# PARADIGMA

# Prediction of Infant Mortality Using The Decision Tree and Genetic Algorithm Methods

**Suhardjono[1], Adjat Sudradjat[2*], Bilal Abdul Wahid[3], Hari Sugiarto[4], Hafis Nurdin[5]**

[1,2,3,4]Universitas Bina Sarana Informatika, Jakarta, Indonesia
[5]Universitas Nusa Mandiri, Jakarta, Indonesia

| ARTICLE INFORMATION | ABSTRACT |
|---|---|
| | *One of the things that play a role in reducing infant mortality is the government. Based on infant mortality data in Jakarta in 2018 that has been previously tested with the decision tree algorithm, the update in this study is to use the genetic algorithm. The purpose of the update is to increase the accuracy of the results to be maximized. From the test results with the DT algorithm optimized by GA, the maximum accuracy value is 100%, and each attribute has a weight value of 1 where the value is the maximum value. After obtaining maximum results, the data will be used to reduce infant mortality, especially in Jakarta.* |

*Corresponding Author:*

Suhardjono,
Sistem Informasi Kampus Sukabumi,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No.98, Senen, Central Jakarta, Indonesia
Email: suhardjono@bsi.ac.id

## INTRODUCTION

Living things, especially humans, are bound to undergo life and death, because the process has been destined by God. Human death can cause sadness to other humans, especially here the death of a baby which will cause a mother to become prolonged sadness, stress and will even cause mental disorders. Sudden infant death is a major factor caused by syndromes that can trigger an increase in infant mortality.

The infant mortality rate is getting higher day by day with the number of deaths for every thousand births of babies there are 24 babies who die. The death of the baby after a mother gives birth to a baby and has not reached the age of the baby 1 year, with the death rate described by the head of the national population and family planning agency which can be abbreviated to (BKKBN). The Indonesian government is making efforts to prevent or reduce infant mortality, because the birth of a baby is something that every human wants, especially newlyweds (Kohno et al., 2020). Infant deaths in this case are infant deaths in Jakarta

caused by several factors, especially diseases in pregnancy (Salina et al., 2019). Birth complications and congenital diseases from a mother to the baby (Kusumawardani & Handayani, 2018). Even nutritional status, age of mother and baby and antenatal care, especially the baby's weight (Lengkong, G.T., Langi, F.L.F.G dan Posangi, 2020). And the position when the baby sleeps that is not normal can result in infant death because it causes breathing problems (Abirami & Karthika, 2018). By studying infant mortality in Jakarta, which is different, it takes a long time and errors often occur, especially the large number of Jakarta residents.

Most of the cases of infant mortality can be prevented early with evidence of high-quality data, which in this study used government data. Research on infant mortality data has been carried out with the same data by classifying the data first into four parts, namely very many, many, few and very few. Data that has been grouped will be predicted using the decision tree (DT) algorithm and produces high accuracy (Ridwansyah et

al., 2022). Because testing with the DT method can produce high accuracy (Priyatama & Satya, 2022).

The main objective in this study is to reduce the number of infant deaths by increasing accuracy to the maximum. In this research, we are interested in improving the accuracy of infant mortality data obtained from data.go.id by using the best algorithm called Genethic Algorithm (GA). Existing parameters can affect accuracy with the GA algorithm (Ridwansyah et al., 2020). By comparing the final results, namely the accuracy results in DT with the help of the GA algorithm and using cross validation.

**RESEARCH METHODOLOGY**

1. Flow of Methods used

The research was conducted based on the following steps:
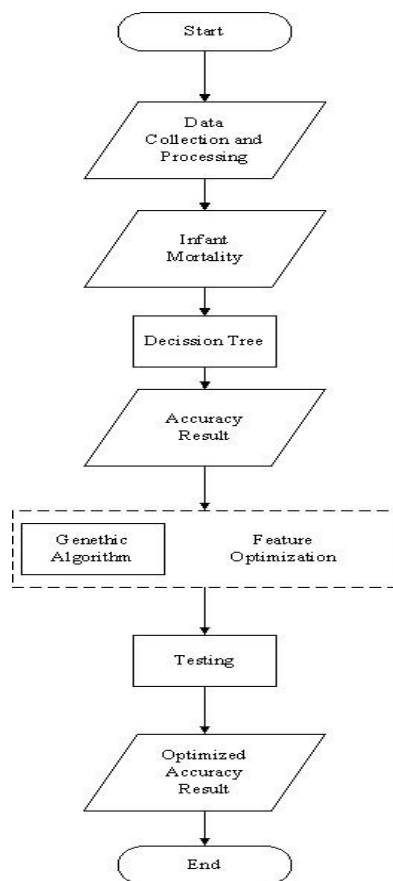


Source: Research Results
Figure 1. Flow diagram of optimization

In the research, the first thing to do is the process of obtaining data and preparing valid data by discarding invalid data, so that the data obtained is ready to be used by obtaining infant mortality data. From the infant death data, the next process is to process the data with the decision tree method to get the results of the method's accuracy. From the results of the decision tree method, a feature optimization process

will be carried out using the genethic algorithm optimization method in testing. From the test results, we get optimal accuracy.

A. Infant Mortality Database

The infant mortality database is obtained from the database on the official government website data.go.id in conducting this research the data on the website is taken based on the city of Jakarta in 2018. This data includes 6 attributes and 476 records, of these data only 5 attributes and 403 records are processed because there is missing value data, so it must be removed manually. The attributes contain the city, sub-district, village, gender and number of deaths in the infant mortality data.

B. Decission Tree

Of the several algorithms that are often used to process data, namely decision tree (DT) (Riyanto et al., 2019). DT is a supervised learning algorithm (Purnama et al., 2020). And the data must vary, if the data complexity is not diverse then it cannot be done in the classification process. After the data is diverse, it will be tested using the decision tree algorithm, after testing, the accuracy results of infant mortality data using the DT algorithm are obtained.

C. Feature Optimization

From several studies, there are many sophisticated optimization algorithms that are often used from several algorithms including PSO, GA, ABC, WA, CSA and others (Elhariri et al., 2020). Algorithms used to find and solve problems from features (Hammami et al., 2019) optimization because the performance process of complex problems becomes more efficient (Elhariri et al., 2020). The feature used to optimize is the genetich algorithm. Infant mortality data after using the DT algorithm will be carried out using the optimization model, namely GA and will produce optimized accuracy.

**RESULTS AND DISCUSSION**

In conducting research related to data processing experiments to investigate and evaluate the performance of decision trees and genethic algorithms proposed for feature optimization with an approach using rapid miner tools, the following results are obtained.

A. Data Infant Mortality

The data in this study have 5 attributes and 1 class which can be seen in Figure 2.

| kota | kecamatan | kelurahan | jenis kelamin | jumlah | class |
|---|---|---|---|---|---|
| KAB.ADM.KEP.SERIBU | KEP. SERIBU UTR | P. HARAPAN | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | GAMBIR | PETOJO UTARA | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | GAMBIR | DURI PULO | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | SAWAH BESAR | PASAR BARU | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | SAWAH BESAR | GUNUNG SAHARI UTARA | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | SAWAH BESAR | MANGGA DUA SELATAN | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | KEMAYORAN | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | KEBON KOSONG | Laki-Laki | 3 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | HARAPAN MULIA | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | SERDANG | Laki-Laki | 4 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | GUNUNG SAHARI SELATAN | Laki-Laki | 4 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | CEMPAKA BARU | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | SUMUR BATU | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | KEMAYORAN | UTAN PANJANG | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | SENEN | PASEBAN | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | SENEN | KRAMAT | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | SENEN | KWITANG | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | CEMPAKA PUTIH | CEMPAKA PUTIH TIMUR | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | MENTENG | MENTENG | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | MENTENG | PEGANGSAAN | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | MENTENG | KEBON SIRIH | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | TANAH ABANG | KARET TENGSIN | Laki-Laki | 2 | Sangat Banyak |
| JAKARTA PUSAT | TANAH ABANG | PETAMBURAN | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | TANAH ABANG | KEBON MELATI | Laki-Laki | 1 | Sangat Banyak |
| JAKARTA PUSAT | TANAH ABANG | KEBON KACANG | Laki-Laki | 1 | Sangat Banyak |

Source: Research Results

Figure 2. Infant mortality data set

The data shown in Figure 2 is clean and ready to be processed because the process of deleting incomplete data has been carried out and has no missing values or data so that the process becomes effective during data classification. The existing data is data with city, sub-district and urban village attributes in the capital city of DKI Jakarta, sex of babies and the number of infant deaths that occurred in the capital city of DKI Jakarta. And 1 class attribute which has 4 categories, namely: very low, low, many and very many. From the data that has been described, it will be tested using different methods.

B. Decission Tree

At this stage, infant mortality data is processed using the decision tree algorithm and produces accuracy results in the form of a confusion matrix which can be seen in table 1.

Table 1. DT Accuracy results

|  | True Very much (%) | True Very Little(%) | True Many(%) | True a Little(%) | Class Precission (%) |
|---|---|---|---|---|---|
| Pred Very much | 275 | 0 | 0 | 0 | 100 |
| Pred Very Little | 0 | 20 | 0 | 0 | 100 |
| Pred Many | 0 | 0 | 87 | 0 | 100 |
| Pred a Little | 1 | 0 | 0 | 20 | 95,24 |
| Class Recall | 99,64 | 100 | 100 | 100 |  |

Source: Research results

From table 1 it can be said that what does not match the prediction results is 1 data where the data is stated to be very much but becomes a little. From these prediction errors, it can produce an accuracy of 99.75% and with a kappa value of 0.995%. from this accuracy, an accuracy matrix diagram can be drawn as follows.
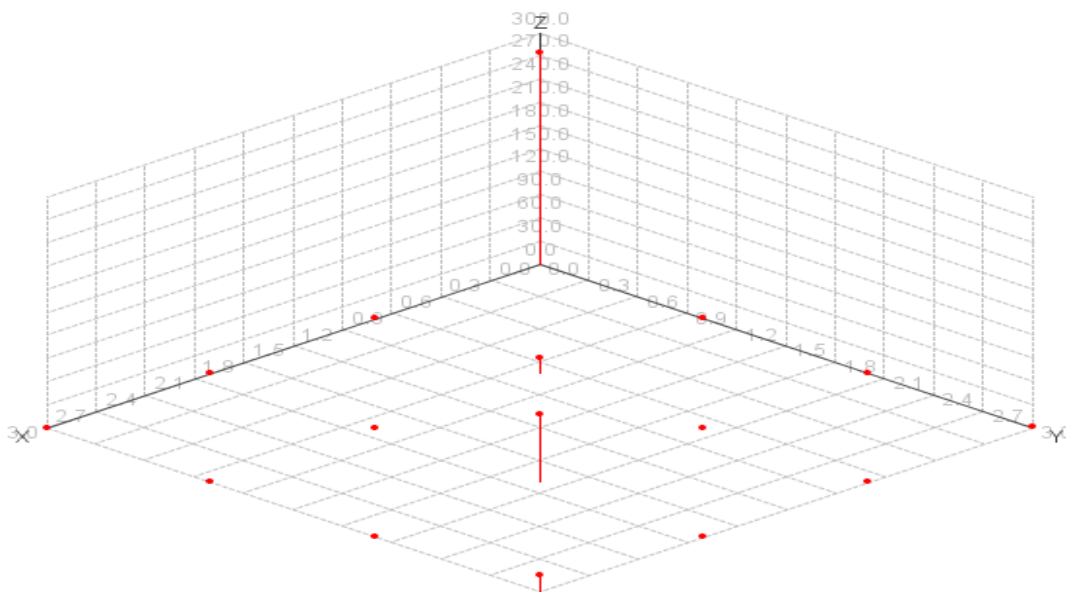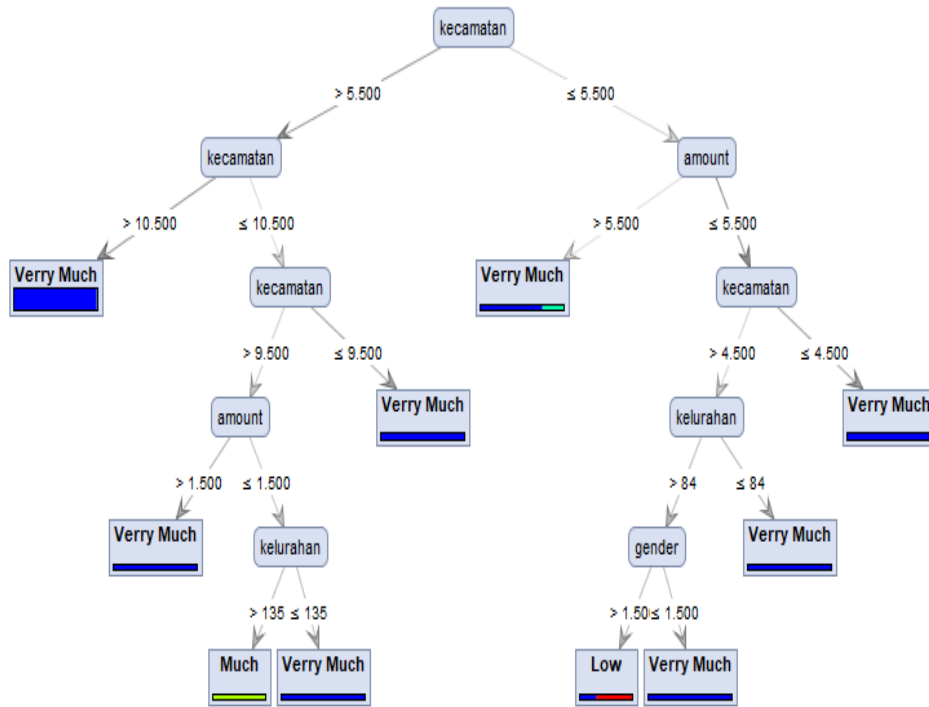


Source: Research Results

Figure 3. Accuracy matrix diagram for decision tree

From fig 3. It can be seen that the accuracy produced in the image shows the highest accuracy point from the X axis to the Y axis with the symbol Z 275 which is the appropriate prediction. From the accuracy matrix diagram, a decision tree is also formed to predict infant mortality data as follows.

Source: Research Results

Figure 4. Infant mortality data with decision tree

From the decision tree, the root node is the urban village level with 99.75% accuracy and will be improved with GA optimization features.

| | | |
|---|---|---|
| Accuracy | 99,75 | 100 |
| Kappa | 0,995 | 1,000 |

Source: Research Results

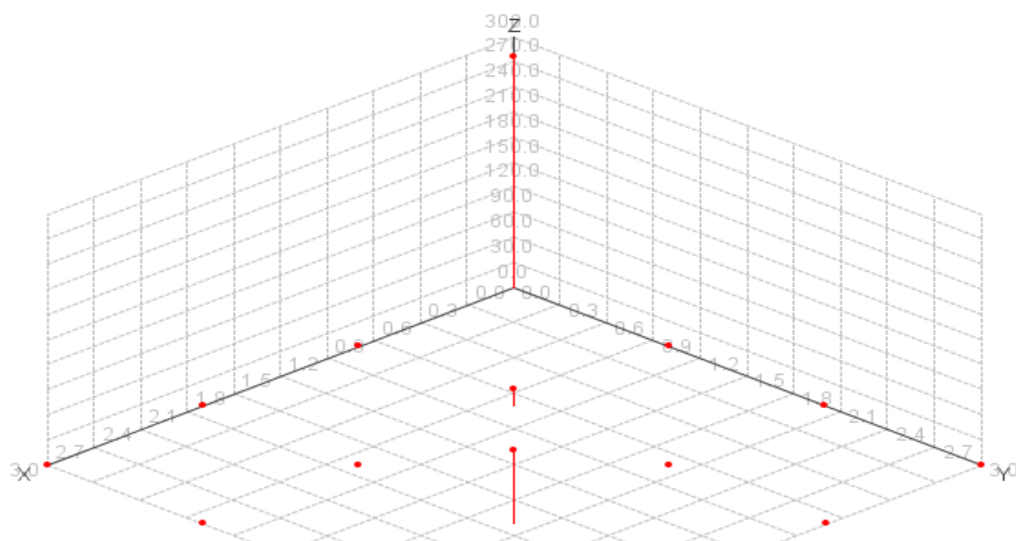**C. Feature Optimization With Genethic Algorithm**

From the results of the DT algorithm, this stage will be tested using DT optimized by GA. And and can produce the following accuracy results.

Tabel 2. Hasil perbandingan setelah optimasi

| DT (%) | DT + GA (%) |
|---|---|

From table 2, we can see a comparison of the accuracy results previously without optimization of only 99.75% after optimization to 100.00% and can also see the matrix diagram obtained from the DT accuracy results with GA optimization.



Source: Research Results

Figure 4. Accuracy matrix diagram decision tree with genethic algorithm

From the accuracy matrix diagram, it can be said that the weight of each attribute has become the

optimal attribute after optimization in table 3.

Table 3. Result attribute weight

| Attribute | Weight |
|---|---|
| City | 1 |
| Subdistrict | 1 |
| Ward | 1 |
| Gender | 1 |
| Amount | 1 |

From this table that each attribute has maximized its weight so that it becomes the optimal attribute.

After testing by optimizing attributes using the genetic algorithm, what needs to be considered is the kelurahan level first, where the kelurahan level is the attribute that is the root of the infant mortality problem, so that the government can utilize this research data in reducing infant mortality in Jakarta.

## CONCLUSION

In using the genetic algorithm model, the attributes on infant mortality data obtained from uci data are optimal attributes with an accuracy of 100 percent so that the research results become a source of processing for testing infant mortality data which can facilitate local governments or communities and can be a source of data in reduce the infant mortality rate in Jakarta.

## REFERENCES

Abirami, E., & Karthika, S. (2018). Design of infant monitoring system to reduce the risk of sudden infant death syndrome. *Proceedings of the 2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017*, *2018-Janua*, 180–183. https://doi.org/10.1109/ICCSP.2017.8286802

Elhariri, E., El-Bendary, N., & Taie, S. A. (2020). Using Hybrid Filter-Wrapper Feature Selection with Multi-Objective Improved-Salp Optimization for Crack Severity Recognition. *IEEE Access*, *8*, 84290–84315. https://doi.org/10.1109/ACCESS.2020.2991968

Hammami, M., Bechikh, S., Hung, C. C., & Ben Said, L. (2019). A Multi-objective hybrid filter-wrapper evolutionary approach for feature selection. *IEEE Congress on Evolutionary Computation (CEC)*, *11*(2), 193–208.

https://doi.org/10.1007/s12293-018-0269-2

Kohno, A., Techasrivichien, T., Pilar Suguimoto, S., Dahlui, M., Nik Farid, N. D., & Nakayama, T. (2020). Investigation of the key factors that influence the girls to enter into child marriage: A meta-synthesis of qualitative evidence. *PLoS ONE*, *15*(7 July), 1–20. https://doi.org/10.1371/journal.pone.0235959

Kusumawardani, A., & Handayani, S. (2018). Karakteristik Ibu dan Faktor Risiko Kejadian Kematian Bayi di Kabupaten Banjarnegara. *Jurnal Promosi Kesehatan Indonesia*, *13*(2). https://doi.org/https://doi.org/10.14710/jpki.13.2.168-178

Lengkong, G.T., Langi, F.L.F.G dan Posangi, J. (2020). Faktor – Faktor yang Berhubungan dengan Kematian Bayi di Indonesia. *Jurnal KESMAS*, *9*(4), 41–47. https://doi.org/https://ejournal.unsrat.ac.id/v3/index.php/kesmas/article/view/29482

Priyatama, I. M. D., & Satya, B. (2022). Klasifikasi Anak Berkebutuhan Khusus Tunagrahita. *Paradigma*, *24*(1), 90–95. https://doi.org/https://doi.org/10.31294/paradigma.v24i1.1087 Paradigma,

Purnama, J. J., Nawawi, H. M., Rosyida, S., Ridwansyah, & Risandar. (2020). Klasifikasi Mahasiswa Her Berbasis Algortima Svm Dan Decision Classification of Her Students Based on Svm Algorithm and. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(6), 1253–1260. https://doi.org/10.25126/jtiik.202073080

Ridwansyah, R., Riyanto, V., Hamid, A., Rahayu, S., & Purnama, J. J. (2022). Grouping Data in Predicting Infant Mortality Using K-Means and Decision Tree. *Paradigma - Jurnal Komputer Dan Informatika*, *24*(2), 168–174. https://doi.org/10.31294/paradigma.v24i2.1399

Ridwansyah, R., Wijaya, G., & Purnama, J. J. (2020). Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students. *Jurnal Pilar Nusa Mandiri*, *16*(1), 53–58. https://doi.org/10.33480/pilar.v16i1.1180

Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, *2*(2), 1–9. https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424

Salina, F. H., Almeida, I. A. de, & Bittencourt, F. R. (2019). *Renewable Energy and Sustainable Buildings* (A. Sayigh (ed.)). Springer. https://doi.org/https//doi.org/10.1007/978-030-1848-9