PREDICTION OF PUBLIC SERVICE SATISFACTION USING C4.5 AND NAÏVE BAYES ALGORITHM

Yuyun Umaidah[1*)] ; Ultach Enri[2]

[1,2] Teknik Informatika
Universitas Singaperbangsa Karawang
www.unsika.ac.id [1, 2]
yuyun.umaidah@staff.unsika.ac.id[1], ultach@staff.unsika.ac.id[2]
(*) Corresponding Author

Abstract*—One of the things that has often been questioned lately is in the field of public services, especially in terms of the quality or service quality of government agencies to the community, the Manpower and Transmigration Office of Kab. Karawang is a government agency in charge of public services. where one of the tasks is to make an AK.1 card (yellow card), based on this problem the Manpower and Transmigration Office of Kab. Karawang Regency. Karawang seeks to improve service quality in order to satisfy consumers by distributing questionnaires to every consumer who is making an AK card.1. In this study, we will apply the C4.5 and Naïve Bayes algorithms to predict the satisfaction of public services with the nominal type of dataset used. The evaluation is done based on a comparison of the level of accuracy, precision, recall, and F-Measure using a confusion matrix. From the research that has been carried out, the Naïve Bayes algorithm with 70% training data distribution and 30% testing is able to provide better predictive results than the C4.5 algorithm as evidenced by the accuracy value = 96.89%, precision = 95.50%, recall = 95.00% and f-measure = 94.60%.*

Keywords: *accurasy, precision, recall, f-measure*

Abstrak— Salah satu hal yang belakangan ini sering dipermasalahkan adalah dalam bidang pelayanan publik, terutama dalam hal kualitas atau mutu pelayanan instansi pemerintah kepada masyarakat, dinas tenaga kerja dan transmigrasi Kab. Karawang adalah instansi pemerintahan yang bertugas dalam bidang pelayanan publik. dimana salah satu tugasnya adalah pembuatan kartu AK.1(kartu kuning), berdasarkan permasalahan tersebut dinas tenaga kerja dan transmigrasi Kab. Karawang Kab. Karawang berupaya untuk meningkatkan kualitas pelayanan agar dapat memuaskan konsumen dengan cara menyebarkan kuesioner kepada setiap konsumen yang sedang membuat kartu AK.1. Pada penelitian ini akan menerapkan algoritma C4.5 dan Naïve bayes untuk memprediksi kepuasan pelayanan publik dengan dataset yang digunakan bertipe nominal. Evaluasi dilakukan berdasarkan pada perbandingan tingkat akurasi, Precision, Recall, dan F-Measure dengan menggunakan confusion matrix. Dari penelitian yang telah dilakukan, algoritma Naïve Bayes dengan pembagian data training 70% dan testing 30% mampu memberikan hasil prediksi yang lebih baik dibandingkan dengan algoritma C4.5 terbukti dengan nilai akurasi = 96,89%, Presisi = 95,50%, recall= 95,00% dan f-measure= 94,60%.

Kata Kunci*:* akurasi, presisi, recall, f-measure.

INTRODUCTION

The government as a service provider must be able to provide good and quality services for the community. Both in terms of facilities, reliability, response, assurance, and empathy. In the era of regional autonomy, the quality of service from government officials will be increasingly challenged to be optimal and able to answer the increasingly high demands of society (Angraeni, 2014). The public as consumers of service users often have the view that convoluted bureaucracy, uncomfortable places, unprofessional, slow, and unresponsive employees, and unsatisfactory performance tend to create a negative image of the quality of service itself.

If this is not addressed, it will lead to a crisis of confidence in government agencies. With this condition of society, it is hoped that government agencies will further improve the quality of services so that what the community wants can be fulfilled. Therefore, efforts to improve service quality cannot be separated from efforts to measure consumer (community) satisfaction, the results of which are input for efforts to improve and improve the quality of service itself (Irmaini & Sugiarti, 2009).

Department of Manpower and Transmigration Karawang Regency as one of the government agencies in charge of public service, one of its duties is to make an AK.1 card or better known as a yellow card which is a job seeker's card which is used as one of the requirements to apply for a job in a company or company. institutions,

either public or private. Every month the manufacture of yellow cards is increasing, so that institutions are more required to be able to provide the best service to increase community satisfaction which will automatically improve the image of public services from the Department of Manpower and Transmigration Karawang Regency.

In helping to improve community services, it can be started by knowing how big the level of community satisfaction is with the services provided by the agency by exploring input from the community in the form of information related to their assessment of the facilities, reliability, response, guarantee and empathy available at the Manpower Office. and transmigration Karawang Regency. Prediction to find out community satisfaction with public services is one way that will be done to help deal with these problems.

The appropriate field of science in dealing with this problem is Data Mining, which is a field of science that aims to utilize data in databases by processing it to produce useful new information. (Prasetyo, 2014). There are various methods used in data mining in processing data in general, such as Clustering, Classification, Association, Estimation, prediction. (Santosa, 2007)

The C4.5 algorithm is one of the algorithms of the classification method that can produce a decision tree. The decision tree is a part of a strong and well-known classification and prediction method, the decision tree performs processing / changes to a large data into a decision tree in which there are rules that are easy to understand (Mochamad Rizki Ilham, 2016). The C4.5 algorithm is a form of ID3 development. Some of the developments carried out by C4.5 include the potential for advanced data processing and pruning. (Rezkika et al., 2021) There are several algorithms that are suitable for making predictions, such as research conducted by (Tarigan et al., 2017) using the KDD method and the C4.5 algorithm in a study entitled the implementation of the C4.5 algorithm on customer satisfaction resulting in an accuracy of 91%. Furthermore, (Rohman, 2015) in his research taking the topic of predicting customer satisfaction at Perum DAMRI, the study used the C4.5 algorithm which obtained data from 90 samples of DAMRI customers. From this research, predictions were obtained for the level of customer satisfaction with an accuracy value of 93%. A study using the Naïve Bayes algorithm was also conducted by (Hakim, 2016) with the title Predicting Restaurant Customer Satisfaction Levels Using the Naive Bayes Algorithm, resulting in an accuracy value of 94.67%.

Naïve Bayes Classifier is a technique classification of probabilistic and statistical methods (Farid et al., 2021). Previous research has also been conducted by (Risqianti & Ismanto, 2017) regarding the comparison of the C4.5 and Naïve Bayes algorithms for student graduation times. The results of this study indicate that the C4.5 algorithm has a better accuracy value than the Naïve Bayes algorithm. C4.5 84.95% and Naïve Bayes 83.36% with a difference in accuracy of 1.59%. (Listiana et al., 2015) conducted a study on the comparison of the Decision Tree (C4.5) and Naïve Bayes algorithms for identifying the growth and development of children under five. The results of this study indicate that Naïve Bayes is better to use than the decision tree method. Accuracy value is 75.66% decision tree and 76.97% nave Bayes.

From previous studies, it can be explained that the C4.5 and Naïve Bayes algorithms are very good for predicting the level of customer satisfaction. Therefore, in this study, the Prediction of Public Service Satisfaction Using the C4.5 Algorithm and Naïve Bayes with Case Study: Manpower and Transmigration Office of Karawang Regency will be carried out.

The purpose of this study is expected to help the Manpower and Transmigration Office of Karawang Regency in improving the AK.1 card for a better service, by applying the best algorithm from C4.5 or Nave Bayes to analyze consumer satisfaction in making AK.1 cards at the Manpower Office and transmigration in Karawang Regency.

MATERIALS AND METHODS

The research uses data collection techniques and utilizes existing methods in data mining to process data, namely KDD (Knowledge Discovery in Databases) based on (Han et al., 2012):

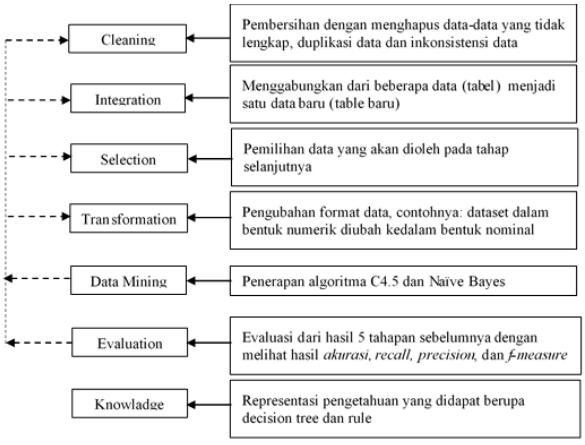

Figure 1. Research design

The steps in completing this research consist of data cleaning, data integration, data selection, data transformation, data mining, evaluation and knowledge as can be seen in Figure1.

At this stage the determination of the sample and the amount of data from consumer or community data who will make an Ak1 card uses the slovin method in determining the number of residents. (Nugroho & Haryati, 2015). The following is the process of determining the number of samples using the slovin method, which is based on consumer data on making AK.1 card in May 2017 with a total of 1414, and an error of 5%. Then the number of samples to be taken are:

$$n = \frac{N}{1+N\,e^2} \quad \text{.....(1)}$$

$$n = \frac{1414}{1 + 1414\,.\,(0.05)^2}$$

$$n = \frac{1414}{1 + 1414\,.\,(0.0025)}$$

$$n = \frac{1414}{4.535} = 311.79713$$

From the results of these calculations, it is rounded up to 312 data or consumers which are used for sampling material. With this method, 312 data were obtained, namely in the form of questionnaires filled out by the community as consumers of making yellow cards at the Manpower and Transmigration Office of Karawang Regency.

Data integration is the process of combining data from the results of 312 questionnaires into a new table. The new table or dataset of the results of this questionnaire comes from one source, namely the Manpower and Transmigration Office of Karawang Regency especially in the making of yellow cards (AK.1). After the data is combined, the average data is taken from the facility data, reliability, responsiveness, assurance, and empathy. The data that has been averaged will be processed in the next stage.

Data selection is the stage of selecting data to be carried out, only data that is suitable for analysis.

Table1. Selected dataset

Attribute	Description
Facility	It is a data of convenience and supporting facilities in the service of making yellow cards (AK.1)
Reliability	Represents data regarding the reliability, ability and understanding of the duties and functions of service personnel.
Response	Represents data relating to punctuality and a sense of responsibility.
Guarantee	It is data on service security and politeness and good communication in providing services.
Empathy	Represents data regarding the impartiality of service personnel and the suitability of the service hours specified.

In the table 1 is the attributes that will be used for this research, there are 312 data and 5 class attributes, namely facilities, reliability, responsiveness, assurance, empathy and 1 predictor attribute, namely results.

The transformation stage is the stage of changing the shape of the dataset which is initially numeric and then transformed into nominal so that it is easy to proceed to the data mining process. The changes were carried out by provisions based on the Likert scale categorization (Sugiyono, 2015):

Strongly Agree = 5
Agree = 4
Disagree = 3
Disagree = 2
Strongly Disagree = 1

Table 2. Numeric type dataset

Facility	Reliability	Response	Guarantee	Empathy	Results
5	4	4	4,8	4,3	Satisfied
5	4	4	4,8	4,3	Satisfied
5	4,8	4,8	4,8	4,5	Satisfied
4,5	4,5	4	3,8	4,5	Satisfied
4,5	4,8	4,3	4	4,5	Satisfied
3,8	3,8	4	4	4	Satisfied
4,5	4,8	4,8	5	4,8	Satisfied
5	5	5	5	5	Satisfied
5	5	5	5	5	Satisfied
2,3	3	2	3,3	1,8	Not Satisfied
3,3	3,5	3	2,8	3	Not Satisfied

In table 2 is the data before it is transformed. Furthermore, the numerical data will be transformed into nominal data.

Table 3. Nominal Dataset

Facility	Reliability	response	Guarantee	Empathy	Results
strongly agree	agree	agree	strongly agree	strongly agree	Satisfied

strongly agree	agree	agree	strongly agree	strongly agree	Satisfied
strongly agree	strongly agree	strongly agree	strongly agree	strongly agree	Satisfied
strongly agree	strongly agree	agree	agree	strongly agree	Satisfied
strongly agree	strongly agree	strongly agree	agree	strongly agree	Satisfied
agree	agree	agree	agree	agree	Satisfied
strongly agree	strongly agree	strongly agree	strongly agree	strongly agree	Satisfied
strongly agree	strongly agree	strongly agree	strongly agree	strongly agree	Satisfied
strongly agree	strongly agree	strongly agree	strongly agree	strongly agree	Satisfied
disagree	disagree	do not agree	agree	do not agree	Not satisfied
agree	agree	disagree	disagree	don't agree	Not satisfied

Table 3 contains a dataset that has been transformed in the form of nominal data and then the data will be processed to the data mining stage

RESULT AND DISCUSSION

In this study, predictions of public service satisfaction will be carried out using the C4.5 and Naïve Bayer algorithms. This study conducted several tests on the distribution of datasets. The test is carried out by dividing the training and testing data into 70%-30%, 80%-20% and 90%-10% of the total dataset of 312 data and the form of data is nominal data.

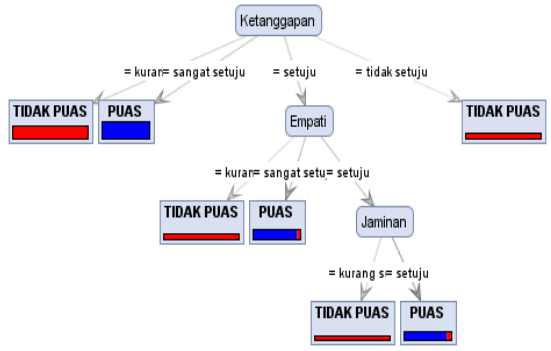

Figure 2. Decision Tree C4.5

```
Ketanggapan = kurang setuju: TIDAK PUAS {PUAS=0, TIDAK PUAS=21}
Ketanggapan = sangat setuju: PUAS {PUAS=33, TIDAK PUAS=0}
Ketanggapan = setuju
|   Empati = kurang setuju: TIDAK PUAS {PUAS=0, TIDAK PUAS=4}
|   Empati = sangat setuju: PUAS {PUAS=15, TIDAK PUAS=1}
|   Empati = setuju
|   |   Jaminan = kurang setuju: TIDAK PUAS {PUAS=0, TIDAK PUAS=3}
|   |   Jaminan = setuju: PUAS {PUAS=10, TIDAK PUAS=1}
Ketanggapan = tidak setuju: TIDAK PUAS {PUAS=0, TIDAK PUAS=6}
```

Figure 3. Rule C4.5

In the Figure 2 and Figure 3 shows the results of the best decision tree and rule from the C4.5 algorithm with the distribution of training and testing data 70%-30%

Attribute	Parameter	PUAS	TIDAK PUAS
Fasilitas	value=sangat setuju	0.603	0.028
Fasilitas	value=setuju	0.379	0.250
Fasilitas	value=kurang setuju	0.017	0.472
Fasilitas	value=tidak setuju	0.000	0.222
Fasilitas	value=sangat tidak setuju	0.000	0.028
Fasilitas	value=sagat setuju	0.000	0.000
Fasilitas	value=unknown	0.000	0.000
Kehandalan	value=setuju	0.327	0.444
Kehandalan	value=sangat setuju	0.672	0.111
Kehandalan	value=kurang setuju	0.000	0.361
Kehandalan	value=tidak setuju	0.000	0.083
Kehandalan	value=sangat tidak setuju	0.000	0.000
Kehandalan	value=sagat setuju	0.000	0.000
Kehandalan	value=unknown	0.000	0.000
Ketanggapan	value=setuju	0.431	0.250
Ketanggapan	value=sangat setuju	0.569	0.000
Ketanggapan	value=tidak setuju	0.000	0.167
Ketanggapan	value=kurang setuju	0.000	0.583
Ketanggapan	value=sangat tidak setuju	0.000	0.000

Figure 4. Distribution Model Naïve Bayes

```
PerformanceVector:
accuracy: 95.89% +/- 5.04% (mikro: 95.74%)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   56      2
TIDAK PUAS:     2       34
AUC: 0.992 +/- 0.019 (mikro: 0.992) (positive class: TIDAK PUAS)
precision: 95.50% +/- 9.07% (mikro: 94.44%) (positive class: TIDAK PUAS)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   56      2
TIDAK PUAS:     2       34
recall: 95.00% +/- 10.00% (mikro: 94.44%) (positive class: TIDAK PUAS)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   56      2
TIDAK PUAS:     2       34
f_measure: 94.60% +/- 6.67% (mikro: 94.44%) (positive class: TIDAK PUAS)
ConfusionMatrix:
True:   PUAS    TIDAK PUAS
PUAS:   56      2
TIDAK PUAS:     2       34
```

Figure 5. Performance Vector Naïve Bayes

In the Figure 4 and Figure 5 shows the process of using the C4.5 and Naïve Bayes algorithms, where from the test results with several data sharing, it can be seen which algorithm is better for predicting public service satisfaction at the Manpower and Transmigration Office of Karawang Regency.

Evaluation

After doing several tests using the C4.5 and Nave Bayes algorithms at the Data mining stage, a model will be obtained where from the model an evaluation stage is carried out to measure the performance of the C4.5 and Nave Bayes algorithms by looking at the level of *accuracy*, *precision*, *recall* and *f-measure*. The results of the evaluation can be seen in table 4 and table 5.

Table 4. Evaluation Results C4.5

C4.5. Data Sharing Results				
Data Sharing		**70% - 30%**	**80% - 20%**	**90% - 10%**
Training	Accuracy	77,06%	78,40%	94,67%
	Precision	100%	97,44%	97,83%
	Recall	39,02%	41,30%	87,50%
	F-Measure	56,14%	58,46%	92,37%
Testing	Accuracy	95,89%	90,71%	86,67%
	Precision	97,50%	97,50%	100%
	Recall	92,50%	85,83%	70%
	F-Measure	93,81%	88,57%	83,33%

In table 4**,** it can be seen from the best testing value, the distribution of 70% *training* data and 30% *testing* is the best data division by producing *accuracy* = 95.89%, *precision* = 97.50%, *recall* = 92.50% and *f-measure* = 93 ,81%.

Table 5 Evaluation Result Naïve Bayes

Naïve Bayes Data Sharing Results				
Data Sharing		**70% - 30%**	**80% - 20%**	**90% - 10%**
Training	Accuracy	94,95%	95,60%	96,09%
	Precision	96,12%	95,45%	96,04%
	Recall	90,24%	92,37%	93,27%
	F-Measure	93,09%	93,81%	94,61%
Testing	Accuracy	96,89%	95,24%	93,33%
	Precision	95,50%	97,50%	100%
	Recall	95%	94,17%	85%
	F-Measure	94,60%	95,14%	92,31%

In table 5, it can be seen from the best testing value, the distribution of 70% *training* data and 30% *testing* is the best data division by producing *accuracy* = 96.89%, *precision* = 95.50%, *recall* = 95.00% and *f-measure* = 94.60%.

Table 6. Comparison Results of C4.5 and Naïve Bayes

Algorithm		**C4.5**	**Naïve Bayes**
Training	Accuracy	77,06%	94,95%
	Precision	100%	96,12%
	Recall	39,02%	90,24%
	F-Measure	56,14%	93,09%
Testing	Accuracy	95,89%	96,89%
	Precision	97,50%	95,50%
	Recall	92,50%	95%
	F-Measure	93,81%	94,60%

Based on the tests that have been carried out, using the C4.5 and Nave Bayes algorithms with data sharing of 70%-30%, 80%-20% and 90%-10%, the best value is obtained in the Naïve Bayes Algorithm with 70% *training* data and 30% *testing* with *accuracy* = 96.89%, *precision* = 95.50%, *recall* = 95.00% and *f-measure* = 94.60% as it can be seen in Table 6.

Knowledge

At this stage, a research report will be made which later on from the results of this research is expected to help the Department of Manpower and Transmigration of Karawang Regency in improving the service of making AK.1 card.

CONCLUSION

From the results of the research that has been done, it can be concluded that in this study the Naïve Bayes algorithm is able to provide better prediction results than the C4.5 algorithm as evidenced by the results of *accuracy*, *precision*, *recall* and *f-measure*, respectively as follows: algorithm C4.5 *accuracy* = 95.89%, *precision* = 97.50%, *recall* = 92.50% and *f-measure* = 93.81%. Nave Bayes algorithm *accuracy* = 96.89%, *precision* = 95.50%, *recall* = 95.00% and *f-measure* = 94.60%. The distribution of the dataset with 70% *training* data and 30% *testing* data is a better dataset distribution if it is used to predict public service satisfaction in the form of nominal data with a total of 312 data.

REFERENCES

Angraeni, R. (2014). *Hubungan Kualitas Pelayanan Pembuatan Kartu Kuning Dengan Kepuasan Pencari Kerja Pada Dinas Tenaga Kerja*.

Farid, F., Enri, U., & Umaidah, Y. (2021). Sistem Pendukung Keputusan Rekomendasi Topik Skripsi Menggunakan Naïve Bayes Classifier. *JOINTECS (Journal of Information Technology and Computer Science)*, *6*(1), 35. https://doi.org/10.31328/jointecs.v6i1.2076

Hakim, A. N. (2016). Prediksi Tingkat Kepuasan Pelanggan Rumah Makan Menggunakan Algoritma Naive Bayes. *Fik*, *1*(1), 1–2.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining Concepts and Techniques. In *Data Mining*. https://doi.org/10.1016/b978-0-12-381479-1.00001-0

Irmaini, Z., & Sugiarti. (2009). Aplikasi Importance-Performance Analysis Dalam Menilai Kualitas Pelayanan Pembuatan Kartu Ak.1 Pada Dinas Tenaga Kerja Dan Transmigrasi Kabupaten Cilacap. *STIE SATRIA Purwokerto*, 1–15.

Listiana, M., Sujalwo, S., & Gunawan, D. (2015). Perbandingan Algoritma Decision Tree (C4.5) Dan Naïve Bayes Pada Data Mining Untuk Identifikasi Tumbuh Kembang Anak Balita

(Studi Kasus Puskesmas Kartasura) [Universitas Muhammadiyah Surakarta]. In *Institusional Repository*. http://eprints.ums.ac.id/36124/

Mochamad Rizki Ilham, P. (2016). Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti. *Simplementasi Data Mining Menggunakan Algoritma C4.5 Untuk Prediksi Kepuasan Pelanggan Taksi Kosti*, *Vol. 4*, *No*(5), 11.

Nugroho, Y. S., & Haryati, S. N. (2015). Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali. *Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika*, *1*(1), 1. https://doi.org/10.23917/khif.v1i1.1175

Prasetyo, E. (2014). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Andi Publisher.

Rezkika, F., Sari, B. N., & Irawan, A. S. Y. (2021). Klasifikasi Masa Tunggu Alumni Untuk Mendapatkan Pekerjaan Berdasarkan Kompetensi Menggunakan Algoritma C4.5 (Studi Kasus : Fasilkom Unsika). *Progresif: Jurnal Ilmiah Komputer*, *17*(2), 95–106. http://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/view/652

Risqianti, & Ismanto, B. (2017). Analisis Komparasi Algoritma Naive Bayes Dan C4-5 Untuk Waktu Kelulusan Mahasiswa. *IC-Tech*, *XII*(1), 33–38.

Rohman, I. F. (2015). Penerapan Algoritma C.45 Pada Kepuasan Pelanggan Perum Damri. *Ilmu Komputer*, 1–14.

Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu.

Sugiyono. (2015). Metode Penelitian Pendidikan, Pendekatan Kuantitatif, Kualitatif, dan R&D. In *Penerbit Alfabeta Bandung*. https://doi.org/10.1103/PhysRev.47.506

Tarigan, H. D., Destiawati, F., & Fitriansyah, A. (2017). Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan. *Ethos (Jurnal Penelitian Dan Pengabdian Masyarakat)*, *Vol. 6*, *No*, 80–86.