

Machine Learning Predicts the Level of Disease Spread

Dhio Saputra^{1*}, Irzal Arief Wisky¹, Sarjon Defit¹

¹ Universitas Putra Indonesia YPTK Padang, Padang, Indonesia.

Received: January 24, 2024

Revised: March 14, 2024

Accepted: April 25, 2024

Published: April 30, 2024

Corresponding Author:

Dhio Saputra

dhiosaputra.mkom@gmail.com

DOI: [10.29303/jppipa.v10i4.7070](https://doi.org/10.29303/jppipa.v10i4.7070)

© 2024 The Authors. This open access article is distributed under a (CC-BY License)



Abstract: The aim of the research is predictive analysis of the spread of disease. Variable analysis at the population level in a region and the total disease events detected in the community. These variables can show the accuracy and certainty of the status of the resulting analysis. The concept of Machine Learning analysis is proposed to develop previous analysis models. The methods used include the K-Means cluster, Naïve Bayes, and Decision Tree (DT). There are two stages in the analysis process: pre-processing and classification. The discussion presented by K-Means provides a classification analysis pattern. The patterns obtained will be passed on to the classification process using Naïve Bayes and DT. Naïve Bayes results provide quite significant results with an accuracy rate of 83.33%. DT can also describe the results of information and knowledge analysis in the form of decision trees. DT produces decision trees that can provide knowledge and information analysis. The DT results provide an accuracy rate of 91.76% so these results can be used as consideration in decision making. The resulting information and knowledge can be used as a guide in making policies for handling health in the community.

Keywords: Decision tree; Machine learning; Naive bayes; Spread of disease

Introduction

The public health management approach makes use of the degree of illness spread as one of its tools (Carroll et al., 2014). According to Wiesinga et al. (2020), this category of disease typically includes tuberculosis, diarrhea, hypertension, influenza, and other illnesses. These illnesses have a rather high percentage of 79.93% that result in agony, paralysis, and even death. Since the spread rate is widespread across the country, it plays a significant role in the present health issues (Bertozzi et al., 2020). The optimal alternative option can be found by using the predictive analysis technique to actively create a model (Sghir et al., 2023; Lepeniotti et al., 2020).

To achieve the intended outcomes, predictive analysis has been extensively developed for a variety of challenges (Lopes et al., 2020). To do prediction analysis, these different models employ a variety of techniques (Toma & Wei, 2023; Shipe et al., 2019). The idea of machine learning (ML) incorporates the analytical model. According to Petropoulos et al. (2022), this model has proven to be quite effective in carrying out the

prediction procedure. When machine learning operates at its best, it produces reasonably accurate output. Growing research on machine learning has produced important graphics that address global health issues. The identification and prediction method reveals this issue. Here, machine learning will also be utilized to conduct a prognostic analysis procedure about the state of illness types' propagation throughout society (Santangelo et al., 2023; Tuli et al., 2020).

K-Means clusters, Naïve Bayes, and Decision Tree (DT) are the techniques that will be employed in the machine learning-based prediction analysis process (Elbasi et al., 2023). This approach has the potential to provide the intended outcomes more successfully. A technique for classifying data based on mathematical computations is the K-Means cluster (Wang et al., 2022; Li et al., 2023). A well-liked technique in the big data notion is K-Means (Mussabayev et al., 2023; Gul & Rehman, 2023). To generate prediction patterns, this method relies on pre-processing. It is possible to demonstrate that the patterns found are useful for completing the predictive analysis procedure. The

How to Cite:

Saputra, D., Wisky, I. A., & Defit, S. (2024). Machine Learning Predicts the Level of Disease Spread. *Jurnal Penelitian Pendidikan IPA*, 10(4), 1714–1722. <https://doi.org/10.29303/jppipa.v10i4.7070>

outcomes of the K-Means cluster pre-processing will be used in the Naïve Bayes idea analysis.

The most recent prediction analysis method described in the analysis model is offered in this study. Preprocessing steps and a prediction method were used to create the model. Furthermore, a two-way strategy for updating the analytical model is described, depending on the population distribution and the number of disease cases in the community. Depending on the results produced, this model can offer updates. In general, the analysis's findings will quantify the degree of precision and error to evaluate sensitivity and performance. In doing so, connected parties can take into account the advantages derived from the analysis process' output when managing public health.

Method

Pre-processing and prediction analysis are the two phases of the machine learning concept-based prediction analysis process. El-Hussein et al. (2021) state that the K-Means cluster, Naïve Bayes, and Decision Tree (DT) are the techniques and algorithms employed. The description of the research stages can be seen in Figure 1.

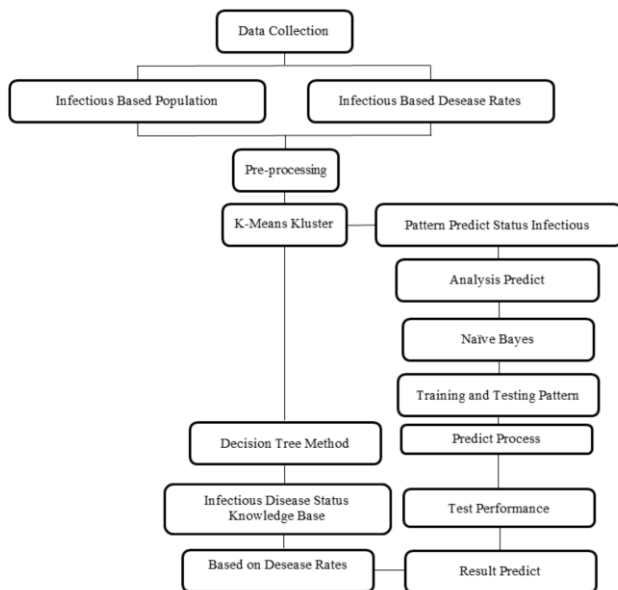


Figure 1. Research stages

The prediction analysis procedure is outlined in Figure 1 and starts with data analysis based on infection cases and population size. To obtain patterns for classification analysis, the K-Means cluster technique is used for pre-processing before moving on to the classification step. Naïve Bayes will be used to carry out the prediction process to produce the analysis pattern, to achieve the highest number of prediction results. The

Decision Tree (DT) approach will be used to gather data and knowledge as the prediction analysis stage progresses.

Data Collection

Disease infection data was taken from 2019 to 2021, Pesisir Selatan District Health Service. Indicators for variable testing are population and number of diseases (Table 1).

Table 1. Variable Analysis of the Infectious Disease Spread Classification Status

Population	Variable	The type of disease	Variable
Population	X1	Ispa	X8
Male	X2	Influenza	X9
Female	X3	Gastritis	X10
(1-12 Year)	X4	Hipertensi	X11
(1-30 Year)	X5	Diarrhea	X12
(31-45 Year)	X6	Rheumatism	X13
> 45 Year	X7	Fever	X14
		Common Cold	X15
		Asthma	X16
		Dengue Fever	X17
		Tuberculosis	X18
		Dyspepsia	X19
		Skin Allergies	X20

K-Means Cluster

K-Means Cluster works by using information and knowledge to identify data patterns. K-Means Cluster is used in machine learning in formula 1.

$$he\ the\ \sum_{j=1}^k \sum_{xi \in} ||Xi - \mu||_2^2 \tag{1}$$

Where:

- K = index cluster
- J = number of clusters
- Xi = Cluster data

Naïve Bayes

The Naive Bayes method, sometimes known as Bayes' Theorem, forecasts future possibilities based on past performance. The Naïve Bayes Classifier's primary feature is its extremely strong (naïve) assumption of each condition's independence, which enables us to generate an exhaustive review. By putting this idea into practice, predictive analysis can learn more effectively and produce better results (Herodotou et al., 2019). Equation 2, displays the Naive Bayes Algorithm equations.

$$P(C|X) = \frac{p(x|c)P(c)}{P(x)} \tag{2}$$

Where:

- x : Data with unknown class
- c : Data hypothesis is a specific class
- P(c | x): Probability of a hypothesis based on conditions (posterior probability)
- P(c) : Hypothesis probability (prior probability)
- P(x | c): Probability based on the conditions in the hypothesis
- P(x) : Probability c

Decision Tree (DT)

DT uses data filters in making decisions and testing validation. DT in the form of decision trees plays a role in helping complex data and generating insights and information. DT works with mathematical principles in making decisions (Parra et al., 2023). The DT equation can be seen in equation 3.

$$\text{entropy}(S) = -\sum_{i=1}^c PS(c_i) \log PS(c_i) \tag{3}$$

Where:

- S = Set
- A = Features
- c i = Partition
- PS = Set Proposition

Result and Discussion

In machine learning, the Naïve Bayes algorithm is a popular concept. According to Taye (2023) and Sarker (2021a), this approach is a supervised learning idea that can yield outcomes with a respectable level of accuracy. Alzubaidi et al. (2021) and Ciaburro et al. (2022) define training data as pre-existing data, also known as original data, and training data as data that will be evaluated on pre-existing data to produce output. Naïve Bayes performs analysis based on these two types of data. Since many issues have been effectively overcome, this concept is still evolving. Naïve Bayes is predicted to perform optimally when examining predictions of disease transmission. The learning process will be

optimized to generate output by maximizing the training and testing phases (Li et al., 2022). This will yield the desired results. The degree of accuracy and error in the output presentation shows the naïve Bayes performance. The approach will yield output results that will then undergo additional analysis to furnish the necessary information and understanding (Ahmed et al., 2023).

The Decision Tree (DT) idea will be employed in the analytic process going forward (Lin & Fan, 2019; Purwanto et al., 2022). It is envisaged that machine learning, which uses a variety of ways to conduct analysis, would be able to offer a structured process and produce output that is exact and correct (Javaid et al., 2022; Aldoseri et al., 2023).

Machine learning can be used to create predictive models in physical science, such as predicting the path of particles in nuclear experiments (Huang et al., 2023). Machine Learning has become an important part of many industries and sectors, including the nuclear field. In today's digital era, a lot of data is generated from various nuclear systems and equipment, and the use of Machine Learning can help in collecting, analyzing, and utilizing that data to optimize nuclear operations and improve safety. In the nuclear field, Machine Learning can be used to develop nuclear accident prediction models, incorporate radiation levels, optimize the use of nuclear fuel, and even in the development of new technologies for nuclear reactors (Sandhu et al., 2023).

Pre-Processing Analysis

Optimizing the prediction process is the goal of the pre-processing analysis stage. superior and more organized analysis can be obtained with this technique, leading to superior output outcomes. The K-Means cluster technique is employed in this pre-processing analysis. Data can be grouped by this method according to how closely related the data are to one another. Table 2 displays the findings of the pre-processing analysis performed with K-Means clusters.

Table 2. K-Means Cluster Preprocessing Results

Region	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Y
Silaut	15.09	7.88	7.20	3.32	2.52	2.09	564	218	144	154	165	46	56	208	96	26	94	3	0	71	High
Lunang	21.53	10.97	10.55	4.73	3.51	3.06	805	26	75	48	87	57	121	165	34	67	87	4	0	56	High
Basa Ampek Balai Tapan	13.79	6.84	6.94	3.03	2.19	2.01	516	32	156	86	71	10	298	89	35	74	66	9	0	28	High
Ranah Ampek Hulu Tapan	15.15	7.58	7.57	3.33	2.42	2.19	567	326	65	168	92	55	156	0	102	67	0	0	0	0	High
Pancung Soal	26.49	13.43	13.05	5.82	4.30	3.78	991	520	220	375	290	0	315	540	240	46	45	4	0	187	High
Airpura	16.01	8.07	7.94	3.52	2.58	2.30	599	71	73	147	166	0	68	174	0	0	0	0	0	0	High

Region	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	Y
Linggo Sari Baganti	45.18	22.58	22.59	9.94	7.22	6.55	1.69	302	134	111	117	52	121	0	0	0	0	0	0	0	Low
Ranah Pesisir	30.31	14.74	15.56	6.66	4.71	4.51	1.13	730	0	268	285	73	254	45	49	88	34	8	34	292	High
Lengayang Sutera	52.59	25.72	26.87	11.57	8.23	7.79	1.96	801	340	693	544	86	637	38	0	130	56	5	0	254	Middle
Batang Kapas	50.51	25.20	25.31	11.11	8.06	7.34	1.88	862	0	412	231	108	362	220	0	98	48	6	0	177	Middle
IV Jurai Bayang	31.48	15.47	16.00	6.92	4.95	4.64	1.17	570	311	562	419	0	302	0	91	0	0	0	0	112	Low
IV Nagari Bayang Utara	46.50	23.02	23.48	10.23	7.36	6.80	1.73	367	233	287	194	38	114	180	0	24	0	0	0	0	Low
Koto XI Tarusan	36.78	17.68	19.09	8.09	5.65	5.53	1.37	401	69	264	149	87	317	217	75	81	66	5	16	72	Low
	7.27	3.47	3.80	1.60	1.11	1.10	272	270	0	128	72	56	56	346	87	87	62	12	0	121	High
	48.55	24.07	24.47	10.68	7.70	7.09	1.81	296	56	180	163	79	165	242	47	68	0	0	0	131	Low

According to Table 2, the cluster results offer a pattern for classification analysis based on data groups at the infection rates by illness type status level. A level of distribution with 8 items in high status (C1), 2 items in medium status, and 5 things in low status are displayed by the cluster results. Table 2 shows that there are three status categories—high, medium, and low status—for the transmission of infectious diseases. A classification procedure will be used with the pre-processed data to determine the distribution of infection cases by disease type.

Prediction Analysis

According to the number of afflicted cases, the prediction process in the discussion seeks to determine

the degree of disease type spread (Ahmad et al., 2021; Keshavamurthy et al., 2022). In this instance, the Naïve Bayes method—which can learn with superior output results—is used to start the analytic process with training and test data. Using the same learning concept—using training and test data—Naïve Bayes can also be applied in the situation of disease prediction. There is a reasonable amount of accuracy in the outcomes. To do testing, this approach first learns the patterns that have evolved in the training data. The goal of this learning is to obtain the optimal choices for the prediction analysis procedure. Figure 2 illustrates the steps of the Naïve Bayes prediction method.

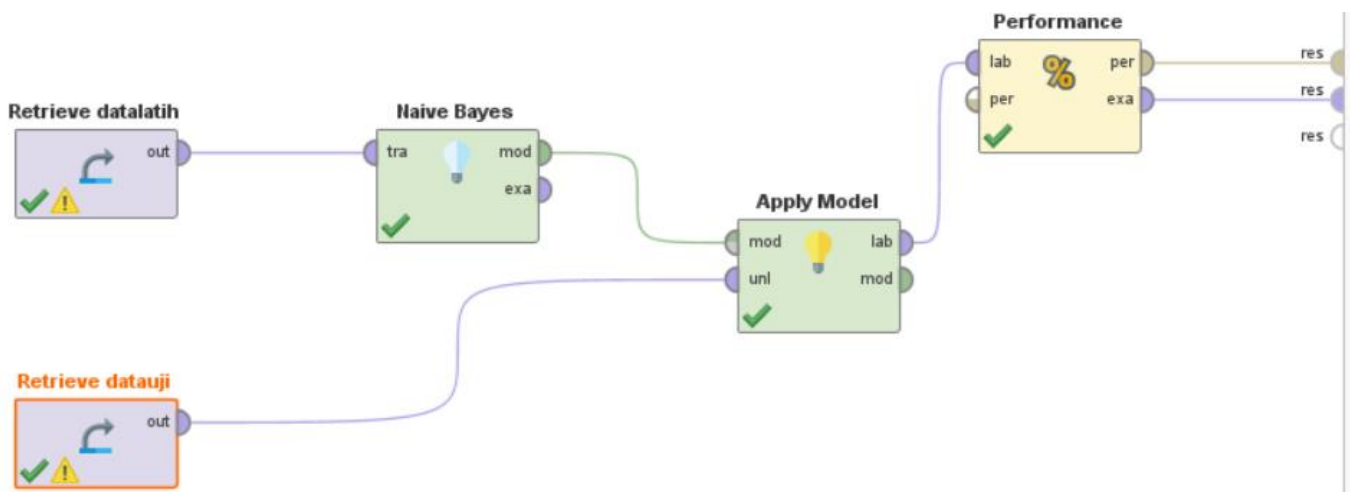


Figure 2. Naïve bayes process

In Figure 2, a naïve Bayes process is depicted. Training data is connected to the process, and it is then connected to an applied model that contains test data already, allowing us to assess the method's performance. Figure 3 shows the outcomes of the Naïve Bayes procedure.

The output of the Naïve Bayes prediction study, which produced quite decent results, is explained in Figure 3. The accuracy number of 83.33% indicates these outcomes, indicating that the Naïve Bayes method approaches maximum performance in carrying out the prediction process. A plot view that shows the distribution of the data being processed can also be used

to demonstrate the effectiveness of Naïve Bayes output. These findings are sufficient to demonstrate that Naïve Bayes is capable of performing a predictive analytic

procedure on the state of the various disease types' rate of spread.

Table View Plot View

Table View Plot View

Confusion Matrix (x: true class, y: pred. class, z: counters)

accuracy: 83.33%

	true Yes	true No	class precision
pred. Yes	3	0	100.00%
pred. No	1	2	66.67%
class recall	75.00%	100.00%	

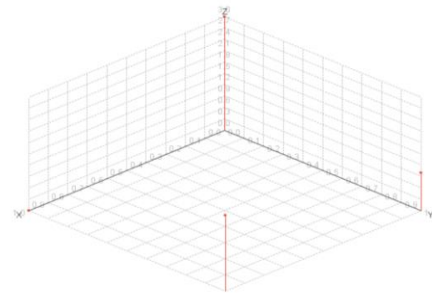
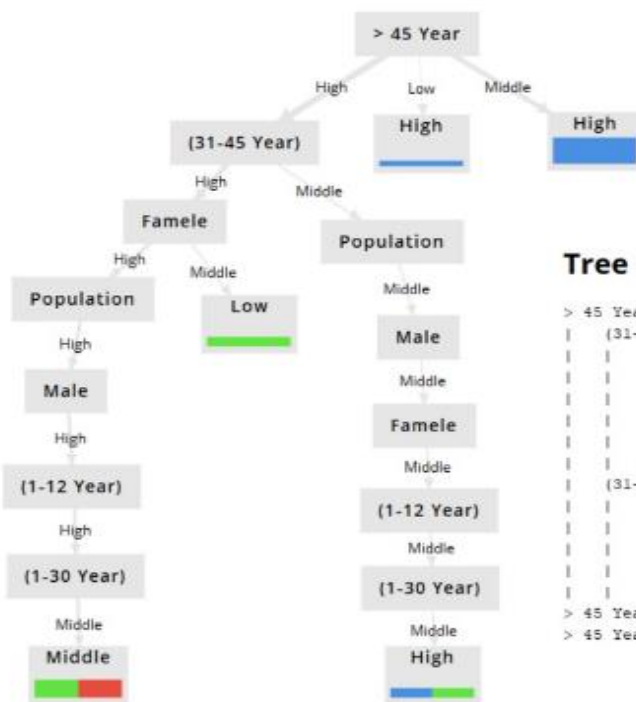


Figure 3. Naïve bayes results

The output of the Naïve Bayes prediction study, which produced quite decent results, is explained in Figure 3. The accuracy number of 83.33% indicates these outcomes, indicating that the Naïve Bayes method approaches maximum performance in carrying out the prediction process. A plot view that shows the distribution of the data being processed can also be used to demonstrate the effectiveness of Naïve Bayes output. These findings are sufficient to demonstrate that Naïve Bayes is capable of performing a predictive analytic procedure on the state of the various disease types' rate of spread.

The next step in the analysis process will be to investigate information derived from Naïve Bayes-analyzed prediction patterns. A knowledge-based format for output can be presented by the Decision Tree (DT) approach. According to its concept, DT analyzes data to uncover knowledge and hidden information. The DT idea will be applied to a classification analysis procedure that will concentrate on two dimensions: population level and number distribution numbers (Rupp et al., 2024; Ishaque et al., 2023). This two-way study aims to identify data and knowledge as perceived from many angles.



Tree

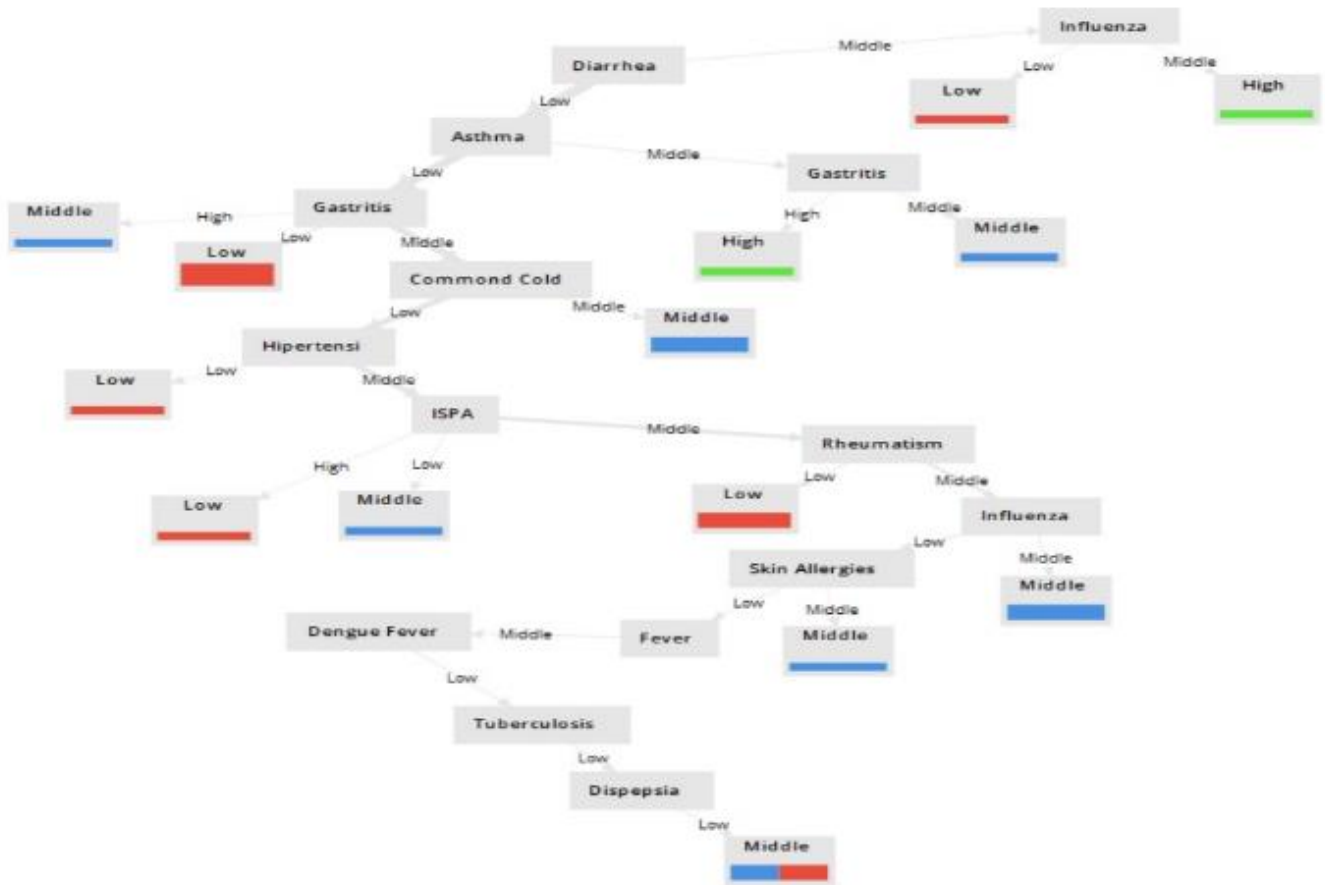
```

> 45 Year = High
| (31-45 Year) = High
| | Famele = High
| | | Population = High
| | | | Male = High
| | | | | (1-12 Year) = High
| | | | | | (1-30 Year) = Middle: Middle (High=0, Low=2, Middle=2)
| | | | | | Famele = Middle: Low (High=0, Low=2, Middle=0)
| | | | | | (31-45 Year) = Middle
| | | | | | Population = Middle
| | | | | | | Male = Middle
| | | | | | | Famele = Middle
| | | | | | | | (1-12 Year) = Middle
| | | | | | | | (1-30 Year) = Middle: High (High=1, Low=1, Middle=0)
| | | | | | | | (1-30 Year) = Middle: High (High=1, Low=0, Middle=0)
| | | | | | | | (1-30 Year) = Middle: High (High=6, Low=0, Middle=0)
    
```

Figure 4. Results of decision tree analysis based on population level

In Figure 4, DT takes the form of a decision tree that produces knowledge and information. Residents aged > 45 years have a greater chance of contracting the disease. Moderate probability for populations < 30 years old and

populations aged 31-45 years. Analytical techniques are used to estimate the level of disease spread (Figure 4). An illustration of the analysis of the findings can be seen in Figure 5.



Tree

```

Diarrhea = Low
| Asthma = Low
| | Gastritis = High: Middle (Middle=1, High=0, Low=0)
| | Gastritis = Low: Low (Middle=0, High=0, Low=3)
| | Gastritis = Middle
| | | Common Cold = Low
| | | | Hipertensi = Low: Low (Middle=0, High=0, Low=1)
| | | | Hipertensi = Middle
| | | | | ISPA = High: Low (Middle=0, High=0, Low=1)
| | | | | ISPA = Low: Middle (Middle=1, High=0, Low=0)
| | | | | ISPA = Middle
| | | | | Rheumatism = Low: Low (Middle=0, High=0, Low=2)
| | | | | Rheumatism = Middle
| | | | | Influenza = Low
| | | | | | Skin Allergies = Low
| | | | | | Fever = Middle
| | | | | | Dengue Fever = Low
| | | | | | Tuberculosis = Low
| | | | | | Dispepsia = Low: Middle (Middle=1, High=0, Low=1)
| | | | | | Skin Allergies = Middle: Middle (Middle=1, High=0, Low=0)
| | | | | | Influenza = Middle: Middle (Middle=2, High=0, Low=0)
| | | | | Common Cold = Middle: Middle (Middle=2, High=0, Low=0)
| Asthma = Middle
| | Gastritis = High: High (Middle=0, High=1, Low=0)
| | Gastritis = Middle: Middle (Middle=1, High=0, Low=0)
Diarrhea = Middle
| Influenza = Low: Low (Middle=0, High=0, Low=1)
| Influenza = Middle: High (Middle=0, High=1, Low=0)
    
```

Figure 5. Results of decision tree analysis based on the rate of spread of disease types

The results are shown in Figure 5 using information consistent with the analysis's findings. Based on the quantity of infection cases submitted, these outcomes are shown. In this instance, machine learning-based predictive analysis can characterize the disease-specific distribution status. The ID3 and C4.5 algorithms are Decision Tree (DT) methods for determining electrical power in physics (Tanjung et al., 2016).

Conclusion

Predictive analysis using machine learning concepts provides quite good results in presenting information about the spread of cases of disease types. The two ways that the analytical process's result is presented are through classification based on information about the population's distribution and the number of infected cases. These results were obtained through initial pre-processing using K-Means clusters to produce classification analysis patterns which will then become an illustration of Naïve Bayes learning. Where from these results Naïve Bayes provides a percentage of output accuracy or an accuracy level of 83.33%. The prediction results obtained can also be described in the form of a decision tree using the decision tree method with an average level of accuracy of 91.76%. The resulting decision tree contains a knowledge base that can be used as a control medium to handle spikes in the number of cases of disease spread. In this way, the analysis results obtained can be used as choices in decision-making.

Acknowledgments

The author would like to thank all forms of participation and support in writing the journal. We are also aware that there are still many shortcomings and errors, so we need suggestions and input for the perfection of this article.

Author Contributions

Conceptualization, D. S., I. A. W., and S. D., methodology, D. S.; validation, I. A. W.; formal analysis, S. D.; investigation, D. S.; resources, I. A. W. and S. D.; data curation, D. S.; writing – original draft preparation, I. A. W. and S. D.; writing – review and editing, D. S.; visualization, I. A. W. and D. S. All authors have read and agreed to the published version of the manuscript.

Funding

This research was independently funded by researchers.

Conflicts of Interest

The authors declare no conflict of interest.

References

Ahmad, A., Garhwal, S., Ray, S. K., Kumar, G., Malebary, S. J., & Barukab, O. M. (2021). The

Number of Confirmed Cases of Covid-19 by using Machine Learning: Methods and Challenges. *Archives of Computational Methods in Engineering*, 28(4), 2645–2653. <https://doi.org/10.1007/s11831-020-09472-8>

- Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modeling techniques: Current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, 56(11), 13521–13617. <https://doi.org/10.1007/s10462-023-10466-8>
- Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. *Healthcare*, 10(3), 541. <https://doi.org/10.3390/healthcare10030541>
- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, 13(12), 7082. <https://doi.org/10.3390/app13127082>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *Journal of Electrical Systems and Information Technology*, 10(1), 40. <https://doi.org/10.1186/s43067-023-00108-y>
- Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117(29), 16732–16738. <https://doi.org/10.1073/pnas.2006520117>
- Bukhari, S. N. H., Webber, J., & Mehbodniya, A. (2022). Decision tree-based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates. *Scientific Reports*, 12(1), 7810. <https://doi.org/10.1038/s41598-022-11731-6>
- Carroll, L. N., Au, A. P., Detwiler, L. T., Fu, T., Painter, I. S., & Abernethy, N. F. (2014). Visualization and analytics tools for infectious disease epidemiology: A systematic review. *Journal of Biomedical Informatics*, 51, 287–298. <https://doi.org/10.1016/j.jbi.2014.04.006>
- Chakri, P., Pratap, S., Lakshay, & Gouda, S. K. (2023). An exploratory data analysis approach for analyzing financial accounting data using machine learning.

- Decision Analytics Journal*, 7, 100212. <https://doi.org/10.1016/j.dajour.2023.100212>
- Ciaburro, G., & Iannace, G. (2022). Machine-Learning-Based Methods for Acoustic Emission Testing: A Review. *Applied Sciences*, 12(20), 10476. <https://doi.org/10.3390/app122010476>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences*, 13(16), 9288. <https://doi.org/10.3390/app13169288>
- Elhussein, M., & Brahimi, S. (2021). Clustering as a feature selection method in spam classification: Uncovering sick-leave sellers. *Applied Computing and Informatics*. <https://doi.org/10.1108/ACI-09-2021-0248>
- Gul, M., & Rehman, M. A. (2023). Big data: An optimized approach for cluster initialization. *Journal of Big Data*, 10(1), 120. <https://doi.org/10.1186/s40537-023-00798-1>
- Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). Large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development*, 67(5), 1273–1306. <https://doi.org/10.1007/s11423-019-09685-0>
- Ishaque, S., Khan, N., & Krishnan, S. (2023). Physiological Signal Analysis and Stress Classification from VR Simulations Using Decision Tree Methods. *Bioengineering*, 10(7), 766. <https://doi.org/10.3390/bioengineering10070766>
- Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Gonzalez, E. S. (2022). Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability. *Sustainable Operations and Computers*, 3, 203–217. <https://doi.org/10.1016/j.susoc.2022.01.008>
- Keshavamurthy, R., Dixon, S., Pazdernik, K. T., & Charles, L. E. (2022). Predicting infectious disease for preparedness and response: A systematic review of machine learning and deep learning approaches. *One Health*, 15, 100439. <https://doi.org/10.1016/j.onehlt.2022.100439>
- Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50, 57–70. <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>
- Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology, an International Journal*, 29, 101021. <https://doi.org/10.1016/j.jestch.2021.06.001>
- Li, M., Frank, E., & Pfahringer, B. (2023). Large-scale K-means clustering using GPUs. *Data Mining and Knowledge Discovery*, 37(1), 67–109. <https://doi.org/10.1007/s10618-022-00869-6>
- Lin, C.-L., & Fan, C.-L. (2019). Evaluation of CART, CHAID, and QUEST algorithms: A case study of construction defects in Taiwan. *Journal of Asian Architecture and Building Engineering*, 18(6), 539–553. <https://doi.org/10.1080/13467581.2019.1696203>
- Lopes, J., Guimarães, T., & Santos, M. F. (2020). Predictive and Prescriptive Analytics in Healthcare: A Survey. *Procedia Computer Science*, 170, 1029–1034. <https://doi.org/10.1016/j.procs.2020.03.078>
- Mussabayev, R., Mladenovic, N., Jarboui, B., & Mussabayev, R. (2023). How to Use K-means for Big Data Clustering? *Pattern Recognition*, 137, 109269. <https://doi.org/10.1016/j.patcog.2022.109269>
- Parra, X., Tort-Martorell, X., Alvarez-Gomez, F., & Ruiz-Viñals, C. (2023). Chronological Evolution of the Information-Driven Decision-Making Process (1950–2020). *Journal of the Knowledge Economy*, 14(3), 2363–2394. <https://doi.org/10.1007/s13132-022-00917-y>
- Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., & Ziel, F. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*, 38(3), 705–871. <https://doi.org/10.1016/j.ijforecast.2021.11.001>
- Purwanto, A. D., Wikantika, K., Deliar, A., & Darmawan, S. (2022). Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia. *Remote Sensing*, 15(1), 16. <https://doi.org/10.3390/rs15010016>
- Rupp, N., Ries, R., Wienbruch, R., & Zuchner, T. (2024). Can I benefit from laboratory automation? A decision aid for the successful introduction of laboratory automation. *Analytical and Bioanalytical Chemistry*, 416(1), 5–19. <https://doi.org/10.1007/s00216-023-05038-2>
- Saha, D., & Manickavasagan, A. (2021). Machine learning techniques for analysis of hyperspectral images to determine the quality of food products: A review. *Current Research in Food Science*, 4, 28–44. <https://doi.org/10.1016/j.crfs.2021.01.002>
- Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D., & Cedrone, F. (2023). Machine Learning and

- Prediction of Infectious Diseases: A Systematic Review. *Machine Learning and Knowledge Extraction*, 5(1), 175–198.
<https://doi.org/10.3390/make5010013>
- Sarker, I. H. (2021a). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Sarker, I. H. (2021b). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(3), 160. <https://doi.org/10.1007/s42979-021-00592-x>
- Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A Decade systematic review (2012–2022). *Education and Information Technologies*, 28(7), 8299–8333. <https://doi.org/10.1007/s10639-022-11536-0>
- Shipe, M. E., Deppen, S. A., Farjah, F., & Grogan, E. L. (2019). Developing prediction models for clinical use using logistic regression: An overview. *Journal of Thoracic Disease*, 11(S4), S574–S584. <https://doi.org/10.21037/jtd.2019.01.25>
- Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*, 12(5), 91. <https://doi.org/10.3390/computers12050091>
- Toma, M., & Wei, O. C. (2023). Predictive Modeling in Medicine. *Encyclopedia*, 3(2), 590–601. <https://doi.org/10.3390/encyclopedia3020042>
- Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of the COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, 11, 100222. <https://doi.org/10.1016/j.iot.2020.100222>
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 281. <https://doi.org/10.1186/s12911-019-1004-8>
- Wang, C.-L., Chan, Y.-K., Chu, S.-W., & Yu, S.-S. (2022). R-Reference points based k-means algorithm. *Information Sciences*, 610, 204–214. <https://doi.org/10.1016/j.ins.2022.07.166>
- Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J., & Prescott, H. C. (2020). Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, 324(8), 782. <https://doi.org/10.1001/jama.2020.12839>