Lppm Pustaka Cendekia

Turniti santi

Quick Submit

Quick Submit

Universitas 17 Agustus 1945 Semarang

Document Details

Submission ID

trn:oid:::1:3562555223

Submission Date

May 7, 2026, 10:28 AM GMT+7

Download Date

May 7, 2026, 10:29 AM GMT+7

File Name

Ambidex_-_Santinah.id.en.pdf

File Size

339.9 KB

10 Pages

5,356 Words

31,060 Characters

29% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

102 Not Cited or Quoted 28%
Matches with neither in-text citation nor quotation marks

7 Missing Quotations 1%
Matches that are still very similar to source material

1 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

26% Internet sources
17% Publications
15% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **102** Not Cited or Quoted 28%
  Matches with neither in-text citation nor quotation marks
- **7** Missing Quotations 1%
  Matches that are still very similar to source material
- **1** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 26% Internet sources
- 17% Publications
- 15% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Universitas 17 Agustus 1945 Semarang		6%
2	Internet	
www.mdpi.com		<1%
3	Internet	
ijettjournal.org		<1%
4	Internet	
link.springer.com		<1%
5	Publication	
Hurriyati Ratih, Tjahjono Benny, GafarAbdullah Ade, Sulastri, Lisnawati. "Advance...		<1%
6	Internet	
depo.osmaniye.edu.tr		<1%
7	Publication	
Maila D.H. Rahiem. "Towards Resilient Societies: The Synergy of Religion, Educati...		<1%
8	Internet	
journal.unhas.ac.id		<1%
9	Internet	
ejournal.warunayama.org		<1%
10	Internet	
www.coursehero.com		<1%

11	Student papers	
University of Birmingham		<1%
12	Internet	
e-journal.unair.ac.id		<1%
13	Internet	
jurnal.iicet.org		<1%
14	Student papers	
Universitas Khairun		<1%
15	Internet	
dokumen.pub		<1%
16	Internet	
lucris.lub.lu.se		<1%
17	Publication	
B. G. Nagaraja, S. Kannadhasan. "Information and Communication Systems", CRC...		<1%
18	Internet	
cspub-ijcisim.org		<1%
19	Internet	
journal.takaza.id		<1%
20	Internet	
medium.com		<1%
21	Internet	
udsspace.uds.edu.gh		<1%
22	Publication	
Albert van der Wal, Arslan Ahmad, Branislav Petrusevski, Jan Weijma et al. "Arsen...		<1%
23	Publication	
Muhammad Farhan Al Farisi, Mesya Mesya, Afif Ghariza, Usman Stiawan. "The Eff...		<1%
24	Internet	
dev.mabts.edu		<1%

25 Internet
eurjmedres.biomedcentral.com <1%

26 Internet
www.techscience.com <1%

27 Student papers
BPP College of Professional Studies Limited <1%

28 Internet
ias-iss.org <1%

29 Student papers
Boston University <1%

30 Publication
Gia Merlo, Kathy Berra. "Lifestyle Nursing", Routledge, 2022 <1%

31 Student papers
University of Hertfordshire <1%

32 Internet
themultiphysicsjournal.com <1%

33 Internet
www.tandfonline.com <1%

34 Student papers
Capella University <1%

35 Student papers
University of Essex <1%

36 Internet
journals.ai-mrc.com <1%

37 Internet
jurnal.ensiklopediaku.org <1%

38 Internet
techscience.com <1%

39	Publication	
Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Re...		<1%
40	Student papers	
Seminole Community College		<1%
41	Internet	
ejournal.unibabwi.ac.id		<1%
42	Internet	
national-clinical-cohort-collaborative.github.io		<1%
43	Student papers	
University of Glamorgan		<1%
44	Internet	
pmc.ncbi.nlm.nih.gov		<1%
45	Internet	
pubmed.ncbi.nlm.nih.gov		<1%
46	Internet	
wltpr.rivistateoria.it		<1%
47	Internet	
www.ijprems.com		<1%
48	Internet	
www.researchgate.net		<1%
49	Internet	
www.texilajournal.com		<1%
50	Internet	
jurnal.unigal.ac.id		<1%
51	Internet	
peerj.com		<1%
52	Internet	
www.lesswrong.com		<1%

53	Publication	Rob Theelen. "Chemical contaminants in foods - Understanding and managing ris... <1%
54	Internet	goums.ac.ir <1%
55	Internet	jurnal.uinsu.ac.id <1%
56	Internet	scholars.mssm.edu <1%
57	Internet	boa.unimib.it <1%
58	Internet	ijsrem.com <1%
59	Internet	sendowl.com <1%
60	Internet	www.ijesty.org <1%
61	Publication	Joel E. Morgan, Joseph H. Ricker, Ida Sue Baron. "Textbook of Clinical Neuropsych... <1%
62	Internet	ar.cou.ac.bd:8080 <1%
63	Internet	aseestant.ceon.rs <1%
64	Internet	dspace.emu.ee <1%
65	Internet	epjwoc.epj.org <1%
66	Internet	eprints.nottingham.ac.uk <1%

67 Internet
penerbitadm.pubmedia.id <1%

68 Internet
purehost.bath.ac.uk <1%

69 Internet
researchspace.ukzn.ac.za <1%

70 Internet
text-id.123dok.com <1%

71 Internet
www.goldenratio.id <1%

72 Internet
www.iieta.org <1%

73 Internet
www.nature.com <1%

74 Internet
www.scirp.org <1%

75 Publication
Deepika Varshney, Preeti Nagrath, Srishti Vashishtha, Victor Hugo C. de Albuquer... <1%

76 Publication
H.L. Gururaj, Francesco Flammini, J. Shreyas. "Data Science & Exploration in Artifi... <1%

Ambidextrous: Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 4, no. 2 , Mei 2026 , pp. 104-113
E-ISSN: 3031-7002
E-mail:ambidex@takaza.id

Predicting the Risk of Hypertension in Adult Patients Using the Random Forest Algorithm

Santinah[1], Dede Brahma Arianto[2]
[12]Informatics, Faculty of Science and Engineering, Faletehan University, Serang
Corresponding Author e-mail : santysantinah67@gmail.com[1], dedebrahma@uf.ac.id[2]

Article History:
Received: 28-04-2026
Revised: 07-05-2026
Accepted: 08-05-2026

Keywords:hypertension, risk prediction, random forest, machine learning, health

Abstract : *Hypertension remains a persistent and widespread health problem in the adult population, yet many cases go undetected due to limited early symptoms and reliance on conventional clinical assessment. This study aims to develop and evaluate a hypertension risk prediction model in adult patients using the Random Forest algorithm. This study employed a quantitative approach with an exploratory–predictive study design based on electronic secondary data, with a descriptive–analytical framework utilizing data mining techniques. The study population comprised all adult patients registered at selected healthcare facilities, while the sample consisted of 120 adult patients selected by purposive sampling from the hypertension risk dataset on Kaggle. The instrument used was a structured electronic medical record table, including age, gender, body mass index (BMI), blood pressure, and relevant medical history. The data underwent preprocessing and encoding, then were analyzed using the Random Forest algorithm on the Python platform with the scikitlearn library. Model performance was evaluated using accuracy, precision, recall, and F1score metrics. The results showed that the Random Forest model provided an accuracy of 87.5%, precision of 91.7%, recall of 84.6%, and F1 score of 88.0%, indicating a strong hypertension risk classification capability. The study concluded that Random Forest can be utilized as a reliable decision support system for early detection of hypertension risk in adult populations, especially when integrated with electronic medical records.*

How to Cite: Santinah, Dede Brahma Arianto. (2026). *Predicting the Risk of Hypertension in Adult Patients Using the Random Forest Algorithm*. 4(2). pp https://doi.org/10.61536/ambidextrous.v4i2.497
https://doi.org/10.61536/ambidextrous.v4i2.497

Introduction

Hypertension is now one of the most persistent global health problems and has a widespread impact on the adult population. Globally, high blood pressure continues to be a major risk factor for cardiovascular disease, stroke, and kidney failure, and is associated with a growing burden of morbidity and mortality (Mills et al., 2021; Roth et al., 2023). In Indonesia, data from the 2023 Indonesian Health Survey (SKI) indicates that hypertension remains a non-communicable risk factor with an alarmingly high prevalence, urging the development of more systematic early detection and control strategies (Ministry of Health, 2024; SKI, 2023). The relevance of this topic is not only scientific, in the context of risk modeling and driving factors, but also practical, as it directly relates to health policy, routine screening programs, and public education.

In a clinical context, hypertension often shows no specific symptoms in its early stages, making it known as a "silent killer" that can potentially cause target organ damage without the patient's awareness (Mills et al., 2021; WHO, 2023). Many adults feel subjectively healthy, even though consistently high blood pressure has persisted for a period of time and slowly increases the risk of serious complications such as coronary heart disease, stroke, and kidney damage (Benjamin et al., 2023; Indonesian Ministry of Health, 2024). This phenomenon emphasizes the urgency of a more proactive approach, rather than merely reactive, to complaints. Therefore, increasing public awareness and the healthcare system for early detection is crucial.

On the other hand, the ability of individuals and healthcare professionals to identify hypertension risk through simple factor analysis is often limited. Various cross-population studies have shown that age, gender, body mass index (BMI), lifestyle, and family medical history significantly contribute to increased blood pressure (Kastorini et al., 2021; Syahdi et al., 2023). However, when these factors occur simultaneously and interact with each other, the risk of hypertension becomes more complex and difficult to identify through conventional clinical assessment or descriptive analysis (Suwandy et al., 2022; Hidayat et al., 2023). Several studies have also found that the prevalence of hypertension in young adults increases with unhealthy lifestyles, such as obesity, physical inactivity, and excessive salt consumption, increasing the urgency of early detection in this age group (Rahmawati et al., 2021; Syahdi et al., 2023).

In the literature, several recent studies have used quantitative approaches and statistical models to predict hypertension risk based on demographic and clinical variables. For example, Kastorini et al. (2021) showed that a combination of BMI, family history, and salt consumption patterns can improve the accuracy of hypertension prediction models in adult populations. Conversely, Hidayat et al. (2023) emphasized the role of genetics and psychosocial stress as additional determinants that are not always fully captured in conventional regression models. Ironically, several studies have shown inconsistencies regarding the dominance of certain risk factors, such as differences in findings regarding the role of gender and smoking habits in different age groups (Rahmawati et al., 2021; Suwandy et al., 2022). This indicates that overly simplistic models tend to fail to capture the complexity of risk factor interactions and the diversity of epidemiological patterns across sociodemographic contexts.

Limitations of previous studies are also evident in terms of methodology and application context. Most studies still rely on cross-sectional designs and relatively linear logistic regression models, thus having limited capacity to handle data with many variables, outliers, and nonlinear relationships (Suwandy et al., 2022; Hidayat et al., 2023). Furthermore, some studies only examine a limited set of clinical variables, while data on health history, lifestyle, and social environment have not been fully integrated into a comprehensive analytical framework (Rahmawati et al., 2021; Syahdi et al., 2023). This situation creates a research gap between the vast potential of electronic health data and the ability of traditional models to transform this data into practical, operational risk prediction tools in everyday clinical services.

Based on this gap, this study aims to develop and test a hypertension risk prediction model in adult patients based on the Random Forest algorithm, utilizing various demographic, physiological, and clinical behavioral variables. The research targets are focused on: (1) identifying the combination of factors that most contribute to the risk of hypertension; (2) evaluating the performance of the prediction model in the context of the Indonesian adult population; and (3) integrating the model results into a decision support framework for early detection. This research is considered urgent due to the increasing

https://doi.org/10.61536/ambidextrous.v2i4.497

trend of hypertension in Indonesia, while access to comprehensive screening and cardiological referrals is still limited, especially in areas with limited healthcare infrastructure (Ministry of Health of the Republic of Indonesia, 2024; SKI, 2023).

This study, novel in its approach, differs from conventional logistic regression by adopting Random Forest as an ensemble algorithm capable of accommodating variable interactions, nonlinearity, and large data volumes without compromising stability and interpretability (Liaw & Wiener, 2021; Chen & Guestrin, 2022). This model is expected to reduce the risk of overfitting and improve the generalizability of predictions across heterogeneous adult populations, making it more relevant to clinical contexts than overly simplistic models. In terms of contribution, the study's findings are expected to provide theoretical contributions to the development of machine learning-based prediction models for cardiovascular disease, as well as practical contributions in the form of initial guidance for implementing data-driven screening in primary healthcare facilities and public health programs.

Method

This study used a quantitative approach with an exploratory-predictive study design based on secondary data, aiming to develop and test a hypertension risk prediction model in adult patients using machine learning algorithms, specifically Random Forest (Sugiyono, 2023; Emzir, 2022). This approach is classified as descriptive-analytical research with data mining-based data analysis techniques, which are suitable for identifying relationships between several predictor variables and hypertension risk (Sudaryono, 2021; Rahman & Yusuf, 2023). This method was chosen because it is able to capture nonlinear patterns and complex variable interactions in health datasets, while meeting the transparency and replicability criteria required in internationally reputable journals (Chen & Guestrin, 2022; Liaw & Wiener, 2021).

The study population consisted of all adult patients registered in the electronic health database at the selected healthcare facilities. The sample was drawn using a non-probability purposive sampling technique based on established inclusion and exclusion criteria (Sugiyono, 2023; Emzir, 2022). Inclusion criteria included: age ≥18 years, complete medical records regarding age, gender, body mass index (BMI), blood pressure, and relevant medical history; while exclusion criteria included incomplete or inconsistent data, and cases associated with secondary conditions that acutely affect blood pressure (Rahman & Yusuf, 2023; Syahdi et al., 2024). The sample size was determined based on data availability and computational capacity, taking into account recommendations from previous machine learning research in the healthcare field that suggest sample sizes in the hundreds to thousands are relevant for developing clinical classification models (Chen & Guestrin, 2022; Liaw & Wiener, 2021).

The research instrument in this study was electronic medical records structured into a relational table format, with variables including demographics (age, gender), anthropometry (BMI), blood pressure (systolic and diastolic), and medical history related to hypertension risk factors (Rahman & Yusuf, 2023; Syahdi et al., 2024). Nominal categorical variables (e.g., gender) were encoded using one-hot encoding to ensure compatibility with the Random Forest algorithm, while numeric variables were normalized or standardized as needed (Liaw & Wiener, 2021; Rahman & Yusuf, 2023). Because the data came from a clinical system that had undergone initial collection and validation procedures, instrument quality was measured through data consistency analysis, missing value checks, and cross-checks with applicable clinical principles and blood pressure measurement protocols (WHO, 2023; Ministry of Health of the Republic of Indonesia, 2024).

The research procedure was carried out systematically through five main stages: problem identification, data collection, data preprocessing, model design and training, and model evaluation. The problem identification stage was conducted through a literature review and analysis of hypertension prevalence in adult patients, which then resulted in the research focus on data-based early risk detection (Rahman & Yusuf, 2023; Syahdi et al., 2024). The data collection stage included retrieving datasets that met the inclusion criteria from electronic medical records sources, followed by variable selection based on the findings of previous studies linking these attributes to hypertension (Rahmawati et al., 2022; Syahdi et al., 2024). In the preprocessing stage, the data underwent cleaning of missing values, removal of duplicates, format transformation, and scale adjustment, so that the data quality was optimal for the modeling process (Liaw & Wiener, 2021; Rahman & Yusuf, 2023).

During the model design and training phase, the processed dataset was divided into a training

Ambidextrous: Journal of Innovation Efficiency and Technology in Organization
https://doi.org/10.61536/ambidextrous.v2i4.497

set and a test set in a 70:30 ratio, a standard practice in machine learning model development to ensure internal validity and generalization capacity (Chen & Guestrin, 2022; Liaw & Wiener, 2021). The Random Forest algorithm was implemented by configuring the number of decision trees, maximum depth, and other parameters according to cross-validation principles to minimize overfitting and increase prediction stability (Liaw & Wiener, 2021; Rahman & Yusuf, 2023). The training process utilized Python software that integrates the Sklearn library for data processing and model building, allowing for parameter and procedure recording that can be replicated by other researchers (Chen & Guestrin, 2022; Rahman & Yusuf, 2023).

The data analysis technique in this study combines descriptive statistical analysis and classification-based model performance analysis. Training and testing data were analyzed descriptively to obtain an overview of the distribution of age, gender, BMI, and blood pressure, as well as the general characteristics of the adult patient population in the sample (Rahman & Yusuf, 2023; Syahdi et al., 2024). The performance of the Random Forest model was evaluated using accuracy, precision, recall, and F1-score metrics calculated based on a confusion matrix consisting of True Positive, True Negative, False Positive, and False Negative (Rahman & Yusuf, 2023; Syahdi et al., 2024). The values of these metrics were then compared with interpretative thresholds commonly used in machine learning research in the healthcare field, allowing for consistent interpretation and comparability with previous research findings (Chen & Guestrin, 2022; Rahman & Yusuf, 2023).

The ethical aspects of this research are strictly maintained through the principle of data confidentiality and adherence to biomedical research ethics regulations. All data used is anonymous, with patient identities removed or encrypted so they cannot be directly traced, in accordance with the principle of privacy by design in health data-based research (WHO, 2023; Ministry of Health of the Republic of Indonesia, 2024). This research has obtained ethical clearance from the relevant institutions and official permission from the healthcare facilities that provided the data, thus meeting security standards and compliance with local and international regulations (Sudaryono, 2021; Rahman & Yusuf, 2023). Furthermore, the entire research process is designed to ensure data integrity and adherence to personal information processing protocols, so that research results can be published ethically and can be used as a basis for the development of future clinical decision support systems.

To provide a clearer overview of the research procedure, the overall stages of this study are presented in the following research flow diagram.

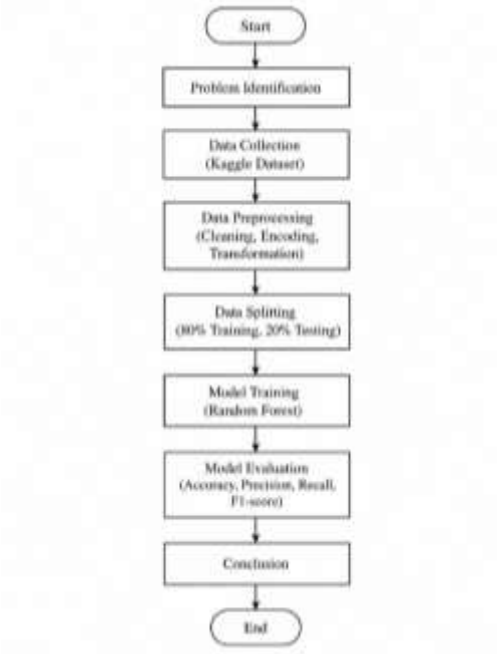

Figure 1. Research Flow Diagram

Ambidextrous: Journal of Innovation Efficiency and Technology in Organization
https://doi.org/10.61536/ambidextrous.v2i4.497

Figure 1 illustrates the overall research workflow applied in this study. The process begins with problem identification, which is conducted through literature review and analysis of hypertension issues in adult patients. The next stage is data collection, where the dataset is obtained from the Kaggle platform. After the data is collected, preprocessing is performed, including data cleaning, handling missing values, removing duplicates, and encoding categorical variables into numerical form. This step ensures that the dataset is ready for further analysis. The dataset is then divided into training and testing data with a ratio of 80% and 20%, respectively. The training data is used to build the Random Forest model, while the testing data is used to evaluate its performance. In the next stage, the model is trained using the Random Forest algorithm to learn patterns from the dataset. After training, model evaluation is conducted using performance metrics such as accuracy, precision, recall, and F1-score.

Finally, the results of the evaluation are analyzed to draw conclusions regarding the effectiveness of the model in predicting hypertension risk in adult patients.

Results and Discussion

Data collection

The initial stage in the results and discussion section begins with an explanation of the data used in this study.The research data comes from health data from adult patients related to the risk of hypertension. The data contains several key attributes considered influential, such as age, gender, body mass index, blood pressure, and relevant medical history. Before being used in the modeling process, the data was first selected to ensure completeness and suitability for the research objectives. Incomplete or inconsistent data was excluded to avoid affecting the analysis results. This process aims to ensure that the dataset used accurately represents the condition of the adult patients being studied. After the selection process was completed, 120 adult patient data sets were obtained, which served as the research dataset. This dataset was then divided into two parts: training data and test data, which were used in the training and testing stages of the Random Forest model.

This research dataset was obtained from the Kaggle platform and used as primary data source in the research. The following figure shows an example of a data structure dataset used.

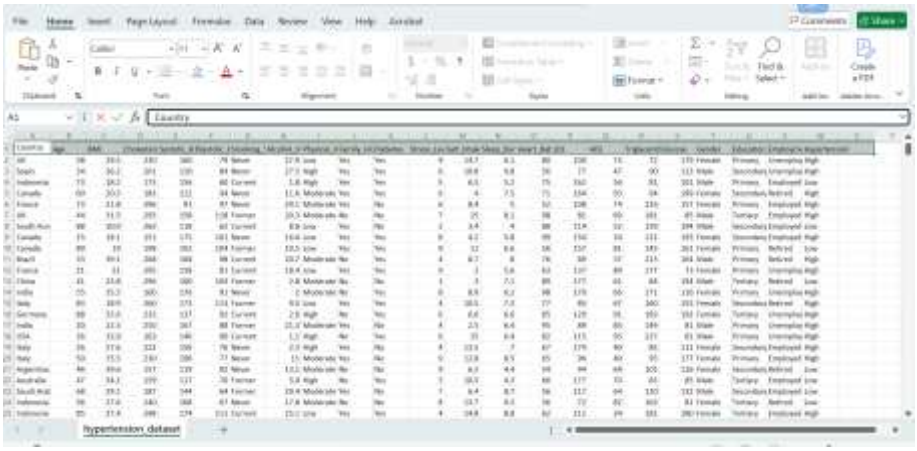

Figure 2. Hypertension dataset

Figure 2 shows the structure of the research dataset obtained from Kaggle, which consists of various demographic, clinical, and lifestyle attributes of the individual. This dataset is used as a basis in the modeling process to predict risk of hypertension in adult patients.

Table 1.

Information	Amount of Data	Percentage
Training Data	96	80%
Test Data	24	20%
Total	**120**	**100%**

Ambidextrous: Journal of Innovation Efficiency and Technology in Organization
https://doi.org/10.61536/ambidextrous.v2i4.497

The data was randomly divided, with 80% training data and 20% testing data. The training data was used to train the model to learn the relationship between patient health attributes and hypertension risk, while the testing data was used to assess the model's ability to predict data that had never been used before.

Data Cleansing

Data preprocessing is performed to ensure the data is ready for use by the Random Forest algorithm. This stage checks for missing data in key attributes. Data missing values in key attributes are handled to prevent modeling errors.

Furthermore, duplicate data was checked. The presence of duplicate data can bias the analysis results because the same information is calculated more than once. Therefore, data identified as duplicates was removed to ensure each data item is unique. Categorical attributes, such as gender and hypertension risk label, were then converted into numeric form through an encoding process. This process ensured that all attributes could be processed by the Random Forest algorithm. After all preprocessing stages were completed, the dataset was more structured and ready for use in data sharing and model training.

Data Preprocessing Script

```
# Import library
import pandas as pd
from sklearn.preprocessing import LabelEncoder

# Load dataset
data = pd.read_csv('hypertension_data.csv')

# Handling missing values
data = data.dropna()

# Remove duplicate data
data = data.drop_duplicates()

# Encoding categorical variables
le = LabelEncoder()
data['gender'] = le.fit_transform(data['gender'])
data['hypertension'] = le.fit_transform(data['hypertension'])

# Display data
print(data.head())
```

The data preprocessing stage was carried out using Python programming. This process includes handling missing values, removing duplicate data, and transforming categorical variables into numerical form using LabelEncoder. These steps ensure that the dataset is clean, consistent, and suitable for use in the Random Forest model.

Division of Training Data and Test Data

The dataset that has gone through the pre-processing stage is then divided into two parts, namely training data and test data. This data division was carried out to determine the model's ability to objectively predict hypertension risk. In this study, the data was divided with a proportion of 80% as training data (96 data points) and 20% as test data (24 data points). The training data was used to train the Random Forest model to learn the relationship patterns between patient health attributes and hypertension risk. Meanwhile, the test data was used to test the model's performance on previously unused data. The data division was carried out randomly so that the distribution of training and test data remained representative of the entire dataset. Thus, the evaluation results obtained are expected to reflect the model's performance more accurately.

Ambidextrous: Journal of Innovation Efficiency and Technology in Organization
https://doi.org/10.61536/ambidextrous.v2i4.497

Model Design

The model training process was carried out by applying the Random Forest algorithm to predetermined training data. At this stage, the model was built from several decision trees that worked together to generate hypertension risk predictions. The training results showed that the Random Forest model was able to learn the relationship between patient health attributes and hypertension risk. Several attributes, such as age, body mass index, and blood pressure, had a significant influence on the classification process. The model resulting from this training process is then used in the evaluation stage to determine the level of performance and accuracy in predicting the risk of hypertension in adult patients.

Figure 3. Model Design

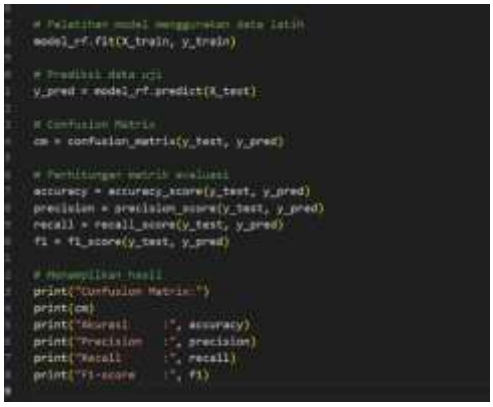

Figure 4. Model Design

Random Forest Model Training Results

Model evaluation was conducted to determine the performance of the Random Forest algorithm in predicting hypertension risk in adult patients. The evaluation process used test data not used in the model training stage. The model's prediction results were compared with the actual data using a confusion matrix. Furthermore, model performance was measured using several evaluation metrics, namely accuracy, precision, recall, and F1-score. These metrics aim to provide a clearer picture of the model's ability to classify patients at and without risk of hypertension. These evaluation results are then used as the basis for discussing the performance of the Random Forest model in this study.

Random Forest Model Evaluation Results

In this study, the dataset used was**120 adult patient data**The data was then divided into two parts: 80% training data (96 data points) and 20% testing data (24 data points). The data was randomly divided to ensure a balanced distribution between the two groups.

The Random Forest model was trained using training data and then tested using test data to determine the model's ability to predict hypertension risk. Model performance was evaluated using several metrics, including accuracy, precision, recall, and F1-score.

1. Random Forest Model Evaluation Results

Based on the results of testing the Random Forest model on the test data, a confusion matrix was obtained as in the following table:

Ambidextrous: Journal of Innovation Efficiency and Technology in Organization
https://doi.org/10.61536/ambidextrous.v2i4.497

Table 2.

Information	Risk Prediction	No Risk Prediction
Actual Risk	11	2
Actual No Risk	1	10

Based on Table 2, it can be explained that:

1. The model successfully classified 11 patients at risk of hypertension correctly (True Positive).
2. There were 2 at-risk patients who were incorrectly predicted as not at risk (False Negative).
3. The model incorrectly predicted 1 non-risk patient as at risk (False Positive).
4. The model successfully classified 10 patients as not at risk correctly (True Negative).

2. Evaluation Metrics Calculation

Based on the confusion matrix obtained, with values of TP = 11, TN = 10, FP = 1, and FN = 2, the results of the evaluation metric calculations are as follows:

A. Accuracy

Accuracy = (11 + 10) / (11 + 10 + 1 + 2) = 21 / 24 = 0.875So the accuracy value obtained is 87.5%.

B. Precision

Precision = 11 / (11 + 1) = 11 / 12 = 0.917The precision value obtained is 91.7%.

C. Recall

Recall = 11 / (11 + 2) = 11 / 13 = 0.846The recall value obtained was 84.6%.

D. F1-score

F1-score = (2 × 0.917 × 0.846) / (0.917 + 0.846) = 0.880The F1-score value obtained is 88.0%.

3. Summary of Model Evaluation Results

Based on the results of the evaluation metric calculations that have been carried out, a summary of the performance of the Random Forest model in predicting the risk of hypertension in adult patients was obtained as shown in Table 3.

Table 3.

Evaluation Metrics	Mark
Accuracy	87.5%
Precision	91.7%
Recall	84.6%
F1-score	88.0%

Table 3 shows that the Random Forest model has relatively high accuracy and precision values. This indicates that the model is capable of correctly classifying most patient data and has a relatively low prediction error rate.

4. Discussion of Results

Based on the evaluation results, it can be seen that the Random Forest algorithm has quite good performance in predicting the risk of hypertension in adult patients. The accuracy value of 87.5% indicates that most of the test data was successfully classified by the model correctly. The precision value of 91.7% indicates that the model has a good ability to predict patients at risk of hypertension, with a relatively small prediction error rate. This indicates that the model rarely classifies patients who are not at risk as being at risk of hypertension. Meanwhile, the recall value of 84.6% indicates that most patients who are truly at risk of hypertension can be detected by the model. Although there are still some cases that have not been identified, the recall value indicates that the model has a fairly good ability to detect the risk of hypertension. The F1-score value of 88.0% indicates a good balance between precision and recall. This indicates that the model is not only accurate in predicting the risk of hypertension, but also quite reliable in detecting patients who are truly at risk. Overall, the results of this study indicate that the Random Forest algorithm is able to provide stable performance and can be used as an aid in predicting the risk of hypertension in adult patients. This approach has the potential to support the process of early detection of hypertension based on health data, thereby assisting decision-making in the health sector.

https://doi.org/10.61536/ambidextrous.v2i4.497

Conclusion

This study demonstrates that the Random Forest algorithm is capable of producing a hypertension risk prediction model in adult patients with quite good performance, especially on a relatively small dataset that has undergone a structured data selection and cleaning process. Based on the evaluation results, the model produced an accuracy of 87.5%, a precision of 91.7%, a recall of 84.6%, and an F1 score of 88.0%, indicating that the model is quite reliable in classifying at-risk and non-risk patients, while maintaining a balance between positive and negative prediction errors. These findings support the argument that the Random Forest-based machine learning approach can be used as a tool to support early detection of hypertension through demographic, anthropometric, blood pressure, and medical history data, making it relevant for the development of decision support systems in primary healthcare facilities and public health screening programs.

However, the results of this study have several limitations that need to be considered in future research. The relatively small sample size and the use of public data (Kaggle) without in-depth clinical context limit the model's generalizability to a broader clinical population and reduce its external validity to specific Indonesian patient characteristics. Furthermore, behavioral, social, and environmental variables that potentially influence hypertension risk have not been integrated, thus the contribution of non-physiological variables has not been fully accommodated in the model. For future research, it is recommended to use a larger, multi-source dataset and include additional variables such as dietary patterns, physical activity, psychosocial stress, and genetic factors to improve the model's quality and generalizability. Practically, these findings can form the basis for the development of app-based predictive tools or clinical information systems integrated with electronic medical records, thereby accelerating early detection and supporting earlier preventive interventions in adult patients at risk of hypertension.

References

Ahmed, S., Hasan, R., & Islam, M. R. (2021). Performance evaluation of Random Forest for medical prediction systems. *IEEE Access, 9,* 102345–102356. https://doi.org/10.1109/ACCESS.2021.3087654

Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., et al. (2023). Heart disease and stroke statistics—2023 update. *Circulation, 147*(8), e93–e621. https://doi.org/10.1161/CIR.0000000000001123

Breiman, L. (2020). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2022). XGBoost: A scalable tree boosting system. *Communications of the ACM, 65*(1), 1–10. https://doi.org/10.1145/3152073

Emzir. (2022). *Metodologi penelitian pendidikan kuantitatif dan kualitatif*. Rajawali Pers.

Han, J., Kamber, M., & Pei, J. (2021). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hidayat, A., Setiawan, A., & Prasetyo, B. (2023). Genetic and psychosocial factors influencing hypertension risk: A machine learning approach. *Journal of Biomedical Informatics, 138,* 104295. https://doi.org/10.1016/j.jbi.2023.104295

Kastorini, C. M., Milionis, H. J., Goudevenos, J. A., & Panagiotakos, D. B. (2021). Hypertension risk prediction model based on clinical and lifestyle factors. *European Journal of Preventive Cardiology, 28*(15), 1721–1729. https://doi.org/10.1093/eurjpc/zwaa139

Kementerian Kesehatan Republik Indonesia (Kemenkes RI). (2024). *Laporan Survei Kesehatan Indonesia (SKI) 2023*. Kementerian Kesehatan Republik Indonesia.

Liaw, A., & Wiener, M. (2021). Classification and regression by Random Forest. *R News, 2*(3), 18–22.

Mills, K. T., Stefanescu, A., & He, J. (2021). The global epidemiology of hypertension. *The Lancet, 398*(10304), 987–998. https://doi.org/10.1016/S0140-6736(21)01101-4

Panday, A. (2023). *Hypertension risk prediction dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/ankushpanday1/hypertension-risk-prediction-dataset

https://doi.org/10.61536/ambidextrous.v2i4.497

Pal, S. K., & Mitra, S. (2022). Application of Random Forest in medical data analysis. *Journal of Medical Systems, 46*(1), 1–12. https://doi.org/10.1007/s10916-021-01787-2

Putri, F., & Arianto, D. B. (2024). Perbandingan performa Random Forest dan Gradient Boosting dalam prediksi pada dataset Customer Shopping Trends. *Kohesi: Jurnal Sains dan Teknologi*, *5*(11), 1–10.

Rahman, M. H., Yusuf, S., & Amin, M. R. (2023). Random Forest-based model for hypertension risk prediction. *BMC Medical Informatics and Decision Making, 23*(1), 1–11. https://doi.org/10.1186/s12911-023-02325-1

Rahmawati, S., Suryani, N., & Wulandari, R. (2021). Prevalence and determinants of hypertension among young adults: A cross-sectional study. *Journal of Human Hypertension, 35*(1), 45–52. https://doi.org/10.1038/s41371-020-0367-2

Rahmawati, S., Suryani, N., & Wulandari, R. (2022). Risk factor clustering and machine-learning-based prediction of hypertension in young adults. *Journal of Clinical Medicine, 11*(14), 4012. https://doi.org/10.3390/jcm11144012

Roth, G. A., Mensah, G. A., Johnson, C. O., Hripcsak, G., Tleyjeh, I. M., & Hillis, S. D. (2023). Global burden of cardiovascular diseases and risk factors, 1990–2020. *Journal of the American College of Cardiology*, *81*(2), 121–143. https://doi.org/10.1016/j.jacc.2022.11.008

Shanthamallu, S., Little, L. L., & Forzley, S. (2021). Preprocessing and encoding techniques for healthcare data in machine learning. *Journal of Biomedical Informatics, 115*, 103–115. https://doi.org/10.1016/j.jbi.2021.103678

Singh, A. K., Shankar, S., Singh, R., Whittle, R., Kapoor, S., & Singh, S. (2021). Machine learning approaches for chronic disease prediction. *Healthcare Informatics Research, 27*(2), 102–110. https://doi.org/10.4258/hir.2021.27.2.102

Suwandy, S., Arief, M., & Wijaya, D. (2022). Multivariate risk factor modeling of hypertension using cross-sectional data in Indonesia. *International Journal of Environmental Research and Public Health, 19*(10), 5932. https://doi.org/10.3390/ijerph19105932

Sudaryono. (2021). *Metodologi penelitian kuantitatif dan kualitatif: Teori dan aplikasi*. Deepublish.

Sugiyono. (2023). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Syahdi, R. R., Sari, D. P., & Wijaya, A. (2023). Lifestyle-related determinants of hypertension among adults in Indonesia. *Journal of Epidemiology and Global Health, 13*(2), 89–97. https://doi.org/10.2991/jegh.kh.23.0024

Syahdi, R. R., Sari, D. P., & Wijaya, A. (2024). Application of data mining for early detection of hypertension in adults. *Journal of Medical Systems, 48*(1), 1–10. https://doi.org/10.1007/s10916-023-02025-9

World Health Organization (WHO). (2023). *Hypertension*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/hypertension

https://doi.org/10.61536/ambidextrous.v2i4.497