C03

Proposing K-Means Clustering on Air Quality governed by AI-Mining Prototype

Happy Alyzhya Haay
Program Studi Teknik Fisika, Fakultas Sains dan Teknologi
Universitas Internasional Papua
email: happyhaay@iup.ac.id

ABSTRACT

Along with the rapid development of technology, making various sectors create intelligent AI-MINING systems that can be used to unify quality using sensor data acquisition. These sensors generate data in real-time to unify air quality. Then the data will be grouped using the K-means clustering algorithm and the results obtained from five cluster points for humidity and temperature data.

Keywords: K-means clustering, humidity, sensor, and temperature.

INTRODUCTION

Air quality and indoor temperature are very important for human life because almost 80% of human life is spent indoors. Therefore, indoor air quality must be maintained. One way is to use a tool that regulates and maintains indoor air quality automatically. AI-Mining intelligent system to store monitored data to measure air quality.

Previously, previous research has been carried out aimed at grouping using the K-Means algorithm for data analysis of air quality pollution. In this study, K-Means clustering is proposed to be improved so that it can be compared with Possibilistic Fuzzy C-Means in terms of accuracy and execution time. Thus, the results obtained show that the K-Means which has been improved by the clustering algorithm provides an air quality index value with higher accuracy (Kingsy et al. 2017)

Another related research is on the multi-task clustering of air quality monitoring in Turkey. The impact of air pollution on modern life today that has an impact on human health makes air quality very important throughout the world, therefore many studies have carried out classification, grouping, and association rules in mining applications for air pollution. This study aims to model an area with similar characteristics determined and placed into the same cluster using a new algorithm called Majority Voting based Multi-Task Clustering (MV-MTC). The results of the research study obtained show that the new method used is superior to the other five algorithms, namely K-Means, Expectation-Maximization, Canopy, Farthest First, and Hierarchical clustering methods (Tuysuzoglu, Birant, and Pala 2019). Based on previous studies, this study will apply the K-means clustering method to air quality measurement sensor data.

RESEARCH METHOD

Data

The used data in this research was obtained from the AI-Mining system for the acquisition of air quality sensor data., the AI-Mining team developed gas acquisition sensors such as CO, NH3, NO2, Temperature, humidity, Methane, Butene, LPG, Carbon Monoxide, and Hydrogen Sulfide.

Elbow Method

The Elbow method is a way for generating information to discover the best number of clusters by comparing the percentage of clusters that will form an angle at a given place. This method generates ideas/ideas by selecting cluster value and then adding value to the cluster as a data model for determining the best cluster. Aside from the percentage (Putu, Merliana, and Santoso n.d.), The comparison of the number of clusters added is the result of the calculation. Using a graph as the source of information, different percentages of each cluster value can be shown. If the angle in the graph is formed by the value of the first cluster and the value of the second cluster, or if the value of the second cluster has declined the most, then the value of the cluster is the best.Calculate the SSE (Sum of Square Error) of each cluster value to get a comparison. Because the SSE value will decrease as the number of K clusters increases. K-Means and the SSE formula:

$$SSE = \sum_{K=1}^{K} \sum_{X_i \in S_K} \|X_i - C_K\|_2^2$$

In presenting the K cluster on the K-mean, the elbow method algorithm is provided as follows:

1. Start
2. Initialization of K value
3. Raise the K value
4. Calculate the sum of square error results of each K value
5. See SSE results from a down K value
6. Set an elbow-shaped K value
7. Done

DHT Sensor

The DHT11 sensor is a sensor module that functions to sense temperature and humidity objects that have an analog voltage output that can be further processed using a microcontroller. This sensor module is classified as a resistive element such as a temperature measuring device such as an NTC. DHT11 sensors in general have a calibration feature for temperature and humidity readings that are quite accurate. The calibration data is stored in the OTP program memory which is also known as the calibration coefficient. This sensor has 4 pins. Using proprietary digital signal acquisition technique and temperature & humidity sensing technology, it ensures high prominence and excellent long term stability.

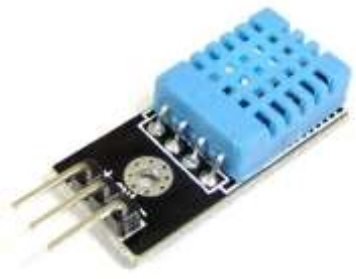

Figure 1. DHT Sensor

MQTT

Message Stream Telemetry Transport is a wired or wireless machine-to-machine communication protocol. MQTT is a TCP/IP-based messaging protocol that has been widely adopted as part of the

Internet of Things (IoT) (Internet of things). MQTT is a lightweight protocol (Hakim and Nurwarsito 2019). Because it can transfer messages fast and consumes little power. MQTT is made up of numerous concepts, including Publish/subscribe. The publish/subscribe concept is used in the MQTT protocol to refer to users who publish messages and users who subscribe to topics. Subscribers to specific linked subjects will receive all messages published on those topics. Users, on the other hand, can publish messages to topics in such a way that they are accessible to all subscribers. The concept of topics and subscribers, on the other hand, indicates that MQTT publishers send messages to subjects that might be deemed the subject of a message.

CONCLUSION

The K-Means Clustering algorithm can be applied and can be grouped using data from data acquisition sensors in unifying air quality. For further research, it can be used with more data and can add variables to be used in order to obtain more accurate results. In addition, for further research, different machine learning algorithms can be used

REFERENCE

Hakim, Fathul, and Heru Nurwarsito. 2019. "Sistem Pemantauan Detak Jantung Dan Suhu Tubuh Menggunakan Protokol Komunikasi MQTT." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 3(11):10705–11.

Kingsy, Grace R., R. Manimegalai, Devasena M. S. Geetha, S. Rajathi, K. Usha, and Baseria N. Raabiathul. 2017. "Air Pollution Analysis Using Enhanced K-Means Clustering Algorithm for Real Time Sensor Data." *IEEE Region 10 Annual International Conference, Proceedings/TENCON* (April 2019):1945–49. doi: 10.1109/TENCON.2016.7848362.

Putu, Ni, Eka Merliana, and Alb Joko Santoso. n.d. "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means." 978–79.

Tuysuzoglu, Goksu, Derya Birant, and Aysegul Pala. 2019. "Majority Voting Based Multi-Task Clustering of Air Quality Monitoring Network in Turkey." *Applied Sciences (Switzerland)* 9(8):1–21. doi: 10.3390/app9081610.