



Accuracy Analysis of Community Satisfaction in Population Administration Services Using the C4.5 Algorithm and Naïve Bayes Method

Muhammad Andhika Maulana¹, Agung Triayudi², Ira Diana Sholihati³

Sistem Informatika

Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jalan Sawo Manila, Pasar Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12520

Email: m.andhikamaulana@yahoo.com, agungtriyudi@civitas.unas.ac.id, iradiana2803@gmail.com

ARTICLE INFO

ABSTRACT

Article history:

Received: 04/04/2020

Revised: 20/04/2020

Accepted: 30/05/2020

Keywords:

Naïve Bayes,
C4.5,
Population,
Public Service,
Rapidminer.

The Directorate General of Population and Civil Registration needs indicators for benchmarks in improving the quality of services for the community, one of which is conducting a survey of community satisfaction by using several attributes including requirements, mechanisms, completion times, product specifications, officer competencies, officer attitudes, reporting and advice as well as facilities and infrastructure where the satisfaction results can be classified using data mining. In this study the authors tested two data mining classification methods, namely the C4.5 Algorithm method and the Naïve Bayes method using RapidMiner software. From the test method, the accuracy of 88.00% was obtained using the Naïve Bayes method and the C4.5 algorithm was obtained with an accuracy of 80.40%.

Copyright © 2020 Jurnal Mantik.
All rights reserved.

1. Introduction

Every citizen is required to have a population document. The process of granting Population Documents by the Population Service is included in the process of public service, Good and bad service of the community is a determining factor for the success of the government in providing the best service for the community. The quality of the service includes the efficiency of service time, providing accuracy and legal certainty[1].

In order to create excellent public services, benchmarks are needed to optimize public services. Based on this, this research aims to classify the indicators / elements that cause the level of community satisfaction with the services provided. To find out this, testing is done by testing two methods, namely the C4.5 algorithm method and the Naïve Bayes method, which will compare the performance of the two algorithms in the classification of community satisfaction that has been carried out by the Directorate General of Civil Registration.

Data mining techniques function to classify target variables of type category. Data mining patterns test a number of records, and each record contains a target variable and a set of input or predictor variables [2], this can be utilized in classifying the results of a survey that is satisfied or unsatisfied by business actors in conducting public services. C4.5 algorithm is a problem solving solution and is often used in classification techniques from many other methods [3]. C4.5 algorithm output in the form of a decision tree which means a hierarchy that is used to break a group of large data into a collection of data records that are smaller in size by implementing the arrangement of decision rules. In addition, Naïve Bayes is also used as a classification method in data mining. Naïve Bayes is classifying two methods namely the probablitias method and the statistical method. Which means it can provide future probability predictions based on previous experience.

2. Study of Literature

Some previous studies that applied the C4.5 Algorithm were research conducted by Erlin Elisa (2017) entitled "Analysis and Application of C4.5 Algorithm in Data Mining to Identify Factors that Caused



PT. Arupadhatu Adisesanti's Construction Work Accident" C4.5 algorithm and tested using Data Mining Weka Gui Chooser software, the results were identified that the factors of the causes of workplace accidents at PT. Arupadhatu Adisesanti is workers and how they work, the environment of the workplace, and personal protective equipment[4].

The study was conducted by Ita Yulianti (2019) entitled "Comparative Analysis of C4.5 Algorithm Classification and Naive Bayes on Prediction of Software Reuse Success", the accuracy obtained is 95% [5].

Senna Hendrian (2018) conducted a study entitled "Data Mining Classification Algorithm to Predict Students in Obtaining Educational Data Assistance", the results obtained for the Accuracy Algorithm value of C4.5 were 98.80%, where the method was the best technique for predicting students in getting Education funding[6].

The study was conducted by Agung Triayudi, Riska Susilawati (2017) entitled "Classification of Prospective Funding Customers at Pt Sinar Mitra Sepadan Finance Using the C4 Algorithm. 5 ". Accurately identifying prospective financing customers will be easier using the C4.5 algorithm method, the data are analyzed using the decision tree strategy method, namely C4.5 algorithm, the method can find out which data from consumers are included in the current class group and data from which consumers are included. class groups are not smooth. Then the grouping can also be used to predict consumers who will join, so that the company can make the decision to accept or reject these consumer candidates [7].

Hartatik and Dewi Lestrari (2019) conducted a study "Implementation of the Naive Bayes Algorithm to Determine the Level of Public Satisfaction of Public Services (Case Study: Institute for Agricultural Technology Study in Yogyakarta Special Region)", Test results obtained by the Naive Bayes Method can classify 43 data from 54 data which was tested with an accuracy level of 79,62962963% [8].

Castaka Agus Sugianto and Tri Herdiawan Apandi (2019) in their study entitled "Naive Bayes Algorithm for Predicting Satisfaction of e-KTP Recording Services", obtained research data that is the accuracy of the Naive Bayes method by 91.70% [9].

A. Mukminin and D. Riana (2017) conducted a study "Comparison of C4.5 Algorithms, Naive Bayes and Neural Networks for Soil Classification" obtained the results of an accurate method namely C4.5 Algorithm. The classification is carried out on two classes, three classes and seven classes in a row achieving the highest accuracy value of 98.45%; 93.21% and 83.40%. Thus it can be concluded in this study that the C4.5 algorithm can be referred to the classification of residential development land[10]

RH Pambudi, BD Setiawan, and Indriati (2018) with the research "Application of C4.5 Algorithm to Predict High School Graduation Grades Based on External Factors" gets the results of analysis and testing that the Decision Tree C4.5 Algorithm is accurate and can be implemented to provide predictive value high school student graduation with a grade of 60% accuracy[11]

WD Septiani (2017) in his research entitled "Comparison of C4.5 Algorithm and Naive Bayes Data Mining Classification Methods for Hepatitis Prediction" and obtained an accuracy of 77.29% using the C4.5 Algorithm and 83.71% accuracy using the Naïve Bayes Method[12]

Limitation of the problem in this study with a case study of research at the Directorate General of Duckcapil Ministry of Home Affairs. The sample data was obtained from the community who had made arrangements at the respective domicile service and got a reference to arrange the population documents at the Directorate General of Civil Registration. The indicators used were 8 (eight) indicators, namely requirements, mechanism, service time, service specifications, competency of officers, attitude of officers, handling complaints / input and suggestions as well as facilities and infrastructure. Analyze the results of the comparison of the accuracy of the C4.5 algorithm method and the Naïve Bayes method.

3. Research Methods

3.1 Research Stages

The steps of this research are Data Collection, Testing Method C4.5 Algorithm, Testing Naïve Bayes Method, Evaluation of the method and Implementation of the Application can be seen in accordance with Figure 1 below.

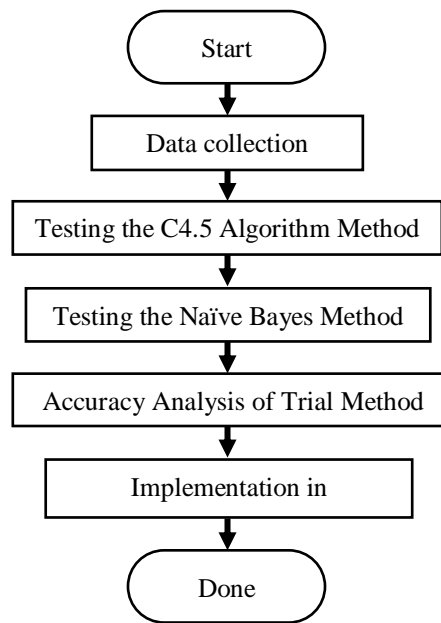


Fig 1. Research stages

3.2 Data collection

Collecting data in this study, used a questionnaire of 250 respondents who were filled out by the community who had taken care of population documents at the Directorate General of Civil Registration.

In this study 8 (eight) indicators were used as input attributes, namely:

1. Requirements are conditions that must be met when handling services, both administratively or technically.
 2. Mechanisms, Procedures and Systems are fundamental service norms for service recipients and implementers.
 3. Service Time in question is the span of time required to complete the service process of each type of service.
 4. Specifications for Service Product Types are outputs of services according to the provisions as they should.
 5. Officer Competency is the level of capacity of implementing officers including skills, knowledge, expertise, and experience.
 6. The attitude of the officer is the behavior of the executor in performing services.
 7. Handling of Reports and Suggestions is orderly in processing the follow-up to complaints.
 8. Means and Prassas. Means are all the objects that function as a tool to achieve certain goals and objectives. Infrastructure is all the mediums that can support the implementation of a process.
- Each attribute is divided into 4 (four) values, namely: Strongly Agree, Agree, Disagree, and Disagree.

3.3 C4.5 Algorithm

C4.5 algorithm of data mining methods used to carry out an estimation and classification task. Algorithm C4.5 is called a tree model like a decision tree which means that each node has an internal node (not including leaves) that can describe each attribute, the branch parts produce an overview of the attributes that have been tested and the parts of the leaf produce a picture from the class, by getting the gain value and the entropy value. The formula for determining the gain value:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \quad (1)$$

Entropy (Si)

Information:

S : set

A : attribute

n : the total partition attribute A

Si : total cases from i-th partition

S : total cases from S

Looking for valueentropy:



$$\text{Entropy}(s) = \sum_{i=1}^n - p_i \times \log_2 p_i \quad (2)$$

Information:

- S : set
- n : total of attribute section S
- A : the features
- p_i : proportion of S_i to S

Stages of testing the C4.5 algorithm method With RapidMiner software, it starts from entering the questionnaire dataset and then cross-validates it for classification and evaluation analysis and produces a decision tree structure as shown in Figure 2.

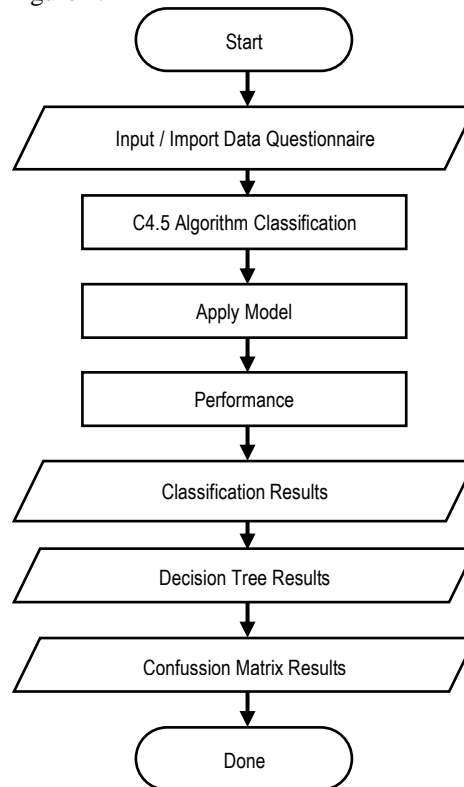


Fig 2. Stages of Testing C4.5 Algorithm Using RapidMiner

3.4 Naïve Bayes

Naive Bayes is a classification algorithm in data mining by implementing the theory of probability used in getting the most appropriate likelihood.

Based on the Bayes theorem the calculation is to classify Naive Bayes predictions:

$$P(C|X) = \frac{P(X|C)P(C)}{P(C)} \quad (3)$$

Information:

- $P(C|X)$: Hypothesis probability based on conditions (posteriori probability)
- $P(X|C)$: Probability according to the circumstances of the hypothesis
- $P(X)$: Probability C
- $P(C)$: Hypothesis probability (prior probability)
- C : Data hypothesis from class Specific
- X : Data class unknown

Stages of method testing Naïve Bayes with RapidMiner software starts from entering the questionnaire dataset and then cross validation is processed for classification and evaluation analysis as shown in Figure 3

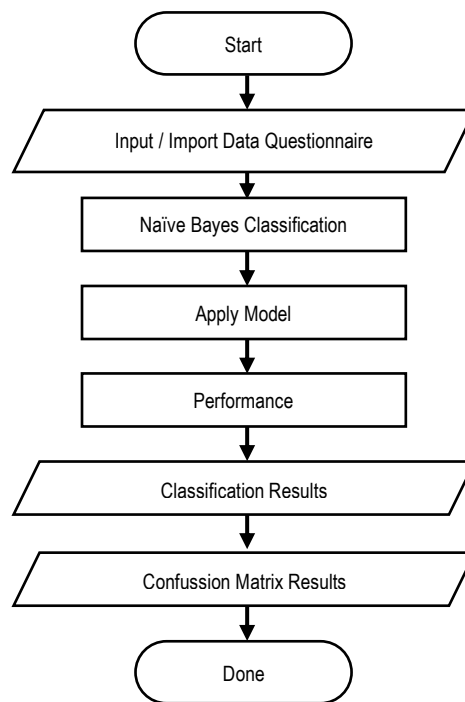


Fig 3. Stages of Naïve Bayes Testing Using RapidMiner

3.5 Analysis of Trial Method

Testing to provide accuracy of data is done by using the confusion matrix technique. Several evaluation methods can be derived from the term confusion matrix such as accuracy, recall, and precision[13].

Table 1
confusion matrix

Prediction Value		True value	
		TRUE	FALSE
TRUE		TP (True Positive)	FP (False Positive)
FALSE		FN (False Negative)	TN (True Negative)

Source: Data Mining Implementation Using Weka
Information:

- TP : the results of the estimates are in accordance with the results actually the same is true.
- FP : Estimated results are correct, but results realit's wrong.
- TN : approximate results and actual results sama - same wrong.
- FN : Estimated results are wrong, but results actually true.

Accuracy is the degree of correlation between the predicted value and the actual value, how to calculate all true-value predictive data then divided by the sum of all data.

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} \tag{4}$$

Precision (Positive Prediction Value) is the level of correctness of the information the user wants on the results presented by the system, the calculation of the total of all positive predictive values stated correctly divided by the total of all class predictions stated true.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$



Recall (Sensitivity) is the level of success of the system to rediscover an information, calculated from the total positive prediction stated correctly divided by the total of all positive stated classes.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

Testing confusion matrix in this study conducted using RapidMiner software.

3.6 Implementation in Application

After analyzing the testing method, the next step is to implement it in an application by implementing the method with the best accuracy value.

4. Results and Discussion

The dataset in this experiment is questionnaire data with a total of 250 questionnaire data. Data filling in this questionnaire is carried out by the people who have taken care of the service in their respective domicile and get a reference to arrange the population documents at the Directorate General of Civil Registration by using 8 (eight) questions that represent the attributes used.

4.1 C4.5 Algorithm

At this stage a data mining trial was conducted, Figure 4 shows the main process of the trial using the C4.5 algorithm method with RapidMiner software by entering the questionnaire dataset then cross validating the classification process and evaluating the analysis and producing a decision tree structure, shown in Figure 5.

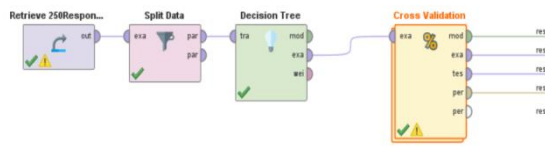


Fig 4. The main process of testing the C4.5 algorithm in rapidminer

Based on the results of a decision tree in Figure 3, the results of the Officer Competency attribute are obtained as the root of the decision tree and the following rules are obtained:

Rules Satisfied

- IF Officer Competency = Strongly Agree THEN Results = Satisfied
- IF Officer Competency = Agree AND Terms = Agree AND Completion Time = Strongly Agree THEN Result = Satisfied
- IF Officer Competency = Agree AND Terms = Agree AND Time Completion = Agree THEN Result = Satisfied
- IF Officer Competency = Agree AND Terms = Agree AND Settlement Time = Disagree THEN Result = Satisfied
- IF Officer Competency = Agree AND Terms = Agree AND Time of Completion = Disagree Agree AND Service Mechanism = Strongly Agree THEN Result = Satisfied

- d) IF Officer Competency = Agree AND Terms = Strongly Agree AND Service Specifications = Disagree AND Service Mechanism = Agree THEN Result = Dissatisfied
- e) IF Staff Competency = Agree AND Terms = Agree AND Service Time = Disagree AND Service Mechanism = Disagree AND Facilities & Infrastructure = Disagree THEN Results = Not Satisfied
- f) IF Officer Competency = Agree AND Terms = Disagree AND Attitude Officer = Disagree THEN Result = Dissatisfied
- g) IF Officer Competency = Agree AND Terms = Disagree AND Attitude Officer = Agree AND Service Specifications = Disagree THEN Result = Dissatisfied
- h) IF Officer Competency = Disagree AND Reporting & Suggestions = Disagree THEN Result = Dissatisfied
- i) IF Officer Competency = Disagree And Reporting & Suggestions = Agree AND Facilities & Infrastructure = Disagree THEN Results = Dissatisfied
- j) IF Officer Competency = Disagree AND Reporting & Suggestions = Agree AND Facilities & Infrastructures = Agree AND Settlement Time = Disagree THEN Result = Dissatisfied
- k) IF Officer Competency = Disagree AND Reporting & Suggestions = Agree AND Facilities & Infrastructure = Agree AND Time Completion = Disagree THEN Result = Dissatisfied
- l) IF Officer Competency = Disagree AND Reporting & Suggestions = Agree AND Facilities & Infrastructure = Disagree AND Service Specifications = Disagree THEN Results = Dissatisfied
- m) IF Officer Competency = Disagree AND Reporting & Suggestions = Agree AND Facilities & Infrastructure = Disagree AND Service Specifications = Strongly Agree THEN Results = Dissatisfied
- n) IF Officer Competency = Disagree AND Reporting & Suggestions = Disagree AND Service Mechanism = Agree THEN Results = Dissatisfied
- o) IF Officer Competency = Disagree And Reporting & Suggestions = Disagree AND Service Mechanism = Disagree THEN Results = Dissatisfied

From the testing of the C4.5 algorithm the following confusion matrix table is also produced:

Table 2
Confusion matrix C4.5 test algorithm

	True satisfied	True not satisfied
Pred. satisfied	146	32
Pred. not satisfied	17	55

Source: Data Processing

In the confusion matrix above obtained positive data clarified correctly numbered 146, positive data but clarified incorrectly as many as 17, then negative data recorded and correctly classified as many as 32 and negative data but classified with incorrect results as many as 55.

In the next stage of the confusion matrix above will be calculated accuracy, recall, and precision.

$$\text{Akurasi} = \frac{146 + 55}{146 + 55 + 17 + 32} \times 100\% = 80.40\%$$

From the calculation of the accuracy obtained 80.40% accuracy results using the C4.5 algorithm

$$\text{Recall} = \frac{146}{146 + 17} \times 100\% = 89.57\%$$

For the calculation of recall results obtained a value of 89.57%

$$\text{Precision} = \frac{146}{146 + 32} \times 100\% = 82.02\%$$

And for the precision calculation results obtained by 82.02%

4.2 Naïve Bayes

Figure 6 shows the main process of testing the Naïve Bayes method with RapidMiner, in this method a data mining trial is performed by calculating P (X) on each attribute. In the dataset used, the target attribute (label) is expressed in 2 (two) categories, namely PUAS and NOT SATISFIED.

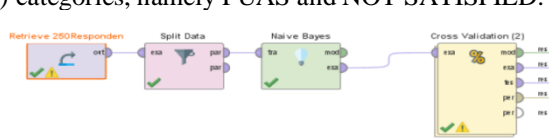


Fig 6. The main process of testing the Naive Bayes method in Rapidminer

For example the calculation will use 1 (one) testing data which will be calculated manually, namely:

Table 3
Testing data

Attribute	Score
Requirements	Not agree
Mechanism	Agree
Service time	Not agree
Product Specifications	Agree
Officer competence	Agree
Attitude of the Officer	Agree
Reporting & advice	Not agree
Infrastructure	Agree
Results	?

Source: Data Processing

Information:

C1 : Results = Satisfied

C2 : Result = Not Satisfied

X1 : Requirements

X2 : Mechanism

X3 : Service Time

X4 : Product Specifications

X5 : Officer Competency

X6 : Attitude of the Officer

X7 : Reporting & Suggestions

X8 : Infrastructure

Here are the steps for calculating the Naïve Bayes method in calculating the class probability for C1 = Satisfied and C2 = Dissatisfied.

$$P(C1 = Satisfied) = 163/250 = 0.652$$

$$P(C2 = Dissatisfied) = 87/250 = 0.348$$

Then the steps to calculate the value of $P(X | Ci)$ for attribute values based on table 3:

$$P(X1 = Disagree | C1 = Satisfied) = 19/163 = 0.11656$$

$$P(X1 = Disagree | C2 = Dissatisfied) = 41/87 = 0.47126$$

$$P(X2 = Agree | C1 = Satisfied) = 98/163 = 0.60123$$

$$P(X2 = Agree | C2 = Not Satisfied) = 44/87 = 0.50575$$

$$P(X3 = Disagree | C1 = Satisfied) = 54/163 = 0.33129$$

$$P(X3 = Disagree | C2 = Dissatisfied) = 40/87 = 0.45977$$

$$P(X4 = Agree | C1 = Satisfied) = 99/163 = 0.60736$$

$$P(X4 = Agree | C2 = Not Satisfied) = 38/87 = 0.43678$$

$$P(X5 = Agree | C1 = Satisfied) = 102/163 = 0.62577$$

$$P(X5 = Agree | C2 = Not Satisfied) = 40/87 = 0.45977$$

$$P(X6 = Agree | C1 = Satisfied) = 86/163 = 0.52761$$

$$P(X6 = Agree | C2 = Not Satisfied) = 22/87 = 0.25287$$

$$P(X7 = Disagree | C1 = Satisfied) = 39/163 = 0.23926$$

$$P(X7 = Disagree | C2 = Dissatisfied) = 44/87 = 0.50575$$

$$P(X8 = Agree | C1 = Satisfied) = 74/163 = 0.45399$$

$$P(X8 = Agree | C1 = Not Satisfied) = 28/87 = 0.32184$$

After that the calculation of $P(X | Ci) P(Ci)$ is as follows:

$$P(X | C1 = Satisfied) = 0.11656 \times 0.60123 \times 0.33129 \times 0.60736 \times 0.62577 \times 0.52761 \times 0.23926 \times 0.45399 = 0.000505711$$

$$P(X | C2 = Not Satisfied) = 0.47126 \times 0.50575 \times 0.45977 \times 0.43678 \times 0.45977 \times 0.25287 \times 0.50575 \times 0.32184 = 0.000905773$$

$$P(X | C1) P(C1) = 0.000505711 \times 0.652 = 0.000329724$$

$$P(X | C2) P(C2) = 0.000905773 \times 0.348 = 0.000315209$$

The results of these calculations can be explained that $P(X | C1 = Satisfied) > P(X | C2 = Not Satisfied)$, which means the results of the testing data are PUAS.

The following is the Naïve Bayes method confusion matrix table



Table 4
Confusion matrix naïve bayes testing

	True satisfied	True not satisfied
Pred. satisfied	151	18
Pred. not satisfied	12	69

Source: Data Processing

In the confusion matrix above obtained positive data clarified correctly as many as 151, positive data but clarified incorrectly as many as 12, then negative data that was clarified correctly as many as 18 and negative data but clarified incorrectly as many as 69.

Furthermore, from the confusion matrix above will be calculated accuracy, recall, and precision.

$$\text{Akurasi} = \frac{151 + 69}{151 + 69 + 12 + 18} \times 100\% = 88.00\%$$

From the calculation of accuracy, the accuracy of 88.00% is obtained using the Naive Bayes method

$$\text{Recall} = \frac{151}{151 + 12} \times 100\% = 92.64\%$$

For the calculation of recall obtained value of 92.64%

$$\text{Precision} = \frac{151}{151 + 18} \times 100\% = 89.35\%$$

And for the precision calculation results obtained by 89.35%

4.3 Analysis of Trial Method

Based on the results of trials on the method C4.5 and Naïve Bayes algorithms obtained the comparison as follows:

Table 5
Comparison of the results of testing the C4.5 algorithm method and the naïve Bayes method

Method	Accuracy	Recall	Precision
C4.5 Algorithm	80.40%	89.57%	82.02%
Naive Bayes	88.00%	92.64%	89.35%

Source: Data Processing

The table above shows the Naïve Bayes method has high accuracy, recall and precision values and is better than the C4.5 algorithm method

4.4 Implementation in Application

Based on the analysis results of the test, the naïve bayes method gets the best accuracy value, so that in implementing it into the application the naïve bayes method will be used as a classification method for community satisfaction in population document services at the Directorate General of Civil Registration

In this application users can add training data by importing documents in Excel format or adding new questionnaire data such as Figure 7 and Figure 8.



Fig 7. Data import page



Fig 8. Data added page

After that the user can process the questionnaire data by clicking on the Naïve Bayes process button and immediately obtain the results as in Figure 9 below.

The screenshot shows a table with the following columns: No, Nama, Pekerjaan, Pekerjaan, Waktu, Spesifikasi, Ketersediaan, Waktu, Pengisian, Status, and Hasil. The table contains 10 rows of data representing individual survey responses.

Fig 9. Results page

5. Conclusion

Based on the results of research that has been done, the accuracy of the classification of public satisfaction with population document services at the Directorate General of Civil Registration, the Naïve Bayes method gets a better accuracy value with an accuracy value of 88.00% compared to using the Algoritma C4.5 method which only obtains an accuracy value of 80.40%, the naïve bayes method is also fast and efficient and easy to understand, implementation in the application can also provide information to the Directorate General of Civil Registration for community satisfaction when serving population documents so that it can improve the quality of service to the community better in the future.

6. Reference

- [1] A. Mukminin and D. Riana, “Komparasi Algoritma C4 . 5 , Naïve Bayes Dan Neural Network Untuk Klasifikasi Tanah,” J. Inform., 2017.
- [2] Agung Triayudi, Riska Susilawati. 2017. Klasifikasi Calon Nasabah Pembiayaan Pada Pt Sinar Mitra Sepadan Finance Menggunakan Algoritma C4. 5. ProTekInfo (Pengembangan Riset dan Observasi Teknik Informatika). Vol. 2: 59-62.
- [3] E. Elisa, “Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti,” J. Online Inform., 2017, doi: 10.15575/join.v2i1.71.
- [4] Hartatik dan Lestari Dewi.2019. Implementasi Algoritma Naive Bayes Untuk Menentukan Tingkat Kepuasan Masyarakat Terhadap Pelayanan Publik (Studi Kasus: Balai Pengkajian Teknologi Pertanian Daerah Istimewa Yogyakarta). FAHMA. Jurnal Teknologi dan Ilmu Komputer. 17(2): 37-46
- [5] I. Yulianti, “Analisis Komparasi Klasifikasi Algoritma C4.5 Dan Naïve Bayes Pada Prediksi Keberhasilan Software Reuse,” Swabumi, 2019, doi: 10.31294/swabumi.v7i1.5637.
- [6] Kemenpan, “Peraturan Menteri Pendayagunaan Aparatur Negara Dan Reformasi Birokrasi Republik Indonesia Nomor 14 Tahun 2017 Tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik,” Pedoman Penyusunan Survei Kepuasan Masy. Unit Penyelenggara Pelayanan Publik, 2017.
- [7] Muflikhah, Lailil, dkk. 2018. Data Mining. Malang: UB Press.
- [8] Nofriansyah, Dicky dan Nurcahyo, GW. 2019. Algoritma Data Mining dan Pengujian. Yogyakarta: Penerbit Deepublish.
- [9] R. H. Pambudi, B. D. Setiawan, and Indriati, “Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal,” J. Pengemb. Teknol. Inf. dan Ilmu Komput., 2018.
- [10] S. Adinugroho and Y. A. Sari, “Implementasi Data Mining Menggunakan Weka,” in Pengantar Data Mining, 2018.
- [11] S. Hendrian, “Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan,” Fakt. Exacta, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [12] T. H. Apandi and C. A. Sugiarto, “Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP,” JUITA J. Inform., 2019, doi: 10.30595/juita.v7i2.3608.
- [13] W. D. Septiani, “Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis,” None, 2017, doi: 10.33480/pilar.v13i1.149.

