IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.16, No.4, October 2022, pp. 435~446
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.78198

Naive Bayes Method and C4.5 in Classification of Birth Data

Asep Afandi*[1], Noviana[2], Deti Nurdianah[3]
[1,2]Information Systems, STMIK Dian Cipta Cendikia Kotabumi, Lampung, Indonesia
[3]Puskesmas Candi Rejo, Lampung, Indonesia
e-mail: *[1]**Asepafandi189@gmail.com**, novianagita1989@gmail.com,
detinurdianah@gmail.com

Abstrak

Data kelahiran dan umur produktif seorang ibu untuk hamil di Lampung masih tinggi. untuk mengetahui perbandingan umur produktif ibu hamil dan apakah sudah memenuhi syarat minimal dan maksimal seorang ibu untuk hamil, dan Kriteria bayi yang dilahirkan. Dimana hasil pengolahan data tersebut akan dijadikan sebagai sumber data bagi ibu-ibu penyuluhan khususnya bagi warga desa Banjar Kertahayu. Pengolahan data tersebut memerlukan metode khusus agar hasilnya menjadi tolak ukur suatu keputusan nantinya, seperti Data Mining. Metode yang digunakan untuk pengolahan data yang digunakan adalah Naive Bayes dan Algoritma C4.5. Data yang digunakan adalah data kelahiran tahun 2017-2021, sumber data dari Bidan Desa Banjar-Kabupaten Lampung Tengah. Hasil Penelitian Metode C 4.5 Umur paruh baya memiliki nilai kategori Umur dominan sebesar 0,3324138. dimana nilai tertinggi pada tahun 2017, dan akurasi 100 persen dari data tahun 2017- 2021. Kriteria berat badan bayi menggunakan Metode Naïve Bayes Kelas memiliki nilai kategori Paruh baya dominan sebesar 0,09675, nilai tertinggi pada tahun 2017, Hasil akurasi selama 5 tahun memiliki akurasi sebesar 92,84% berdasarkan data kelahiran 2017-2021.

Kata kunci*—Algoritma C4.5, Naif Bayes, Kategori Age Dominan, Berat Badan Bayi.*

Abstract

Data on the birth and productive age of a mother to get pregnant in Lampung is still high. to find out the comparison of the productive age of pregnant women and whether they have met the minimum and maximum requirements for a mother to become pregnant, and the criteria for babies born. Where the results of data processing will be used as a source of data for counseling mothers, especially for residents of Banjar Kertahayu village. The data processing requires a special method so that the results become a benchmark for a decision later, such as Data Mining. The method used for data processing used is Naive Bayes and C4.5 Algorithm. The data used is birth data in 2017-2021, the source of data from the Banjar Village Midwife-Central Lampung Regency. Research Results Method C 4.5 Middle age has a dominant age category value of 0.3324138. where the highest value is in 2017, and accuracy is 100 percent from the 2017-2021 data. The baby weight criterion using the Naïve Bayes Class Method has a dominant Middle-aged category value of 0.09675, the highest value in 2017, The results of accuracy for 5 years have accuracy of 92.84% based on 2017-2021 birth data

Keywords— *C4.5 Algorithm, Naïve Bayes, Python, Dominant Age Category, The Baby's Weight*

Received October 5th,2022; Revised October 31th, 2022; Accepted October 31th, 2022

1. INTRODUCTION

Pregnancy is the development of the fetus starting from conception and ending until delivery. At the productive age, the chance of getting pregnant is higher than in the elderly in terms of health and physical condition. Young age is the ideal age for pregnancy there is no specific benchmark for the best age to get pregnant. However, generally, a woman's fertility will decrease with age. In addition, pregnancy at an advanced age is also at risk of causing complications in pregnant wome [1].

Based on calculated estimates, the crude birth rate in Lampung Province decreased from 29.5 per 1000 population in 1990-1995 to 26.6 per 1000 population in 1995-2000. The population pyramid in 2019 shows interesting characteristics, including: first; male and female population ratio/sex ratio: 104,87 [2]. The number of deaths of Neonatal, infants, and toddlers in Lampung Province in 2019 was 404, 79, and 27 cases. The neonatal mortality rate in Province Lampung has decreased in the last 3years, from 3.35 to 2.7 contrast to the infant mortality rate which continues to increase. Meanwhile, the under-five mortality rate fluctuated from 2017 to 2019 [3]. From these data, the birth and age of pregnant women in Lampung are still high. The data will be processed using the Data Mining method.

Data mining is a data storage pattern to help the data analysis process [4]. Data Mining Consists of analysis classification, grouping, and association rules. Classification is a type of prediction that builds a pattern based on the training set and uses that pattern to classify the testing set [5]. Therefore, the researcher wants to raise the Classification of Childbirth Data Processing using the Naive Bayes method and the C4.5 Algorithm. The results of the study can be used as material for consideration of the health data of Banjar Kertahayu Village for data on counseling pregnant women. The data used is birth data in 2017-2021, the source of data from the Banjar Village Midwife-Central Lampung Regency. Where is the Research Result Method C4.5 To measure the age classification based on the results of the dominant age decision tree algorithm and the accuracy of the 2017-2021 data. As in previous research shows that the decision tree algorithm provides new knowledge[6] and the experimental results C4.5 also confirms that the most influential instant data attribute[7].

While the criteria for infant weight use the Naïve Bayes method based on dominant age values and accuracy values for 5 years based on 2017-2021 birth dataas in previous studies using the Naive Bayes method on health topics is very large because of the availability of health problems to make decisions[8], and importance of the main causes of heart disease. to consider almost all the important features that influence heart disease and perform step analysis predicting heart disease[9].

2. METHODS

2.1 Data mining

Data mining is a logical process used to search and find patterns through large amounts of data [10]. The goal is to find previously unknown patterns. In data mining, this pile of past data is considered a mine that can be processed to produce valuable knowledge. Data mining is an integral part of knowledge discovery in databases (KDD)[11]. knowledge discovery in databases is a nontrivial process to identify valid, new, potentially useful, and ultimately understandable patterns in data [12]. The whole KDD process for converting raw data into useful information is Input data can be stored in various formats such as flat files, spreadsheets, or relational tables, and can occupy a centralized or distributed data repository in many places. The purpose of preprocessing is to convert the raw input data into a suitable format for further analysis. The steps involved in data preprocessing include combining data from multiple sources, cleaning the data to remove noise and duplicate observations, and selecting records and features relevant to data mining work [13].

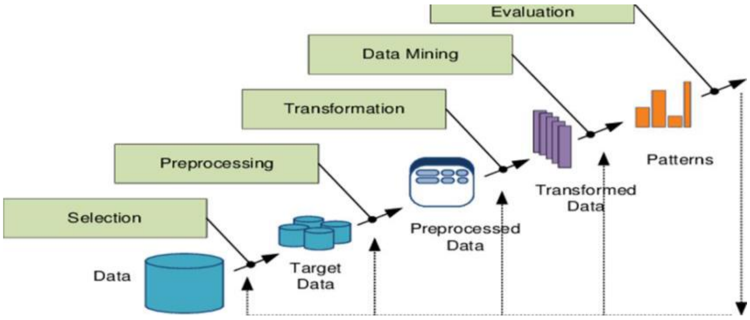

Figure 1. KDD [11].

2.2 Naive Bayes

One of the methods in Data Mining is data classification, which is mapping (classifying) data into one or several predetermined classes. One of the methods in data classification is the Naïve Bayes Classifier [14]. Naïve Bayes Classifier is a machine learning method that utilizes probability and statistical calculations, predicting future probabilities based on previous experience [15], and predicting future opportunities based on experience known as Bayes' theorem [16]. The basis of Naïve Bayes used in programming is the Bayes formula [17]:

P (A|B) = (P(B|A) * P(A))/P(B)

Nave Bayes Classifier/Multinomial Nave Bayes is a simplification model of the Bayes method for classifying text or documents [18].

$$. E_{MAP} = {}^{argmax}_{vj\ EV} P(Vj) \prod \ i, P(ai|Vj)$$

2.3 C 4.5

The C4.5 method is a basic classifier model using a hierarchical structure. The main idea of the Hierarchical structure is to transform data into a rooted tree graph as a decision rule [17]. In general, the C4.5 method for building a decision tree is as follows [19]:

1. Select attribute as root
2. Create a branch for each value
3. Divide cases in branches
4. Repeat the process for each until all have the same class

The selection of attributes as root is on the highest gain value of the existing[20]. The following formula:

$$Gain(S, A) = Entropy(s) \quad + \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si)$$

Before getting the Gain value find the Entropy value. Entropy used to determine how informative an attribute input is to produce an attribute[21]. The basic formula for Entropy is as follows:

$$Entropy(S) = \sum_{i=1}^{n} - Pi * \log_2 Pi$$

2.4 Phyton

Python is a multi-platform interpretive programming language with a design philosophy focused on code readability and is one of the popular languages related to Data Science, Machine Learning, and the Internet of Things (IoT). Python library for processing textual data. Provides a consistent API for analyzing natural language processing (NLP) tasks such as tagging parts of speech, noun extraction, and sentiment analysis [22]. Researchers chose python because it is a programming language that is very easy to understand. And python has many libraries that support machine learning and statistical analysis and are used to analyze data and get approximate data [23].

3. RESULTS AND DISCUSSION

This study perform data mining calculations through several stages, namely:

3.1. Data Sample

From the results of the study, the data samples from 2017-2021 as follows:

Table 1. Data Sample

No	Date	Patient's name	Age	Adress	Baby Birth Date	Baby Genders	Baby Weight	Panjang Bayi	Gender	no. hp
1	#######	Dwi Lestari		BKH dsn 6	01/07/2017	7,15	3400	50	woman	0823 0627 8253
2	13/1/2017	Reni Pauji		BKH dsn 2	13/1/2017	22,30	4000	50	man	0822 9908 1857
3	#######	Tari		Banjar ratu	02/01/2017	3,40	2500	45	woman	
4	#######	Dede Rohaeti		BKH dsn 6	02/01/2017	11,00	4000	50	man	
5	27/3/2017	Iyan Susi Kagenah		BKH dsn 2	02/07/2017	10,30	1800	47	man	
6	29/3/2017	Supiah	30	BKH dsn 5	29/3/2017	10,10	2800	49	woman	0853 6956 7819
7	#######	Rahma	20	BKH dsn 3	04/03/2017	4,45	2500	46	man	0823 7762 8477
8	13/4/2017	Siti Khalifah Ulfa	19	BKH dsn 2	13/4/2017	7,20	3000	49	man	
9	24/4/2017	Sri Purwanti	21	Banjar ratu	24/4/2017	17,00	3200	49	man	
10	#######	Triyani	29	BKH dsn 3	05/01/2017	8,30	3200	50	man	0812 7851 6143
11	#######	Fitriyani	28	Banjar ratu	05/06/2017	12,30	3000	49	woman	0823 7269 3630
12	#######	Rina	19	BKH dsn 3	05/08/2017	3,00	3000	42	man	0852 6737 0348
13	#######	Sutini	24	tanggerang	05/09/2017	18,15	3200	48	man	
14	#######	Devi		Banjar ratu	05/10/2017	3,30	3000	48	woman	
15	16/5/2017	Kurniasih		BKH dsn 1	16/5/2017	11,00	3500	46	man	
16	#######	Ati Widi Astuti	22	BKH dsn 1	06/02/2017	16,15	3000	49	woman	0821 2726 3342
17	#######	Tri Nuryanti	24	candirejo	06/07/2017	4,45	3000	48	woman	0823 8849 5756
18	#######	Imas Siti Hajar	19	BKH dsn 3	06/10/2017	0,05	3500	50	woman	0852 1041 1521
19	#######	Wayan Suati	30	Bali agung	06/10/2017	8,00	3400	48	woman	0853 6810 8026
20	#######	Rini	19	BKH dsn 3	06/11/2017	14,30	2400	46	man	0821 7881 3376
21	16/6/2017	Sukarmi	25	spontan	16/6/2017	1,30	3300	48	man	0822 8061 2247
22	29/6/2017	Deti Harianti	19	BKH dsn 3	29/6/2017	22,43	4000	50	man	0853 8480 5079
23	30/6/2017	Ronayah	25	BKH dsn 2	30/6/2017	23,23	2500	45	man	0852 7376 8810
24	#######	Anita	29	BKH dsn 3	07/02/2017	7,30	2800	47	woman	0823 7691 5441
25	#######	Nuning Asriyani	20	BKH dsn 2	07/06/2017	11,30	3500	49	woman	0853 8424 1347
26	20/7/2017	Dewi Setyawati	26	Banjar ratu	20/7/2017	12,30	4000	50	man	0823 0654 2678
27	23/7/2017	Siti Fatimah	27	BKH dsn 4	23/7/2017	8,00	3100	50	woman	
28	23/7/2017	Devi Nurmala	25	BKH dsn 2	23/7/2017	16,30	3800	50	man	
29	#######	Wayan Kerti	32	Bali agung	08/04/2017	10,30	3500	48	woman	0823 0695 7405
30	13/8/2017	Sinta Kartika	18	BKH dsn 1	13/8/2017	17,30	3300	48	woman	0831 3094 2380
31	17/8/2017	Juwariyah	28	BKH dsn 1	17/8/2017	9,30	3200	49	man	
32	21/8/2017	Nengah Parwati	18	Bali agung	21/8/2017	8,05	3700	48	woman	0852 1356 3684
33	26/8/2017	Ningrumiyati	30	BKH dsn 1	26/8/2017	22,15	3400	48	man	0823 7218 4383

3.2. Data Cleaning

From the sample table data, the next step is data cleaning, namely eliminating incomplete data and double data, so that the following data:

Tabel 2. Data Cleaning

No	Patient's name	Adress	Age	Baby Weight	Baby Length	Gender
1	supiah	BKH dsn 5	30	2800	49	Man
2	rahma	BKH dsn 3	20	2500	46	Man
3	siti khalifah ulfa	BKH dsn 2	19	3000	49	Women
4	sri purwanti	Banjar ratu	21	3200	49	Man
5	triyani	BKH dsn 3	29	3200	50	Man
6	fitriyani	Banjar ratu	28	3000	49	Man
7	rina	BKH dsn 3	19	3000	42	Women
8	sutini	tanggerang	24	3200	48	Man
9	ati widi astuti	BKH dsn 1	22	3000	49	Man
10	tri nuryanti	candirejo	24	3000	48	Man
11	imas siti hajar	BKH dsn 3	19	3500	50	Women
12	wayan suati	Bali agung	30	3400	48	Man
13	rini	BKH dsn 3	19	2400	46	Women
14	sukarmi	spontan	25	3300	48	Man
15	deti harianti	BKH dsn 3	19	4000	50	Women
16	ronayah	BKH dsn 2	25	2500	45	Man
17	anita	BKH dsn 3	29	2800	47	Man
18	nuning asriyani	BKH dsn 2	20	3500	49	Man
19	dewi setyawati	Banjar ratu	26	4000	50	Man
20	siti fatimah	BKH dsn 4	27	3100	50	Man
21	devi nurmala	BKH dsn 2	25	3800	50	Man
22	wayan kerti	Bali agung	32	3500	48	Man
23	sinta kartika	BKH dsn 1	18	3300	48	Women
24	juwariyah	BKH dsn 1	28	3200	49	Man
25	nengah parwati	Bali agung	18	3700	48	Women
26	ningrumiyati	BKH dsn 1	30	3400	48	Man

3.3. Data Transformation

Table 3. Data Transformation

No	Patient's name	Baby Length	Baby Weight	Age Mother	Description of age
1	Supiah	49	2800	30	Mature
2	Rahma	46	2500	20	Teenager
3	Siti Khalifah Ulfa	49	3000	19	Teenager
4	Sri Purwanti	49	3200	21	Teenager
5	Triyani	50	3200	29	Mature
6	Fitriyani	49	3000	28	Mature
7	Rina	42	3000	19	Teenager
8	Sutini	48	3200	24	Teenager
9	Ati Widi Astuti	49	3000	22	Teenager
10	Tri Nuryanti	48	3000	24	Teenager
11	Imas Siti Hajar	50	3500	19	Teenager
12	Wayan Suati	48	3400	30	Mature
13	Rini	46	2400	19	Teenager
14	Sukarmi	48	3300	25	Mature
15	Deti Harianti	50	4000	19	Teenager
16	Ronayah	45	2500	25	Mature
17	Anita	47	2800	29	Mature
18	Nuning Asriyani	49	3500	20	Teenager
19	Dewi Setyawati	50	4000	26	Mature
20	Siti Fatimah	50	3100	27	Mature
21	Devi Nurmala	50	3800	25	Mature
22	Wayan Kerti	48	3500	32	Middle-Aged
23	Sinta Kartika	48	3300	18	Teenager
24	Juwariyah	49	3200	28	Mature
25	Nengah Parwati	48	3700	18	Teenager
26	Ningrumiyati	48	3400	30	Mature

3.4 Implementation of Data Mining Techniques

3.4.1. Testing Data

Tabel 4. Testing Data

No	Patient's name	Baby Length	Baby Weight	Mother Age	Age Description
1	Supiah	49	2800	30	Mature
2	Rahma	46	2500	20	Teenager
3	Siti Khalifah Ulfa	49	3000	19	Teenager
4	Sri Purwanti	49	3200	21	Teenager
5	Triyani	50	3200	29	Mature
6	Fitriyani	49	3000	28	Mature
7	Rina	42	3000	19	Teenager
8	Sutini	48	3200	24	Teenager
9	Ati Widi Astuti	49	3000	22	Teenager
10	Tri Nuryanti	48	3000	24	Teenager
11	Imas Siti Hajar	50	3500	19	Teenager
12	Wayan Suati	48	3400	30	Mature
13	Rini	46	2400	19	Teenager
14	Sukarmi	48	3300	25	Mature
15	Deti Harianti	50	4000	19	Teenager
16	Ronayah	45	2500	25	Mature
17	Anita	47	2800	29	Mature
18	Nuning Asriyani	49	3500	20	Teenager
19	Dewi Setyawati	50	4000	26	Mature
20	Siti Fatimah	50	3100	27	Mature
21	Devi Nurmala	50	3800	25	Mature
22	Wayan Kerti	48	3500	32	Middle-Aged
23	Sinta Kartika	48	3300	18	Teenager
24	Juwariyah	49	3200	28	Mature
25	Nengah Parwati	48	3700	18	Teenager

3.4.2 Variable Value

Table 5. Variable Value

No.	Rating Attribute		Rating Range	
	Code	Description	Range	Rating
1.	A1	Mother Age	17-24	4
			25-31	3
			32-39	2
			40-47	1
2	A2	Baby Length	38-40	4
			41-44	3
			45-48	2
			49-51	1
3	A3	Baby Weight	1900-2475	4
			2476-3050	3
			3051-3625	2
			3626-4200	1

3.4.3. C4.5. Algorithm Method

Based on the value of information entropy and gain in the data above, the criteria that most influence predicting childbirth in Banjar Kertahayu Village, Way Pengubuan District, Central Lampung Regency show:

Table 6. Knowledge Presentation

Tahun	Kriteria Dominan	Nilai
2017	A3 (Baby Weight)	0,3324138
2018	A3 (Baby Weight)	0,2973251
2019	A3 (Baby Weight)	0,208335758
2020	A1 (Mhoters Age)	0,30103
2021	A3 (Baby Weight)	0,169393

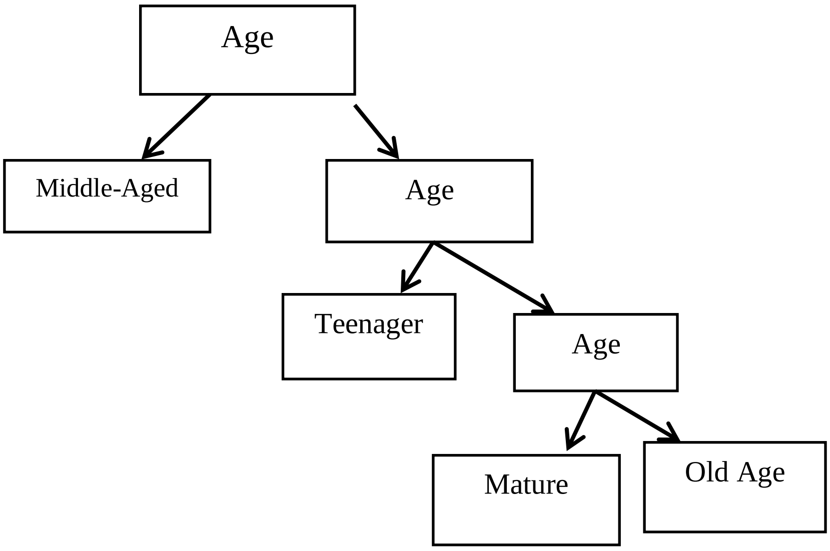

Figure 2. Decision Tree results

The middle-aged attribute has a dominant age category value of 0.3324138. So that the highest value in 2017. Then the average accuracy can be seen in the table below

Tabel 7. average accuracy

Year	Accuracy
2017	100%
2018	100%
2019	100%
2020	100%
2021	100%
Average	100%

The calculation of labor above has 100 percent accuracy from 2017 to 2021 data. The dominant criterion is the baby's weight.

3.4.4. Implementation of Phyton Method C 4.5

```
[ ] Confusion Matrix
    [[3 0]
     [0 5]]
    Tingkat Akurasi Algoritma C4.5
    Akurasi :              precision    recall  f1-score   support

               1       1.00      1.00      1.00         3
               2       1.00      1.00      1.00         5

        accuracy                           1.00         8
       macro avg       1.00      1.00      1.00         8
    weighted avg       1.00      1.00      1.00         8

    Tingkat Akurasi: 100 persen
```

```
[23] from sklearn.tree import export_graphviz
     export_graphviz(tree_dataset, out_file="Tree_Keterangan.dot", class_names=["Remaja","Dewasa","Paruh Baya","Usia Tua"
                     feature_names=atr_dataset.columns, impurity=False, filled=True)
```

```
import graphviz

with open("Tree_Keterangan.dot") as fig:
  dot_graph = fig.read()
  graph = graphviz.Source(dot_graph)

  graph.view()
```

Figure 3. Results and Accuracy C 4.5

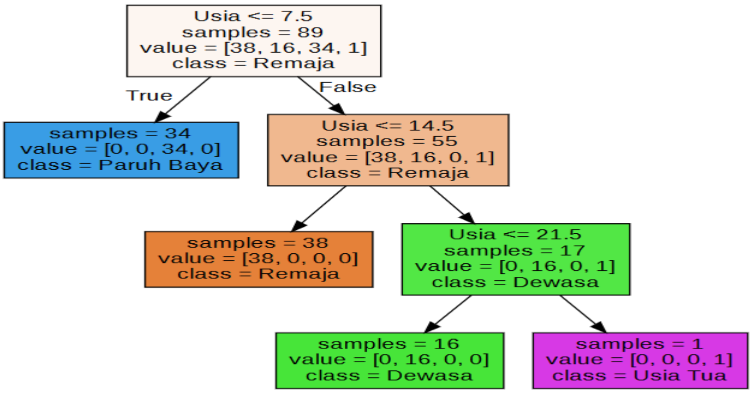

Figure 4. Decision Tree

3.4.5 Naive Bayes method

Based on the probability value in the data above, the most influential criteria in predicting childbirth in Banjar Kertahayu Village, Way Pengubuan District, Central Lampung Regency show:

Table 8. Knowledge Presentation

Year	Kategori Age Dominan	Score
2017	Middle-aged	0,09765
2018	-	0
2019	Middle-aged	0,02775
2020	-	0
2021	Middle-aged	0,333

The middle-aged class has a dominant age category value of 0.09675. So that the highest value in 2017. Then the average accuracy can be seen in the table below.

Tabel 9. Average Accuracy

Tahun	Akurasi
2017	97,56%
2018	95,83%
2019	95,45%
2020	85,71%
2021	95,65%
Average	92,84%

3.4.6 Python implementation of Naive Bayes method

[35]

110	111	ayu handayani	21	3200	48	Remaja
111	112	komang ayu	23	2800	48	Remaja

112 rows × 6 columns

```
[36] from sklearn.preprocessing import LabelEncoder

[37] enc = LabelEncoder()

[38] data['Nama Pasien'] = enc.fit_transform(data['Nama Pasien'].values)
     data['Usia'] = enc.fit_transform(data['Usia'].values)
     data['BB bayi'] = enc.fit_transform(data['BB bayi'].values)
     data['Panjang Bayi'] = enc.fit_transform(data['Panjang Bayi'].values)
     data['Keterangan Usia'] = enc.fit_transform(data['Keterangan Usia'].values)

[39] atr_data = data.drop(columns='Keterangan Usia')
     atr_data.head()
```

	No	Nama Pasien	Usia	BB bayi	Panjang Bayi
0	1	90	13	6	6
1	2	68	3	3	3

Figure 5. Results Naïve Bayes

```
[41] from sklearn.naive_bayes import GaussianNB

[42] xtrain, xtest, ytrain, ytest = train_test_split(atr_data, cls_data, test_size=0.2, random_state=500)
     tree_data = GaussianNB()
     tree_data.fit(xtrain, ytrain)

     GaussianNB()

[43] tree_data.class_count_

     array([38., 16., 35.])

[44] print("Nilai akurasi pada data testing:", tree_data.score(xtest, ytest))

     Nilai akurasi pada data testing: 0.8260869565217391
```

Figure 6. Accuracy Naïve Bayes

3.5 Evaluation

Based on the results of algorithm calculations and python implementations we get:

Tabel 10. Result Conclusion

Method	Year	Category	Score	Accuracy	Accuracy Python
C 4.5	2017	A3 (Baby Weight)	0,3324138	100%	100%
	2018	A3 (Baby Weight)	0,2973251	100%	
	2019	A3 (Baby Weight)	0,208335758	100%	
	2020	A1 (Mohters Age)	0,30103	100%	
	2021	A3 (Baby Weight)	0,169393	100%	
Naïve Bayes	2017	Middle-aged	0,09765	97,56%	82%
	2018	-	0	95,83%	
	2019	Middle-aged	0,02775	95,45%	
	2020	-	0	85,71%	
	2021	Middle-aged	0,333	95,65%	

4. CONCLUSIONS

From the calculation results Method C 4.5, The middle-aged attribute has a dominant age category value of 0.3324138. The highest value was in 2017, with and accuracy of 100 percent, and for Python implementation accuracy of 100% was taken from data from 2017 to 2021. With criteria on Baby Weight. The Naïve Bayes method The middle-aged class has a dominant age category value of 0.09675, the highest value was in 2017, with accuracy for five years accuracy 92.84% based on 2017-2021 training data for Python implementation accuracy of 82%.

ACKNOWLEDGEMENTS

Author thanks for STMIK Dian Cipta Cendikia Kotabumi, and Lampung Lldikti region 2, Directorate of Research and Community Service Director General Strengthening Research and Development Ministry of Research, Technology and Higher Education.

REFERENCES

[1] dr. M. D. C. Pane, “Berapa Usia Terbaik untuk Hamil?,” *https://www.alodokter.com/*, 2022. https://www.alodokter.com/berapa-usia-terbaik-untuk-hamil#:~:text=Saat seorang wanita berada di,an dikatakan ideal untuk hamil.

[2] D. Lampung, “Profil Kesehatan Provinsi Lampung Tahun 2019,” *Pemerintah Provinsi Lampung Dinkes*, no. 44, p. 136, 2019.

[3] 2015-2019 Renstra dinkes lampung, “Rencana Strategis Dinkes Provinsi Lampung Tahun 2015-2019,” no. 46, p. 58 (9), 2019, [Online]. Available: https://dinkes.lampungprov.go.id/wp-content/uploads/2016/07/1.RENSTRA-DINAS-KESEHATAN-PROVINSI-LAMPUNG-2015-2016.pdf

[4] Nurhachita and E. S. Negara, “A comparison between deep learning, naïve bayes and random forest for the application of data mining on the admission of new students,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 2, pp. 324–331, 2021, doi: 10.11591/ijai.v10.i2.pp324-331.

[5] R. Hasan, S. Palaniappan, S. Mahmood, K. U. Sarker, and A. Abbas, “Modelling and predicting student’s academic performance using classification data mining techniques,” *Int. J. Bus. Inf. Syst.*, vol. 34, no. 3, pp. 403–422, 2020, doi: 10.1504/IJBIS.2020.108649.

[6] F. Aziz and A. Lawi, “Increasing electrical grid stability classification performance using ensemble bagging of C4.5 and classification and regression trees,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2955–2962, 2022, doi: 10.11591/ijece.v12i3.pp2955-

2962.
[7] M. Sadikin and F. Alfiandi, “Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk,” *Int. J. Electr. Comput. Eng.*, vol. 8, no. 6, p. 4763, 2018, doi: 10.11591/ijece.v8i6.pp4763-4771.
[8] R. R. K. Al-Taie, B. J. Saleh, A. Y. F. Saedi, and L. A. Salman, “Analysis of WEKA data mining algorithms Bayes net, random forest, MLP and SMO for heart disease prediction system: A case study in Iraq,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 6, pp. 5229–5239, 2021, doi: 10.11591/ijece.v11i6.pp5229-5239.
[9] T. R. Stella Mary and S. Sebastian, “Predicting heart ailment in patients with varying number of features using data mining techniques,” *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2675–2681, 2019, doi: 10.11591/ijece.v9i4.pp2675-2681.
[10] A. Saleh, “Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Citec J.*, vol. 2, no. 2354-5771. 207–217., p. 3, 2015.
[11] A. Mailana, A. A. Putra, S. Hidayat, and A. Wibowo, “Comparison of C4.5 Algorithm and Support Vector Machine in Predicting the Student Graduation Timeliness,” *J. Online Inform.*, vol. 6, no. 1, p. 11, 2021, doi: 10.15575/join.v6i1.608.
[12] Xiaoling Shu, *Knowledge Discovery in the Social Sciences: A Data Mining Approach*. okland, california: University of California Press, 2020.
[13] R. Takdirillah, “Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan,” *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
[14] I. I. Wiratmadja, S. Y. Salamah, and R. Govindaraju, “Healthcare data mining: Predicting hospital length of stay of dengue patients,” *J. Eng. Technol. Sci.*, vol. 50, no. 1, pp. 110–126, 2018, doi: 10.5614/j.eng.technol.sci.2018.50.1.8.
[15] C. V. D, “Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments,” *Int. J. Electr. Comput. Eng.*, vol. 9, no. 5, p. 4452, 2019, doi: 10.11591/ijece.v9i5.pp4452-4459.
[16] R. Amanda and E. S. Negara, “Analysis and Implementation Machine Learning for YouTube Data Classification by Comparing the Performance of Classification Algorithms,” *J. Online Inform.*, vol. 5, no. 1, pp. 61–72, 2020, doi: 10.15575/join.v5i1.505.
[17] N. Kewsuwun and S. Kajornkasirat, “A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2829–2838, 2022, doi: 10.11591/ijece.v12i3.pp2829-2838.
[18] A. R. Isnain, N. S. Marga, and D. Alita, “Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 1, p. 55, 2021, doi: 10.22146/ijccs.60718.
[19] Y. I. Kurniawan, U. M. Surakarta, and N. Bayes, “COMPARISON OF NAIVE BAYES AND C . 45 ALGORITHM IN DATA MINING,” vol. 5, no. 4, pp. 455–464, 2018, doi: 10.25126/jtiik.
[20] D. Y. H. Tanjung, “Analisis perbandingan algoritma id3 dan c4.5 terhadap data pengisian uang atm,” *CSRID J.*, vol. 13, no. 3A, pp. 231–242, 2021.
[21] C. Paramitha Lubis, R. Rosnelly, and Z. Situmorang, “Penerapan Metode Naïve Bayes Dan C4.5 Pada Penerimaan Pegawai Di Universitas Potensi Utama Application of Naïve Bayes and C4.5 Methods in Receiving Employees Inuniversity of Potensi Utama,” *CSRID J.*, vol. 12, no. 1, pp. 51–63, 2020, [Online]. Available: https://www.doi.org/10.22303/csrid.12.1.2020.51-63
[22] R. Maulid, “Belajar Python Otodidak dari Fungsi dan Prosedurnya,” *https://www.dqlab.id/*, 2022. https://www.dqlab.id/belajar-python-otodidak-dari-fungsi-dan-prosedurnya
[23] H. Thamrin, “Analyzing and Forecasting Admission data using Time Series Model,” *J. Online Inform.*, vol. 5, no. 1, pp. 35–44, 2020, doi: 10.15575/join.v5i1.546.