Analysis of Disease Data Patterns in the Elderly with Cardiovascular Patients using the Association Rule Method

M. Fadil Mardiansyah [1*], Rizka Nur Pratama [1]

[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Indonesia
*Corresponding author: fadilmardiansyah@students.unnes.ac.id

ARTICLE INFO

Article history
Received 15 August 2020
Revised 9 September 2020
Accepted 12 October 2020

Keywords
Cardiovascular Disease
Association Rule Mining
FP-Growth
Elderly

ABSTRACT

Cardiovascular disease is a disease associated with modern behavior patterns. This disease is now attacking developed countries and has threatened countries that are heading towards modernization. Some sources say cardiovascular causes are stress due to work, hypothyroidism, heart rate, chronic kidney, and many more. In general, the reason tends to be due to unhealthy lifestyles such as eating lots of fatty foods, not exercising, etc. In today's times, people tend not to have a healthy lifestyle because of an increasingly modern lifestyle. This causes the cardiovascular disease to increase rapidly and is one of the leading causes of human death in the world. Therefore, it is necessary to analyze the pattern of factors that cause cardiovascular disease to prevent or anticipate cardiovascular disease in today's era. Association rules using the FP-GROWTH algorithm are a method that can perform tracing on historical data to identify data patterns based on previously identified properties. The relationship pattern between data can be searched by looking at the correlation variable between patients with cardiovascular disease. This study found that obesity is a determinant factor for cardiovascular disease; even when you do not consume alcohol and do not smoke, cardiovascular sufferers.

1 Introduction

Cardiovascular disease is a disease associated with modern behavior patterns (World Health Organization, 2003). This disease is now attacking developed countries and has threatened countries that are heading towards modernization (Istiyarini & Rosijidi, 2008). The World Health Organization (WHO) (2002) reported that Non-communicable Diseases (NCDs) or non-infectious diseases accounted for 60 percent of mortality and 47 percent of the burden of disease in the world and will continue to increase with predictions that in 2020. Death due to NCDs is 73 percent and constitutes 60 percent of the world's burden of disease (WHO, 2002). The main non-infectious diseases which occupy the highest proportion are coronary heart disease, stroke, diabetes (DM), cancer, and lung disease (WHO, 2003). Based on several studies, it was found that women are more susceptible to cardiovascular disease than men. The burden of risk factors for women's cardiovascular disease is greater than that of men, namely high LDL, high TG, and lack of physical activity (Centers for Disease Control and Prevention, 2013). While the three dominant risk factors for cardiovascular disease in women are age, hypertension, and high cholesterol, the three dominant risk factors for cardiovascular disease in men are hypertension, age, and smoking (Nurhidayat. 2014).

Research on the statin class has found that lowering LDL-cholesterol levels will reduce morbidity and mortality caused by coronary heart disease, whose success is only 20-30% (Degoma, Leeper, &

Heidenreich, 2008). This causes the need for other efforts to reduce morbidity and mortality rates for cardiovascular disease (Rampengan, 2015). There are many facts about the factors that influence cardiovascular disease, including the first one caused by hyperlipidemia (Nelson, 2013) In addition, cardiovascular disease is also caused by radiation obtained by humans in everyday life (Baker, Moulder, & Hopewell, 2011). Some sources mention cardiovascular disease causes are stress due to work, hypothyroidism, heart rate, chronic kidney, and many more (Kivimäki & Kawachi, 2015; Perret-Guillaume, Joly, & Benetos, 2009; Weiner *et al.,* 2004). In general, the cause tends to be due to unhealthy lifestyles such as eating lots of fatty foods, not exercising, etc. Several key parameters, such as plasma HDL-c and LDL-c levels and systolic blood pressure, are currently needed to accurately estimate CVD risk ten years (Balder *et al.*, 2015).

Moderate activity, such as brisk walking for 30 to 60 minutes a day most days of the week, was associated with a significant reduction in cardiovascular disease incidence and mortality (Haennel & Lemire, 2002). In today's times, people tend not to have a healthy lifestyle because of an increasingly modern lifestyle. This causes the cardiovascular disease to increase rapidly and is one of the leading causes of human death in the world. Therefore, it is necessary to analyze the pattern of factors that cause cardiovascular disease to prevent or anticipate cardiovascular disease in today's times. In order to determine the pattern of causative factors for cardiovascular disease, it is necessary to analyze patient data at the hospital. The patient data analyzed were patients suffering from cardiovascular disease.

Many studies conducted using the FP-Growth association rule algorithm. Based on this research, it can be concluded that the FP-Growth algorithm is able to trace historical data to identify data patterns based on previously identified properties and can produce knowledge models in the form of rules with confidence value (Kurniawan, Fujiati, & Saleh, 2014; Ramdhani & Said, 2014).

The knowledge model used in the FP-Growth algorithm is a data mining application that can generate knowledge models in the form of rules with trust values so that they can be used to predict future data trends.

The research objective is to do early prevention if the same indication is found. Indications are leading to the onset of cardiovascular disease later. The FP-Growth algorithm can find the relationship between items in the data so that it can analyze the disease patterns that have in cardiovascular disease (Waruwu, Buulolo, & Ndruru, 2017). Based on the information above, the author raised the title of the study "Analysis of Disease Data Patterns in the Elderly with Cardiovascular Patients Using the Association Rules Method".

2 Methods

2.1 Study Literature

The initial step of this research is to find and study the problem to be studied. After that, determine the scope of the problem, the background of the problem, and find a solution to the problem.

2.2 Data Retrieval the Data

To be processed is a dataset of cardiovascular patient medical records. This dataset was obtained from UCI machine learning.

2.3 Pre-processing Data

In the pre-processing data stage, data normalization was carried out to eliminate missing values in the dataset and the selection of attributes to be used in the study

2.4 Processing Data

Processing in this study used the Weka application to determine rules -rule based on the relationship between cardiovascular patient medical record data variables.

2.5 Evaluations of The Results of The Association Rule

At this stage, a valid best rule is selected and has a value lift ratio of more than one (Suryanto, Proboyekti, & Oetomo, 2013). The lift ratio can be used to determine how important the rule has been based on the value of support and confidence (Triyanto, 2015).

2.6 Analysis of Data Patterns

After evaluating the association rule results, analysis of the rules that have been obtained is carried out to obtain knowledge that can be used to support the analysis of cardiovascular disease or diagnosis of patients.

3 Results and Discussion

This study uses a dataset taken from UCI machine learning. From several attributes in the dataset, data pre-processing is carried out so that only the selected attributes will be processed. The attributes to be used are as follows.

Table 1. Attribute data

Features and description	Final Values
Age (year)	UA (40-50), UB (51-60), UC (>60)
Gender	PR (Woman), LK (Man)
Height (cm)	HA (140-155), HB (156-170), HC (>170)
Weight (Kg)	WA (40-60), WB (61-80), WC (>80)
Systolic blood pressure	APA (50-110), APB (110-120), APC (121-140) APD (>140)
Diastolic blood pressure	APLA (50-79), APLB (80-89), APLC (90-120), APLD (>120)
Cholesterol	Normal, above normal, well above normal
Glucose	Normal, above normal, well above normal
Smoke	Yes, no
Alcohol	Yes, no
Active	Yes, no
Cardiovascular	Yes, no

Based on Table 1, the processed data is imported into the Rapid Miner. Then the attribute is selected will be used to select any attributes to be processed. The next step is to choose FP-Growth for the algorithm that will be used to find itemset that often appears. FP-Growth is proven to have advantages over a priori algorithms to be applied to large datasets. FP-Growth can process large data quickly because it does not need to generate candidate keys, and FP-Growth uses the concept of building a tree for frequent itemset searches.

From 1000 processed data, we get association rules from Rapid Miner software. There are 50 rules generated. Based on the 50 rules generated, the ten best rules are obtained with a lift ratio> 1, which is shown in Table 2. The rules that are formed will then be analyzed to acquire knowledge that can be used to support the analysis of cardiovascular disease or diagnosis of patients.

Table 2. Strong rule

Rule	Support (%)	Confidence (%)	Lift Ratio (%)
Weight more than 80 Kg	99	100	1,1
Weight more than 80 Kg and do not consume alcohol	95,6	100	1,1
Weight more than 80 Kg and do not smoke	90,9	100	1,1
Weight more than 80 Kg, do not consume alcohol and do not smoke	89,1	100	1,1
Weight more than 80 Kg and active	77	100	1,1
Weight more than 80 Kg, do not consume alcohol, and active	74,1	100	1,1
Weight more than 80 Kg and woman	66,4	100	1,1
Weight more than 80 Kg, do not smoke, and woman	65	100	1,1
Weight more than 80 Kg, do not consume alcohol and woman	64,8	100	1,1
Weight more than 80 Kg and normal cholesterol	62,6	100	1,1

After analyzing the existing rules, obesity is a determining factor in cardiovascular disease, even when not consuming alcohol and not smoking, it turns out that there are cardiovascular sufferers. In addition, other supporting factors for cardiovascular disease were obtained, namely female gender and activity. This study's results can be used as a reference for the community to anticipate early cardiovascular disease. Improvement efforts so that more optimal results can be developed in the pre-processing process (normalizing data by adding weight so that the rule becomes simple). Second, in the analysis process by combining predictive methods or determining the pattern of cardiovascular disease factors.

4 Conclusion

Based on the research conducted, it can be concluded that the application of the association rule can be used in the health sector to determine the relationship between variables from the medical records of patients with cardiovascular disease. The FP-Growth algorithm is used to find itemset that appears frequently. The association rules that are formed can be used as a reference for the community to anticipate cardiovascular disease.

References

Baker, J. E., Moulder, J. E., & Hopewell, J. W. (2011). Radiation as a Risk Factor for Cardiovascular Disease. *Antioxidants & Redox Signaling*, 15(7), 1945-1956. doi:10.1089/ars.2010.3742

Balder, J. W., Scholtens, S., De Vries, J. K., Van Schie, L. M., Boekholdt, S. M., Hovingh, G. K., . . . Kuivenhoven, J. A. (2015). Adherence Guidelines to revent cardiovascular diseases: The LifeLines Cohort Study. *The Netherlands Journal of Medicine*, 73(7), 316-323. Retrieved from http://www.njmonline.nl/getpdf.php?t=i&id=186#page=12

Centers for Disease Control and Prevention (CDC). (2013). *Facts on Women and Heart Disease*. Retrieved from https://www.cdc.gov/heartdisease/women.htm on 10 September 2020.

Degoma, E. M, Leeper, N. J., & Heidenreich, P. A. (2008). Clinical Significance of High-Density Lipoprotein Cholesterol in Patients with Low Low-Density Lipoprotein Cholesterol. *Journal of the American College of Cardiology*, 51(1), 49. doi: 10.1016/j.jacc.2007.07.086

Haennel, R. G., & Lemire, F. (2002). Physical Activity to Prevent Cardiovascular Disease. How Much is Enough? *Canadian Family Physician*, 48, 65-71. Retrieved from https://pubmed.ncbi.nlm.nih.gov/11852614/

Istiyarini, Y. D., & Rosijidi, C, H. (2008). *Aktivitas Fisik Penderita Penyakit Kardiovascular [Physical Activity of Patients with Cardiovascular Disease]*. Ponorogo: UNMUH

Kivimäki, M., & Kawachi, I. (2015). Work Stress as a Risk Factor for Cardiovascular Disease. *Current Cardiology Reports*, 17(9), 74. doi:10.1007/s11886-015-0630-8

Kurniawan, H., Fujiati, & Saleh, A. (2014). Analisa Pola Data Penyakit Rumah Sakit dengan menerapkan Metode Association Rule menggunakan Algoritma Apriori [Hospital Disease Data Pattern Analysis by applying the Association Rule Method using the Apriori Algorithm]. *Seminar Nasional Informatika*, 1(1), 195. Retrieved from http://e-journal.potensi-utama.ac.id/ojs/index.php/SNIf/article/view/214

Nelson, R. H. (2013). Hyperlipidemia as a Risk Factor for Cardiovascular Disease. *Primary Care*, 40(1), 195-211. doi:10.1016/j.pop.2012.11.003

Nurhidayat, S. (2014). Faktor Risiko Penyakit Kardiovaskuler pada Remaja di Ponorogo [The Adolescentes Cardiovascular Disease Risk Factors in Ponorogo]. *Jurnal Dunia Keperawatan*, 2(2), 40-47. Retrieved from http://eprints.umpo.ac.id/1295/1/Jurnal%20Dunia%20Keperawatan%20.pdf

Perret-Guillaume, C., Joly, L., & Benetos, A. (2009). Hearth Rate as a Risk Factor for Cardiovascular Disease. *Progress in Cardiovascular Diseases*, 52(1), 6-10. doi: 10.1016/j.pcad.2009.05.003

Ramdhani, N., & Said, B. (2014). Analisis Pola Asosiasi dan Sekuensial Data Rekam Medis RSUD DR. H. Slamet Martodirdjo Pamekasan dengan Teknik Data Mining Menggunakan Algoritma Apriori [Analysis of Association Patterns and Sequential Data of Medical Records RSUD DR. H. Slamet Martodirdjo Pamekasan with Data Mining Techniques using the Apriori Algorithm]. In *Seminar Nasional Sistem Informasi Indonesia (SESINDO) 2014*, 339-345. Retrieved from http://is.its.ac.id/pubs/oajis/index.php/file/download_file/1414

Rampengan, S. H. (2015). Edema Paru Kardiogenik Akut [Acute Cardiogenic Pulmonary Edema]. *Jurnal Biomedik (JBM)*, 6(3). Retrieved from https://ejournal.unsrat.ac.id/index.php/biomedik/article/download/6320/5840

Suryanto, E., Proboyekti, U., & Oetomo, B. S. D. (2013). Implementasi Customer Relationship Management dengan Market Basket Analysis pada Toko Buku Online. Studi Kasus: Toko Buku Toga Mas [Implementation of Customer Relationship Management with Market Basket Analysis at Online Bookstores. Case Study: Toga Mas Bookstore]. *Jurnal Eksplorasi Karya Sistem Informasi dan Sains*, 6(2), 12-20. Retrieved from https://www.neliti.com/publications/78802/implementasi-customer-relationship-management-dengan-market-basket-analysis-pada

Triyanto, W. A. (2015) Algoritma K-Medoids untuk Penentuan Strategi Pemasaran Produk [K-Medoids Algorithm for Product Marketing Strategy Determination]. *Jurnal SIMETRIS*, 6(1), 183-188. doi:10.24176/simet.v6i1

Waruwu, F. T., Buulolo, E., & Ndruru, E. (2017). Implementasi Algoritma Apriori pada Analisa Pola Data Penyakit Manusia yang Disebabkan oleh Rokok [Implementation of the Apriori Algorithm in Analysis of Data Patterns of Human Diseases Caused by Smoking]. *Konferensi Nasional Teknologi Informasi dan Komputer,* 1(1), 176. doi:10.30865/komik.v1i1.492

Weiner, D. E., Tighiouart, H., Amin, M. G., Stark, P. C., MacLeod, B., Griffith, J. L., . . . Sarnak, M. J. (2004). Chronic Kidney Disease as a Risk Factor for Cardiovascular Disease and All-Cause Mortality: A Pooled Analysis of Community-based Studies. *Journal of the American Society of Nephrology*, 15(5), 1307-1315. doi:10.1097/01.ASN.0000123691.46138.E2

World Health Organization (WHO). (2002). *The World Health report 2002 - reducing risk, promoting healthy life*. Geneva: WHO

World Health Organization (WHO). (2003). *The WHO STEPwise approach to Surveillance of NonComunication Disease (STEPS)*. Geneva: WHO