Implementation of Naïve Bayes algorithm in IoT-based water cleanliness monitoring system

K Manaf1*, F M Kaffah2, E Mulyana3 and N Agnia3

1Information Systems, Faculty of Engineering, Universitas Sangga Buana, Jalan PHH Mustofa No.68, Bandung 40124, Indonesia.
2Informatics, Faculty of Engineering Universitas Informatika dan Bisnis Indonesia, Jl. Soekarno-Hatta St No.643, Bandung 40285, Indonesia.
3Informatics, Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung, Jl. A.H. Nasution No.105, Bandung 40614, Indonesia.

*khaerul.manaf@usbypkp.ac.id

Abstract. Internet of Things or commonly referred to as IoT is a concept that has the aim to expand the utilization of the internet. The application of IoT can be done in various fields, both in the fields of science, industry, health, and geographical. One example of the application of IoT in this study is for net AI classification, according to Government Regulation No. 82 of 2001 the classification of water quality is divided into 4 classes, namely classes one, two, three and four. By utilizing IoT and implementing the Naïve Bayes algorithm as a basis for classification, it can facilitate monitoring of water itself whether it is feasible or not for use in everyday life. In addition to applying the Naïve Bayes algorithm as an algorithm for classification, it also designs and builds tools for collecting water data, using temperature, pH, and turbidity sensors. The results obtained are the accuracy of 99.16% for the temperature sensor, 96.89% for the pH sensor and 100% for the water turbidity sensor.

1. Introduction
Technology and Information Systems continue to experience rapid development from time to time along with the emergence of various kinds of problems and existing needs, including actual information needs that are very important for the community [1]. In the current technological era, humans can hardly be separated from the internet, that is because by utilizing the internet, humans can search for various kinds of information they want and need and can communicate and transact whenever and wherever. Internet of Things or commonly referred to as IoT is a concept that has the aim to expand the utilization of the internet. The main challenge of the existence of IoT is how IoT can bridge between the physical world and also the world of information, for example how to process data obtained from electronic equipment through an interface between the user and the electronic device [2].

To determine the class classification of water based on parameters of pH, temperature and turbidity, the Naïve Bayes algorithm is used, which is one of the classification algorithms. Naïve Bayes’s own algorithm according to a British scientist named Thomas Bayes uses probability and statistical methods [3,4]. Besides being easy to make or apply, Naïve Bayes algorithm has a good accuracy value and superior performance in previous studies conducted on the classification of Indonesian language.
2. Research methods

2.1. Data collection techniques

There are two stages that are carried out in data collection, namely:

2.1.1. Observation. Data collection by direct observation to the research location, the intended location includes ponds, rivers, and springs.

2.1.2. Literature study. Data collection by collecting various kinds of related sources and support this research, both sources from books, journals, and literature.

2.2. Software development stage

The method of development carried out in this Final Project research is the prototype development method. The following are the steps in the Prototype method implemented in this study:

2.2.1. Listen to customers. It is a process whereby designing system requirements in accordance with the wishes of the customer [8].

- Functional requirements analysis. Functional requirements are requirements that contain services or processes that must be provided by the system, whether it is how the system must react to an input or how the system behaves in a situation.
- Proposed system model. The proposed system to be built in the form of a merger between the Internet of Things, Arduino, and also the Website as an interface page that contains the Naïve Bayes algorithm for the classification process, a system that works like a researcher and also an expert in the laboratory.
- Data analysis. In this study, the processed data is data from rivers, springs, and ponds.

Based on the explanation of the statistical formula above, it has been obtained the class and the range of data available in this study, the following is the data from the water classification obtained:

Table 1. Classification results of water classes.

| No | Research Object               | Water pH | Water temperature | Turbidity | Water class |
|----|-------------------------------|----------|-------------------|-----------|-------------|
| 1. | Citarum River                 | 7.7      | 27.6              | 1.9       | 2           |
| 2. | Cirebu River                  | 7.6      | 27.7              | 2.0       | 2           |
| 3. | Citarik River                 | 7.8      | 30.3              | 1.8       | 2           |
| 4. | Cipamokolan River             | 7.3      | 28.1              | 2.1       | 2           |
| 5. | Cidurian River                | 7.6      | 27.9              | 2.1       | 2           |
| 6. | Cimaherang Springs            | 6.4      | 30.1              | 0.4       | 1           |
| 7. | Captering Springs             | 6.9      | 25.6              | 0.4       | 1           |
| 8. | Batu Ampar Curug Spring       | 7.6      | 26.6              | 0.1       | 1           |
| 9. | Caringin Springs              | 7.1      | 25.2              | 1.5       | 2           |
| 10. | Cilengkrang waterfall         | 7.8      | 30.2              | 1.3       | 1           |
Table 1. Cont.

|   |                |    |    |    |
|---|----------------|----|----|----|
| 11.| Manisi Fishpond I | 7.2 | 25.1 | 3.1 | 3  |
| 12.| Manisi Fishpond II | 7.3 | 28.3 | 2.9 | 3  |
| 13.| Ibun Fishpond     | 7.6 | 25.7 | 2.2 | 2  |

However, the range of the data is not standard meaning it will change according to the time of the study, this is caused by turbidity sensors that affect the time of the study. The value of the turbidity sensor will be different in the morning, afternoon, evening, and the night because the turbidity sensor influences the presence of light. Analysis Algorithm Naïve Bayes

2.2.2. Use Case diagram. Is a modelling that will illustrate the behaviour of the system to be made, in addition to the use case, it can be seen what functions are on the system and who has the right to carry out these functions [9,10]. Here are some functions of the system:
- Can see water data information
- View the history of water data classification
- Run the naïve Bayes algorithm
- Record water data

2.2.3. Building and improving prototype
This is the stage where the prototype was made and designed using a system architecture, which is explained below [11,12]:
- Clean water is the object of research, where data from water in the form of pH, temperature and turbidity are used as indicators of research.
- PH sensor as a tool to detect and obtain pH from water that is used as the object of research.
- Temperature sensor is a sensor that works as a temperature detector on water which is used as an object of research.
- As the name implies turbidity sensor, this sensor works to obtain data about the turbidity of water that is the object of research.
- The ATmega328P microcontroller as a core tool, which contains program codes and will process data obtained from the sensors that have been described previously, then send the data to the database.
- Is a tool that is a link between Arduino and the internet by utilizing a Wi-Fi network and sending data to the API.
- Internet As a bridge or communication network between Arduino and the classification of water hygiene systems.
- Web Services to conduct or transfer user request files through a communication protocol that has been determined as such.
- Water Hygiene Classification Website Used to display data that has been obtained and for water data processing.

3. Results and discussion

3.1. Turbidity sensor accuracy
The turbidity sensor test is carried out to determine the accuracy of the reading of the data through the turbidity sensor by using Aquades water as a comparison material, which has been known to have a turbidity of less than 1.
Table 2. Testing turbidity sensors.

| No | Sample Sensor   | Test Turbidity Sensor | Error Sensor | Information |
|----|-----------------|-----------------------|--------------|-------------|
| 1  | Aquades water   | 0.62                  | 0.0%         | <=1         |
|    |                 | 0.51                  | 0.0%         | <=1         |
|    |                 | 0.74                  | 0.0%         | <=1         |
|    |                 | 0.28                  | 0.0%         | <=1         |
|    |                 | 0.64                  | 0.0%         | <=1         |
|    |                 | 0.18                  | 0.0%         | <=1         |
|    |                 | 0.39                  | 0.0%         | <=1         |
|    |                 | 0.62                  | 0.0%         | <=1         |
|    |                 | 0.05                  | 0.0%         | <=1         |
|    |                 | 0.16                  | 0.0%         | <=1         |
|    |                 |                       |              | Average error | 0.0 % |

From Table 2 above, it can be concluded that the accuracy value obtained from the turbidity sensor is 100% with the following calculation:

Turbidity sensor accuracy = 100% - average error
Turbidity sensor accuracy = 100% - 0.0% = 100%

3.2. Temperature sensor accuracy

Temperature sensor testing is done by comparing the data obtained from the temperature sensor and the Digital Thermometer, the following is a test table of the temperature sensor:

Table 3. Testing temperature sensors.

| No | Test sample       | Data to Temperature Sensor | Digital Thermometer | Sensor Error |
|----|-------------------|---------------------------|---------------------|-------------|
| 1  | Fishpond          | 1                         | 31.2                | 31.4        | 0.32%       |
|    |                   | 2                         | 31.2                | 31.4        | 0.64%       |
|    |                   | 3                         | 31.2                | 31.4        | 0.64%       |
|    |                   | 4                         | 31.2                | 31.4        | 0.64%       |
|    |                   | 5                         | 31.1                | 31.4        | 0.64%       |
| 2  | Cimaherang Springs| 6                         | 30.1                | 30.1        | 0%          |
|    |                   | 7                         | 30.1                | 30.1        | 0%          |
|    |                   | 8                         | 30.1                | 30.1        | 0%          |
|    |                   | 9                         | 30.2                | 30.1        | 0.33%       |
|    |                   | 10                        | 30.2                | 30.1        | 0.33%       |
| 3  | Citarik River     | 11                        | 30.5                | 30.7        | 0.65%       |
|    |                   | 12                        | 30.3                | 30.7        | 1.3%        |
|    |                   | 13                        | 30.3                | 30.7        | 1.3%        |
|    |                   | 14                        | 30.3                | 30.7        | 1.3%        |
|    |                   | 15                        | 30.3                | 30.7        | 1.3%        |
| 4  | Cilengkrang waterfall | 16                  | 30.3                | 30.7        | 1.3%        |
|    |                   | 17                        | 30.3                | 30.7        | 1.3%        |
|    |                   | 18                        | 30.2                | 30.7        | 1.63%       |
|    |                   | 19                        | 30.2                | 30.7        | 1.63%       |
|    |                   | 20                        | 30.2                | 30.7        | 1.63%       |
|    |                   |                           |                      | Average Error | 0.84% |

Based on Table 3 above, it can be concluded that the accuracy value obtained by the temperature sensor is 99.16% with the following calculation:

Temperature sensor accuracy = 100% - average error
Temperature sensor accuracy = 100% - 0.84% = 99.16%

3.3. **PH sensor accuracy**

The pH sensor test is carried out to determine the accuracy of the pH sensor used in this thesis, by comparing the data obtained from the pH sensor with the data obtained from the pH meter. The following is the data table obtained:

| No | Test sample            | Data to pH | Sensor pH | pH Meter | Sensor Error |
|----|------------------------|------------|-----------|----------|--------------|
| 1. | Cilengkrang waterfall  | 1          | 7.83      | 7.7      | 1.68%        |
|    |                        | 2          | 7.91      | 7.7      | 2.72%        |
|    |                        | 3          | 7.91      | 7.7      | 2.7%         |
|    |                        | 4          | 7.96      | 7.7      | 3.38%        |
|    |                        | 5          | 7.96      | 7.7      | 3.38%        |
| 2. | Cimaherang Springs     | 6          | 6.47      | 6.6      | 1.97%        |
|    |                        | 7          | 6.61      | 6.6      | 0.15%        |
|    |                        | 8          | 6.5       | 6.6      | 1.52%        |
|    |                        | 9          | 6.63      | 6.6      | 0.45%        |
|    |                        | 10         | 6.69      | 6.6      | 1.36%        |
| 3. | Fishpond               | 11         | 7.58      | 7.2      | 5.28%        |
|    |                        | 12         | 7.58      | 7.2      | 5.28%        |
|    |                        | 13         | 7.64      | 7.2      | 6.11%        |
|    |                        | 14         | 7.64      | 7.2      | 6.11%        |
|    |                        | 15         | 7.67      | 7.2      | 6.53%        |
| 4. | Citarik River          | 16         | 7.4       | 7.7      | 3.89%        |
|    |                        | 17         | 8.05      | 7.7      | 4.55%        |
|    |                        | 18         | 7.7       | 7.7      | 0%           |
|    |                        | 19         | 7.8       | 7.7      | 1.29%        |
|    |                        | 20         | 7.99      | 7.7      | 3.77%        |

Average Error = 3.11%

Based on Table 3.3 the accuracy of the pH sensor used is 96.89 with the following calculation:

Sensor accuracy pH = 100% - average error

Sensor accuracy pH = 100% - 3.11% = 96.89%

4. **Conclusion**

Based on research on the implementation of the Naïve Bayes algorithm for the classification of water cleanliness based on the Internet of Things can be concluded as follows:

- The system can monitor pH, temperature and turbidity of water based on the Internet of Things then classify it using the Naïve Bayes algorithm.
- The performance of the sensors used in the process of taking water data has accuracy for each sensor, including 99.16% for temperature sensors, 96.89% for pH sensors and 100% for water turbidity sensors. Besides Naïve Bayes algorithm that is used has an accuracy of 100% this is because the data used is a number so that the training data and the testing data are appropriate.

**Acknowledgment**

Special thanks to the Department of Informatics of UIN Sunan Gunung Djati Bandung and Department System Information USB YPKP that give full support for our published research.
References

[1] Kaffah F M, Slamet C, Rahman A B A, Manaf K and Subaeki B 2019 Framework Index security information to support evaluation adoption and improvement of methods in developing variables In *Journal of Physics: Conference Series* **1402**(7) 077035

[2] Junaidi A 2015 Internet of things sejarah teknologi dan penerapannya *Jurnal Ilmiah Teknologi Infomasi Terapan* **1**(3)

[3] Manaf K, Uriawan W, Agustian W, Gerhana Y A, Jumadi J and Ramdhani M A 2019 Designing futsal match finder application with floyd-warshall algorithm In *Journal of Physics: Conference Series* **1280**(2) 022024

[4] Suyanto D 2017 *Data Mining untuk klasifikasi dan klasterisasi data* (Bandung: Informatika Bandung)

[5] Devita R N, Herwanto H W and Wibawa A P 2018 Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia *J Teknol Inf dan Ilmu Komput*

[6] Alam C N, Manaf K, Atmadja A R and Aurum D K 2016 Implementation of haversine formula for counting event visitor in the radius based on Android application In *2016 4th International Conference on Cyber and IT Service Management* 1-6

[7] Arnomo R A 2017 *Implementasi Algoritma K-Nearest Neigboruntukidentifikasi Kualitas Air (Studi Kasus: PDAM Kota Surakarta)* (Doctoral dissertation) (Surakarta: STMIK Sinar Nusantara Surakarta)

[8] Shalahuddin M and Rosa A S 2013 *Rekayasa perangkat lunak terstruktur dan berorientasi objek* (Bandung: Informatika)

[9] Janner S 2010 *Rekayasa perangkat lunak* (Yogyakarta: Penerbit Andi)

[10] Widodo P P and Prabowo H 2011 *Menggunakan UML* (Bandung: Informatika) p 393-403

[11] Pressman R 2002 *Rekayasa Perangkat Lunak Pendekatan Praktisi* (Yogyakarta: Andi)

[12] Manaf K, Pitara S W, Subaeki B and Gunawan R 2019 Comparison of Carp Rabin Algorithm and Jaro-Winkler Distance to Determine The Equality of Sunda Languages In *2019 IEEE 13th International Conference on Telecommunication Systems Services and Applications (TSSA)* 77-81