crossref site. The keywords used in the



literature search were "Machine Learning Prediction", "Productive Waters", and "Analysis of Capture Fisheries Production". After the literature is collected, literature selection is carried out. Literature is selected based on relevance and year of publication. The title of the literature is considered whether it is in accordance with the themes of "prediction with machine learning" and "analysis of capture fisheries production". Selection of the year of literature publication is limited from 2002 to 2022.

After selecting the literature by title and year of publication, the remaining literature was reviewed. The literature review includes its suitability with the research discussed and the availability of full text articles. Selected literature will be discussed in this study. Selected literature to be discussed is a minimum of 15 articles. The focus to be analyzed is the relevant method to be used and the determination of the research model variables.

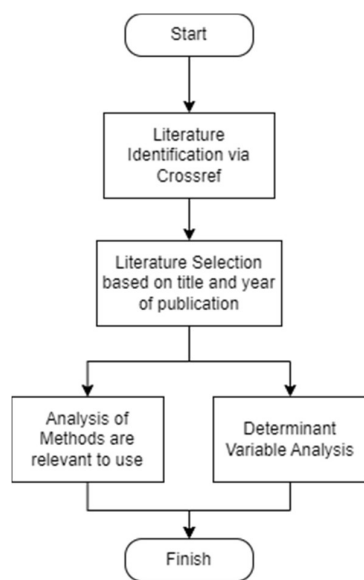


Figure. 1 Research Stages.

III. RESULTS AND DISCUSSION

Relevant primary literature that was successfully collected is shown in Table 1. This literature was chosen because it met the criteria in terms of relevance and year of publication. The total number of selected articles is 20 articles. There are 11 articles related to the machine learning prediction method. While the articles relating to the analysis of capture fisheries production are as many as 9 articles.

A. Machine Learning Method

The results of identifying the machine learning method for predictive analysis based on the collected literature are shown in Table 2. The results of identification based on the relevant literature collected show that the Neural Network method is most widely used as a tool for predicting various things with its determining variables, 10 times. After that follow the Logistic Regression method 2 times. While the Decision Tree, Linear Regression, Random Forest, SVM, and Naïve Bayes methods are also used as Machine

Learning methods for predictive analysis. This literature study also shows the results that the Random Forest and Linear Logistics methods produce a better level of accuracy compared to Neural Networks [10] and [12]. A summary of the identification results of these methods is shown in Table 3.

The following is a brief review of the identified machine learning methods.

a. Neural Network

Neural Networks are a type of machine learning process, called deep learning that uses nodes or neurons connected to each other in a layered structure that resembles the human brain. In this way, Neural Networks give computer programs the ability to recognize patterns and solve problems. Currently, this neural network model can be developed into Artificial Neural Networks, Convolution Neural Networks, and Recurrent Neural Networks [21].

b. Support Vector Machine

Support Vector Machine (SVM) is a method in supervised learning which is usually used for classification (such as Support Vector Classification) and regression (Support Vector Regression). In classification modeling, SVM has a more mature and clearer concept mathematically compared to other classification techniques. SVM can also solve classification and regression problems with linear and non-linear [22].

c. Logistic and Linier Regression

Regression models, including logistic regression, are data analysis methods that explain the correlation between a dependent variable and one or more independent variables. The goal of developing a regression model is to find the best-explaining relationship between these variables. The thing that distinguishes the logistic regression model from the simple linear regression model is that in logistic regression, the output variable is binary or dichotomous, so that the value of the dependent variable from the calculation results in the model will be mapped with a function to a binary value [22].

d. Random Forest

Random forest is a prediction algorithm which is a combination of several decision trees. This algorithm is quite computationally efficient because each decision tree randomly selects the variables used and does not use all the variables. The random forest model was built using the concept of bagging (bootstrap aggregation), namely collecting random sample observations into a container or called a bag. Some bags are composed of observations randomly selected from the original observational data in the training dataset. The dataset in the bag is formed from taking with returns which allows an observation result to have the opportunity to be re-elected [10].

e. Naive Bayes

Naive Bayes is a suitable method for binary and multiclass classification. This method, also known as Naive Bayes Classifier, applies supervised object classification techniques in the future by assigning class labels to instances/records using conditional probabilities. Conditional probability is a measure of the probability of an event occurring based on other events that have (with assumptions, presumptions, assertions, or proven) occurred [22].



Table 1. List of Primary Literature.

Author	Theme Category	Method
[1]	Fisheries Production Results	Regression Analysis
[2]	Sustainability of Capture Fisheries	Rapfish Technique (Rapid Appraisal for Fisheries).
[4]	Capture Fisheries Prediction with Data Mining.	Comparison of Decision Tree, Naive Bayes, SVM, and Neural Network.
[5]	Yellow Fin Tuna Fishing Area.	Remote Sensing, Geographic Information Systems, and Linear Regression.
[6]	Mackerel Fishing Area	Remote Sensing and Geographic Information Systems.
[7]	Fisheries Production	Cobb-Douglas production function.
[8]	Stock Predictions with Deep Learning.	Hidden Layer Neural Network.
[9]	Comparison of Exchange Rate Prediction Models During the Pandemic	Comparison of Neural Networks, Genetic Algorithms, and Particle Swarm Optimization
[10]	Fish Catchment Zone Prediction	Remote Sensing and Geographic Information Systems
[11]	Prediction of Students Graduated Not on Time	Neural Network Backpropagation
[12]	Delivery Prediction with Machine Learning	Comparison of Random Forest, artificial neural network (ANN), and logistic regression.
[13]	Prediction of Number of Aircraft Passengers	Neural Network Backpropagation
[14]	Capture Marine Fishery Production	Cobb-Douglas production function.
[15]	Stock Prediction	Neural Networks, and Particle Swarm Optimization
[16]	Rice Production Prediction with Artificial Neural Network.	Artificial Neural Network Backpropagation
[17]	Hotel Cancellation Predictions	Comparison Logistic Regression and Artificial Neural Network
[17]	House Price Prediction with Machine Learning.	Linear Regression
[19]	Pelagic Fisheries Capture Production.	Multiple Linear Analysis.
[20]	Capture Fisheries Production	CPUE Analysis, Scoring, Location Quotation.
[21]	Stock Price Prediction with Recurrent Neural Network	Recurrent Neural Network

f. Decision Tree

Decision tree is a structure that is used to assist the decision-making process. Called a "tree" because this structure resembles a tree complete with roots, trunks, and branches. In data science, a decision tree structure can help make effective decisions while still paying attention to the possible results and consequences [22].



Table 2. The identification results of the machine learning method.

Author	Machine Learning Method	Result
[4]	Decision Tree, Naive Bayes, SVM, and Neural Network	The application of the Decision Tree, Naive Bayes, SVM, and Neural Network algorithms in the case of predicting an increase in the average volume of capture fisheries is quite good. The Neural Network Algorithm has the highest accuracy value.
[8]	Hidden Layer Neural Network	The challenge in processing Deep Learning (DL) data is determining the epoch parameter values to produce high accuracy predictions.
[9]	Comparison of Neural Networks, Genetic Algorithms, and Particle Swarm Optimization	The results showed that the performance of the optimization algorithm (Genetic Algorithm, and Particle Swarm Optimization) in increasing the Neural Network error rate was the same, namely 0.020 +/- 0.006.
[11]	Neural Network Backpropagation	Data mining techniques using neural network backpropagation can produce a precise and accurate prediction from previous research to determine students graduating not on time at State Vocational High School 1 Kertak Hanyar.
[12]	Random Forest, artificial neural network (ANN), and logistic regression.	The random forest method produces a better accuracy value when compared to the logistic regression and artificial neural network (ANN) method, which is 76.6%, while the ANN method and logistic regression are 73.81% and 72.84%.
[13]	Artificial Neural Network Backpropagation	The level of accuracy reaches 88.14% or with a relatively low error rate of 11.86%.
[15]	Neural Network Backpropagation	Predictions using the Backpropagation Neural Network method produce good predictions with a MAPE of 19.77%. The prediction for the number of passengers in the next period, namely May 2022, is 1,060,500 passengers.
[16]	Neural Networks (NN), and Particle Swarm Optimization (PSO)	By using NN, the results of the experiments carried out with the Neural Network 500 training cycle, 3 Hidden layers, Momentum 0 and Learning rate 0.2 get 0.466 rmse. Meanwhile, the results of the NN experiment with PSO or PSO-NN feature selection yielded 0.373 rmse. This PSO-NN-based research is able to predict more accurately.
[17]	Logistic Regression and Artificial Neural Network	From the five trials, the fifth experiment (logistic regression with GridSearchCV) is the most optimal method for predicting hotel booking cancellations, with an accuracy value of 79.77%, a precision value of 85.86% and a recall value of 55.07%.
[18]	Linear Regression	Using 80% of the dataset for training and 20% of the dataset used for testing produces output values with an accuracy rate of 88% in predicting.
[21]	Recurrent Neural Network	Prediction of seven variable features with RNN produces an accuracy of 94% for training data and 55% for test data. Accuracy is obtained after training using 1218 data.

Table 3. Summary of Machine Learning Method Identification Results.

No	Method	Quantity
1	Neural Network	10 (ten)
2	Logistic Regression	2 (two)
3	Linear Regression	1 (one)
4	Random Forest	1 (one)
5	SVM	1 (one)
6	Naïve Bayes	1 (one)
7	Decision Tree	1 (one)



B. Fisheries Production Model Variables

The results of the identification of the capture fisheries production model variables based on the collected literature are shown in Table 4. The identification results based on the collected relevant literature managed to find 12 variables that determine the capture fisheries production model (Table 5). The number of boats, the number of fishermen, and fishing gear are the most frequently encountered variables, the more boats and also the number of fishermen in an area, of course, will also increase capture fisheries production. Fishing gear also affects fisheries production, adequate fishing gear technology will make it easier for capture fisheries.

Sea Surface Temperature and Chlorophyll-a also affect the productivity of the waters. Usually productive waters have low surface temperatures and high amounts of chlorophyll a, this is related to the upwelling phenomenon. Illegal fishing practices and legal aspects can also affect capture fisheries production in an area. Other factors that may be used as a model variant are type of boats time at sea, fishermen's income, wind speed, amount of rainfall and number of rainy days.

Table 4. The identification results of the machine learning method.

Author	Independent Variable	Dependent Variable	Result
[1]	Number of Vessels, Number of Fishermen	Fisheries Production Results	Number of Vessels and Number of Fishermen together affect Fisheries Production.
[2]	Ecology (Catch), Social (Quantity and Quality of Fishermen), Economic (Revenue of Fishermen), Technology (Vessels and Fishing Gear), Law and Institutions (Illegal Fishing, Law Enforcement Personnel, Regulations).	Sustainability of Capture Fisheries	Ecological conditions are the worst aspects, technological aspects are the most differentiated, and integration between or across aspects is an important thing.
[5]	Sea Surface Temperature, Chlorophyll-a	Yellowfin Fish Catches.	The distribution of SST and chlorophyll-a did not significantly affect the catch of yellowfin tuna in the waters of Aceh Province.
[6]	Sea Surface Temperature, Chlorophyll-a	Mackerel Catches	Bangka waters are suitable as an estimator of potential mackerel fishing grounds. The distribution of fishing grounds is not only in waters close to the fishing base (PPN Sungailiat), but also in waters far enough from the fishing base.
[7]	Fuel, time, distance, number of ships, number of crew, distance, fishing gear.	Fisheries Production	Significant factors are time, vessel and fishing gear. Meanwhile, the insignificant factors were diesel fuel, number of crew members, and distance.
[10]	Sea Surface Temperature, Chlorophyll-a	Fish Catch	The fishing potential zone in Aceh Jaya waters was only detected in November and December. This is due to the high distribution of chlorophyll-a and the presence of sea surface temperatures suitable for fishing in that time period.
[14]	Types of Boats, Number of Boats, Types of Fishing Gear	Capture Fisheries Production	Adequate marine fishing fleets and equipment are needed to increase marine capture fisheries production in Aceh Jaya District.
[19]	Wind Speed, Total Rainfall, Number of Rainy Days.	Catches of Cob, Mackerel, Squid.	Wind speed is significant for squid and cob catches. The amount of rainfall and the number of rainy days are significant for mackerel catches.
[20]	Historical Fish Catches, Fishing Equipment, Illegal Fishing.	Capture Fisheries Production	Strategic steps needed: increasing the capacity and reach of the fishing fleet, effective efforts to prevent illegal fishing practices, construction of fishing ports that are directed to become the basis of fishing efforts and improving the quality of human resources.



Research on machine learning methods requires the availability of relatively large amounts of historical or time series data. This is supported by the results of research [21] which stated that good accuracy in his research was achieved after training using 1218 data. In further technical research, perhaps these determining variables need to be re-selected based on the availability of time series data.

Table 5. Results Summary of the Identification of Fisheries Production Model Variables.

No	Independent Variables / Determinants	Quantity
1.	Number of Boats	4 (four)
2.	Fishing Gear	4 (four)
3.	Number of Fishermen	3 (three)
4.	Sea Surface Temperature	3 (three)
5.	Chlorophyll-a	3 (three)
6.	Illegal Fishing and Other Legal Aspects	2 (two)
7.	Types of Boats	1 (one)
8.	Time to Go to Sea	1 (one)
9.	Fishermen's Income	1 (one)
10.	Wind velocity	1 (one)
11.	Number of Rainy Days	1 (one)
12.	Total Rainfall	1 (one)

IV. CONCLUSION

The results of the study show that the most widely used method for prediction models with Machine Learning is Neural Network. The Random Forest and Linear Logistics methods also need to be considered for further technical research in this regard, because they provide better accuracy results than the Neural Network method. It is hoped that further technical research will be able to use these methods as tools and also conduct comparative studies of some of these methods on capture fisheries production prediction models.

The results of the study also succeeded in finding 12 determining variables that could be used for the capture fisheries production prediction model with Machine Learning. These variables include: Number of Boats, Fishing Gear, Number of Fishermen, Sea Surface Temperature, Chlorophyll-a, Illegal Fishing and Legal Aspects, Types of Boats, Time to Go to Sea, Fishermen's Income, Wind Speed, Number of Rainy Days, and Total Rainfall. Further technical research, these determining variables need to be re-selected based on the availability of time series data.

REFERENCES

- [1] Sofiyanti, N., Suartini, S. (2016). Pengaruh Jumlah Kapal Perikanan dan Jumlah Nelayan terhadap Hasil Produksi Perikanan di Indonesia. *Journal of Accounting and Finance*, 1 (1), 49-61. DOI: <https://doi.org/10.35706/acc.v1i01.442>
- [2] Nababan, B.O., Sari, Y.D., Hermawan, M. (2017). Analisis Keberlanjutan Perikanan Tangkap Skala Kecil di Kabupaten Tegal, Jawa Tengah. *Jurnal Bijak dan Riset Sosek KP*, 2 (2), 137-158. DOI: <https://doi.org/10.15578/jsekp.v2i2.5868>
- [3] Kusdiantoro, Fahrudin, A., Wisudo, S. H., dan Juanda, B. (2019). Perikanan Tangkap di Indonesia: Potret dan Tantangan Keberlanjutannya. *Jurnal Sosek KP*, 14 (2), 145-162. DOI: <http://dx.doi.org/10.15578/jsekp.v14i2.8056>
- [4] Andriani, A. (2015). Prediksi Kenaikan Rata-Rata Volume Perikanan Tangkap Dengan Teknik Data Mining. *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan*, 117-121. Source: <http://conference.dinus.ac.id/index.php/SEMANTI/K/SEMANTIK2015/paper/view/79>
- [5] Bahri, S., Simbolon, D., dan Mustaruddin, M. (2017). Analisis Daerah Penangkapan Ikan Madidihang (*Thunnus albacares*) Berdasarkan Suhu Permukaan Laut dan Sebaran Klorofil-a di Perairan Provinsi Aceh. *Jurnal Teknologi Perikanan Dan Kelautan*, 8 (1), 95-104. DOI: <https://doi.org/10.24319/jtpk.8.95-104>
- [6] Bukhari, B., Adi, W., Kurniawan, K. (2017). Pendugaan Daerah Penangkapan Ikan Tenggiri Berdasarkan Distribusi Suhu Permukaan Laut dan Klorofil-a di Perairan Bangka. *Jurnal Perikanan Tangkap*, 1 (3), 1-22. Source: <https://ejournal2.undip.ac.id/index.php/juperta/article/view/1871>
- [7] Damayanti, H.O. (2018). Faktor-Faktor yang Mempengaruhi Produksi Ikan pada Nelayan Kecil (Studi di Desa Pecangaan Kabupaten Pati). *Jurnal Litbang*, 8 (2), 83-92. DOI: <https://doi.org/10.33658/jl.v12i2.39>
- [8] Hastomo, W., Karno, A.S.B., Kabuana, N., Bisfiani, E., Lussiana. (2021). Optimasi Deep Learning untuk Prediksi Saham di Masa Pandemi Covid-19. *Jurnal Edukasi dan Penelitian Informatika*, 7 (2), 133-140. DOI: <https://doi.org/10.26418/jp.v7i2.47411>
- [9] Ikhsan, A.N., Arsi, P., Suhuman, J. (2021). Komparasi Model Prediksi Kurs Pada Masa Pandemi Covid-19 Menggunakan Neural Network



- Berbasis Genetic Algorithm dan Particle Swarm Optimization. *Jurnal Infotekmesin*, 13 (01), 74-79. DOI: <https://doi.org/10.35970/infotekmesin.v13i1.938>
- [10] Mursyidin. (2019). Prediksi Zona Tangkapan Ikan Menggunakan Citra Klorofil-a dan Citra Suhu Permukaan Laut Satelit Aqua MODIS Di Perairan Aceh Jaya. *Jurnal Ilmiah Pendidikan Teknik Elektro*, 3 (1), 11-18. DOI: <http://dx.doi.org/10.22373/crc.v3i1.3657>
- [11] Noor, A. (2017). Prediksi Siswa Lulus Tidak Tepat Waktu Menggunakan Backpropagation Neural Network. *Jurnal Humaniora Teknologi*, 1 (3), 8-15. DOI: <https://doi.org/10.34128/jht.v3i1.28>
- [12] Pambudi, H.K., Kusuma, P.G.A., Yulianti, F., Julian, K.A. (2020). Prediksi Status Pengiriman Barang Menggunakan Metode Machine Learning. *Jurnal Ilmiah Teknologi Informasi Terapan*, 6 (2), 100-109. DOI: <https://doi.org/10.33197/jitter.vol6.iss2.2020.396>
- [13] Pitriyani, D., Permasari, Y. (2022). Prediksi Jumlah Penumpang Pesawat dengan Backpropagation Neural Network. *Jurnal Riset Matematika*, 2 (2), 129-136. DOI : <https://doi.org/10.29313/jrm.v2i2.1327>
- [14] Pradana, R.S. (2019). Faktor Penentu Produksi Perikanan Laut Tangkap di Kabupaten Aceh Jaya. *Jurnal Ekonomi dan Pembangunan*, 10 (1), 51-62. DOI: <https://doi.org/10.22373/jep.v10i1.48>
- [15] Pramuntadi, A. (2017). Model Prediksi Rentet Waktu Neural Network berbasis Particle Swarm Optimazion untuk Prediksi Saham. *Jurnal Informatika dan Teknologi Informasi*, 14 (2), 100-106. DOI: <https://doi.org/10.31315/telematika.v14i2.2097.g1834>
- [16] Putra, H., Walmi, N.U. (2020). Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. *Jurnal Nasional Teknologi dan Sistem Informasi*, 6 (2), p. 100-107. DOI: <https://doi.org/10.25077/TEKNOSI.v6i2.2020.100-107>
- [17] Putra, M.S.T., Azhar, Y. (2021). Prediksi Model Logistic Regression dan Artificial Neural Network pada Prediksi Pembatalan Hotel. *Jurnal Informasi Sunan Kalijaga*, 6 (1), 29-37. DOI: <https://doi.org/10.14421/jiska.2021.6i1-04>
- [18] Saiful, A., Andryana, S., Gunaryati, A. (2021). Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression. *Jurnal Teknik Informatika dan Sistem Informasi*, 8 (1), 41-50. DOI: <https://doi.org/10.35957/jatisi.v8i1.701>
- [19] Sari, M., Wiyono, E.S., Zulkarnain. (2021). Pengaruh Cuaca terhadap Pola Musim Penangkapan Ikan Pelagis di Perairan Teluk Lampung. *ALBACORE Jurnal Penelitian Perikanan Laut*, 5 (3), 277-289. DOI: <https://doi.org/10.29244/core.5.3.277-289>
- [20] Solihin, I., Widodo, S.H., Haluan, J., Martianto, D. (2011). Pengembangan Produksi Perikanan Tangkap di Wilayah Perbatasan (Kasus Kabupaten Nunukan Kalimantan Timur). *Buletin PSP*, 19 (2), 9-18. Source: <https://journal.ipb.ac.id/index.php/bulpsp/article/view/4180>
- [21] Suyudi, M.A.D., Djamal, E.C., Maspupah, A. (2019). Prediksi Harga Saham Menggunakan Metode Recurrent Neural Network. *Seminar Nasional Aplikasi Teknologi Informasi*, A33-A38. Source: <https://journal.uin.ac.id/Snati/article/view/13398>
- [22] Kubat, M. (2017). *An Introduction to Machine Learning Second Edition*. Cham: Springer International Publishing.