Mantik Journal, 8 (1) (2024), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Determining recipients of uninhabitable house rehabilitation program assistance using the classification method

Chilmiatus Silfiyah[1], Ririen Kusumawati[2], Cahyo Crysdian[3]

[1,2,3]Master of Informatics, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang, Indonesia

ARTICLE INFO

Article history:

Received Apr 25 2024
Revised May 15, 2024
Accepted May 30, 2024

Keywords:

Assistance program eligibility determination system
Classification;
Naïve Bayes;
Uninhabitable House.

ABSTRACT

The data used in this study amounted to 15182 datasets consisting of 14 variables. Existing variables are divided into basic variables and additional variables. The basic variables consist of 5 variables namely Home ownership, Roof type, Wall type, Floor type, Defecation facilities. While the additional variables consist of 9 variables, namely employment, having money / livestock / jewelry deposits and others, welfare deciles, education, recipients of non-cash food assistance, recipients of productive assistance for micro enterprises, recipients of cash social assistance, recipients of family hope programs, and recipients of basic necessities. Using the naïve bayes algorithm classification method, the values of accuracy, precision, recall, and f-measure are 67.61%, 67.97%, 93.71% and 78.79%. The addition of additional variables to the basic variables resulted in an accuracy of 68.29% in the additional variables of education. This shows that by adding additional variables, the accuracy results are higher than using only basic variables, so that this study can be used as a recommendation in decision making on the implementation of determining the beneficiaries of the rehabilitation program for uninhabitable houses.

Corresponding Author:

Chilmiatus Silfiyah,
Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang,
[1]Probolinggo 3st Vocational School, [2]State Islamic University,
UIN Campus, Jl. Gajayana No. 50 Malang, Postal Code 65144 Indonesia,
Tel. +62 341-558933, Fax. +62 341-558933
Email: 200605210010@student.uin-malang.ac.id

1. INTRODUCTION

Developing countries such as Indonesia every year have the problem of increasing the number of poor people, thus having an impact on the sustainability of life's welfare. For the poor or underprivileged, housing as a basic need becomes something that is very difficult to meet, even many live in illegal and uninhabitable locations. This is the responsibility of the government as stipulated in Law Number 1 of 2011 concerning housing and settlements in order to help accelerate development through the empowerment of each region, so that people can live in a livable place.

The Uninhabitable House Rehabilitation Program is a form of social service carried out by the government so that it has a direct impact on community welfare (Fakhrun Nisa & Setyadharma, 2020). This program is held in a decentralized manner by involving the role of the community (Maulana Rosyid, 2017). In line with the opinion

Megawati et al., (2020) the involvement of citizens refers to the participation of citizen involvement in public affairs. The policy regarding Uninhabitable Houses in Probolinggo Regency is regulated in Regent Regulation Number 7 of 2020 concerning Guidelines for the Implementation of Uninhabitable House Assistance, from 24 sub-districts in Probolinggo Regency, East Java, the number of Uninhabitable House units has been surveyed as many as 15182 units. In the implementation of handling rehabilitation activities, it is considered not optimal because the recording of the results of handling uninhabitable houses is still manual or still in the form of a matrix so that the location and prospective recipients of assistance are not clearly mapped. Seeing these conditions, strategic steps are needed to optimize the implementation of handling uninhabitable houses so that they can know the progress of implementation, and the right decision making process to determine the beneficiaries of the uninhabitable house handling program.

The classification method is used to produce the determination of the eligibility of beneficiaries of the Uninhabitable House handling program, because so far the determination and selection of prospective beneficiaries has been done manually by the implementing committee or the Probolinggo Regency Housing and Land Settlement Office (DPKPP_Dinas Perumahan Kawasan Permukiman dan Pertanahan) from data obtained based on the results of social service surveys in the form of Social Poverty Data, and there is no appropriate method to determine the eligibility of candidates recipients of Uninhabitable Houses for repairs or refurbishment in Probolinggo Regency.

In the study, the random forest data mining classification method was applied to classify creditworthiness status. By applying the random forest algorithm, WITH attributes, namely: marital status, dependent data, age, status of residence, home ownership, occupation, employment status, company status, income, down payment, last education, length of stay, and home conditions used in the research conducted as a feature input, the results of an accuracy rate of 78.60% (Pahlevi et al., 2023)

The classification method was chosen based on research aimed at evaluating the performance of the combination of Naïve Bayes and Chi-Square algorithms in classifying occupancy status as part of home rehabilitation and reconstruction efforts. the accuracy achieved by the combination of Naïve Bayes and Chi-Square algorithms is 89.59% (Wijaya, 2020).

In a study that aims to investigate how classification techniques are used in assisting decision makers of The Federal Emergency Management Agency during disasters using machine learning decision-making models based on logistic regression, decision trees, and K-nearest neighbors to determine a person's eligibility to receive temporary housing assistance from FEMA because about 90% of natural disasters in the United States caused by floods obtain the highest accuracy in the KNN method, it is 70%, while the decision tree is 68.83%, and logistics regression is 65.97% (Afkhamiaghda & Elwakil, 2021).

The study, which aims to predict the distribution of social assistance funds on target to residents who deserve social assistance in Jatipancur village using the Naïve Bayes algorithm, obtained an accuracy of 91. 10% with presicison class and reccal class of 91. 85% (Surahman & Hayati, 2023).

Research by Song et al., (2023), aims to build an efficient Bayesian classifier in evaluating home security and habitability risks, so that it can be used as an assessment to reduce the level of home security risk by being checked regularly and at low cost, using chi square, The samples collected in this study came from the southeast coast of China with a sample number of 864, The accuracy of the classification of the level of home security was 94% and the habitability rate was 92%.

The purpose of this study is 1) to produce a classification model used in determining prospective recipients of uninhabitable housing assistance and can be reviewed based on the parameters of accuracy, precision, recall, f-measure. 2) Analyze additional variables that influence the performance of classification methods in determining prospective recipients of uninhabitable housing assistance.

Based on the above description, the research on determining the recipients of assistance for the uninhabitable house rehabilitation program using the classification method is expected to be a reference for the Probolinggo Regency Government and the Residential and Land Housing Office to recommend the eligibility of prospective recipients of uninhabitable house assistance in the repair or restoration of low-income non-governmental houses in Probolinggo Regency.

2. RESEARCH METHOD

The research design to be carried out can be shown in figure 1:

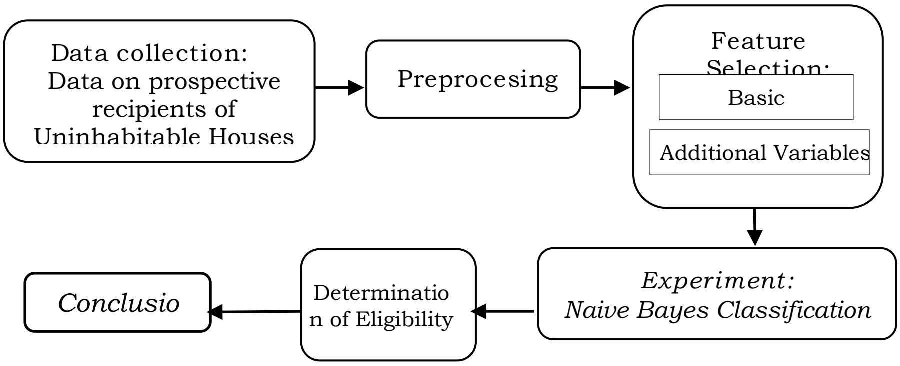

figure 1 Flowchart Design Reseach

a. Data Collection

In this study, the data used were secondary data on Prospective Uninhabitable House Assistance Recipients from Probolinggo Regency, obtained from the Housing Office, Probolinggo Regency Settlement Area managed by the implementing official of the activity. The data amounted to 15182 in the form of P3KE (Targeting the Acceleration of Extreme Poverty Eradication) data in 2022.

b. Preprocessing

The data obtained does not all meet the variables of data adequacy so preprocessing is necessary. Preprocessing is the processing of raw data into ready-to-use data that can be more easily understood by the system (Yustina et al., 2024). The data obtained are combined in a single data unit (dataset). The process of merging data is adapted to the same format.

c. Feature Selection

Feature selection in this study is used to determine basic variables and additional variables. In this study, the basic variables that will be used in the calculation are guided by Probolinggo Regent Regulation Number 7 of 2020 concerning Guidelines for the Implementation of Assistance for Uninhabitable Houses, at research conducted by Mubarak & Hidayat, (2019), and Siahaan & Situmorang, (2023) namely: Home ownership, Roof type, Wall type, Floor type, Defecation facilities. The basic variables are described in table 1.

Table 1. Basic Variables.

Symbols	Variable Name	Criteria
VD1	Home ownership	Own Hitchhike Contract/ lease Rent-free Service Other
VD2	Types of roofs	Asbestos / Zinc Bamboo Concrete Tile Straw/Weeds/Leaves Wood/ Shingle Other
VD3	Wall type dinding	Bamboo Wood/board Zinc Wall Other
VD4	Floor type	Ceramics/Granite/Marble/Tiles/Tiles/Terraces Cement Soil Other
VD5	Defecation Facility /MCK /MCK	Yes, with septic tank Yes, without septic tank No, public/shared latrines Other

Meanwhile, additional variables were obtained from inventory data on uninhabitable houses from the Housing, Settlement Area, and Land Office in the form of data on the Self-Help Housing Stimulant Assistance program) on handling extreme poverty and stunting based on INPRES Number 4 of 2022 concerning the Acceleration of Extreme Poverty Eradication and the research by Furqan et al., (2023). Additional variables totaling 9 are shown in table 2.

Tabel 2. Additional Variables

Symbols	Variable name
VT1	Work
VT2	Have money / livestock / jewelry and other deposits
VT3	Desil Welfare
VT4	Education
VT5	Recipients of non-cash food assistance
VT6	Recipients of productive assistance for micro-enterprises
VT7	Recipients of cash social assistance
VT8	Family Hope Program Recipients
VT9	Food Recipients

d. Experiment

Researchers use the Naive Bayes method as in figure 1 of the Research System Design. Data testing using naïve bayes is a machine learning algorithm using the concept of probability (Heliyanti Susana, 2022), probability P(c|x) is the probability of class c in x or other languages of P(c|x) is a percentage with the number of c's in x, P(c|x) is the probability of X in C, P(c) is the probability of C and P(X) is the prior probability of X, Based on the Naïve Bayes algorithm. The classification process can be clearly seen in figure 2:

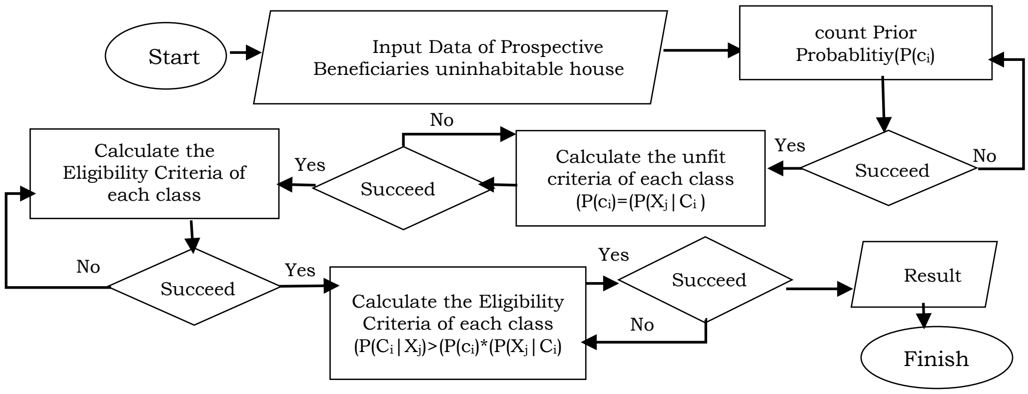

Figure 2 Flowchart Naïve Bayes

The Naive Bayes process is a machine learning approach proposed by Thomas Bayes that utilizes probability and statistical calculations to predict future probabilities based on previous experience (Sinaga & Simanjuntak, 2020). The calculation of the naïve bayes algorithm is the equation for calculating the value of the probability p(X|C). shown in equation 1:

$$P(X_j|C_i) = \frac{\sum\{X_j|C_i\}}{\sum c_i} \tag{1}$$

Keterangan:

$P(c_i)$	: Probability of hypothesis c (*prior probability*)
C_i	: Probability with feature i
$\sum\{X_j|C_i\}$	: Number of hypotheses c based on condition x (posteriori probability)
$P(X_j | C_i)$	: Probability x based on conditions on hypothesis c

This equation then became the basis for the naïve bayes method, because each attribute is assumed to be conditionally independent, the equation can be expressed at the posterior probability in equation 2:

$$P(C_i|X_j) = P(C_i)\prod P(X_j|C_i) \tag{2}$$

Tests are performed to measure the accuracy of the results of each proposed model. Accuracy is defined as the degree of closeness between the predicted value and the actual value. By using the confusion matrix method, the performance value of the classification method can be known (Marga, 2022). Performance Evaluation uses a Confusion Matrix that declares the correct classification of the amount of test data and the incorrect amount of test data (Normawati & Prayogi, 2021). The Confusion Matrix can be seen in table 3.

Table 3 *Confusion Matrix*

Prediction class	Actual class	
	Positif	Negatif
Positif	FP	TP
Negatif	TN	FN

Description :

TP (True positive) = If the actual value of the results determined by the DPKPP implementing official, the rehabilitation program of uninhabitable houses in Probolinggo Regency is declared eligible to receive assistance for the rehabilitation program of uninhabitable houses and the predicted value of the classification method is eligible to receive assistance for the rehabilitation program of uninhabitable houses

FP (False positive) = If the actual value of the results determined by the DPKPP implementing official, the rehabilitation program of uninhabitable houses in Probolinggo Regency is declared unfeasible and the predicted value of the prediction from the classification method is feasible

FN (False negative) = If the actual value of the results determined by the implementing officer of DPKPP rehabilitation program of uninhabitable houses in Probolinggo Regency is declared feasible and the predicted value of the classification method is not feasible

TN (True negative) = If the actual value of the results determined by the implementing official of DPKPP rehabilitation program of uninhabitable houses in Probolinggo Regency is declared unfeasible and the predicted value of the classification method is not feasible

After obtaining the TP, FP, FN and TN values, the accuracy, precision, recall and f1-score values can be calculated with the equations 3, 4, 5, dan 6:

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}\ x\ 100\% \tag{3}$$

$$Precision = \frac{\text{TP}}{\text{TP+FP}}\ x\ 100\% \tag{4}$$

$$recall = \frac{\text{TP}}{\text{TP+FN}}\ \text{x}\ 100\% \tag{5}$$

$$f - measure = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \tag{6}$$

3. RESULTS AND DISCUSSIONS

The next stage is to test the application of the naïve bayes algorithm. From the application of these strategies, the performance value of the naïve bayes algorithm (accuracy, precision, recall and f-measure) will be generated from each experimental strategy. The testing strategy is carried out with two scenarios, the first scenario with data processing using 5 basic variables and the second scenario using 5 basic variables combined with 1 variable each of 9 additional variables so that the second scenario has 9 combinations of tests.

3.1 The First Scenario (Using Basic Variables)

At this stage, the performance testing of the naïve bayes algorithm in determining the eligibility of prospective recipients of uninhabitable house assistance against basic variables (home ownership; roof type; wall type; floor type; defecation facilities based on table 4.

Table 4 Basic variable set data of prospective recipients of uninhabitable housing assistance

NIK	Home ownership	Type Roof	Wall Type	FloorType	Has defecation facilities	Actual
3513116110400002	One's own	Rooftile	Zinc	Land	No, Public/Shared Toilet	Worthy
3513081603850002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513120107830054	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513110610730002	One's own	Rooftile	Zinc	Cement	Yes, with Septic Tank	Notfeasible
3513085507590011	One's own	Rooftile	Wall	Cement	Yes, without a Septic Tank	Notfeasible
3513080702850001	One's own	Rooftile	Wall	Cement	Yes, without a Septic Tank	Worthy
3513121006670002	One's own	Rooftile	Wall	Land	Other	Notfeasible
3513104107630150	One's own	Rooftile	Wood/Planks	Land	No, Public/Shared Toilet	Worthy
3513120104830001	One's own	Rooftile	Wall	Land	Other	Notfeasible
3513130107650279	One's own	Rooftile	Wall	Land	Other	Notfeasible
3513121506840003	One's own	Rooftile	Wall	Cement	Other	Notfeasible

NIK	Home ownership	Type Roof	Wall Type	FloorType	Has defecation facilities	Actual
3513120108660001	One's own	Rooftile	Wall	Land	Other	Notfeasible
3513104107560202	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Worthy
3513081405700002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Worthy
3513116507850001	One's own	Rooftile	Wood/Planks	Cement	Other	Worthy
3513112209900001	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513100107510071	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513120401750002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513124710680001	One's own	Rooftile	Wall	Land	Other	Notfeasible
3513080206670001	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513151702790002	One's own	Rooftile	Wall	Cement	No, Public/Shared Toilet	Notfeasible
3513111705850002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513110107510075	One's own	Rooftile	Zinc	Land	No, Public/Shared Toilet	Worthy
3513120404690007	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Worthy
3513100107730042	One's own	Rooftile	Wood/Planks	Land	Other	Worthy
3513161011820003	One's own	Rooftile	Wall	Cement	Yes, without a Septic Tank	Worthy
3513121302790001	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Worthy
3513102005870004	Hitchhiking	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513121407880002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513104107700239	One's own	Rooftile	Wood/Planks	Land	Other	Notfeasible
3513171703900001	Hitchhiking	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
3513104701850001	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Worthy
3513152205790002	One's own	Rooftile	Wall	Cement	Yes, with Septic Tank	Notfeasible
.....	....	....	.....	...	...	....
.....	....	....	.....	...	...	....
.....	....	....	.....	...	...	....
3513152304820002	One's own	Rooftile	Zinc	Land	Other	Worthy
3513142912910001	Hitchhiking	Rooftile	Zinc	Land	Other	Worthy

Furthermore, testing the Naive Bayes algorithm in table 4, with a dataset of prospective recipients of uninhabitable houses with a total dataset of 15182 rows was tested. The testing of the classification method using naïve bayes was carried out with several tests, the first test divided 90% of training data and 10% of testing data and used split data techniques with confusion matrix, the second test used 80% training data and 20% testing data, the third test used 70% training data and 30% testing data, and the fourth test used 60% training data and 40% testing data.

Testing was done with the rapidminer tool. For the trial process can be seen in figure 3

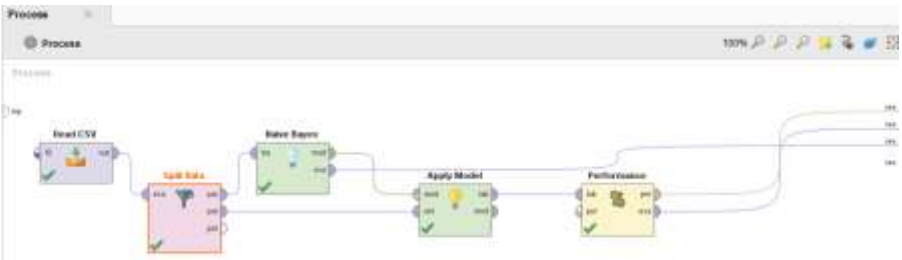

Figure 3. Testing process

For the first experiment using 90:10 division of training data and testing data resulted in 66.07% accuracy, 66.69%, precision 92.38% recall, and 77.46% f-measure. The results of the classification of the first experiment can be seen in figure 4.

accuracy: 66.07%			
	true Worthy	true Notfeasible	class precision
pred. Worthy	118	73	61.78%
pred. Notfeasible	442	885	66.69%
class recall	21.07%	92.38%	

Figure 4. Use 90:10 split data

Furthermore, the second experiment using 80:20 division of training data and testing data resulted in 67.19% accuracy, 67.76% precision, 92.95% recall, and 78.38% f-measure. The classification results can be seen in figure 5.

accuracy: 67.19%			
	true Worthy	true Notfeasible	class precision
pred. Worthy	235	137	63.17%
pred. Notfeasible	859	1805	67.76%
class recall	21.48%	92.95%	

Figure 5 Use 80:20 split data

Furthermore, the third experiment using 70:30 division of training data and testing data resulted in 67.61% accuracy, 67.97% precision, 93.71% recall, and 78.79% f-measure. The classification results can be seen in Figure 6.

accuracy: 67.61%			
	true Worthy	true Notfeasible	class precision
pred. Worthy	339	184	64.82%
pred. Notfeasible	1291	2740	67.97%
class recall	20.80%	93.71%	

Figure 6. Use 70:30 split data

The fourth experiment using 60:40 division of training data and testing data resulted in 67.56% accuracy, 67.96% precision, 93.59% recall, and 78.74% f-measure. The classification results can be seen in Figure 7.

accuracy: 67.56%			
	true Worthy	true Notfeasible	class precision
pred. Worthy	454	250	64.49%
pred. Notfeasible	1720	3648	67.96%
class recall	20.88%	93.59%	

Figure 7. Use 60:40 split data

After testing with 4 experiments, Naïve Bayes performance based on accuracy value, precisión value, recall value, and f-measure value can be seen in table 5.

Table 5. Naïve bayes performance with basic variables

Trial	Split data	BASIC VARIABLE NAÏVE BAYES PERFORMANCE			
		Accuracy	*Precision*	*Recall*	*f-measure*
1st trial	90:10	66,67%	66,69%	92,38%	77,46%
2nd trial	80:20	67,19%	67,76%	92,95%	78,33%
3rd trial	70:30	67,61%	67,97%	93,71%	78,79%
4th trial	60:40	67,56%	67,96%	93,59%	78,74%

In table 5, it is known that the highest accuracy, precision, recall, and f-measure values were obtained in the third experiment at 70:30 division, namely 67.61% and 67.97%, while the highest recall and f-measure values were obtained 93.71% and 78.79%. So that in the next study the distribution of data in the trial combination of basic variables and additional variables uses the 70:30 data set division because it has the highest level of performance measured based on accuracy, precision, recall, and f-measure. This is based on the opinion of Halim & Anraeni, (2021) that accuracy is defined as the level of closeness between the predicted value and the actual value, precision can be interpreted as a match between a request for information and an answer to the request. Precision is used to measure the extent to which classification models provide positively correct predictions (Febrisa Sidabutar et al., 2023). Recall is defined as the ratio of the selected relevant items to the total number of relevant items available, recall is used to measure the extent to which the classification model can identify all actual positive cases (Argina, 2020). While F- measure is the harmonic mean between precision and recall values (Muhammad Daffa Al Fahreza et al., 2024), so it is defined that f-measure is a combined parameter between precision and recall, this parameter measures how well the model can predict positive classes by considering the balance between precision and recall. The 70:30 division is also based on the opinion of Gholamy et al., (2018) the division of training data and test data generally uses 70-80% of training data and 30-20% of available data.

3.2 Second Scenario (Combining Basic and Additional Variables)

This test was carried out using basic variables by adding 1 additional variable from 9 additional variables so that 9 combinations of tests were obtained, the additional variables were work (VT1), having money/livestock/jewelry deposits and others (VT2), welfare deciles (VT3), Education (VT4), recipients of non-cash food assistance (VT5), recipients of micro business productive assistance (VT6), recipients of cash social assistance (VT7), recipients of the Family Hope Program (VT8), and recipients of basic necessities (VT9), The test was carried out with 9 scenarios. The addition of additional variables is described in table 6..

Table 6. Added additional variable columns

NIK	Work	Have savings ofmoney/ jewelry /livestock /Other	Prosperity Decile	Education	BPNT recipient	BPUM recipients	BST recipients	PKH recipients	SEMBAKO recipie
3513 1161 1040 0000	Farmer	No	1	Not finished elementary school/ equivalent	Yes	No	Yes	Yes	Yes
3513 0816 0385 0000	Self-employed	Yes	2	Graduated from high school/ equivalent	Yes	Yes	Yes	No	Yes
3513 1201 0783	Self-employed	Yes	2	High school students / equivalent	No	No	No	No	No

NIK	Work	Have savings ofmoney/ jewelry /livestock /Other	Prosperity Decile	Education	BPNT recipient	BPUM recipients	BST recipients	PKH recipients	SEMBAKO recipie
0050									
3513 1106 1073 0000	Private employees	Yes	3	Graduated from high school/ equivalent	Yes	No	Yes	Yes	Yes
3513 0855 0759 0010	Farmer	Yes	3	Not finished elementary school/ equivalent	Yes	No	Yes	Yes	Yes
3513 0807 0285 0000	Farmer	No	1	Not finished elementary school/ equivalent	Yes	No	Yes	Yes	Yes
3513 1210 0667 0000	Farmer	No	1	Completed elementary school/ equivalent	Yes	Yes	Yes	No	Yes
3513 1041 0763 0150	Farmer	Yes	3	No/ not yet at school	Yes	No	Yes	No	Yes
3513 1201 0483 0000	Farmer	No	1	Completed elementary school/ equivalent	Yes	No	No	No	Yes
3513 1301 0765 0270	Freelancer	Yes	2	Not finished elementary school/ equivalent	No	Yes	No	No	No
3513 1215 0684 0000	Trader	No	1	Completed elementary school/ equivalent	No	Yes	No	No	No
3513 1201 0866 0000	Farmer	Yes	1	Completed junior high school/ equivalent	No	Yes	No	No	No
3513 1041 0756 0200	Farmer	Yes	1	Completed elementary school/ equivalent	Yes	No	Yes	No	Yes
3513 0814 0570 0000	Farmer	Yes	2	Completed elementary school/ equivalent	No	No	No	No	No
3513 1165 0785 0000	Farmer	No	1	Completed junior high school/ equivalent	Yes	No	Yes	Yes	Yes
3513 1122 0990 0000	Farmer	Yes	2	Graduated from high school/ equivalent	Yes	Yes	Yes	Yes	Yes
3513 1001 0751 0070	Farmer	Yes	3	Not finished elementary school/ equivalent	No	No	No	No	No
3513 1204 0175 0000	Worker Free	Yes	1	Not finished elementary school/ equivalent	Yes	No	Yes	Yes	Yes

NIK	Work	Have savings ofmoney/ jewelry /livestock /Other	Prosperity Decile	Education	BPNT recipient	BPUM recipients	BST recipients	PKH recipients	SEMBAKO recipie
3513124710680000	Worker Free	No	2	Not finished elementary school/ equivalent	Yes	No	Yes	No	Yes
3513080206670000	Self-employed	Yes	2	Graduated from high school/ equivalent	No	No	No	No	No
3513151702790000	Farmer	Yes	1	Graduated from high school/ equivalent	No	No	No	No	No
3513111705850000	Farmer	No	1	Completed junior high school/ equivalent	No	No	No	No	No
3513110107510070	Farmer	Yes	1	Completed elementary school/ equivalent	Yes	Yes	Yes	Yes	Yes
3513120404690000	Farmer	Yes	3	Completed junior high school/ equivalent	Yes	No	Yes	Yes	Yes
3513100107730040	Farmer	Yes	2	Completed elementary school/ equivalent	No	No	No	No	No
3513161011820000	Worker Free	Yes	1	Completed junior high school/ equivalent	Yes	Yes	Yes	Yes	Yes
3513121302790000	Farmer	Yes	2	Graduated from high school/ equivalent	No	Yes	Yes	No	No
3513102005870000	Farmer	Yes	3	Graduated from high school/ equivalent	No	No	No	No	No
3513121407880000	Self-employed	Yes	3	Graduated from high school/ equivalent	Yes	No	Yes	Yes	Yes
3513104107700230	Trader	No	3	Completed elementary school/ equivalent	Yes	No	Yes	No	Yes
3513171703900000	Worker Free	Yes	1	Graduated from high school/ equivalent	Yes	No	Yes	Yes	Yes
3513104701850000	Farmer	Yes	1	No/ not yet at school	No	Yes	No	Yes	No
3513152205790000	Farmer	Yes	2	Completed elementary school/ equivalent	Yes	Yes	Yes	Yes	Yes
..	....	..	....	....	...	...	...	...	...

NIK	Work	Have savings ofmoney/ jewelry /livestock /Other	Prospe rity Decile	Education	BPNT recipie nt	BPUM recipi ents	BST reci pien ts	PKH reci pien ts	SEMBA KO recipie
..	....	..	....	....	...	...	...	...	...
..	....	..	....	....	...	...	...	...	...
3513 1523 0482 0000	Freelanc er	No	1	Graduated from high school/ equivalent	Yes	No	Yes	Yes	Yes
3513 1429 1291 0000	Farmer	Yes	2	Completed elementary school/ equivalent	No	No	No	No	No

Based on table 6, the test in the second scenario was carried out with 9 combinations. The results of the tests are described in table 7

Table 7. Test Results of Scenario 2

Combination of Variables	Accuracy	Presisi	Recall	F-measure
VD+VT1	68,07%	68,90%	91,42%	78,62%
VD+VT2	67,41%	67,82%	93,71%	78,69%
VD+VT3	67.63%	68.15%	93.09%	78.69%
VD+VT4	68.29%	68.90%	92.27%	78.89%
VD+VT5	68.25%	68.74%	92.72%	78.95%
VD+VT6	67.55%	67.92%	93.71%	78.76%
VD+VT7	67.46%	67.89%	93.57%	78.69%
VD+VT8	67.57%	67.94%	93.71%	78.77%
VD+VT9	68.16%	68.63%	92.85%	78.92%

Table 7. shows the results of the trial in the second scenario on testing the accuracy of the basic variable combined with 1 variable from 9 additional variables. The results of the trial showed that the combination of all basic variables added VT4 (education) at number 4 obtained the highest accuracy of 68.29%.

4. CONCLUSION

In determining prospective recipients of uninhabitable housing assistance using the classification method with the naïve bayes algorithm, recommendations can be used as recommendations in determining eligibility for the implementing officials of the uninhabitable house assistance program at the Housing, Settlement and Land Office of Probolinggo district. Based on tests conducted using 2 scenarios, accuracy results increased by 0.78%. The accuracy results in scenario 1 using basic variables with a comparison of 70:30 data sets, accuracy of 67.61%, accuracy values, while precision, recall, and f-measure are 67.97%, 93.71% and 78.79%. And the accuracy results in scenario 2 by adding an additional variable to the base variable of 68.29% by adding an additional variable of education level. This shows that the addition of variables has an effect. The results of this study can be used as a consideration in determining the eligibility of prospective recipients of uninhabitable house rehabilitation assistance by adding a variable of education level. The results of this study are expected to help the implementing officials of the rehabilitation of uninhabitable houses of the Probolinggo Regency Housing, Settlement Area, and Land Office (DPKPP) in determining the eligibility of prospective recipients of assistance given to the people of Probolinggo Regency, so as to minimize errors on the part of the Housing, Settlement Area, and Land Office.

REFERENCES

Afkhamiaghda, M., & Elwakil, E. (2021). Machine learning-based FEMA Transitional Shelter Assistance (TSA) eligibility prediction models. *Journal of Emergency Management*, *19*(6), 561–

573. https://doi.org/10.5055/jem.0595

Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, *1*(2), 29–33. https://doi.org/10.33096/ijodas.v1i2.11

Data P3KE (Penyasaran Percepatan Penghapusan Kemiskinan Ekstrem) tahun 2022 Kabupaten Probolinggo (Inventaris RTLH DPPKP (Dinas Perumahan Kawasan Permukiman dan Pertanahan) Kabupaten Probolinggo.

Fakhrun Nisa, A., & Setyadharma, A. (2020). Benefit Incidence Analysis of Uninhatitable Houses Rehabilitation Program in Indonesia. *KnE Social Sciences*. https://doi.org/10.18502/kss.v4i6.6690

Febrisa Sidabutar, A., Habibi, R., & Isti Rahayu, W. (2023). Perbandingan Metode Klasifikasi Untuk Pengelompokan Risiko Magang Mahasiswa. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(3), 2071–2076. https://doi.org/10.36040/jati.v7i3.7026

Furqan, M., Hasibuan, M. S., & Sapitri, B. (2023). *Jurnal Mantik Application of the support vector machine algorithm in the classification of livable houses*. *7*(3).

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 Or 80/20 Relation Between Training And Testing Sets : A Pedagogical Explanation. *Departmental Technical Reports (CS)*, *1209*, 1–6.

Halim, A. A. D., & Anraeni, S. (2021). Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN). *Indonesian Journal of Data and Science*, *2*(1), 01–12. https://doi.org/10.33096/ijodas.v2i1.23

Heliyanti Susana. (2022). Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *4*(1), 1–8. https://doi.org/10.52005/jursistekni.v4i1.96

INPRES Nomor 4 Tahun 2022 tentang Percepatan Penghapusan Kemiskinan Ekstrem.

Marga, N. S. (2022). Sentimen Analisis Tentang Kebijakan Pemerintah Terhadap Kasus Corona Menggunakan Metode Naive Bayes. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *2*(4), 453–463. https://doi.org/10.33365/jatika.v2i4.1602

Maulana Rosyid, E. (2017). Implementasi Peraturan Menteri Perumahan Rakyat Republik Indonesia Nomor 06 Tahun 2013 Tentang Pedoman Pelaksanaan Bantuan Stimulan Perumahan Swadaya Di Kelurahan Mondokan Kecamatan Tuban Kabupaten Tuban. *Publika*, *5*(5).

Megawati, S., Ma'ruf, M. F., Fanida, E. H., Niswah, F., & Oktariyanda, T. A. (2020). Strengthening Family Resilience through Financial Management Education in Facing the Covid-19 Pandemic. *Journal La Bisecoman*, *1*(5), 8–15. https://doi.org/10.37899/journallabisecoman.v1i5.246

Mubarak, M. F., & Hidayat, N. (2019). Rekomendasi Perbaikan Rumah Tidak Layak Huni Menggunakan Metode TOPSIS Studi Kasus Badan Keswadayaan Masyarakat Di Kelurahan Bekasi Jaya. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(4), 3390–3395.

Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, & Michael Indrawan. (2024). Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z. *Journal of Applied Computer Science and Technology*, *5*(1), 16–25. https://doi.org/10.52158/jacost.v5i1.715

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, *5*(2), 697–711.

Pahlevi, O.-, Amrin, A.-, & Handrianto, Y.-. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. *Jurnal Infortech*, *5*(1), 71–76. https://doi.org/10.31294/infortech.v5i1.15829

Peraturan Bupati Nomor 7 Tahun 2020 tentang Pedoman Pelaksanaan Pemberian Bantuan Rumah Tidak Layak Huni.

Siahaan, M. L., & Situmorang, Z. (2023). Algoritma Naïve Bayes Dalam Penentuan Bantuan Renovasi Rumah Di Desa Sialang Buah. *KAKIFIKOM (Kumpulan Artikel Karya …*, *05*(02), 71–81. https://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3102%0Ahttps://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3102/2508

Sinaga, A. S. R., & Simanjuntak, D. (2020). Sistem Pakar Deteksi Gizi Buruk Balita Dengan Metode Naïve Bayes Classifier. *Jurnal Inkofar*, *1*(2). https://doi.org/10.46846/jurnalinkofar.v1i2.110

Song, M., Zhu, Z., Wang, P., Wang, K., Li, Z., Feng, C., & Shan, M. (2023). An Alternative Rural Housing Management Tool Empowered by a Bayesian Neural Classifier. *Sustainability*

(Switzerland), *15*(3). https://doi.org/10.3390/su15031785

Surahman, A., & Hayati, U. (2023). Implementasi Algoritma Naïve Bayes Untuk Prediksi Penerima Bantuan Sosial. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 347–352. https://doi.org/10.36040/jati.v7i1.6302

Wijaya, N. (2020). Evaluation of Naïve Bayes and Chi-Square performance for Classification of Occupancy House. *International Journal of Informatics and Computation*, *1*(2), 46. https://doi.org/10.35842/ijicom.v1i2.20

Yustina, E., Hariyadi, M. A., & Crysdian, C. (2024). Recommendation of Prospective Construction Service Providers in Government Procurement Using Decision Tree. *International Journal of Advances in Data and Information Systems*, *5*(1), 39–48. https://doi.org/10.59395/ijadis.v5i1.1316

Undang -Undang Nomor 1 Tahun 2011 tentang Perumahan dan Permukiman bahwa dalam penyelenggaraan perumahan dan kawasan permukiman, setiap orang berhak menempati, menikmati, dan atau memiliki atau memperoleh rumah yang layak dalam lingkungan yang sehat, aman, serasi, dan teratur.