International Journal of Health, Economics, and Social Sciences (IJHESS)
Vol. 8, No. 3, 2026, pp. 2365~2370
DOI: 10.56338/ijhess.v8i3.9816
Website: https://jurnal.unismuhpalu.ac.id/index.php/IJHESS

Analysis of Risk Factors for Hypertension in Urban Adolescents Using the Rapidminer Data Mining Method with Decision Tree C4.5 Operator

Nur'ain Manasai[1], Laksmyn Kadir[2], Ramly Abudi[3]
[1,2,3]Jurusan Kesehatan Masyarakat, Fakultas Olahraga dan Kesehatan, Universitas Negeri Gorontalo

***Corresponding Author:** nurainmanasai@gmai.com

Article Info

Article history:
Received 26 Apr, 2026
Revised 17 May, 2026
Accepted 25 Jun, 2026

Keywords:
Adolescent Hypertension, Risk Factors, Data Mining, RapidMiner, Decision Tree C4.5

ABSTRACT

Hypertension in adolescents is an increasing health problem due to lifestyle changes. This study aims to identify risk factors for hypertension in urban adolescents using the data mining method with the Decision Tree C4.5 algorithm on RapidMiner. This quantitative research involved 130 adolescents aged 15–19 years in the working area of the Dumbo Raya Health Center, Gorontalo City. The variables analyzed included genetic factors, obesity, stress, smoking, excessive salt consumption, coffee consumption, sleep duration, and hypertension status. The data was processed through the preprocessing stage, the formation of the C4.5 Decision Tree model, and evaluation using a confusion matrix with an 80:20 training and testing data division. The results showed that sleep duration was the most dominant factor, followed by coffee consumption, genetic factors, obesity, excessive salt consumption, smoking, and stress. The model produces an accuracy of 88.46% and is able to form easy-to-understand classification rules. The Decision Tree C4.5 algorithm is effectively used to support early detection of hypertension in adolescents.

INTRODUCTION

Hypertension is one of the public health problems that continues to increase in various countries, including Indonesia. This disease is the main risk factor for cardiovascular diseases such as coronary heart disease, heart failure, stroke, and chronic kidney failure. The World Health Organization (WHO) reports that about 1.28 billion people in the world over the age of 18 live with hypertension and most are in low- and middle-income countries. Hypertension also contributes to about 45% of deaths from heart disease and 51% of deaths from stroke (WHO, 2023).

Hypertension is no longer only found in the adult and elderly age groups, but has begun to occur more in adolescent groups. The results of Riskesdas in 2018 show that the prevalence of hypertension in the Indonesian population is 34.11%, with around 20.1% occurring in the age group of 18–24 years (Jayanti et al., 2022).

In Gorontalo Province, hypertension is still one of the non-communicable diseases with a fairly high number of cases. Data from the Gorontalo City Health Office shows that the number of hypertension cases has increased from 4,408 cases in 2020 to 11,537 cases in 2024. Based on the 2024 report of the Dumbo Raya Health Center, there are 2,594 people with hypertension and as many as 130 cases occur in adolescents aged 15-19 years. The data shows that hypertension in adolescents has become a health problem that needs serious attention.

Hypertension is a multifactorial disease that is influenced by both non-modifiable and modifiable factors. Factors that cannot be modified include genetics or a family history of hypertension, while factors that can be modified include obesity, stress, smoking, excessive salt consumption, coffee consumption, and sleep duration (Sabillah & Aidha, 2023). In adolescents, changes in modern lifestyles such as staying up late, excessive use of social media, unhealthy diet, and consumption of caffeinated beverages are factors that have the potential to increase blood pressure (Mardianti et al., 2022).

The development of information technology allows the use of data mining in the health sector to identify patterns and relationships between variables that are difficult to find through conventional analysis. Data mining is the process of extracting information or hidden patterns from a large amount of data so that it can be used to support decision-making. One of the algorithms that is often used in classification is the Decision Tree because it is able to generate an easy-to-understand model in the form of a Decision tree (Marpaung et al., 2024).
Decision Tree works by dividing data based on the attributes that have the greatest influence on the target variable, resulting in decision rules that can be used for classification and prediction processes. The advantage of this method is its ability to automatically identify dominant factors and generate visualizations that are easy for users to understand (Aditya et al., 2024).
Based on these problems, this study was conducted to analyze the risk factors for hypertension in urban adolescents using a data mining approach with the Decision Tree algorithm in the RapidMiner application. The results of the study are expected to provide information on the factors that play the most role in the incidence of hypertension in adolescents and support efforts to detect and prevent hypertension from a young age.

RESEARCH METHODS

This research is a quantitative research with a data mining approach using the Decision Tree C4.5 algorithm in the RapidMiner application. The data mining approach was used to identify patterns and factors that play a role in the incidence of hypertension in urban adolescents based on available data.
The research was carried out in the working area of the Dumbo Raya Health Center, Gorontalo City. The population in this study is all adolescents aged 15–19 years who will be diagnosed with hypertension in 2024. The research sample amounted to 130 adolescents obtained based on the available secondary data.
The variables analyzed consisted of genetic factors, obesity, stress level, smoking status, excessive salt consumption, coffee consumption, sleep duration, and hypertension status as target variables. The data used is secondary data that has gone through a process of selection and data cleaning before being analyzed.
The data mining stages are carried out following the Knowledge Discovery in Database (KDD) process, namely:
Data Selection, which is the selection of variables that are relevant to the research objectives.
Data Cleaning, which is removing incomplete data, duplication, and input errors.
Data Transformation, which is converting data into a categorical format that suits the needs of the Decision Tree algorithm.
Data Mining, which is the formation of a classification model using the Decision Tree operator in the RapidMiner application.
Evaluation, which is the evaluation of the model using the Confusion Matrix with the Split Validation method of 80% training data and 20% testing data.
The results of the analysis are displayed in the form of a decision tree, model accuracy value, class precision, class recall, and decision rules (rules) generated from the classification model.

RESULTS AND DISCUSSION

Respondent Characteristics

Table 1 Distribution of Respondents by Gender

Gender	Frequency (f)	Percentage (%)
Male	68	47,69
Women	62	52,31
Total	**130**	**100**

Based on Table 1, it is known that out of 130 respondents, most of the respondents were female as many as 68 respondents (52.31%), while the respondents were male as many as 62 respondents (47.69%).

Table 2. Distribution of Respondents by Sleep Duration

Sleep Duration	n	%
Good	38	29,2
Bad	92	70,8
Quantity	130	100

Based on Table 2, it is known that out of 130 respondents, most of them had a poor sleep duration, namely 92 respondents (70.8%), while respondents who had a good sleep duration were 38 respondents (29.2%).

Table 3 Distribution of Respondents Based on Coffee Consumption

Coffee Consumption	n	%
Sometimes	54	41,5
Frequent	55	42,3
Never	21	16,2

Based on Table 3, it is known that out of 130 respondents, most of them have the habit of consuming coffee with frequent frequencies as many as 55 respondents (42.3%), followed by respondents who consume coffee occasionally as many as 54 respondents (41.5%), and respondents who have never consumed coffee as many as 21 respondents (16.2%).

Table 4. Distribution of Respondents by Genetics

Geneticsn %

Genetics	n	%
No	84	64,6
Ya	46	35,4
Total	130	100

Based on Table 4, it is known that of the 130 respondents, most of them had a genetic history, namely 75 respondents (57.7%), while respondents who did not have a genetic history of hypertension were 55 respondents (42.3%).

Table 5 Distribution of Respondents by Obesity

Obesity	n	%
No	119	91,5
Ya	11	8,5
Total	130	100

Based on Table 5, it is known that of the 130 respondents, most of them were not obese, namely 119 respondents (91.5%), while 11 respondents (8.5%) were obese.

Table 6. Distribution of respondents based on salt consumption

Consume More Salt	n	%
No	52	40,0
Ya	78	60,0
Total	130	100

Based on Table 6, it is known that out of 130 respondents, most of them have more salt consumption habits, namely 78 respondents (60.0%), while respondents who do not have more salt consumption habits are 52 respondents (40.0%).

Table 7 Distribution of Respondents by Smoking

Smoking	n	%
No	84	64,6
Ya	46	35,4
Total	130	100

Based on Table 7, it is known that out of 130 respondents, most of them did not have a smoking habit, namely 84 respondents (64.6%), while respondents who had a smoking habit were 46 respondents (35.4%).

Table 8 Distribution of Respondents by Stress

Stress	n	%
Lightweight	116	89,2
Medium	12	9,2
Weight	2	2,5
Total	130	100

Based on Table 8, it is known that out of 130 respondents, most of them had a mild stress level of 116 respondents (89.2%), followed by moderate stress as many as 12 respondents (9.2%), and severe stress as many as 2 respondents (1.5%)

Table 9 Distribution of Respondents by Hypertension

Hypertension	n	%
Prehypertension	60	46,2
Lightweight	70	53,8
Total	130	100

Based on Table 9, it is known that out of 130 respondents, most of them were in the category of mild hypertension, namely 70 respondents (53.8%), while respondents in the prehypertension category were 60 respondents (46.2%).

Grafik Decision Tree

Modeling using Decision Tree C4.5 allocrythm in the RapidMiner application. With a data sharing of 80% for training data and 20% for testing. The results of the Decision tree are as follows:

Model Evaluation Results

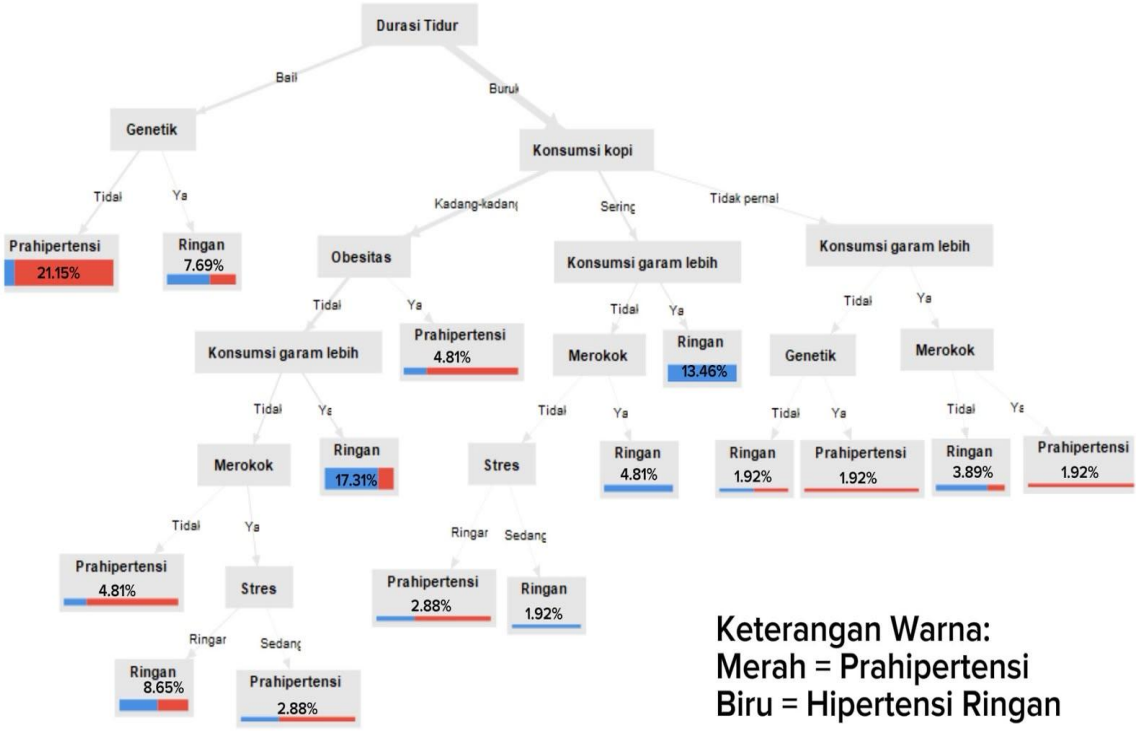

Table 10 Confusion Matrix Results

Actual Classification/ Model Prediction	True Lightweight	True Prehypertension	Class Precision
Lightweight Prediction	14	3	82.35%
Prehypertension Prediction	0	9	100.00%
Class Recall	100.00%	75.00%	**Model Accuracy** 88,46%

Based on table 10, an accuracy value of 88.46% was obtained. The recall value for the mild hypertension class is 100% which means that all mild hypertension data is correctly classified. The recall value for the pre-hypertension class was 75%, which means that from all pertention data, as many as 75% were successfully recognized by the

model and 25% were not recognized. The precision value for the mild hypertension class was 82.35% and for the hypertension class was 100%.

DISCUSSION

Dominant Factors in the Incidence of Hypertension in Adolescents

The modeling results using the Decision Tree C4.5 algorithm show that the sleep duration variable becomes the root node of the decision tree. This shows that sleep duration is the most dominant factor in the classification of hypertension in urban adolescents. Based on Table 2, it is known that out of 130 respondents, there were 92 respondents (70.8%) who had a poor sleep duration and 38 respondents (29.2%) who had a good sleep duration. This condition is supported by research by Oematan and Oematan which shows that the majority of adolescents experience lack of sleep due to the use of gadgets, watching television, and other activities at night (Oematan & Oematan, 2021).

In the first branch of the Decision Tree, it was seen that adolescents with good sleep duration were subsequently influenced by genetic factors. Adolescents who do not have a genetic history of hypertension tend to be classified as prehypertension by 21.15%, while adolescents who have a genetic history of hypertension tend to be classified as mild hypertension by 7.69%. These results suggest that hereditary factors still play a role in increasing blood pressure in adolescents despite having a good sleep pattern.

In the group of adolescents with poor sleep duration, the next variable that appeared was coffee consumption. Based on Table 3, it is known that 55 respondents (42.3%) consume coffee frequently, 54 respondents (41.5%) consume coffee occasionally, and 21 respondents (16.2%) never consume coffee. The emergence of the coffee consumption variable at the second branching level shows that coffee consumption is an important factor after sleep duration in determining hypertension status in adolescents. The caffeine content in coffee can increase blood pressure through stimulation of the sympathetic nervous system. Caffeine works by binding to adenosine receptors thus causing narrowing of blood vessels and increased peripheral resistance which ultimately increases blood pressure. This mechanism explains why coffee consumption is one of the factors that appear in the decision tree (Safitri & Muwakhidah, 2024).

In the group of adolescents who consumed coffee occasionally, the next factor that played a role was obesity. Based on Table 5, it is known that there are 11 respondents (8.5%) who are obese and 119 respondents (91.5%) who are not obese. The results showed that adolescents who were obese tended to be classified as prehypertensive by 4.81%, while in adolescents who were not obese, the classification of hypertension was still influenced by excessive salt consumption and smoking behavior. Various bioactive substances are released by fat tissue, such as hormones, adioxins, and proinflammatory cytokines, which together increase sodium retention, vasoconstriction, as well as peripheral resistance. This combination of processes makes obesity one of the important risk factors in the development of hypertension (Badriyah, 2024).

The variable of excessive salt consumption appeared in several branches of the Decision Tree which showed that this factor contributed to the incidence of hypertension. Excessive salt consumption can lead to fluid retention, increase plasma volume, and ultimately raise blood pressure (Rahmadayanti, 2023). Based on Table 46, it is known that 78 respondents (60.0%) have excessive salt consumption habits and 52 respondents (40.0%) do not have such habits. In the group that consumed coffee occasionally and was not obese, respondents with excess salt consumption were more likely to have mild hypertension by 17.31%, while those who did not consume excess salt were more often classified as prehypertension.

The smoking variable was also found in several branches of the decision tree. Based on Table 7, it is known that there are 46 respondents (35.4%) who smoke and 84 respondents (64.4%) who do not smoke. In certain groups, adolescents who smoke tended to be classified as mild hypertension by 8.65%, while in other groups smoking behavior was still influenced by the level of stress in determining the classification of hypertension. These results are in line with the findings Diana (2023) which showed a significant association between smoking habits, high salt intake, and increased blood pressure in adolescents living in urban areas.

In addition, stress variables appear at the final branching of the Decision Tree. Based on Table 8, it is known that respondents with moderate stress amounted to 12 people (9.2%), 116 people with mild stress (89.2%), and severe stress 2 people (2.5%). Even in the final branch, stress still contributes to the status of hypertension. The results showed that adolescents with mild stress tended to be classified as mild hypertension by 8.65%, while moderate stress was more classified as prehypertension by 2.88%. Under stressful conditions, the production of catecholamines tends to increase, which further stimulates increased secretion of renin, angiotensin, and aldosterone. The increase in these hormones directly contributes to the occurrence of an increase in blood pressure (Ridho et al., 2021).

Genetic factors also reappeared in adolescent branches who never consumed coffee. Based on Table 4, it is known that 46 respondents (35.4%) have a family history of hypertension and 84 respondents (64.6%) do not have a family history of hypertension. In this group, adolescents with a genetic history tended to be classified as prehypertensive by 1.92%, while those without a genetic history were more classified as mild hypertension by 1.92%. This is in line with research by Setiani & Wulandari (2023) states that individuals who have a family history of hypertension have a greater chance of experiencing hypertension than individuals without a family history of hypertension. Genetic

factors are known to play a dominant role through monogenic and polygenic mechanisms and contribute more than 95% to a person's susceptibility to hypertension.
Overall, the structure of the Decision Tree shows that the order of factors most influential on the classification of hypertension in adolescents is sleep duration, coffee consumption, genetics, obesity, excessive salt consumption, smoking, and stress. Factors that appear at the initial branching level show a greater influence on model formation than factors that appear at subsequent branching. These findings indicate that lifestyle factors have a more dominant contribution than hereditary factors in the incidence of hypertension in urban adolescents.

Evaluate Decision Tree Model Performance
Based on the results of the evaluation using the Confusion Matrix, the Decision Tree model produces an accuracy rate of 88.46%. These values indicate that the model is able to classify hypertension status with a high degree of accuracy. Of the total 26 test data, 23 data were correctly classified, while 3 data experienced misclassification.
The precision value for the mild hypertension category was 82.35%, while the prehypertension category reached 100%. Meanwhile, the recall value for the mild hypertension category is 100% and the prehypertension category is 75%. The results show that the model has an excellent ability to recognize cases of mild hypertension and is quite good at identifying cases of prehypertension.
The high level of accuracy shows that the Decision Tree C4.5 algorithm is able to produce a good classification model in identifying risk factors for hypertension in adolescents. In addition to providing high classification performance, this model also produces easy-to-understand decision rules that can be used as a basis for determining the priorities of hypertension prevention programs in adolescents.

CONCLUSION
Based on the results of the analysis using the Decision Tree C4.5 algorithm on the RapidMiner application, a hypertension classification model in adolescents was obtained with an accuracy rate of 88.46%, which shows that the model has a good ability to identify hypertension risk factors. The modeling results showed that sleep duration was the most dominant factor in the classification of hypertension, followed by coffee consumption, genetics, obesity, excessive salt consumption, smoking, and stress. In addition, the results of the model evaluation resulted in a precision value of 82.35% for mild hypertension and 100% for prehypertension, as well as a recall value of 100% for mild hypertension and 75% for prehypertension. Thus, the data mining method using the Decision Tree C4.5 algorithm can be used to identify risk factors for hypertension in adolescents and support efforts to prevent and detect hypertension early.

BIBLIOGRAPHY
Aditya, M. F. R., Azizah, N. L., & Uce, I. (2024). Prediction of Hypertension Disease Using Decision Tree and Random Forest methods. Scientific Journal of Computing, 23.
Diana, T. S. (2023). The Effect of Lifestyle on Hypertension in Adolescents in the City of Surabaya. Indonesian Journal of Public Health, 18(2), 145–153.
Jayanti, A., Mulyati, D., & Atika, S. (2022). Treatment of hypertension in late adolescents: a case study. Case Study. JIM Fkep, 1(1), 160–168.
Lulu'ul Badriyah, R. I. R. P. (2024). Ghidza: Journal of Nutrition and Health Relationship between Obesity and Hypertension Incidence and Association of Obesity with Hypertension and Hyperglycemia in Indonesia. Ghidza: Journal of Nutrition and Health, 8(1), 33–38.
Mardianti, F., Rachmawati, D., & Suprajitno. (2022). Risk Factors for Hypertension in Adolescents. Hesti Wira Sakti Health Journal, 10. https://doi.org/10.47794/jkhws
Marpaung Hamid, P., Darwis, M., M, K., & Pardamean, M. (2024). RapidMiner: Data Mining Learning Solutions (R. Huda (ed.)). CV. Alfa Pustaka.
Oematan, G., & Oematan, G. (2021). Sleep Duration and Sedentary Activity as Risk Factors for Obese Hypertension in Adolescents. Indonesian Nutrition Sciences, 04, 147–156.
Organization, W. H. (2023). Global Report on Hypertension 2023. WHO. https://www.who.int/publications/i/item/9789240082937
Rahmadayanti, F. (2023). Clustering of hypertension data based on lifestyle using the K-Means method. Journal of Data Science and Health, 2(3), 77–85.
Ridho, M., Frethernety, A., & Widodo, T. (2021). Literature review of the relationship between stress and the incidence of hypertension. IX(2), 1366–1371.
Sabillah, V. S., & Aidha, Z. (2023). Analysis of Hypertension Risk Factors in Adolescents at Langsa Lama Health Center. Health Information Research Journal, 15. https://myjurnal.poltekkes-kdi.ac.id/index.php/hijp
Safitri, A., & Muwakhidah. (2024). The Relationship between Coffee Consumption and the Incidence of Hypertension in the Working Area of the Polanharjo Klaten Health Center. 1–18.
Setiani, R., & Wulandari, S. A. (2023). Relationship between Genetic Factors and Hypertension : Scoping Review. 5(1), 60–66.