

Prediction of hospital intensive patients using neural network algorithm

Rizka Rosita¹, Dwika Ananda Agustina Pertiwi², Oktaria Gina Khoirunnisa³

¹Department of Information System, Universitas Ahmad Dahlan, Indonesia

^{2,3}Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Feb 10, 2022

Revised Feb 20, 2022

Accepted Mar 5, 2022

Keywords:

Hospital intensive patient,
Electronic health record,
Hospitalization decision

ABSTRACT

This study aims to predict whether the patient deserves to be inpatient or outpatient by comparing several machine learning techniques, namely, logistic regression, decision tree, neural network, random forest, gradient boosting. The research method uses three stages of research, namely data collection, data preprocessing, and data modeling. Implementation of program code using google colab and python programming language. The dataset used as the research sample is Electronic Health Record Predicting data. Based on the accuracy results generated in this study, the use of the Neural Network machine learning algorithm to predict hospitalization decisions for patients has proven to be a machine learning algorithm that has the highest accuracy rate reaching 74, 47% compared to other comparison machine learning algorithms, namely logistic regression, decision tree, neural network, random forest, gradient boosting.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rizka Rosita
Department of Information System,
Universitas Ahmad Dahlan.
Email: rizka1800016083@webmail.uad.ac.id

1. INTRODUCTION

In the current pandemic era, of course, the role of health services is very important. There are several kinds of health services in Indonesia, one of which is a hospital. In accordance with Law number 44 of 2009, the hospital is one of the health service institutions that provides complete individual health services that provide various services such as outpatient, inpatient and emergency services. In its operation, the hospital has three pillars of authority including the owner, the pillar of health professionals and the pillar of management [1]. The hospital itself has an important role because every community who experiences complaints of health problems certainly needs proper care and treatment. Along with increasing population growth in Indonesia, the number of existing hospitals also increases. Reporting from the website of the Directorate General of the Ministry of Health, there are currently 2925 hospitals in Indonesia with various categories [2]. With this, of course, it is necessary to improve the quality of service.

The services in question are such as improving facilities or increasing the efficiency of services to patients. In this pandemic era, hospitals certainly experience a surge in patient admissions. This of course makes the hospital must be responsive in serving patients, one of which is in determining whether the patient must undergo inpatient or outpatient treatment. Hospitalization itself is the most expensive and valuable asset for a health organization [3]. In contrast to inpatient care, outpatient is a medical service to a patient where the

patient does not need to stay in the hospital [4]. With this prediction, of course, it will speed up the existing services at the hospital.

Related research that has been done is to compare machine learning random forest techniques, neural networks, and support vector machines [5]. These three techniques are used to assist in the process of diagnosing appendicitis or better known as acute appendicitis in a hospital. The study used 180 patients who were pooled, 135 used for training and 45 for testing. The results of this study indicate that random forest can predict acute appendicitis with good accuracy and used appropriately, can be an effective tool in clinical decision making.

In making these predictions, data from the patient's laboratory test results will be used which will later be classified whether the patient will then undergo inpatient or outpatient treatment. In the classification itself, there are various ways including logistic regression, decision tree, neural network, random forest, gradient boosting [6], [7]. This study aims to compare the accuracy of the results of several machine learning techniques.

2. METHOD

In this study, there were 3 stages used which were carried out during the research, namely data collection, data preprocessing, and data modeling [9]. The stages of the research are shown in Figure 1.

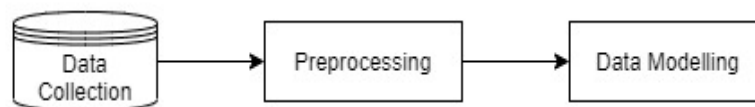


Figure 1. The stages of research

1) Data Collection

The research dataset is in the form of Electronic Health Record Predicting data collected from private hospitals in Indonesia. The data contains the results of the patient's laboratory tests which are used to determine the next patient's treatment, whether the patient is inpatient or outpatient. There are 10 available attributes, namely the results of laboratory tests for patients with hematocrit, laboratory test results for hemoglobin patients, laboratory test results for patients with erythrocytes, laboratory test results for leukocytes, laboratory test results for platelets, laboratory test results for MCH patients, laboratory test results for MCHC patients, laboratory test results. MCV patients, patient age, gender.

2) Data Preprocessing

Data preprocessing is the process of cleaning datasets from data that does not support the analysis process [8], [9]. Data that does not support the analysis process are duplicated data, empty data, and abnormal data. Dataset features that are not relevant to the analysis process are removed. Dataset features that have various data meanings are divided into additional features. This is done to improve the quality of the dataset being analyzed.

3) Data Modeling

Data modeling is the process of predicting patient hospitalization decisions in hospitals using previously cleaned datasets [10], [11]. The prediction process divides the dataset into training data and testing data with the result in the form of the accuracy of the predictions made. The machine learning algorithm used is Neural Network. Neural Network Algorithm is one of the methods in machine learning developed from Multi Layer Perceptron (MLP) which is designed to process two-dimensional data. Neural Network is included in the type of Deep Neural Network because of the depth of the network level and is widely implemented in image data [12]. Neural Network model serves for non-linear data processing [13].

The data processing, the author uses Google Colab using the Python programming language. Google Colab [14] is an open-source service provided by Google to all Gmail account owners. Google colab also provides GPUs for research related to machine learning. The Google Colab service provides 12.72 GB of RAM and 358.27 GB of hard disk space at a time. Each runtime lasts for 12 hours after which the runtime is reset, and the user must make the connection again. This is to ensure that people do not use GPU services for cryptocurrency mining and other illegal purposes. Python is a multipurpose interpretive programming language with a design philosophy that focuses on code readability [15]. Python is claimed to be a language that

combines capabilities, capabilities, with a very clear code syntax, and is equipped with a large and comprehensive standard library functionality.

3. RESULTS AND DISCUSSIONS

The results of this study compare machine learning logistic regression algorithms, decision trees, neural networks, random forests, gradient boosting as predictive models for determining patient decisions, whether inpatient or outpatient, which is adjusted to the data from laboratory tests. The result after implementing the program code in Google Colab shows a comparison of the percentage levels of accuracy of each algorithm. The accuracy results can be seen in Figure 2. And the F1-Score results are shown in Table 1.

Figure and format tables is using center alignment. Each of figures and tables are given number and description, as well as referred to the writing. Number and figure title is placed below the image, as shown in Figure 1.

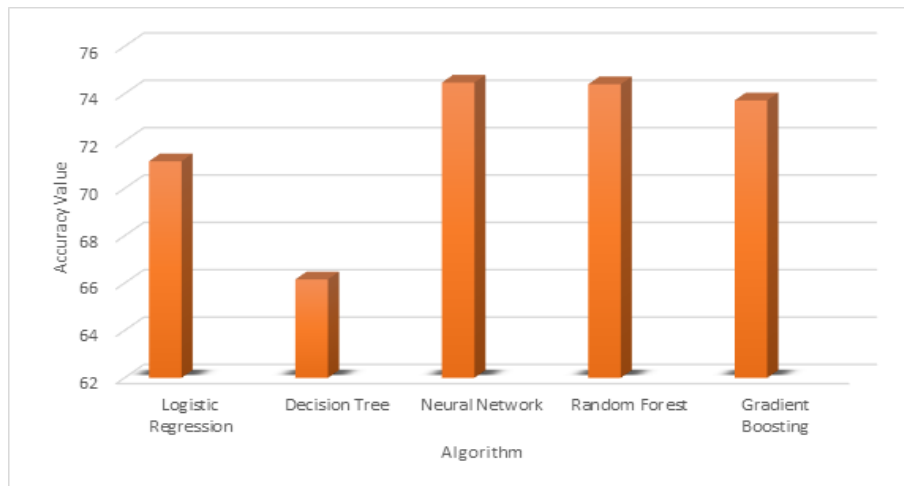


Figure 2. Accuracy comparison results

Table 1. F-1 Score Results

Algorithm	F1-Score
Logistic Regression	0.59705
Decision Tree	0.59273
Neural Network	0.67248
Random Forest	0.66797
Gradient Boosting	0.64848

In Figure 2, the neural network is superior to the comparison algorithm with an accuracy value of 74.47%. Ranked second in the machine learning random forest algorithm with an accuracy range of 74.40%. Furthermore, the accuracy of the comparison algorithm on gradient boosting is 73.72%, logistic regression is 71.15%, and the lowest accuracy is the decision tree with a percentage of 66.16%. The accuracy of the neural network algorithm is 0.07% superior to the random forest algorithm.

Based on the accuracy results generated in this study, the use of the machine learning algorithm Neural Network to predict hospitalization decisions for patients has proven to be a machine learning algorithm that has the highest level of accuracy compared to other comparison machine learning algorithms, namely logistic regression, decision tree, neural network, random forest, gradient boosting.

4. CONCLUSION

Research on predicting patient hospitalization decisions in hospitals uses a machine learning algorithm Neural Network through several stages, namely data collection, data preprocessing and data modeling. The dataset that has been cleaned in the preprocessing data is continued to the data modeling stage. Data modeling aims to analyze the prediction accuracy of patient hospitalization decisions in hospitals. Our study succeeded in using the Neural Network algorithm to predict patient hospitalization decisions at the hospital with an accuracy rate of 74.47%.

REFERENCES

- [1] A. Royani, "Lingkungan kerja dan kompetensi perekam medis terhadap kualitas sistem rekam medis di rs al islam bandung," *J. Teras Kesehat.*, vol. 2, no. 1, pp. 1–13, 2019.
- [2] Ditjen Layanan Kesehatan, "Grafik rumah sakit by kelas," <http://sirs.yankes.kemkes.go.id/fo/>, 2020.
- [3] L. He, S. C. Madathil, G. Servis, and M. T. Khasawneh, "Neural network-based multi-task learning for inpatient flow classification and length of stay prediction," *Appl. Soft Comput.*, vol. 108, 2021.
- [4] A. Abdulaziz Mohsen, M. Alsurori, B. Aldobai, and G. Abdulaziz Mohsen, "New approach to medical diagnosis using artificial neural network and decision tree algorithm: application to dental diseases," *Int. J. Inf. Eng. Electron. Bus.*, vol. 11, no. 4, pp. 52–60, 2019.
- [5] C. H. Hsieh, R. H. Lu, N. H. Lee, W. T. Chiu, M. H. Hsu, and Y. C. Li, "Novel solutions for an old disease: Diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks," *Surgery*, vol. 149, no. 1, pp. 87–93, 2011.
- [6] H. Wang, C. Liu, and L. Deng, "Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.
- [7] A. Beucher, A. B. Møller, and M. H. Greve, "Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark," *Geoderma*, vol. 352, no. September 2017, pp. 351–359, 2019.
- [8] S. A. N. Alexandropoulos, S. B. Kotsiantis, and M. N. Vrahatis, "Data preprocessing in predictive data mining," *Knowl. Eng. Rev.*, vol. 34, 2019.
- [9] M. Sam'an and Y. N. Ifriza, "Performance comparison of support vector machine and gaussian naive bayes classifier for youtube spam comment detection," *J. Soft Comput. Explor.*, vol. 2, no. 2, pp. 93–98, 2021.
- [10] N. L. W. S. R. Ginantra, A. H. Arifah, F. N., Wijaya, R. S. Septarini, N. Ahmad, D. P. Y. Ardiana, and E. S. Negara, *Data mining dan penerapan algoritma*. Yayasan Kita Menulis, 2021.
- [11] I. G. A. Suciningsih, M. A. Hidayat, and R. A. Hapsari, "Comparison analysis of naïve bayes and decision tree C4.5 for caesarean section prediction," *J. Soft Comput. Explor.*, vol. 2, no. 1, pp. 46–52, 2021, doi: 10.52465/joscex.v2i1.25.
- [12] R. Ridwan, H. Lubis, and P. Kustanto, "Implementasi algoritma neural network dalam memprediksi tingkat kelulusan mahasiswa," *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 286–293, 2020.
- [13] X. Yao *et al.*, "Land use classification of the deep convolutional neural network method reducing the loss of spatial features," *Sensors*, vol. 19, no. 12, p. 2792, 2019.
- [14] P. Kanani and M. Padole, "Deep learning to detect skin cancer using google colab," *Int. J. Eng. Adv. Technol. Regul. Issue*, vol. 8, no. 6, pp. 2176–2183, 2019.
- [15] A. N. Syahrudin and T. Kurniawan, "Input dan output pada bahasa pemrograman python," *J. Dasar Pemrograman Python Stmik*, vol. January, pp. 1–7, 2018.