

Decision Support System for Receiving Waste Retribution at the Housing and Sanitation Services Uses the Naïve Bayes Algorithm

Rachmat¹, Abdul Ibrahim²

¹Department of Electrical Engineering, Polytechnic Informatic National (POLINAS), Indonesia

²Department of Computer Management, STMIK Dipanegara Makassar, Indonesia

Article Info

Article history:

Received Jan 19, 2021

Revised Mar 11, 2021

Accepted Dec 21, 2021

Keywords:

Settlement and Cleaning Service

Visual Basic

Machine Learning

Naïve Bayes

ABSTRACT

The Department of Settlements and Cleanliness of Pangkep Regency is tasked with maintaining the cleanliness and comfort of Pangkep Regency so that Pangkep City becomes a clean and beautiful city. Duties and responsibilities are on the Department of Settlements and Sanitation to jointly maintain the city of Pangkep so that it remains clean and beautiful. The waste generated by the community can generate retribution for Pangkep Regency which will later be managed into the Pangkep Regency APBD every year, so far the Settlement Service is and the cleanliness of Pangkep Regency in taking waste retribution is still used in traditional ways by using a semicomputer or manual system. So that the receipt of waste retribution is prone to being diverted by irresponsible people, therefore the writer tries to create a system which later can control waste retribution. admin. With a computerized system using a visual basic application program and using the Naïve Bayes algorithm with an accuracy of 77, 50% and an Area Under Curve (AUC) value of 0.464 succeeded in improving the accuracy of previous research, namely 68.09%. where later employees in charge of transporting waste from households, restaurants, hospitals and markets are in accordance with the costs set by the community and government, in this case the Department Settlement and Cleanliness of Pangkep Regency. The employee in charge will distribute tickets to every household, restaurant, hospital and market every day so that how many tickets are carried by the employee will be calculated the remaining tickets multiplied by the ticket price based on the color of the tickets that have been determined by the Pangkep District Office of Settlement and Sanitation.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Name of Corresponding Author,

Department of Computer Science,

Polytechnic Informatic National,

Street address Sultan Alauddin No.250 Makassar City of Indonesia.

Email: rachmat27udinus@gmail.com

1. INTRODUCTION

The Department of Settlements and Cleanliness of Pangkep Regency is tasked with maintaining the cleanliness and comfort of Pangkep Regency so that the city of Pangkep becomes a clean and beautiful city. The duties and responsibilities are with the Department of Settlements and Cleanliness to together with its citizens to keep Pangkep city clean and beautiful [1].

The waste generated by the community can generate retribution for Pangkep Regency which will later be managed into the Pangkep Regency APBD every year, so far the Pangkep Regency

Settlement and Sanitation Service in collecting waste retribution is still using traditional methods using a semicomputer or manual system [2].

So that the receipt of waste retribution is prone to being misled by irresponsible people, therefore the author tries to create a system which later on the receipt of waste retribution can be controlled by the admin.

Where later the employees who are in charge of transporting waste from households, restaurants, hospitals and markets are in accordance with the costs determined by the community and government, in this case the Department of Settlement and Cleanliness of Pangkep Regency [3].

With the research above it is expected to be able to help local governments in increasing the results of waste retribution at each location (there is a Pangkep district area, so that with local revenue it can increase development in these areas it is also hoped that the local government can incentivize the receipt of waste retribution in every area, both markets, shops, stalls and supermarkets.

So that this research can avoid things that are not desirable such as cases of corruption, this system is designed to avoid this. The method used in this study uses the Naive Bayes method, one of the methods used in Datamining to select unused attributes.

This study also uses a visual basic application to create a program that is used to input retribution receipt data which is then processed using a dataset to get the desired results.

2. RESEARCH METHOD

2.1. Types and Sources of Data

Sources of data used to design a decision support system for receiving levies at the Settlement and Sanitation Service are [4][5]:

1. Primary Data

Namely the source of data obtained directly from the source, in this case from observations of information processes in the field of society and interviews with related people. The primary needs are system constraints, information needs at each level of management and decision-making models.

2. Secondary Data

That is, the data source whose collection is not directly carried out by the researcher but the data is obtained from hospital documents and literature or written books related to the object of research.

1.1. TESTING

Tests were carried out using the Naïve Bayes classification method. In general, the greater the search for the training data closest to the test sample, the error rate will increase. This does not appear in the test results because it reduces the error rate when the search area becomes larger, the number of variations in the extracted texture feature values as a classification reference for each category is multiplied by using the larger amount of training data.

Retribution employee data entry form which includes data for officers who perform retribution duties including employee id, employee name, position, location, telephone and address to be inputted into the application.

ID	NAMA PEGAWAI	JABATAN	LOKASI
101	Rudy	Pegawai Kontrak	Panaikang
102	Usman	Pegawai Kontrak	Panaikikkang
103	Rachmat	Pegawai Kontrak	Maccini
104	Rappocini	Pegawai Tetap	Parangtambung
105	Syamsul Bhari	Pegawai Tetap	Nusantara
106	Zainuddin	Pegawai Kontrak	Irian
107	Nurhadi	Pegawai Kontrak	Abd.dg Sirua

Figure 1.1 Employee InputData Form

Entering the data of employees who carry out garbage collection at each levy place so that employees will know who is doing the retribution task, which includes employee ID, employee name, position, location, telephone, address.

ID	URAIAN	JUMLAH	HARGA_KARCIS	JUMLAH	TOTAL
01	Rumah Tangga	10	1000	2	2000
02	Ruko	5	5000	2	10000
03	Pasar	5	2000	1	2000

Figure 1.2 Waste Input Form

Entering waste retribution which is useful to find out how much waste will be inputted based on the ticket code which includes ID_Uraian, Description, Amount, Kode_Karcis, Warna_Karcis, Harga_Karcis, Amount_Bln and Total.

ID	Layanan	JENIS LAYANAN	TARGET REALISASI
01	Rumah Sakit	Rumah Sakit	5000000
02	Ruko	Ruko	20000000
03	Pasar	Pasar	30000000

Figure 1.3 Input Realization Receipt Form

In this form, inputting the realization of the receipt of waste retribution includes Id_Service, Service, Sub_Service, Target.

ID SETORAN	ID PEGAWAI	NAMA PEGAWAI	ID LAYANAN	JUMLAH SETORAN	KETERA
1	101	Rudy	01	Rumah Sakit	12-Febr
10	111	Sitti Aminah	02	Ruko	12-Febr
101	101	Rudy	01	Rumah Sakit	12-Febr
11	102	Usman	01	Rumah Sakit	12-Febr
12	103	Rachmat	02	Ruko	12-Febr

Figure 1.4 Input Deposit Form

In this form, the input of waste retribution deposits is carried out by each employee in each area which includes Id_setoran, Id_pegawai, Id_layanan, Name, Type_service, Tgl_Setor, Total_Stor and Description.

1.2. Using the RapidMiner Application

In these tables where the data from the input in the retribution application is entered in the rapidminer tools, the process is carried out to get the correct input data so that it remains in use when the next step will be carried out.

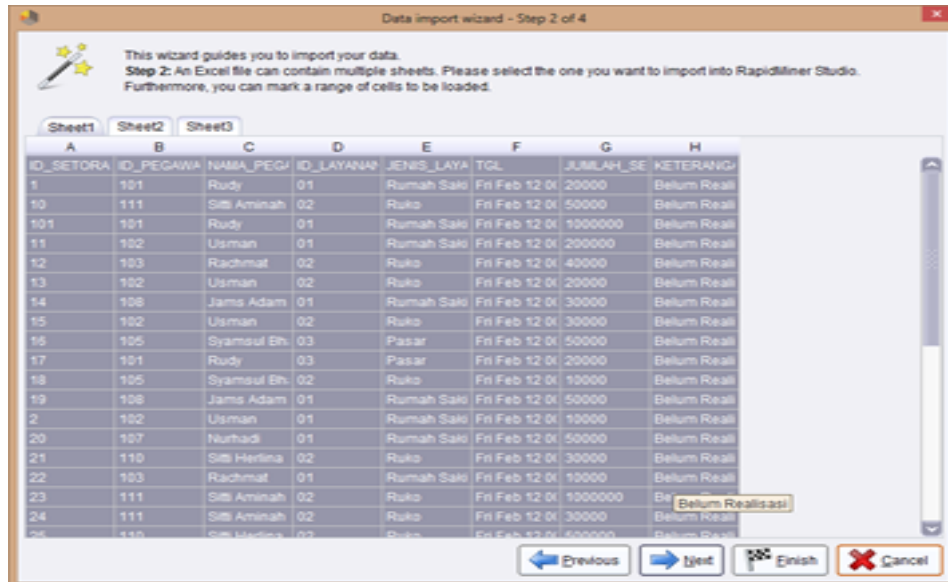


Figure 2.1 ImportDataExcel form

Import the data that has been inputted and applied to Microsoft Excel which will then be entered into the Rapid Miner tools to process the data

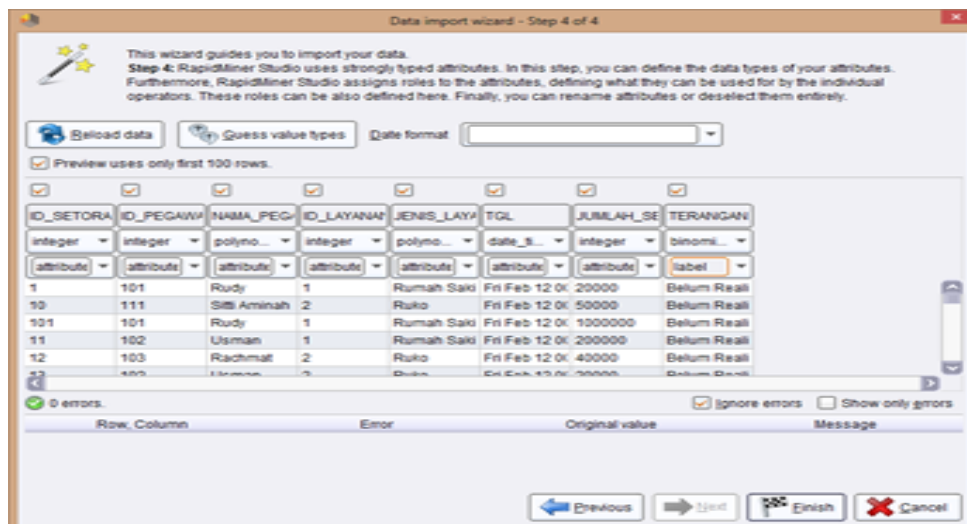


Figure 2.2 Form Attributes

In this section where missing or inappropriate attributes are selected to be discarded so that only the attributes that are used are left

Before testing is carried out, first look for suitable parameters as input attributes to improve accuracy. The parameters used in the test consist of K, Population size, Inertia weight, local best weight, Global best weight, Min weight, Max weight and Number of validation. The results of the Naive Bayes method will be tested using the K-Fold Cross Validation, as shown in Table 1.

Table 1. Testing Weights

	Population Size	Inertia Weight	Local Best Weight	Global Best Weight	Min Weight	Max Weight	Number Of Validation	Method Naïve Bayes
	5	1	1.	1	0	1	10	67,
	10	.0	0	.0	.0	.0	10	52%
	15	1	1.	1	0	1	10	69,
	20	.0	0	.0	.0	.0	10	59%
		1	1.	1	1	1		69,
		.0	0	.0	.0	.0		45%
		1	1.	1	1	1		77,
		.0	0	.0	.0	.0		50%

Where in table 1 the weight testing is tested four times to get maximum results, it can be seen from the following explanation with the weighting value starting from $K = 1$ with a population size 5 where with the Naive Bayes method the results are 67.52%, where $K = 3$ With a population of size 10, using the Naive Bayes method, the results are 69.59%, where $K = 5$ with a population of size 15, then with the Naive Bayes method the results are 69.45% and $K = 7$ with a population of size 20, then by using this method. naive bayes obtained the result of 77.50%.

So the research concludes with $K = 7$, the maximum value with the Naive Bayes method is 77.50%, this means that this study has an accuracy value of 77.50%.

Based on the results of experiments and research evaluations conducted with different datasets in previous studies, it can be concluded that the experiment with the highest level of accuracy in the Naive Bayes method was 77.50% and an Area Under Curve (AUC) value of 0.464 succeeded in improving the accuracy of previous research, namely 68.09%.

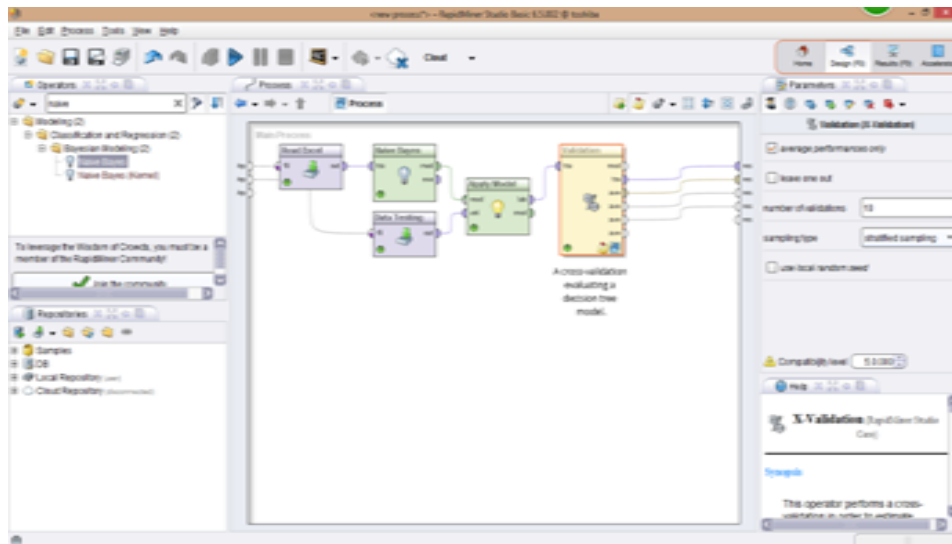


Figure 2.3 Form Using the NaiveBayes Algorithm

In this section, we use the Naive Bayes algorithm in the Rapidminer tool, which is included in this Rapidminer application

	True Return Realized	True Realized	Class precision
pred Return Realized	38	3	80.0%
pred Realized	5	0	0.0%
class recall	85.71%	0.0%	

Figure 2.4 Form AccuracyNaiveBayes

At the end of the rapidminer application, we will get the results of the accuracy stated in the trial data that we entered

3. RESULTS AND DISCUSSION

Naïve Bayes is a statistical classification that can be used to predict the probability of a class membership [6]. Bayes has very high accuracy and speed when applied to databases with large data. Naïve Bayes is an algorithm that can minimize rates compared to all other classifiers. However in practice this is not always the case, due to inaccuracies in the assumptions made for using incomplete classes and a lack of available probability data. Another classifier that does not explicitly use the theorem [7].

In this discussion, the algorithm used by Naive Bayes, so it is necessary to improve its accuracy by using another PSO method to get maximum accuracy or you can also use more than one method so that the results are very good.

In the dataset used is a dataset of input results through an application that is made by yourself using the visual basic programming language, so that the data inputted is not arbitrary and can be accounted for.

This research can also be used as input for local governments in increasing waste retribution for building their respective regions and can also be used to achieve regional revenue and anticipate cases that do not want to occur in the housing office.

The waste that is in the Pangkep Regency area always increases every year, both household waste, and garbage that is in the market every day very much and this can be used as input or regional income in increasing the regional budget.

Waste increases every day whether it is household waste or those found in markets and shops, therefore this can be used as an opportunity for regions to increase regional income or waste retribution.

In the use of the Naive Bayes method, how to classify the attributes used in this study so that the parameters used are in accordance with the desired needs. Both the method used and the algorithm used for classifying the parameters used for each attribute used.

Naïve Bayesian is also a classification method based on the application of the Bayes theorem and is included in the type of supervised learning (classification). For parameter estimation using maximum likelihood with only two parameters used, namely mean and variance [8]. The first step is to find the probability of each attribute for each class using uncertainty which is measured by probability. Bayes' Theorem has the following general formula:

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

Naïve Bayes is a simplification of the Bayes theorem. The following is the formula for naïve Bayes according to [9]

$$P(X|E) = P(E|X)P(X)$$

Information :

X: Data with an unknown class

C: The data hypothesis X is a specific class

P (C | E): Probability of hypothesis H based on condition X (posteriori probability)

P (C): Probability hypothesis H (prior probability)

P (E | C): Probability of X based on the conditions in the hypothesis H

P (E): Probability of X

Some of the advantages of the Naïve Bayes classification algorithm [10], are:

3.1. Uses a very intuitive technique. Bayes classifications, unlike neural networks, do not have some free parameters to set. This greatly simplifies the design process.

3.2. Since classifiers return probabilities, it is simpler to apply these results to various tasks if an arbitrary scale is used.

3.3. Does not require large amounts of data before learning begins The Naïve Bayes Classifier is computationally fast when making decisions.

4. CONCLUSION

With a computerized system using a visual basic application program and using the Naïve Bayes algorithm with an accuracy of 77, 50% where later employees in charge of transporting waste from households, restaurants, hospitals and markets are in accordance with the costs set by the community and government, in this case the Department Settlement and Cleanliness of Pangkep Regency.

As a conclusion, this research can be used as input for local governments, especially related agencies in increasing local revenue from waste retribution in the area. It is also hoped that other regions in Indonesia can do this. Namely by making applications to the relevant agencies to run according to existing or applicable rules

Based on the results of testing the information system for receiving waste retribution at the Department of Settlements and Hygiene, it can be concluded:

1. Decision Support System for receiving waste retribution is a program used to process waste retribution receipts at the Department of Settlements and Sanitation
2. The new decision support system is more optimal than the old system.
3. By using a Data Mining Decision Support System for receiving waste retribution can be controlled to avoid fraud.
4. It is hoped that further researchers to improve the accuracy of this research can use Particelle Swarm Optimization (PSO)

REFERENCES

- [1] H. Naidah, "Pengaruh Sistem Informasi Manajemen (Sim) Terhadap Kinerja Karyawan Pada Pt. Metro Batavia Air Distrik Makassar," *Sist. Inf.*, vol. 1, p. 11, 2009.
- [2] D. Pendukung and D. Rangka, "Katalog Perpustakaan Bidang Farmasi," no. 1, 2017.
- [3] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.

-
- [4] S. Zhang, C. Zhang, and Q. Yang, *Data preparation for data mining*, vol. 17, no. 5–6, 2003.
- [5] B. Green, “Canny edge detection tutorial,” *Retrieved: March*, vol. 6, p. 2005, 2002.
- [6] Z. Lin, M. Hewett, and R. B. Altman, “Using binning to maintain confidentiality of medical data.,” *Proc. AMIA Symp.*, pp. 454–458, 2002.
- [7] A. Mujumdar and V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms,” *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.
- [8] J. Xin, “Feature selection and visualization techniques for network anomaly detector,” 2003.
- [9] L. Sweeney, “Weaving Technology and Policy Together to Maintain Confidentiality,” *J. Law, Med. Ethics*, vol. 25, no. 2–3, pp. 98–110, 1997, doi: 10.1111/j.1748-720X.1997.tb01885.x.
- [10] S. Kazhuparambil and A. Kaushik, “Cooking is all about people: Comment classification on cookery channels using bert and classification models (malayalam-english mix-code),” *arXiv*, pp. 1–23, 2020, doi: 10.20944/preprints202006.0223.v1.