p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 6 June 2024 http://jist.publikasiindonesia.id/

Comparison of K Nearest Neighbor Algorithm with Apriori Algorithm to Analyze Lifestyle Patterns in Hypertensive Patients

Steven[1*], Ricky Wijaya[2], Helbert Yawin[3], Saut Dohot Siregar[4]
Universitas Prima Indonesia, Indonesia
Email: stevenkho159@gmail.com[1*], rickyhwang123@gmail.com[2],
helbertyao@gmail.com[3], sautdohotsiregar@gmail.com[4]

*Correspondence

Keywords: k nearest neighbor algorithm, apriori algorithm, lifestyle patterns, hypertensive patients

ABSTRACT

Hypertension is one of the most influential cardiovascular diseases that can lead to organ disorders such as heart dysfunction or stroke and hypertension is often discovered by chance. This disease can interfere with the work of other organs if left untreated, especially the heart and kidneys. Not paying attention to diet, exercise, stress, smoking, and drinking alcohol can all be causes of increased risk of hypertension. To predict people with hypertension and find out the comparison of behavior and lifestyle patterns with hypertension patients using a priori algorithm in the case study of Sei Semayang Health Center. So the results of rapidminer use the apriori algorithm to analyze the Comparison of K the nearest Neighbor Algorithm with the apriori Algorithm to Analyze Lifestyle Patterns in Hypertensive Patients the results obtained is U1 which means there are people with hypertension aged 25-38 years who have more hypertension and the results are H2 which means that people have but do not control to the doctor with a pattern style such as consuming alcohol, Smoking, and lack of exercise, sugar consumption, consumption of saturated fat and foods that contain a lot of salt and rarely consume vegetables or fruits and foods containing MSG then more and more people who have hypertension with an unhealthy lifestyle.

Introduction

Hypertension is one of the most influential risks of cardiovascular diseases that can lead to organ disorders such as heart dysfunction or stroke and hypertension is often discovered by chance. Hypertension is a condition when blood pressure rises drastically to above-normal limits and forces the intensity of work on the heart to pump blood in order to provide the body's oxygen needs (Agus et al., 2021). This disease can interfere with the work of other organs if left untreated, especially the heart and kidneys. Not

paying attention to diet, exercise, stress, smoking, and drinking alcohol can all be causes of increased risk of hypertension (Lasmadasari et al., 2021).

Based on data collected by researchers at the Sei Semayang Health Center from January to March of 2023, out of 1100 patients, there were 760 people with hypertension. The results of interviews with patients with high blood pressure at Sei Semayang Health Center found that they often ate fast food with added salt, sodium, and fatty foods; they also only did physical activities such as work and never exercised regularly, and 4 of them had a habit of smoking by smoking more than 10 cigarettes a day. According to the information from the regional health center of the Sei Semayang Health Center, the number of sick people at the Sei Semayang Health Center in 2021 was 1100 cases and most of the cases were hypertension.

Based on data on the age of early adulthood (early adult), namely the age of 21-40 years old, with male and female genders in total, and information about the highest incidence of hypertension from several villages in Sei Semayang Health Center, the exact cause of blood pressure is not yet clearly known (Awalullaili, Ispriyanti, & Widiharih, 2023). Experts reveal the cause of high blood pressure, namely environmental factors such as behavior or lifestyle in the form of obesity, lack of activity, and foods high in salt (Khairani, Kamil, & Tahlil, 2020).

The Apriori Algorithm is used in data mining for the itemset of transactional databases. To establish the mining rule association of a database exchange, it takes time to carry out the frequent items process, which is done to find the smallest support value and the smallest confidence value (Melviani, Aryzki, Rahman, Putri, & Riadi, 2022). Apriori uses an iterative approach that uses k-itemset to explore (k+1)-itemset, (k+1)-itemset candidates are obtained from combining two itemsets in the k region. (k+1)-item candidates that have frequencies in subsets that rarely appear or below the threshold will be separated and excluded to determine association rules and one method can be used to classify regional poverty depth levels in the East Java Province using the K-Nearest Neighbor (KNN) Algorithm (Rahmadayanti, Anggraini, & Susanti, 2023). KNN is a method of classification based on objects "close" to each other that will have similar characteristics; That is, if an object is known for its characteristics, then the object although different can also be predicted based on its closest (Matondang, Mayanda, & Nurul, 2019). K-Nearest Neighbor has the advantage that it can process large training data efficiently has robust noise training and can provide accurate data results. The operating principle of the K-Nearest Neighbor Algorithm is to analyze the shortest distance between the data you want to evaluate and the closest (k) to the training data based on the supervised teaching method. The data is then grouped into classes based on type (k) majority.

In the previous study, the identification of patterns in the symptoms of hypertension utilizing the Apriori Algorithm based on a case study at the Rafina Medical Center Clinic which in this study used Orange Software, obtained 3% and a minimum confidence value of 70%, which resulted in 2 association rules (Nurzanah, Alam, & Hermanto, 2022). The application of the Apriori Algorithm to patients with

hypertension was carried out by distributing questionnaires at the health center which found 6 patterns/rules with lift ratios ≥ 1 from testing 300 data on patients with hypertension using a support value of 30% and confidence 85%. There is also a previous research on the Implementation of Apriori Algorithm to Find Frequent items in Shopping Cart using a sample of 100 transactions from point of sales database with Apriori Algorithm resulting in research results in the form of 98.9% of people with hypertension will feel symptoms of headache with a lift ratio value of 2.41; Based on these results, the rules of the association can be said to be valid (Azwanti & Elisa, 2019).

Based on the analysis above, the purpose of the researcher is to conduct a study entitled "Comparison of K-Nearest Neighbor Algorithm with Apriori Algorithm to Analyze Lifestyle Patterns in Hypertensive Patients". Such studies can help gain patterns in behavior and lifestyle in hypertensive patients.

This study aims to analyze cases found in Sei Semayang Health Center in analyzing behavior and lifestyle patterns against sufferers of hypertension using the Apriori Algorithm in the case study of Sei Semayang Health Center and comparing the method with the K-Nearest Neighbor Algorithm to find the most effective method in cases of hypertension that can prevent more hypertension in Sei Semayang people with a better lifestyle pattern.

Research Methods

Types of Research

This study uses quantitative methods with clear data sources. The object of this study is data on hypertension patients at the Sei Semayang Health Center. The object of study will be assessed objectively based on datasets using the K-Nearest Neighbor Algorithm and Apriori Algorithm to analyze lifestyle patterns in hypertensive patients.

Object of Research

The object of research in this study is the Sei Semayang Health Center located on Jl. Medan Krio, Sunggal, Deli Serdang, North Sumatra 20351, with the online website link of https://dinkes.deliserdangkab.go.id/halaman/puskesmas-sei-semayang.html

Method Data Collection

Observation

Observe research subjects to determine effectiveness, development, impact, etc. In this regard, observation can be done in various ways.

Questionnaire

Data was collected by asking respondents a series of questions or written statements. Questions can be open or close ended.

Interview

Interviews are conducted in order to obtain correct information from reliable sources.

Document

A written or printed letter that can be used as evidence to assist in the completion of the research report.

Literature Study

Data is collected by reviewing literature, documents, books, and reports related to the problem to be solved.

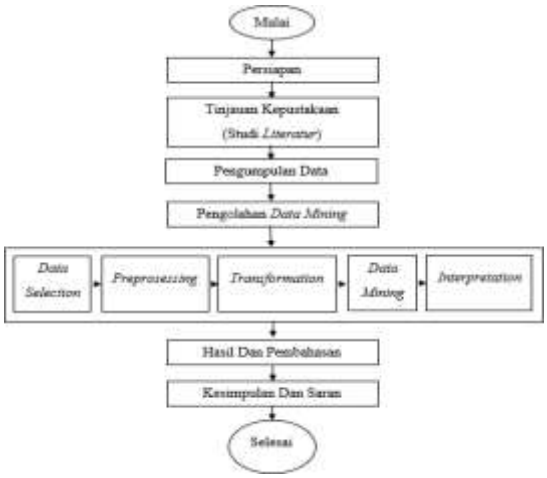

Figure 1 Stages of Research

Stages of Research

The research phase in Figure 1 describes the process to be carried out from the research and the research as a whole. From Figure 1 it can be seen that the stages to be carried out are as follows:

1. Preparation: This stage is the stage of processing research data at the Sei Semayang Health Center. Define and create a research plan.
2. Literature review: conducted by reviewing and examining literature related to the comparison of the K-Nearest Neighbor Algorithm with the Apriori Algorithm to analyze lifestyle patterns in patients with hypertension.
3. Comparison of the K-Nearest Neighbor Algorithm with the Apriori Algorithm to analyze lifestyle patterns in people with hypertension.
4. Data collection: conducted through accounting, observation, and documentary interviews.
5. Mining data processing: The collected information is processed through the results of questions about the analysis of hypertension patients using the Apriori Algorithm in the case study of Sei Semayang Health Center.
6. Results and discussion: Describe the results of the data mining process carried out using the comparison method between Apriori and K-Nearest Neighbor
7. Conclusions research and make proposals for the further development of the company.

Method Algorithm *K-Nearest Neighbor*

The K-Nearest Neighbor method is an algorithm of learning outcomes and monitored queries of new cases classified under part of a large catalog of the K-Nearest Neighbor Algorithm. Proximity is determined by distance metrics. Euclidean distance

or Euclidean metrics is the distance of the straight "normal" line between two points in Euclidean space (Purwono, Dewi, Wibisono, & Dewa, 2022).

$$D_{x,y} = \sqrt{\sum_{t=1}^{n} (x_i - y_i)2}$$

Explanation:

D = Proximity distance of o

x = Training data 1

y = Testing data o

n = Number of individual attributes between 1 to n

Results and Discussion

Dataset

This study uses datasets from data on hypertensive patients in the case study of Sei Semayang Health Center from January to March 2023 which was proposed to predict the analysis of behavior patterns.

Table 1
Data on hypertension patients in the case study of Sei Semayang Health Center Data from January to March of 2023

No	Name	Age	Hypertension	Smoking	Alcohol Consumption	Exercise	Rice Consumption	Sugar Consumption	…	MSG Consumption
1	A. Jalaluddin Alfarizi	25-34 years old	Present but not controlled	Never smoked	Yes	Enough	3 X / Week	1 X / Week	…	Yes
2	A'an Mahendra	> 64 years old	Present but not controlled	Never smoked	Yes	Less	1 X / Week	2 X / Week	…	Yes
3	Aan Pahyubi	55-64 years old	Present and controlled	Never smoked	Yes	Less	Every day	1 X / Week	…	Yes

4	Abdul Roup	> 64 years old	Present but not controlled	Never smoked	Yes	Less	1 X / Week	2 X / Week	…	Yes
5	Abdullah Sopian	34-55 years old	Present and controlled	Never smoked	Yes	Less	1 X / Week	Never	…	Yes
6	Ade Kholilah	55-64 years old	Present and controlled	Never smoked	None	Enough	Every day	1 X / Week	…	Yes
7	Christian Wing	25-34 years old	Present and controlled	Never smoked	None	Less	Every day	2 X / Week	…	Yes
8	Adelia Ekka Princess	25-34 years old	Present and controlled	Never smoked	Not	Less	Every day	2 X / Week	…	Yes
9	Adinda Arijki Islamiati	25-34 years old	Present but not controlled	Never smoked	Yes	Enough	Never	1 X / Week	…	None
10	Adira Cahya Putri	25-34 years old	Present but not controlled	Never smoked	Yes	Enough	2 X / Week	2 X / Week	…	None
11	Aditya Chandra	25-34 years old	Present but not controlled	Never smoked	Yes	Enough	2 X / Week	2 X / Week	…	None
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
761	Alfieri	34-55 years old	Present and control	Never smoked	Yes	Enough	1 X / Week	Never	…	Yes

led

Research Design

This stage starts with an observation, after which it is continued by collecting data, which is then plugged into Excel for processing with the calculation process and following the stages of the linear regression method. The results of the data processing can be used in the rapid-miner application to see accurate results.

Information:

1. The problem analysis analyzed data on people with hypertension in the case study of Sei Semayang Health Center from January to March of 2023.
2. Studying scientific literature requires utilizing various sources that can be used to obtain information for research.
3. Establish a method to resolve the problem. This study used the Apriori Algorithm.
4. Collect data on people with hypertension in a case study of Sei Semayang Health Center from January to March of 2023.
5. Processing data is done using data mining in the Apriori Algorithm method.
6. Testing the data, done using the rapid miner 5.3 tool.
7. Conclusion was included in analyzing data on patients with hypertension in the case study of Sei Semayang Health Center.

Algorithm Method

After equalizing the values, the next step is grouping using the Apriori Algorithm. From this process will be obtained a prediction of hypertension sufferers in the case study of Sei Semayang Health Center using the Apriori Algorithm method. The process of the Apriori Algorithm is as follows:

Collecting Datasets

Below is a table using a dataset from data on hypertension patients in the case study of Sei Semayang Health Center which is proposed to predict data on patients with hypertension in the case study of Sei Semayang Health Center conducted using linear algorithm method.

Table 2
Table of classification data for people with hypertension in the case study of Sei Semayang Health Center

Attribute	**Information**		
Gender	**Male**	**Female**	
Age	25-37 years old (U1)	38-55 years old (U2)	55-64 years old (U3)
Hypertension	Present but not controlled (H1)	Present and controlled(H2)	
Smoking	None (M1)	Yes (M2)	
Alcohol Consumption	Less (KA1)	Enough (KA2)	

Sports	Less (O1)	Sufficient (O2)	
Rice Consumption	Never (KB1)	1 X / Week (KB2)	2X/Week (KB3)
Sugar Consumption	Never (KG1)	1X/Week (KG2)	2 X/Week (KG3)
Consumption of Side Dishes	Never (KLP1)	1 X / Week (KLP2)	2 X/Week (KL3)
Consumption of saturated fat	Never (KLJ1)	1 X/Week (KLJ2)	2 X / Week (KLJ3)
Consumption of Unsaturated Fats	Never (KLTJ1)	1 X/Week (KLTJ2)	2 X / Week (KLTJ3)
Salt Consumption	Never (KG1)	1X/Week (KG2)	2 X/Week (KG3)
Vegetable Consumption	Never (KS1)	1 X/Week (KS2)	2 X/Week (KS3)
Fruit Consumption-Fruits	Never (KB1)	1 X / Week (KB2)	2 X / Week (KB3)
Consumption of Fried Snacks	Never (KJG1)	1X/Week (KJG2)	2 X/Week (KJG3)
MSG Consumption	None (KM1)	Yes (KM2)	

Table 3
Data initializing table

NO	Name	Gender	Age	Hypertension	Smoking	Alcohol Consumption	Exercise	…	MSG Consumption
1	A. Jalaludin Alfarizi	Male	25-37 years old	Present but not controlled	Never smoked	KA2	O2	…	KM2
2	A'an Mahendra	Male	> 64 years old	Present but not *controlled*	Never smoked	KA2	O1	…	KM2
3	Aan Pahyubi	Male	55-64 years old	Present and controlled	Never smoked	KA2	O1	…	KM2
4	Abdul Roup	Male	> 64 years old	Present but not *controlled*	Never smoked	KA2	O1	…	KM2
5	Abdullah Sopian	Male	38-55 years old	Present and controlled	Never smoked	KA2	O1	…	KM2
6	Ade Kholilah	Female	55-64 years old	Present and controlled	Never smoked	KA1	O2	…	KM2
	Adella	Female	25-37	Present	Never smoke			…	

7	Christiyanti		years old	and controlled	d	KA1	O1		KM2
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
760	Andreas Trimurni	Male	25-37 years old	Present and controlled	Never smoked	KA1	O1	…	KM2
761	Andres Setiawan	Male	25-37 years old	Present and controlled	Never smoked	KA1	O1	…	KM2

Apriori Algorithm testing using rapid miner

At this stage, the data normalization process aims to provide data characteristics for a more specific range, so that data can be processed efficiently using Apriori Algorithm methods with rapidminer applications.

Results of Hypertensive Patient Data in the Case Study of Sei Semayang Health Center Using the Rapid Miner Application

The results of export value data using the rapid miner application from hypertensive patient data in the case study of Sei Semayang Health Center from January to March of 2023 obtained the results of analysis of Lifestyle Patterns in Hypertensive Patients

Table 4
Data table of patients with hypertension using the rapid Miner application

Premises	***Conclusion***	***Support***	***Confidence***	***Laplace***	***Gain***	**p-s**	***Lift***	***Condition***
Male	KM2	0.42127 0718	0.83106 267	0.943171 402	-0.592541 436	0.046692 332	1.124653 034	1.545245945
H2	KB3, KM2	0.46685 0829	0.85353 5354	0.948214 286	-0.627071 823	0.107246 421	1.298234 445	2.338731187
KB3, H2	KM2	0.46685 0829	0.91598 916	0.971637 694	-0.552486 188	0.090231 144	1.239581 592	3.107333809
H2	KM2	0.50414 3646	0.92171 7172	0.972321 429	-0.589779 006	0.099966 423	1.247333 145	3.334699697
Female	M1	0.42265 1934	0.85714 2857	0.952821 462	-0.563535 912	0.073944 629	1.212053 571	2.049723757
O2	M1	0.47790 0552	0.86716 792	0.952804 987	-0.624309 392	0.088168 859	1.226229 637	2.20441989
KJG3	KG3	0.44198 895	1	1	-0.441988 95	0.246634 718	2.2625	1.545245945
KLJ3	KG3	0.44198	1	1	-	0.246634	2.2625	2.3387311

		895			0.44198895	718		87
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
KLJ3, KLTJ3	KG3	0.44198895	1	1	-0.44198895	0.24663718	2.2625	2.20441989
KLJ3	KG3, KS3	0.44198895	1	1	-0.44198895	0.24663718	2.2625	1.545245945

Age Analysis Graph of the Case Study of Sei Semayang Health Center

From data on hypertension patients in the case study of Sei Semayang Health Center, data from January to March of 2023, using the rapid miner application, results in the form of U1 were obtained, which means that most people with hypertension are aged 25-38 years old.

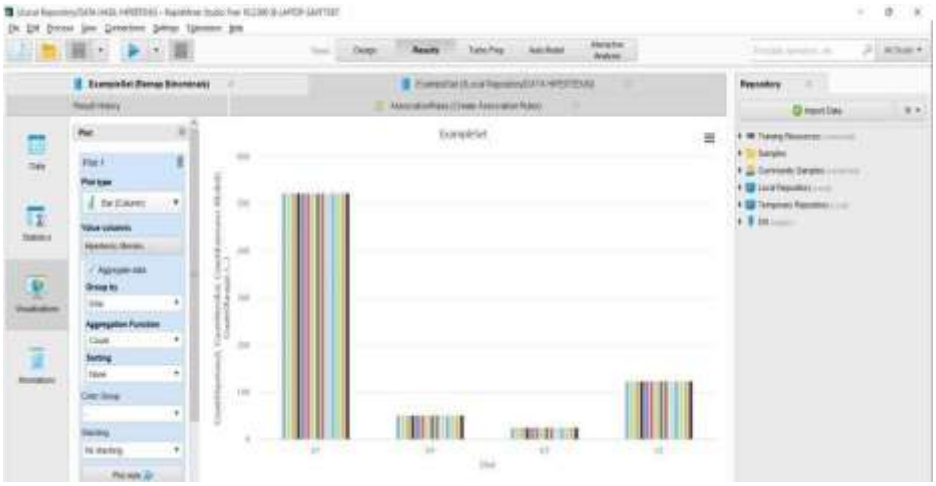

Figure 3.
Graph of age analysis of hypertension sufferers, case study of Sei Semayang Health Center

Hypertension Analysis Graph of the Case Study of Sei Semayang Health Center

From data on hypertension patients in the case study of Sei Semayang Health Center, data from January to January March 2023, the result obtained was H2, which means that people have hypertension but do not visit a doctor.

Figure 4
Hypertension analysis graph of the case study of Sei Semayang Health Center

K-Nearest Neighbor (K-NN) Algorithm Testing Using Rapid Miner

In this stage, we will normalize data which aims to equalize the characteristics of the data to be more specific, so that it can be processed efficiently with the K-Nearest Neighbor Algorithm method in the rapidminer application.

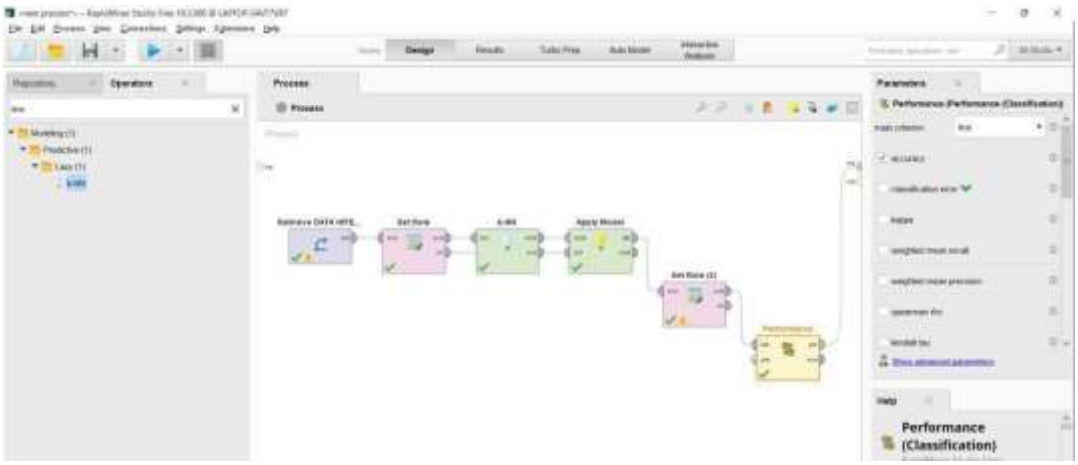

Figure 5. Display of the K-Nearest Neighbor (K-NN) Algorithm testing process

K-3 Cross Validation Test Results Using K-Nearest Neighbor (K-NN) Algorithm

Below are the results of K-3 Cross Validation testing in the case study of Sei Semayang Health Center, which obtained 92.74% class precision for those who have hypertension but are not controlled and 88.84% class precision for those who have hypertension and control.

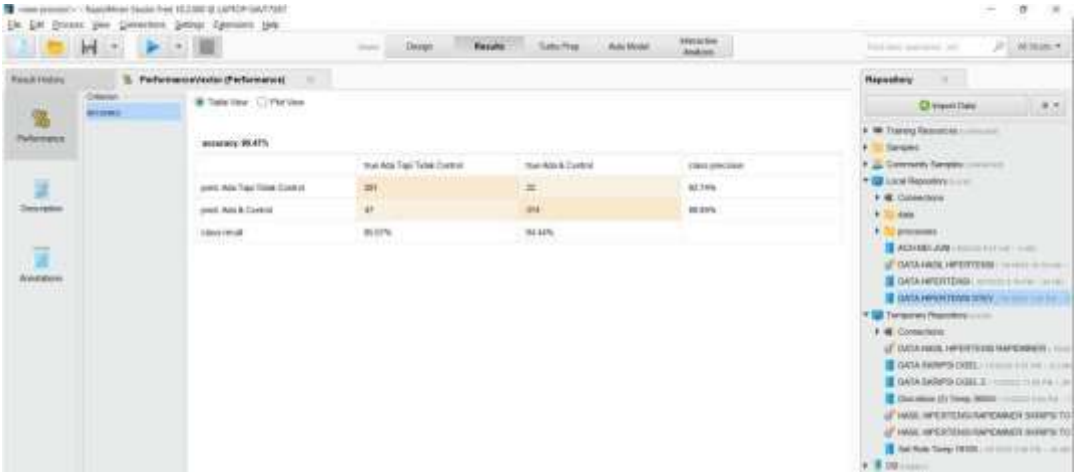

Figure 6. K-3 Cross Validation test process view

Conclusion

Following the analysis using linear regression, the next step was to group the results for comparison. The research utilized the Apriori Algorithm in RapidMiner to study the behavior patterns and lifestyle of hypertension patients at Sei Semayang Health Center. The findings, labeled as U1, revealed that most hypertension patients are aged 25-38 years old. Additionally, the same data analysis identified a pattern, denoted as H2, showing that many hypertensive individuals do not regularly visit doctors and engage in unhealthy behaviors such as consuming alcohol, smoking, lack of exercise, high sugar intake, consumption of saturated fats, MSG, and high-salt foods, and rarely consuming vegetables or fruits. This indicates that an unhealthy lifestyle significantly contributes to hypertension. Furthermore, using the K-Nearest Neighbor (K-NN) Algorithm with K-3 Cross Validation, the study achieved a class precision of 92.74% for uncontrolled hypertension cases and 88.84% for controlled cases, demonstrating the algorithm's effectiveness in accurately classifying hypertension patients.

Bibliography

Agus, Wantor, Admi, Syarif, Khairun, Nisa Berawi, Kurnia, Muludi, Sri, Ratna Sulistiyanti, & Sutyarso, Sutyarso. (2021). Implementasi Metode Pembobotan Berbasis Aturan Dan Metode Profile Matching Pada Sistem Pakar Medis Untuk Prediksi Risiko Hipertensi. *Jurnal Teknoinfo*, *15*(2). Retrieved from http://ejurnal.teknokrat.ac.id/index.php/teknoinfo/article/view/1523

Aprian Hidayat, R. (2019). Identifikasi Pola Data Hipertensi Dengan Implementasi Metode Association Rules Menggunakan Algoritma Apriori (Doctoral dissertation, Universitas Siliwangi).

Aprillia, Y. (2020). Gaya Hidup dan Pola Makan Terhadap Kejadian Hipertensi. Jurnal Ilmiah Kesehatan Sandi Husada, 9(2), 1044-1050.

Awalullaili, Fithroh Oktavi, Ispriyanti, Dwi, & Widiharih, Tatik. (2023). Klasifikasi Penyakit Hipertensi Menggunakan Metode Svm Grid Search Dan Svm Genetic Algorithm (Ga). *Jurnal Gaussian*, *11*(4), 488–498.

Azwanti, Nurul, & Elisa, Erlin. (2019). Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4. 5. *InfoTekJar (Jurnal Nas. Inform. Dan Teknol. Jaringan)*, *3*(2), 220–227. Retrieved from https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/944/pdf

Khairani, Cut, Kamil, Hajjul, & Tahlil, Teuku. (2020). Analisis faktor resiko dalam mengendalikan hipertensi di kabupaten Aceh Besar. *Jurnal Kesehatan Cehadum*, *2*(4), 1–13. Retrieved from https://jkc.puskadokesa.com/jkc/article/view/66

Lasmadasari, Novi, Nilawati, Iin, Sukatemin, Sukatemin, Ester, Ester, Elvionita, Elva, Pralingga, Yogi, & Tiara, Tiara. (2021). Penerapan Algoritma Hipertensi, Menu Dash Dan Latihan Otot Progresif Dalam Asuhan Keperawatan Untuk Menurunkan Tekanan Darah. *Jurnal Riset Media Keperawatan*, *4*(2), 54–62.

Matondang, Nurhafifah, Mayanda, M. S., & Nurul, C. (2019). Deteksi Hipertensi Dengan Metode Artificial Neural Network. *Seminar Nasional Informatika, Sistem Informasi Dan Keamanan Siber (SEINASI-KESI)*, *2*(1), 1–6. Retrieved from https://conference.upnvj.ac.id/index.php/seinasikesi/article/view/184

Melviani, Melviani, Aryzki, Saftia, Rahman, Agenia, Putri, Gina Aima, & Riadi, Setia D. R. (2022). Pengaruh Gaya Hidup Terhadap Penyakit Hipertensi Berdasarkan Teori TPB (Theory Of Planned Behavior). *Journal Pharmaceutical Care and Sciences*, *2*(2), 30–38. Retrieved from https://doi.org/10.33859/jpcs.v2i2.164

Nurzanah, Salsabilla Choerunnisa, Alam, Syariful, & Hermanto, Teguh Iman. (2022). Analisis Association Rule Untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori (Studi Kasus: Klinik Rafina Medical Center). *JIKO (Jurnal Informatika Dan Komputer)*, *5*(2), 132–141. Retrieved from https://ejournal.unkhair.ac.id/index.php/jiko/article/view/4792/3121

Purwono, Purwono, Dewi, Pramesti, Wibisono, Sony Kartika, & Dewa, Bala Putra. (2022). Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network. *Insect (Informatics and Security): Jurnal Teknik Informatika*, *7*(2), 82–90. Retrieved from https://doi.org/10.33506/insect.v7i2.1828

Putra, S. (2022). Pengaruh Gaya Hidup Dengan Kejadian Hipertensi di Indonesia (A: Systematic Review). Jurnal Pendidikan Tambusai, 6(2), 15794-15798.

Putri, H. N. (2022). Penerapan Algoritma Apriori Dalam Menentukan Pola Perilaku Dan Gaya Hidup Terhadap Penderita Hipertensi. Penerapan Algoritma Apriori Dalam Menentukan Pola Perilaku Dan Gaya Hidup Terhadap Penderita Hipertensi, 5(3), 450-458.

Rahmadayanti, Fitria, Anggraini, Inda, & Susanti, Tri. (2023). Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means. *Journal of Information System Research (JOSH)*, *4*(2), 737–741.

Wijaya, Y. F., & Triayudi, A. (2023). Perbandingan Algoritma Klasifikasi Data Mining Pada Prediksi Penyakit Diabetes. Journal of Computer System and Informatics (JoSYC), 5(1), 165-174.