Challenging Analytic Data Opportunities in Smart Health with Algorithm
(Study Case: Bumi Medika Ganesha ITB)

Aristyo Hadikusuma[#1] , Anung Asmoro [#*2] , and Joko Rurianto[#+3]

#School of Computing, *Telkom University*
Bandung, Indonesia

[1] tyokusuma@telkomuniversity.ac.id

[*]*Telkom Indonesia*
[2] anungasmoro@outlook.com

[+]*Telkomsel. Indonesia*
[3] jokorurianto21@gmail.com

Abstract

Weka, a powerful data science tool, offers valuable assistance in the field of data clustering. One of its notable features is the K-means algorithm, which enables efficient clustering analysis for specific data analysis tasks. Clustering analysis is a technique employed to categorize and partition objects into meaningful groups based on their distinct characteristics. Given the vast diversity and abundance of data, utilizing the K-means algorithm in Weka proves to be instrumental in grouping patient temperature data into multiple clusters. This data grouping technique through clustering holds great potential as a strategic approach for informed decision-making processes.

Keywords: Weka, K-means

Abstrak

Weka adalah alat yang membantu ilmu data dalam mengelompokkan data secara clustering. Weka memiliki fitur K-means yang membantu mengelompokkan data untuk analisis tertentu. Analisis clustering adalah teknik untuk mengategorikan dan membagi objek menjadi kelompok. Setiap objek memiliki karakteristik tertentu. Karena data memiliki beragam dan banyaknya variasi, dengan menggunakan algoritma K-Means ini, data suhu pasien yang sudah diperoleh akan dikelompokkan menjadi beberapa kluster. Pengelompokan data melalui clustering diharapkan dapat menjadi strategi dalam pengambilan keputusan.

Kata Kunci: Weka, K-mean

I. Introduction

Health is necessary all of people in worldwide and very important. Bandung of technology institute, one of institute which consideration with health. Bandung of technology institute is supporting health activity

Hadikusuma et al.
Challenging Analytic Data Opportunities in Smart Health with...

2

[1]. The attention of ITB on the health factor is creating an clinic, it called Bumi Medika Ganesha. Nowdays, all of organisation to develop data analytic for their bussiness including the hospital. The hospital is average data consume too much [2]. The purpose of hospital industy use the data is to analyse disease to know how to medical treatment appropriately. The patient always get treatment with the same disease and the medical record difficult to analyse and waste the time to check in database the applicable criteria that follow.

II. Related Work

A. Clustering

Clustering is the process of grouping from different objects into different groups, or it can be called the partitions of a data set [3]. A cluster can consist of different objects or similarities between one another. This clustering algorithm is used in finding clusters sequentially from previously initiated clusters and determining all groups in a given time [4]. The process of this clustering method is to divide the data into several groups. A similar pattern of data will be grouped on the same cluster and different patterns will be in different clusters by breaking apart to form another cluster pattern [5].

B. Arima

ARIMA is a merger of two models namely the Otoregresi Model (AR) and Moving Average (MA). This model does not have different variables with independent variables, but uses information in the same series in some models that will ultimately be useful for forecasting [6]. The AR model refers to the relationship between the Y variable and the independent variable, which is the Y value at the previous time, whereas the MA Model shows the Y independent variable against the residual values at the previous time in sequence. Before performing the analysis using the ARIMA method, the data must be stationary [7]. The stationary meaning in general is data that contains no focus or trend.

The ARIMA model is written with ARIMA notation (p, d, q). p is the degree of process AR, d is the order of distinction, and q is the degree of MA process. The mathematical model of ARIMA can be written as follows.

$$y = \theta_0 + \emptyset_1 y(t-1) + \emptyset_2 y(t-2) + \emptyset_p y(t-p) + \mathcal{E}(t - \theta_1)\mathcal{E}(t - 1 - \theta_2)\mathcal{E}(t - 2 - \dots - \theta_p)\mathcal{E}(t-q) \tag{1}$$

Forecasting the number of tourists using ARIMA method will be done using patient temperature in Bandung.

C. Temperature

Body temperature is a measure of the ability used to produce and what is in the body. Body temperature is one part important to understand, because it can tell if someone has been attacked by serious illness or not. To know the reliable temperature difference in the body, need to be distinguished first as is the usual body in the circumstances normal [8]. Lower body temperature is usually better known as hypothermia. When it's temperature very low body, you must be careful. Because the body temperature can be used a fatal result. Why is that? This low temperature will slow down the system nerves, breathing, and blood circulation. A new low body temperature can refer to as hypothermia freedom of temperature reaches a figure below 35 degrees Celsius [9]. High body temperature is usually known as hyperthermia. On this one condition You will usually be at a temperature of 38.3 degrees Celsius or even more. In hyperthermia, the increasing temperature in the body is different from that of photography temperature at the time of fever. This happens when the fever is in progress are in a controlled temperature control system. While in patients with hyperthermia, temperature It will continue to increase uncontrollably excessive [10-[14]. The very low person in the body, he will report some symptoms. Some things that happen are chills, longer talk does not clear, short and weakened breathing, and awareness of

increasing sufferers long will disappear slowly [15]. Also marked also with the perceived skin more and more long colder and reddish, become limp, not energized, and cry overall continuously without power. 4G symptoms Hyperthermia occurs during heat stings that start from the hot environment the body cannot cool normally and effectively [17]-[18]. Replace, increasing temperature rising in a person's body that can be sustained and also lead to recovery dehydration able to restore organs even permanently.

III. METHODOLOGY

The research method used in this research is Scientific Method method. In Scientfic Method there are two types of research that is quantitative and qualitative methods [16] . This study uses a combination of quantitative and qualitative research. In general, there are six stages (1) Identification of Research Problem, (2) Literature Review, (3) Determining Research Objectives, (4) Collecting Data, (5) Data Analysis and Interpretation, and (6) Research Report and Evaluation. Flow Scientific Method can be seen in Fig. 1 below.
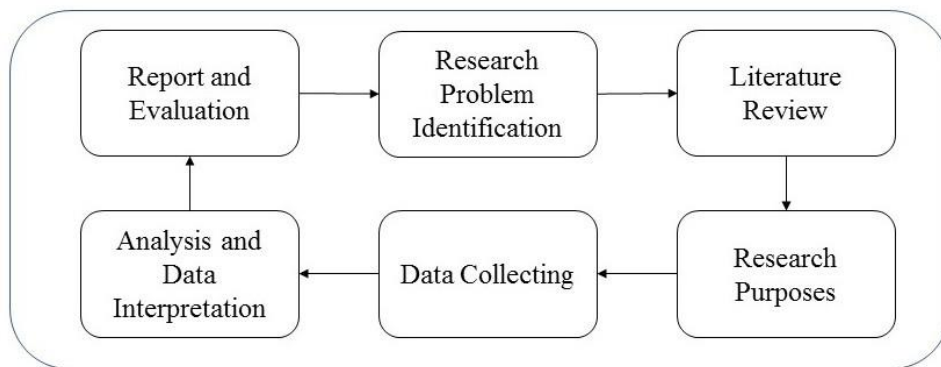

Fig. 1. Scientific Method

A. Research Problem Identification

This is the first stage of identifying research problems. Research problem of this research is to determine the method of data analytic and variable that suitable for the average of bandung city and determine the strategy and policy to be taken related to analytic result.

B. Research Purposes

General purpose in this research is to forecast the patient temperature of Bandung city by involving big data analytic. The specific purpose of this research is to determine the method of data analytic and the right variable for forecasting the patient temperature of bandung city and determine the strategy and policy to be taken.

C. Data Collecting

What is done at this stage is data collection. The data used is the patient temperature of Bandung city from 2010 to December 2017. The data is obtained from the website of the Central Bureau of Statistics. This data will be used to forecast city temperature bandung

D. Report and Evaluation

Evaluation of ARIMA model can be seen with the smallest Mean Square Error result or compare EIC value of each candidate model. The model with the smallest EIC value is the best model.

IV. Results

The next stage in the research involves the clustering process, which utilizes the k-means algorithm to group the data into pre-defined clusters. For this particular study, three clusters have been set as the desired outcome. The output of the k-means clustering is a list of clusters to which the processed data belongs. Once the processing is complete, the subsequent stage involves analyzing and validating the results of the clustering process by examining the formed clusters.

To provide further insight, let's refer to Table I, which presents the temperature data used in the analysis:

TABLE I
FONT SIZE FOR MANUSCRIPTS

No	Temperature (°C)
1	33
2	33.9
3	33.7
4	33.7
5	33
6	33.7
7	31.5
8	31.5
9	33.3
10	33.1
11	33.4
12	33.6
13	33.3
14	33.3
15	33.9
16	33
17	33.4
18	33.3
19	31.8

Using the k-means algorithm [20], these temperature values will be grouped into three distinct clusters. The subsequent analysis and validation process will involve examining the resulting clusters to gain insights and make informed decisions based on the patterns and characteristics observed within each cluster.

After completing the preprocessing phase, the subsequent step involves data processing. In this research, the data processing analysis is conducted using data mining tools, following these steps: applying the k-means algorithm to the patient's temperature test data using Weka, and examining the resulting cluster output.

The k-means algorithm, implemented through Weka (Fig 2), plays a crucial role in this process by grouping the temperature data into distinct clusters based on their similarities. By leveraging the power of data mining tools, this analysis aims to uncover meaningful patterns and relationships within the dataset, facilitating a deeper understanding of the underlying trends in patient temperatures.

The cluster output obtained from the k-means algorithm will provide valuable insights into the grouping of temperature data, enabling further analysis and decision-making based on the identified clusters. This step acts as a bridge between the preprocessing phase and the subsequent analysis, offering a foundation for exploring and interpreting the dataset in a more structured and meaningful manner.t
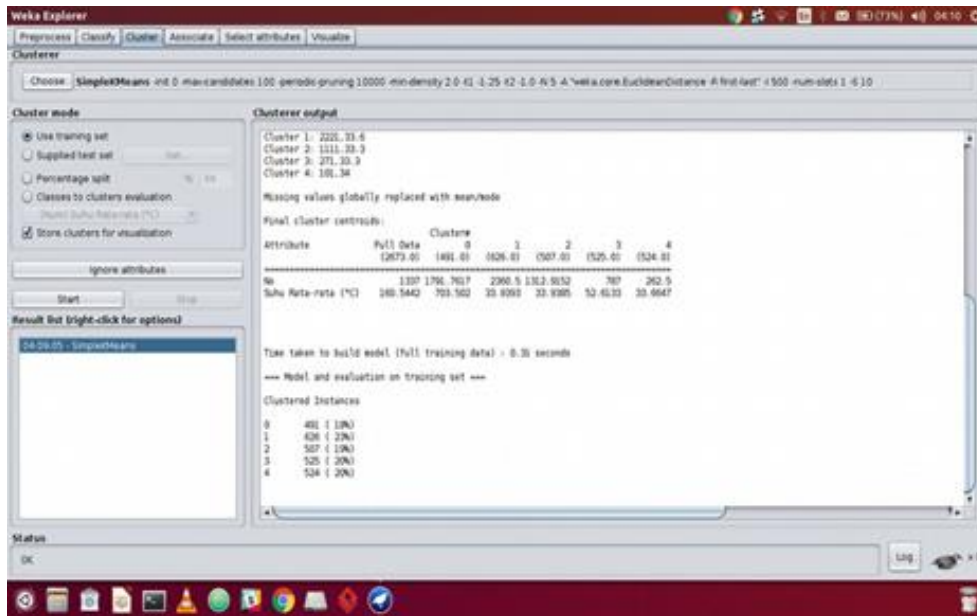
Fig. 2. Result of k-mean with 5 cluster.

After obtaining the information from the final cluster centroids, the results can be effectively showcased through visualization techniques. As depicted in Fig. 3, the visualization of clusters provides a clear representation of the clustering outcome.
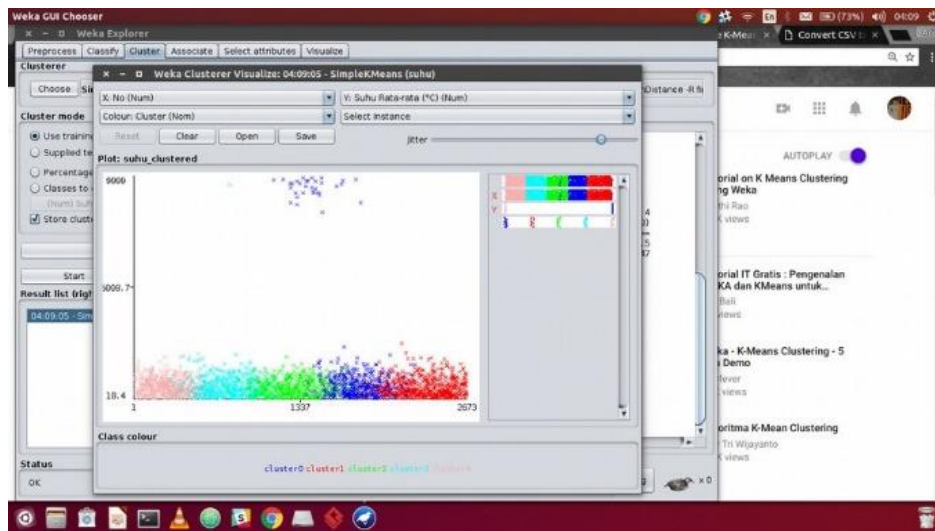

Fig. 3. Visualization of Cluster.

Utilizing the Clustering K-Means method, the data processing stage yields the clustering results after undergoing multiple iterations. The obtained clustering results are as follows:
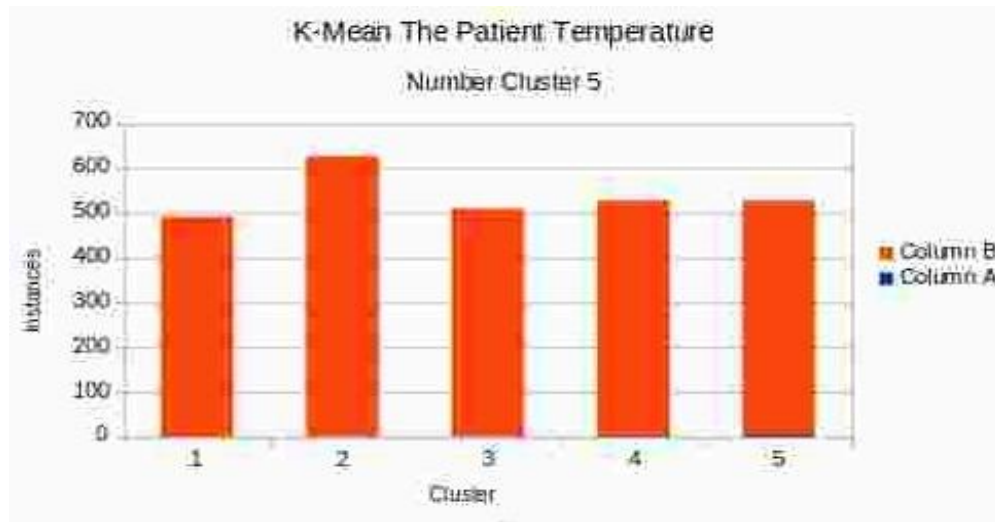
Fig. 4. Chart of Result Cluster.

Figure 5 depicts the results of the k-means clustering algorithm with k=5. The analysis has successfully categorized all instances into five distinct clusters. Cluster 2 emerges as the largest cluster, containing over 600 instances, indicating a significant presence within the dataset. In contrast, cluster 1 exhibits the lowest number of instances, with less than 500. The remaining clusters have a slightly higher number of instances, hovering just above 500. These results provide valuable insights into the distribution and composition of the dataset, highlighting the varying degrees of representation within each cluster.

V. CONCLUSION

K-Means is a typical clustering algorithm it is widely used for clustering sets of data The final result of data processing, information about cluster type is formed. K-Means is a typical clustering algorithm it is widely used for clustering sets of data. with the help of k mean. temperatures can be grouped to take policy decisions. in 5 groupings doctors can predict which diseases are at risk if the temperature pressure increases or low. By using this K-Means algorithm the patient temperature data already obtained are grouped into several clusters. Grouping of data by clustering is expected to be a strategy for decision making.

REFERENCES

[1] Florence, Agnes Grace. Hubungan Pengetahuan Gizi dan Pola Konsumsi dengan Status Gizi pada Mahasiswa TPB Sekolah Bisnis dan Manajemen Institut Teknologi Bandung. Diss. Fakultas Teknik Unpas, 2017.

[2] Kapila, Dhiraj, et al. "IoT Research on Healthcare Data to Control Illness of Oncology." 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE, 2022.

[3] Ezugwu, Absalom E., et al. "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects." Engineering Applications of Artificial Intelligence 110 (2022): 104743.

[4] Liu, Peide, et al. "A clustering-and maximum consensus-based model for social network large-scale

group decision making with linguistic distribution." Information Sciences 602 (2022): 269-297.

[5] Guerreiro Miranda, Bruno, Rogério Galante Negri, and Luana Albertani Pampuch. "Using clustering algorithms and GPM data to identify spatial precipitation patterns over southeastern Brazil." Atmósfera 37 (2023).

[6] Wang, Meng, et al. "ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021." BMC Public Health 22.1 (2022): 1447.

[7] Fan, Yamin, and Ximei Liu. "Auxiliary model-based multi-innovation recursive identification algorithms for an input nonlinear controlled autoregressive moving average system with variable-gain nonlinearity." International Journal of Adaptive Control and Signal Processing 36.3 (2022): 521-540.

[8] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[9] M. Maila and J. Shi, "A random walks view of spectral segmentation," in AI and STATISTICS (AISTATS) 2001, 2001

[10] M. Cocco and A. Tuzzi, "New data collection modes for surveys: A comparative analysis of the influence of survey mode on question- wording effects," Qual. Quant., vol. 47, no. 6, pp. 3135–3152, 2013

[11] Zhang G.P 2001. Time Series Forecasting Using a hybrid ARIMA and Neural Network Model,

[12] Night, John F. Jantung Kuat Bernafas Lega.Bandung:Indonesia Publishing House, 1995

[13] Handoyo, R. 2013. *Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen.Proposal Tugas Akhir* Institut Teknologi Telkom. Bandung

[14] [6] Bholowalia, Kumar., (2014). *A Clustering Techniques based on Elbow Method and K-means in WSN.* International Journal of Computer Application (0975 - 8887)

[15] Hung, C.M., Wu, J., Chang, J.H. & Yang, D.L., 2005. An Efficient k-Means Clustering Algorithm Using Simple Partitioning. Journal of Information Science and Engineering, XXI(1), pp.1157-77

[16] Ediyanto, et.al, 2013. Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster), II(2), pp. 133-136

[17] Tiransri, Loudry Achmad, Favian Dewanta, and Hilal Hudan Nuha. "Analisis Quality of Service Modul NRF24L01 pada Sistem Stasiun Cuaca Lokal." TELKA-Jurnal Telekomunikasi, Elektronika, Komputasi dan Kontrol 9.1 (2023): 44-51.

[18] Ryandana, Farrish Fauzan, Hilal H. Nuha, and Endro Ariyanto. "Query Optimization for Room Temperature Sensor using Principal Component Analysis." 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, 2022.

[19] Fauzy, Rinaldy Achmad, Hilal Hudan Nuha, and Ahmed Hamdi Abo Absa. "Implementation of Monitoring System and Prediction of Room Humidity for Orchid." 2022 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob). IEEE, 2022.

[20] Nuha, H., Liu, B., Mohandes, M., Balghonaim, A., & Fekri, F. (2021). Seismic data modeling and compression using particle swarm optimization. Arabian Journal of Geosciences, 14, 1-11