CLASSIFICATION OF STUNTING STATUS IN TODDLERS USING NAIVE BAYES METHOD IN THE CITY OF MADIUN BASED ON WEBSITE

Ari joko Purnomo[1]; Abdul Rozaq[2*]

Teknik Informatika
Universitas PGRI Madiun
https://unipma.ac.id
arijoko527@gmail.com [1]; rozaq@gmail.com [2*]

Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract— *Stunting in toddlers is a chronic nutritional problem that is being experienced by the health world. Children with stunting have a tendency to decrease their level of intelligence, have speech disorders and have difficulty capturing learning in the usual method. Madiun City still faces challenges in stunting nutrition problems. The prevalence of stunting in 2020 is 10.18 percent or 814 children out of a total of 7,996 measured. The use of data mining can be used in various fields related to large data sets. There are several techniques of working on data mining in retrieval of information, including classification. Generally, the classification of stunting status uses the TB/U index or height compared to age. In this study, the method used is the naive bayes method, which is the method used to predict probability based, the system is built using the python programming language and flask as the framework. The results of the tests carried out show that the naive bayes method can be used in classifying stunting status in toddlers. The Naïve Bayes algorithm implemented has an average performance value of 58% accuracy, 68% precision, and 58% recall from the results of the confusion matrix test with 30% testing data and 70% training data.*

Keywords: *Classification, Naïve Bayes, Data Mining, Stunting, Python.*

Intisari— Stunting pada balita merupakan masalah gizi kronis yang sedang dialami dunia kesehatan. Anak dengan kondisi stunting mengalami kecenderungan penurunan tingkat kecerdasan, gangguan berbicara dan kesulitan dalam menangkap pembelajaran dalam metode yang biasa. Kota Madiun masih menghadapi tantangan dalam permasalahan gizi stunting. Prevalensi angka stunting tahun 2020 sebesar 10,18 persen atau 814 anak dari total 7.996 yang diukur. Penggunaan data mining dapat digunakan dalam berbagai bidang yang berhubungan dengan sekumpulan data yang banyak. Terdapat beberapa teknik pengerjaan data mining dalam pengambilan suatu informasi, diantaranya adalah klasifikasi. Umumnya klasifikasi status stunting menggunakan indeks TB/U atau tinggi badan dibanding usia. Pada penelitian ini, metode yang digunakan adalah metode naive bayes, yakni metode yang digunakan untuk memprediksi berbasis probabilitas, sistem yang dibangun menggunakan bahasa pemrograman python dan flask sebagai framework-nya. Dari hasil pengujian yang dilakukan menunjukkan bahwa metode naive bayes dapat digunakan dalam melakukan klasifikasi terhadap status stunting pada balita. Algoritma Naïve Bayes yang diimplementasikan ini, memiliki performansi nilai rata-rata yaitu akurasi sebesar 58%, precision sebesar 68%, dan recall sebesar 58% dari hasil pengujian confusion matrix dengan 30% data testing dan 70% data training.

Kata Kunci: *Klasifikasi, Naïve Bayes, Data Mining, Stunting, Python.*

INTRODUCTION

Stunting in toddlers is a chronic nutritional problem experienced, especially in the world of health. The intelligence level of children with stunting status tends to be lower, has speech disorders, and has difficulty capturing learning in the usual way. Factors that cause stunted growth in children can be caused during pregnancy, childbirth, breastfeeding or during the puerperium as well as the MPASI factor, which is not enough to feed young children. In addition, the hygiene factor in the poor environment can be a trigger for toddlers to get sick easily. Poor parenting is one of the causes of stunted growth. Poor parenting is

Vol. 19, No. 2 September 2021 | DOI https://doi.org/10.33480/techno.v19i2.3337

often caused by several factors, such as a mother who is too young or a pregnancy factor that is too close together. In this case, the Ministry of Health of the Republic of Indonesia seeks to improve the nutritional status of the community which is one of the priority programs in national health development as stated in the main objectives of the 2015 – 2019 medium-term development plan to reduce the prevalence of stunting in children under five years of age (Prasetiya et al., 2020).

Chronic malnutrition during early growth and development with a growth assessment of a high Z-score for age (TB/U) less than -2SD (standard deviation) of the WHO growth standard. Generally, stunting can affect 1 in 4 young children. The short-term effects caused by stunting are delays in speech ability, limitations, motor and cognitive sensor development, infectious diseases to death. The long-term effects include the risk of degenerative diseases such as coronary artery disease, stroke, high blood pressure and diabetes mellitus. In addition, it can reduce work productivity in adulthood (Zeniarja et al., 2020).

Madiun City is still facing challenges in terms of nutrition (stunting). The prevalence of stunting in 2020 reached 10.18 percent of 814 children out of a total of 7,996 measured. The prevalence of children under five in Madiun City is quite low, even below the standard when compared to the results of Riskesdas, which is 10.2%.

Data mining is a science that combines machine learning, pattern discovery, statistical calculations, databases and visualization to obtain information from a wide range of data. The use of data mining can be used in various fields related to large data sets. There are several data mining techniques for retrieval of information, including regression, clustering, association and classification. In classification, there are several methods that can be used, one of the methods commonly used is the Naive Bayes algorithm. Naive Bayes method is a simple probabilistic classification method from Bayes theory where classification is carried out through efficient construction of a number of data sets. Naive Bayes assumes that a value based on an input attribute in a shared class is independent of using another attribute value (Ismasari Nawangsih & Setyaningsih, 2020).

Related to the explanation above, the authors conducted a study entitled "Classification of Stunting Status in toddlers using the Naive Bayes method in the city of Madiun based on the website". This system was created in order to assist health workers in knowing the classification of stunting status in Madiun City.

.

MATERIALS AND METHODS

I. Literary studies

There are many case studies regarding stunting prediction. The following are some related studies regarding the prediction of stunting status.

According to (Zeniarja et al., 2020) in his research entitled Application of the Naive Bayes Algorithm and Forward Selection in Classifying Stunting Nutritional Status at Pandanaran Health Center Semarang. The purpose of this study is to optimize the accuracy value of the Naive Bayes classification algorithm by removing inappropriate attributes using the Forward Selection feature. The results of testing the NBC solving procedure without using feature selection are 83.33%, while the performance of the NBC solving procedure using the foward selection method has increased by 2.67% to 86.00%, and the suggestion needed is to do a comparison test of output accuracy using classification algorithm procedures in addition to the Naive Bayes classification procedure are also feature selection methods other than foward selection which may be able to form a better accuracy value.

According to (Prasetiya et al., 2020) in his research the Classification of Toddler Stunting Status in Slangit Village Using the K-Nearest Neighbor Method. The purpose of this study is to classify the status of stunting toddlers using the K-Nearest Neighbor method. The study used 300 data to calculate the Euclidean distance and involved age, height and weight parameters. Then the calculation is carried out with the RapidMiner program with the KNN Classification method which produces an accuracy of 98.89% with NORMAL and LESS statuses. it can be concluded that the application of the K-Nearest Neighbor method in classifying the nutritional status of toddlers using the Eucledian distance calculation formulation has a good performance.

Research conducted by (Siregar et al., 2020) The purpose of this study is to facilitate the processing of comment data, so the comment classification process is applied using the Naïve Bayes Classifier method to find out whether the comments were positive or negative. The test results on 50 comment data using the Naive Bayes Classifier method resulted that the output accuracy value was 68%.

a. Data Mining

The definition of data mining is the extraction of information or patterns of retrieving data that is in the database. Data mining is known as Knowledge Discovery in Database (KDD). (Siregar & Puspabhuana, 2020) defines data mining as a set of techniques that are used automatically for full exploration and leading to complex relationships in datasets. The dataset in question is a set of tabulated

data, and is often implemented in relational database technology. However, data mining techniques can be used in other data representations, such as domains, text, and multimedia.

According to (Yenderizal, 2022) a term used to describe the discovery of knowledge in databases. Data mining is the process of using statistical, mathematical, artificial intelligence and machine learning techniques to identify and extract information and related knowledge from databases.

b. Classification

According to (Prasetyo, 2012) classification is a job in assessing data objects to include in certain class categories from a number of existing classes. In classification there are two types of core work carried out, namely the development of a prototype model that is stored as memory and the use of that model in conducting the introduction/classification/prediction of a data object with the aim of knowing in which class the data object is in a stored model.

Classification consists of a two-step process. The first is the training phase, where the classification algorithm is created with the aim of analyzing the training data which is then represented in the form of a classification rule. The second is classification where test data is used to estimate the classification rule. There are several algorithms commonly used in the classification process, including the Naïve Bayes classifier, neural network, statistical analysis, rough sets, support vector machines and many others (Hesananda, 2021).

c. *Naive Bayes Classifier (NBC)*

Bayes theorem is a rule to improve or revise a probability by utilizing more information. This theory was developed by Thomas Bayes (1702-1763). (Daqiqil, 2021) explained that the Naive Bayes Classifier (NBC) is a technique that uses Bayes' theorem in the classification process of data. NBC assumes that the features contained in the data are independent.

According to (Daqiqil, 2021) Bayes' theorem is generally stated as follows:

$$P(H|e) = \frac{P(e|H) \cdot P(H)}{P(e)} \quad \text{.....(1)}$$

Description:

P(H | e) = Possible hypothesis (H) given the evidence (e) observed (Posterior)

P(e | H) = Possible hypothesis (H) given the evidence (e) observed (Posterior)

P(H) = How big is the probability of the hypothesis (H) before observing the evidence (e) (Prior)

P(e) = What is the probability of proof on all hypotheses

d. Stunting

Stunting according to (Tanoto Foundation, 2021)) is a condition of failure to thrive that occurs in toddlers as a result of chronic malnutrition and repeated infections, especially in the First 1000 Days of Life (HPK) period. Physically, stunted children look shorter than their peers. According to (Kementerian Kesehatan Republik Indonesia, 2018) Children are categorized as stunting if their height is below -2 SD (standard deviation) among children their age. Stunting and other nutritional deficiencies that occur in 1,000 HPK do not also cause stunted physical growth and increased susceptibility to disease, but can threaten cognitive development that may affect the current intelligence and productivity of a child in adulthood.

The stunting condition that is often experienced by toddlers as well as children can be caused by several triggering factors, such as maternal nutrition during pregnancy, pain experienced by babies, and lack of nutritional intake in infants to socio-economic conditions of the community. A mother is an important key in the role of determining the development of stunting, considering that the initial development of a child starts from the fetus (Yoto et al., 2020).

e. Confusion matrix

The confusion matrix is a matrix of size N x N where N is the number of predicted classes. So this matrix is good when used in classification problems. (Daqiqil, 2021) explaining the Confusion matrix provides a summary of the prediction results that have been generated by comparing the predicted results with the expected results. The confusion matrix table can be seen in Table 1.

Table 1. Table Confusion Matrix

	Actual = Yes	Actual= No
Predicted= Yes	TP	FP
Predicted= No	FN	TN

Source: (Daqiqil, 2021).

Based on the table above, the confusion matrix is described into 2 classes, namely Yes and No. in the confusion matrix table is divided into 4 categories namely True Positive (TP), False positive (FP), False Negative (FN) and True Negative (TN). TP is the number of correct positive data in the classification process, FP is the number of negative data classified into positive values, FN is the number of positive data classified into positive values & TN is the number of negative data classified into negative values.

Vol. 19, No. 2 September 2021 | DOI https://doi.org/10.33480/techno.v19i2.3337

II. Research methods

a. Thinking framework

The type of research used in this study is an experimental research model. This study aims to evaluate the data mining classification algorithm. This experimental research emphasizes on existing theories. In this study, the type of research taken is a comparative experiment based on a problem-solving framework as shown in Figure 1.

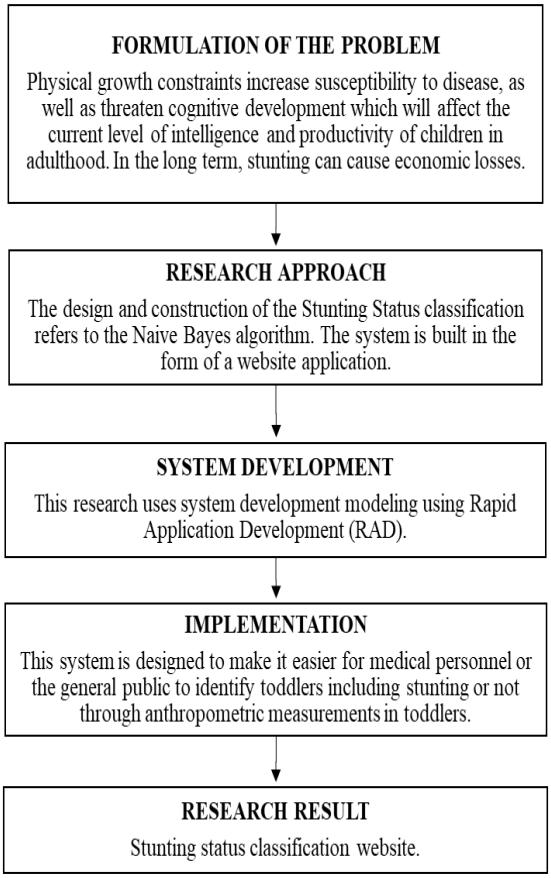

Figure 1. Thinking Framework

The thinking framework of this research begins with the formulation of the stunting status problem which is then made into a model design and system development in the form of a Naive Bayes algorithm to solve the problem.

b. Research design

The research design used can be seen in Figure 2, as follows:

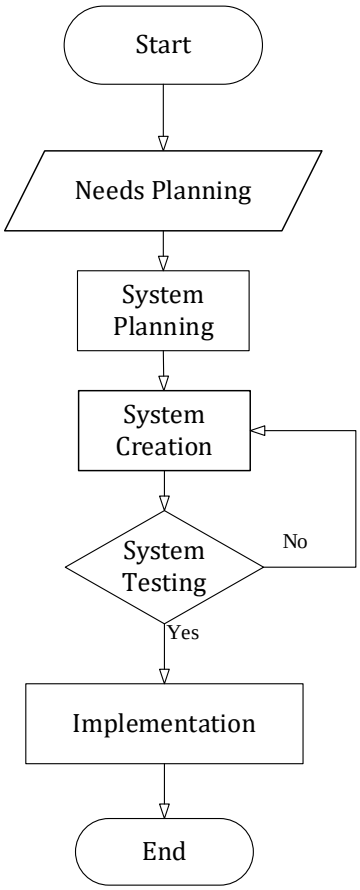

Figure 2. Research Design Flowchart

Description:

1. Needs Planning

This stage is the initial stage, where problem identification is carried out and data collection obtained from users aims to identify the purpose of the system.

2. System Design

At this stage, the system workflow design, system workflow modeling using Unified Modeling Language (UML), calculations using the Naive Bayes algorithm and system User Interface design are carried out.

3. System Creation

At this stage, the program code is written according to the design that has been done. Furthermore, the modules that have been finished are tested whether there are discrepancies in the work.

4. System testing

At this stage the system is tested to ensure the entire workflow of the system works according to its function.

5. Implementation

At this stage the system that has been tested can be implemented by adding hosting so that it gets access to the internet and can be accessed openly.

RESULTS AND DISCUSSION

The data used in this study were taken from the Madiun City Health Office. The data is the result of anthropometric measurements of 673 toddlers. From this data, it will be divided into training data and testing data. From the training data, 70% of the total data are taken, namely 471 data, and 30% of the testing data is taken from the total data, namely 202 data. The stunting data consists of 4 attributes, namely gender, age, height, weight and label. As shown in table 2:

Table 2. Toddler Sample Data

No	Gender	Age (month)	Height (cm)	Weight (kg)	Label
1	F	49	95	14.2	Short
2	F	37	89.7	11.6	Normal
3	M	29	80.5	11.7	Very Short
4	M	45	100	15.6	Normal
5	F	11	69	7	Normal
6	F	2	52.9	4.3	Short
7	M	38	90.7	16.5	Normal
8	M	21	80.4	10.2	Normal
9	M	18	79.3	10.3	Normal
10	M	20	85.3	13	Normal
11	M	38	95	14.6	Normal
12	M	38	90.5	19.3	Normal
13	M	30	85	12	Short
14	M	12	76.7	10	Normal
15	M	57	102.5	16	Normal
16	M	6	69	8.5	Normal
17	M	13	76	9.5	Normal
18	M	24	84.2	12.5	Normal
19	M	52	100.6	14.3	Normal
20	M	42	96	14.9	Tall

Source: Dinas kesehatan (2022)

From table 2 above is a sample of stunting status data which then from the 20 samples above will be used in the naive Bayes calculation data, then the next process is to transform or change the data into a form that is more suitable for the data mining process. The data will be converted into a format that can facilitate the process of predicting stunting status. In this case, in classifying stunting status, namely height according to age (TB/U). Transformation category data can be seen in table 3.

Table 3. Transformation Category Data

Category	Description
Age <=16	Age below or equal to 16 months
Age 17-31	Age between 17-31 months
Age >32	Age above 32 months
Height <=50	Height below or equal to 50 cm
Height 51-100	Height between 51-100 cm
Height >100	Height above 100 cm

The target variable or class will be divided into 4 categories, as shown in table 4.

Table 4. Classes

Variable	Description
1	Very Short
2	Short
3	Normal
4	Tall

1. Naive Bayes Calculation

The training data is then entered into the nave Bayes algorithm model to produce a stunting status prediction model. To test the nave Bayes algorithm, the testing data is entered into the prediction model. In this calculation, training data will be taken from 20 sample data in table 2 which will then be used to determine the classification of the testing data. Below are the steps for nave Bayes using 1 data testing, namely if gender = male, age = 17-31, height = 51-100.

1) It is known that the Label class has 4 classifications, namely:
   C1 = Diagnosis Result = very short
   C2 = Diagnosis Result = Short
   C3 = Diagnosis Result = Normal
   C4 = Diagnosis Result = Tall
   Calculation:
   Count the number of classes.
   The number of each Label class is divided by the total data contained in the training data.
   P(Y=Very Short) = 1/20 = 0,05
   P(Y=Short) = 3/20 = 0,15
   P(Y= Normal) = 16/20 = 0,8
   P(Y=Tall) = 1/20 = 0,05
2) Calculate P(X|Ci), which is the probability of each attribute in data X, then divided by the number of class categories.
   a. Calculating gender class probability
      P(gender = Male | Y=Very short) = 1/1 = 1
      P(gender = Male | Y=Short) = 1/3 = 0,333
      P(gender = Male | Y= Normal) = 14/16 = 0,875
      P(gender = Male | Y= Tall) = 1/1 = 1
   b. Calculating the probability of an age class
      P(Age = 17-31 | Y=Very short) = 1/1 = 1
      P(Age = 17-31 | Y=Short) = 1/3 = 0,333
      P(Age = 17-31 | Y= Normal) = 4/16 = 0,25
      P(Age = 17-31 | Y= Tall) = 0/1 = 0
   c. Calculating the probability of an age class
      P(Height = 51-100 | Y=Very short) = 1/1 = 1
      P(Height = 51-100 | Y=Short) = 3/3 = 1
      P(Height = 51-100 | Y= Normal) = 14/16 = 0,875
      P(Height = 51-100 | Y= Tall) = 0/1 = 0
3) Calculate the total value for each attribute in each classification.

Techno Nusa Mandiri: Journal of Computing and Information Technology
Vol. 19, No. 2 September 2021 | DOI https://doi.org/10.33480/techno.v19i2.3337

P(gender = Male | Y=Very short)* P(Age = 17-31 | Y=Very short)* P(Height = 51-100 | Y=Very short)
= 1 * 1 * 1
= 1
P(gender = Male | Y=Short) * P(Age = 17-31 | Y=Short)* P(Height = 51-100 | Y=Short)
= 0,333*0,333*1
= 0,110889
P(gender = Male | Y=Normal) * P(Age = 17-31 | Y=Normal)* P(Height = 51-100 | Y=Normal)
= 0,875*0,25*0,875
= 0,19140
P(gender = Male | Y=Tall) * P(Age = 17-31 | Y=Tall)* P(Height = 51-100 | Y=Tall)
= 1*0*0
= 0

4) Compare each class result in the classification.

Based on the test data for which the result class is not known, then calculated using the Naïve Bayes method, the data X = (gender = Male, Age = 17-31, Height = 51-100) produces a very short diagnostic value, namely with a score of that is 1.00.

2. Confusion Matrix Calculation

Performance evaluation is carried out using the confusion matrix technique to measure the performance of the data model created. Evaluations carried out using the confuson matrix technique include accuracy, recall precision and f-1 score. The results of the evaluation using the confusion matrix can be seen in Table 5.

Table 5. Confusion Matrix Results

		Predicted Class			
		1	2	3	4
Actual Class	1	40	10	1	0
	2	14	19	21	0
	3	33	5	57	0
	4	1	0	0	1

In table 5, there are actual classes and predicted classes. The actual class is a class whose class/label has been previously determined, while the predicted class is a class that is predicted using the Naive Bayes method.

The results of the calculation of the confusion matrix in the table produce label 1 (very short) there are 40 test data that are classified correctly by the system and 11 data that are classified incorrectly, while on label 2 (short) there are 19 test data that are classified correctly and 35 data that are classified wrong. label 3 (normal) there are 57 test data that are classified correctly by the system and 38 data that are classified incorrectly, then on label 4 (tall) there is 1 test data that is classified correctly and 1 data that is classified incorrectly

Accuracy, precision, recall and f-1 score are each calculated by the confusion matrix module in the sklearn library using jupyter notebook. The calculation of the value of accuracy, precision, recall and f-1 score can be seen in Table 6.

Table 6. Confusion Matrix Result

Label	precision	recall	f1-score
1	0.45	0.78	0.58
2	0.56	0.35	0.43
3	0.72	0.60	0.66
4	1.00	0.50	0.67
accuracy			0.58
macro avg	0.68	0.56	0.58

In Table 6, the average accuracy value is 0.58 or 58%, the average precision value is 0.68 or 68%, the recall average value is 0.56 or 56% and the average value is 56%. the average f-1 score is 0.58 or 58%.

3. System Implementation

The Home page serves as the initial display of the system (dashboard) which contains information such as a navigation bar to carry out the process of moving pages. Then there is speech and information about the classification system as the initial interaction of the system to the user. there is a prediction menu shortcut button that can be classified by the system. The design of the home page can be seen in Figure 3.

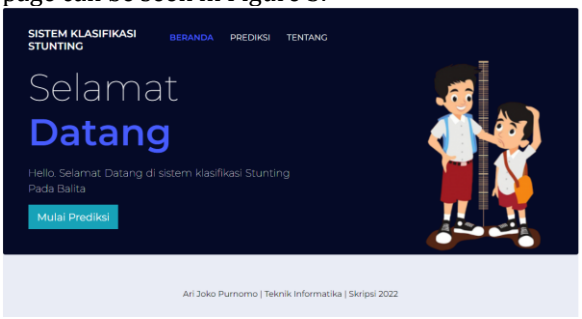

Figure 3. Home page

In the prediction page design, users will be directed to fill out 4 input forms that can be filled with toddler data including gender, age, weight, and height. Then the classification process is carried out on the previously created data model. The prediction process runs by taking data from user data input which is then classified to determine the position of user data belonging to the closest label based on the similarity of the data. The data label

consists of 4 classes, each of which is very short, short, normal, and high. The implementation of the prediction page can be seen in Figure 4.

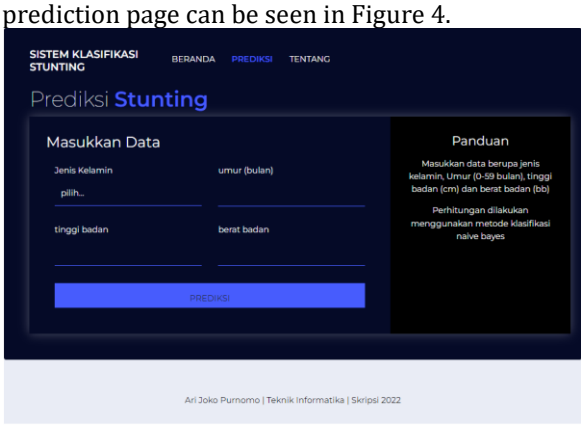

Figure 4. Prediction form page

The prediction results page displays toddler data that is input by the user and has undergone a data classification process. The Stunting data label will appear on the user's screen along with the right solution or steps that the user can take based on the stunting status of toddlers. The implementation of the prediction results page can be seen in Figure 5.

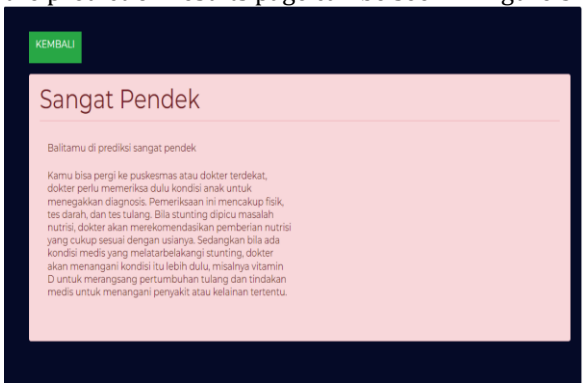

Figure 5. Prediction results page

The guide page contains information about an introduction to stunting and what causes stunting. The implementation of the home page can be seen in Figure 6.

Figure 6. About page

CONCLUSION

Based on the results of the research that has been done, it can be concluded that the outcome of this study is a classification system for stunting status in toddlers that can be used to assist in identifying stunting status in toddlers. Then the calculation of the performance of the naive Bayes classification data model that is made is calculated using the confusion matrix resulting in an accuracy value of 58%, a precision value of 68% and a recall value of 58%. From the results of the confusion matrix test with 30% testing data and 70% training data.

REFERENCE

Daqiqil, I. (2021). *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. Riau: UR Press.

Hesananda, R. (2021). *Algoritma Klasifikasi Bibit Terbaik untuk Tanaman Keladi Tikus*. Pekalongan: Penerbit NEM.

Kementerian Kesehatan Republik Indonesia. (2018). *Pedoman Strategi Komunikasi (Perubahan Perilaku Dalam Percepatan Pencegahan Stunting Di Indonesia)*. Jakarta: Kementerian Kesehatan RI.

Prasetiya, T., Ali, I., Rohmat, C. L., & Nurdiawan, O. (2020). Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, *5*(1), 93.

Prasetyo, E. (2012). *DATA MINING - Konsep dan Aplikasi Menggunkan MATLAB*. Yogyakarta: ANDI Yogyakarta.

Siregar, A. M., & Puspabhuana, A. (2020). *DATA MINING Pengolahan Data Menjadi Informasi dengan RapidMine*. Surakarta: CV Kekata Group

Siregar, N. C., Siregar, R. R. A., & Sudirman, M. Y. D. (2020). *Jurnal Teknologia Implementasi Metode Naive Bayes Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ) Jurnal Teknologia*. *3*(1), 102–110.

Tanoto Foundation. (2021). *Cegah Stunting Sebelum Genting: Peran Remaja Dalam Pencegahan Stunting*. Jakarta: PT Gramedia.

Yenderizal. (2022). *Monograf Algoritma C4.5 Pada Teknik Klasifikasi Penyusutan Volume Pupuk*. Pasaman barat: CV. AZKA PUSTAKA.

Yoto, M., Hadi, M. I., Maghfiroh, I. P. <ila S., Tyastirin, A. Z. M. E., Media, A., Sarweni, A. A. R. K. P., Husnia, Z., Nugraheni, M. E. R., Megatsari, H., & Laksono, A. D. (2020). *Determinan Sosial Penanggulangan Stunting: Riset Aksi*

Partisipatif Desa Sehat Berdaya Fokus Penanggulangan Stunting. Surabaya: Health Advocacy.

Zeniarja, J., Widia, K., & Sani, R. R. (2020). Penerapan Algoritma Naive Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang. *JOINS (Journal of Information System)*, *5*(1), 1–9