Business Prospects Prediction for Waqf Lands Using Naïve Bayes And Apriori Algorithm

Amiq Fahmi*[1], Edi Sugiarto[2], Agus Winarno[3]
[1,2,3]Dian Nuswantoro University, Semarang
{[1]amiq.fahmi, [2]edi.sugiarto, [3]agus.winarno}@dsn.dinus.ac.id

Abstract. *Waqf* is a donation activity of an own property for charity and the general welfare under sharia. The effective waqf empowerment in perspective economic changes the use of waqf from consumptive to productive. Lands are one form of waqf, and they are strategic assets for productive waqf empowerment. This research aims to build a classifier to predict waqf lands as 'productive' or 'not productive' assets for business prospects. The classification used Naïve Bayes with attributes summarised from administrative data of waqf lands. A modified Apriori algorithm was proposed as a new method to improve classification accuracy. A threshold value defined based on a mean value from the classification process by the Naïve Bayes was used to select classification results with a deviation of the posterior value. The value below was to be reclassified using the Apriori algorithm. The proposed method can improve prediction accuracy better than using only one Naïve Bayes classifier.

Keyword: *Apriori algorithm, business prospects, classification, Naïve Bayes, waqf lands.*

1   Introduction

In many studies and literature, data mining (DM) and machine learning (ML) techniques have penetrated all areas, both in social science and computer science. The DM/ML technique has been widely used to assist in overcoming the challenges of the global socio-economic development of society [1]. Especially in developing countries, strategic use of DM/ML can assist increase knowledge of decision-making, systems and policies for socio-economic development and increase effectiveness to avoid financial crises [2]–[4].

One of the diligent currently being carried out by the Ministry of Religion is to empower strategic waqf lands to build the socio-economic life of the people [5]. *Waqf* can be defined as donating property ownership, such as land, which has economic potential and benefits for public welfare based on legislation/statutory regulations and sharia (Islamic religious rules) [6]. Waqf in the form of land is one of the most prominent types in Indonesia, with 421,289 locations with varying land areas up to 55,634.15 hectares [7] which is widespread throughout Indonesia. Among them, there are strategic lands with potential for economic development.

The Ministry of Religion has an operational system in which the registration of waqf property, especially land, is permanently recorded and documented in a waqf commitment (*ikrar wakaf* -in Bahasa), as database processed to produce waqf land reports which contain the total number and area. However, it is still difficult to predict

or determine for planning, developing, and empowering strategic waqf land with economic potential. The development of waqf land for commercial or productive and socio-economic activities is influenced by several factors, such as the location, shape of the waqf land, and the ability to manage nadzir (people who receive waqf property) to manage and develop it along with its designation [8]. Many variables determine strategic waqf land to empowered for business development, such as health services, shopping centres, educational facilities, gas stations, and others. Site selection is a critical strategic decision in starting, expanding, or changing industrial locations. Several criteria consider technical, economic, social, environmental and political issues [9]. It is also related to the performance measures, population structure, economic factor, competition, saturation level, magnetism, and business characteristic [10].

One of the popular algorithms in DM/ML that can be used to make predictions based on probability values is Naïve Bayes. Naïve Bayes is famous for its efficiency when assigned to independent attribute models. Many studies have proposed the Naïve Bayes algorithm with superior prediction accuracy [11]. Another popular algorithm is Apriori, which adopts the idea of parallel, which can dramatically increase its performance efficiency in predictions. The efficiency of the performance of the Apriori algorithm is where each local item set frequency is calculated, and all local items that occur frequently are combined into a global item candidate. Finally, the frequency of itemset that meets the conditions are filtered according to the minimum support threshold [12].

Preliminary research has been carried out by Fahmi et. al. [8] to predict waqf land. They predict waqf land using the Naïve Bayes classification into two categories, productive and not productive waqf. However, the accuracy results still need to be improved to support perfect planning. This study proposes a new method to improve prediction accuracy using the Apriori algorithm. Predictions with high accuracy are expected to help at the planning stage of strategic waqf land empowerment for the community's socio-economic development by the Ministry of Religion.

2   Related Work

A study comparative in waqf land for commercial use in Indonesia [13] concluded that waqf could function for commercial use as long as it supports public needs or the primary part of living. Another research on waqf land management categorizes waqf lands into the right investment category based on a strategic location in the Selangor region of Malaysia [14], [15]. The research was conducted by identifying the waqf land status and the utility of its investment models, such as mosques, graves, schools, charity foundations, and others. The highest percentage of all functions is used to utilize for a mosque. The waqf lands were analyzed based on their location utility to identify appropriate land utility changes if the improper was found. The results were followed by matching the waqf land categories with the proper investment model and looking for investors. This provided allocation resulted in waqf land assets that can be helped investors select types of funding projects for waqf land development.

Using a Bayesian Network with an Apriori algorithm in classifying problems has been conducted in several pieces of research, where the association rules were represented with the Bayesian Network [16], [17]. Like Vedula and Thatavarti (2011), there was a phase added by Xiao et al. (2016), in which the reuse of the association rules was conducted in the final phase. Association rules defined using the Apriori algorithm were represented with Bayesian Network [17]. Association rules were first conducted to select valuable rules as knowledge, and then the ability was sent to Bayesian Network

for probability calculation. Finally, association rules were reused with the probability calculation result for decision-making.

In classifying problems, Naïve Bayes and apriori algorithms have also been conducted in several research types [18]–[21]. A combination between Naïve Bayes and apriori algorithm [20] to identify a context of text documents. Text documents classification was conducted using Naïve Bayes at first, and then the document context identification was performed using the Apriori algorithm. The method used by D'Angelo et al. combines the Naïve Bayes and Apriori algorithm to achieve similarity with humans in decision making. The Apriori algorithm extracts the patterns, and Naïve Bayes is used for the final decision-making based on the user's trust probability [18].

Apriori algorithm and Naïve Bayes were used for text categorization [19]. Apriori algorithm was used to define word set with occurrence frequency, and then the word set was sent to Naïve Bayes for probability calculation. Similar to [19], the Apriori algorithm was used [21] to define frequent itemsets used by Naïve Bayes to calculate the probability for classification. Naïve Bayes and apriori algorithms were used by [22] to build a classifier that can detect spam or ham in SMS. The idea was created based on every word treated as an independent attribute in Naïve Bayes. In contrast, the words were independent of each other in the case of spam or ham detection. Apriori algorithm was used to identify high-frequency words in two separated databases, ham and spam. Further, the classifier was run by considering the frequent value of words in ham and spam databases.

Another classifier was built using Naïve Bayes, and an Apriori algorithm was developed [23]. The classifier was to classify documents into predefined categories based on their contents. Apriori algorithm was used to identify the keywords of data which was then the class probability was measured by Naïve Bayes. The Author [24] sed Naïve Bayes and apriori algorithm to build a text classifier. Apriori algorithm was used to derive a word set, and then Naïve Bayes calculated the probability of the word set. A similar text classifier used Naïve Bayes, and an apriori algorithm was also developed [25]. Naïve Bayes was used to building a text classifier, and the apriori algorithm defined rules that create a content profile.

Before doing this work, the authors [8] have proposed waqf land designation classification to distinguish productive and not productive assets for business development. The classification technique only uses one algorithm, the Naïve Bayes. The crucial attributes of the waqf asset data are analyzed, then these attributes are used to distinguish productive assets from not productive assets. The experimental results show that the proposed method can achieve an accuracy of 91%. In this study, it is proposed to add a new method of Apriori algorithm to improve prediction accuracy in determining productive waqf land for business prospects. Bayes' theorem and its combination with association rules are expected to improve prediction accuracy, which is higher and better than the previous experiment. The most optimal predictions can assist in planning and empowering waqf land for the community's socio-economic development.

3   Proposed Method

Waqf assets are vacant land, agriculture, plantation lands, or land and buildings. In the long term, the allocation of waqf lands can be empowered as productive or not productive assets. Productive assets allocation includes schools, shopping centres, health centres, gas stations, agriculture, plantation, finance/ coop, workshop. The not productive assets allocation, including mosques, mushallah (prayer room), cemeteries, and vacant lands. Figure 1 illustrates a model of productive and not productive waqf land assets allocation.
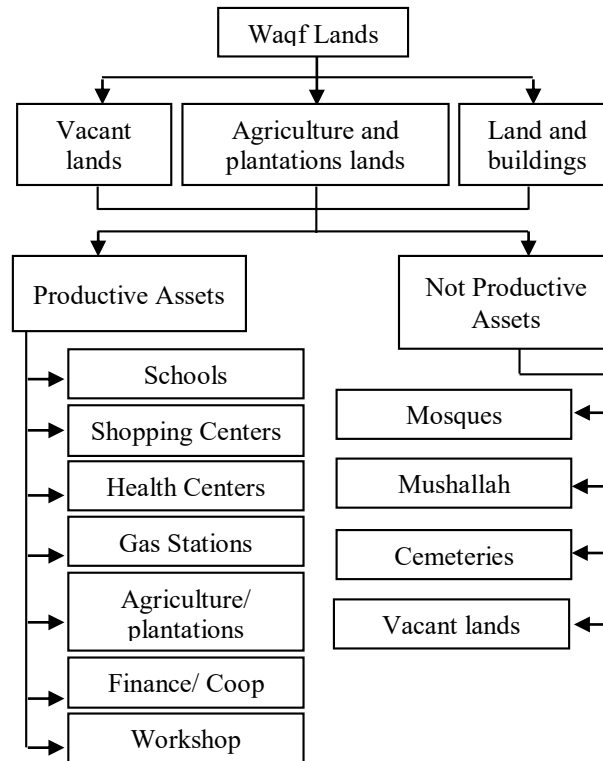
Fig 1. Proposed methode

This paper proposed Waqf land's business prospects classified into asset allocation productive or not productive using Naïve Bayes combined with apriori algorithm. We set a threshold value to divide the classification results by Naïve Bayes into an assumption of belief and confusion. If a result of a productive prior difference and not productive prior were under the threshold, it would be reclassified using the Apriori algorithm. In this paper, the research method is carried out in three stages: data collection, preprocessing classifications, and evaluation.

a. Data Collection and Pre-Processing

Data was collected from two religious affairs offices in Pedurungan, and East Semarang sub-districts, Semarang city, Center of Java, Indonesia. The data sample contained 159 waqf lands, including vacant land, land farms, and land with building contents that identified their origin and designation and matched current conditions. Predictions were built based on land value, so the waqf assets inland farms and lands with building contents were ruled out and considered lands. The selection of independent attributes was conducted by identifying the critical factors based on potential mapping of productive waqf assets that included legality, location, land size, population, type of Waqif, number of Nadzir, and Nadzir education [26]. The independent attributes are described in Table 1.

Table 1. Independent attributes

Attribute		Values
A1	Legality	Certified
		Not certified
A2	Locations	Near the main road
		Near public place
		In the housing area
A3	Size	Very large
		Large
		Medium
		Small
A4	Population	Very densely
		Densely
		Less densely
		Not densely
A5	Type of Waqif	Individual
		Group
		Legal entity
A6	Number of Nadzir	<=3
		>3
A7	Nadzir education	High
		Middle
		Low

b. Classification Using Naïve Bayes

Bayes theorem is a prediction technique based on simple probabilistic using a fundamental statistic approach to pattern recognition. This approach is based on quantification trade-offs between various classification decisions using the probabilities and consequences generated in those decisions [8] [27]. One of the applications of the Bayes theorem in classification is Naïve Bayes. Naïve Bayes works based on the assumption of simplification that attribute values are conditionally independent. In Bayes's theorem, if there are two separate cases, for instance, case A and case B, consequently Bayes's theorem is formulated as follows:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A) \qquad (1)$$

The Bayes theorem can be developed considering the validity of the total probability principle as follows:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{i=1}^{n} P(A_i|B)} \qquad (2)$$

The classification process is required to determine which class is appropriate for sample analysis. Therefore, the Bayes theorem can be adjusted as follows:

$$P(C|F_1 \ldots F_n)^n = \frac{P(C)P(F_1 \ldots F_n|C)}{P(F_1 \ldots F_n)} \qquad (3)$$

Variable C represents a class, while variable F1 ... Fn represents the characteristics of the instructions needed for classification. Opportunities of emergence samples that correspond to the characteristics in class C (posterior) is the opportunity of the emergence class C (prior) multiplied by an opportunity of appearance of the sample characteristic on class C (likelihood) and divided by the opportunity of the emergence of characteristics globally (evidence). So as the formula above can be simplified as follows:

$$Posterior = \frac{prior \; x \; likelihood}{evidence} + \cdots \qquad (4)$$

Table 2. Dataset and Class lable

No	Id	A1	A2	...	A7	Current Alocation	Class
1	001	Certified	Near the Main Road		Middle	Health center	P
2	002	Certified	In the Housing Area		High	School	P
3	003	Not Certified	Near Public Place		Low	Mosque	NP
4	004	Certified	Near Main Road		High	School	P
…	...	....	...		...	...	...
159	0159	Certified	Near Main Road		High	Health center	P

The current allocation of waqf assets data was used to label the sample data class as productive or not productive assets. Table 2 shows examples of a class label of waqf assets in the dataset with A1-A7 stands for independent attributes, current allocation, class label where 'P' stands for 'Productive', and 'NP' stands for 'Not Productive'.

$$P(A|B) \ = \ \frac{P(A) \prod_{i=1}^{n} P(Bi|A)}{P(B)} \qquad (5)$$

P(A|B) is the probability of the data with vector B on class A. P(B) is the probability of initial class(A). $\prod_{i=1}^{q} P(Bi|A)$ is an independent probability of class A of all features in vector B. The value of P(B) is permanently fixed. The prediction calculation has only calculated part $P(A) \prod_{i=1}^{q} P(Bi|A)$ by selecting the largest as the class selected as the prediction results.

A classification of waqf lands as productive and not productive assets was built using Naïve Bayes and apriori algorithm. Naïve Bayes was used to determining decisions by assuming that object attributes are independent. A threshold value calculated based on an average of the most significant false and most minor true of posterior values was defined as a classification parameter for Naïve Bayes. Apriori algorithm was used when the classification process by Naïve Bayes resulting a deviation in the posterior value, which was under the threshold. Apriori algorithm was expected to improve classification accuracy by adding a filter to classify data formerly predicted as not productive assets by Naïve Bayes. Fig. 2 shows the classification model using the Naïve Bayes and Apriori algorithm.

The experiment of classification using Naïve Bayes was randomly divided into 80% for training data and 20% for testing data. The division of these numbers with consideration of optimization on the sample data. The dataset contained 159 lands, so there were 127 lands for training data and 32 for testing data. The training data included 70 lands identified as productive assets and 57 not productive ones. The testing data contained 20 lands identified as productive assets and 12 identified as not productive assets.

The first training phase calculated the probability of an independent attribute using data of 127 lands identified as either productive or not productive assets. The prior value for productive assets/P (productive) was 0.55, and not productive assets/NP (not productive) was 0.45. The probability of the characteristics emergence sample on each class (likelihood) was listed in Table 3.

Then, the posterior data can be calculated using formula $P(A) \prod_{i=1}^{q} P(Bi|A)$. For example, given a piece of information as follows:
A1: Certified; A2: In the housing area; A3: Medium; A4: Densely; A5: Group; A6: >3; A7: Middle

Then the Posterior (Productive) = 0.57042 x 0.27778 x 0.14815 x 0.55556 x 0.16667 x 0.74074 x 0.59259= 0.0053, and the Posterior (Not productive) = 0.34444 x 0.40741 x 0.27778 x 0.38889 x 0.40741 x 0.61111 x 0.40741= 0.0069. The prediction was not productive, since the posterior value of not productive is higher than

productive. Table 4 shows some of the classification results conducted for training data, and Table 5 shows for testing data with productive and not productive.
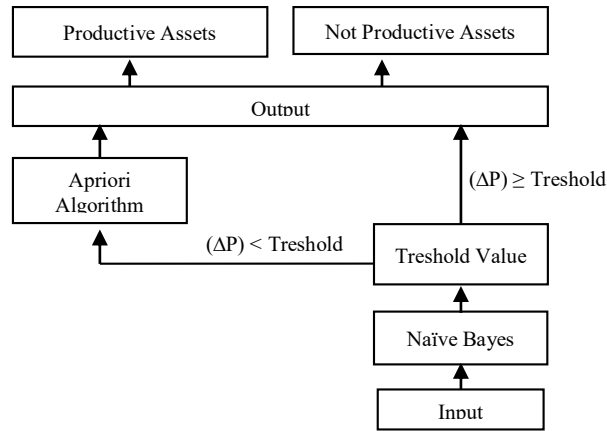

Figure 2. Model of classification using Naïve Bayes and Apriori Algorithm

Table 3. The probability results

Attributes		Productive	Not Productive
A1	Certified	1.03704	0.44444
	Not certified	0.03704	0.61111
A2	Near the main road	0.74074	0.61111
	Near Public place	0.14815	0.33333
	In the housing area	0.22222	0.22222
A3	Very large	0.44444	0.22222
	Large	0.40741	0.27778
	Medium	0.14815	0.16667
	Small	0.14815	0.55556
A4	Very densely	0.40741	0.27778
	Densely	0.44444	0.38889
	Less densely	0.18519	0.16667
	Not densely	0.11111	0.38889
A5	Individual	0.40741	0.55556
	Group	0.44444	0.27778
	Legal entity	0.11111	0.38889
A6	<=3	0.40741	0.55556
	>3	0.44444	0.27778
A7	High education	0.48148	0.33333
	Middle	0.59259	0.77778
	Low education	0.14815	0.55556

Table 4. Classification results of training data

No	Id	Posterior Productive	Not Productive	Class
1	001	0.0494	0.0052	Productive
2	002	0.0120	0.0008	Productive
3	003	0.0003	0.0068	Not Productive
4	004	0.0146	0.0023	Not Productive
5	006	0.0036	0.0099	Not Productive
…	...	...	...	...
127	159	0.0438	0.0031	Productive

Accuracy was measured using the confusion matrix by dividing the classes into positive data (true positive/TP; false positive/FP) and negative data (true negative/TN; false negative/FN). The classification results for training data achieved accuracy at

91%, with 115 correct predictions and 12 wrong predictions. Testing data achieved accuracy at 84%, with 26 accurate predictions and 6 wrong predictions. Table 6 shows an evaluation of classification results by Naïve Bayes.

Table 5. Classification results of testing data

No	Id	Posterior Productive	Not Productive	Class
1	005	0.0183	0.0013	Productive
2	007	0.0111	0.0010	Productive
3	009	0.0368	0.0028	Productive
4	010	0.0001	0.0046	Not Productive
5	018	0.0078	0.0075	Not Productive
…	...	…	…	…
32	052	0.0029	0.0039	Not Productive

Table 6. Accuracy of classification by Naïve Bayes

Data	True Positive	False Positive	True Negative	False Negative	Accuracy
Training	73	8	42	4	0.91
Testing	18	4	8	2	0.84

c.   Improving Classification Using Apriori Algorithm

Apriori is an algorithm that searches frequent itemsets by using the association rule technique [28]. Agrawal and Srikant proposed the algorithm to determine frequent itemset in the rules of boolean associations. This algorithm controls the development of candidate itemset from the regular itemset results with the support-based pruning to eliminate unappealing itemset [29] [30]. This algorithm is also defined as finding all Apriori rules that qualify for support and minimum confidence requirements.

There are two main processes to configure the itemset candidate [31], joining in which each item is combined with other things so that no more combinations can be formed again and pruning in which the results of the items that have been integrated are then pruned by using the determined minimum support. Below is the formula to find the support value of an itemset X:

$$\text{Support(X)} = \frac{Num. \, of \, Trans. \, in \, which \, X \, appears}{Total \, Number \, of \, transactions} \quad (6)$$

While formula to find confidence, value is:

$$\text{Conf} (X \rightarrow Y) = \frac{Supp \, (X \cup Y)}{Supp \, (X)} \quad (7)$$

In this phase, the classification process using the Naïve Bayes was extended by applying rules defined using the apriori algorithm. A threshold value is determined based on a mean value of the false classification's most significant posterior value. The smallest posterior value of the correct category was calculated after conducting a classification experiment using the Naïve Bayes. The Naïve Bayes' classification with a deviation of posterior value below the threshold value was reclassified using an apriori algorithm modified in defining rules.

The deviation of the most significant posterior value of the false classification results at 0.0039, and the smallest posterior value of the accurate classification results at 0.0015, then the threshold value was (0.0039 + 0.0015)/2 = 0.0028. There were 8 lands in data training below the threshold value, and 4 lands were the wrong prediction. While in data testing, 4 lands were below the threshold value, and 2 were the wrong predictions.

Rules defined by the apriori algorithm were expected to increase the prediction accuracy by keeping the 8 prediction results and correcting the wrong prediction results. Itemsets were determined based on seven independent attributes values (A1-A7) used in Naïve Bayes, and data of prediction results by Naïve Bayes were added as attributes and denoted as A8. So there were 15 items collected from seven independent attribute values and two items of productive and not productive used to define association rules as follows:

items = {Certified, Not certified, Near the main road, Near public place, In the housing area, Very large, Large, Medium, Small, Very densely, Densely, Less densely, Not densely, Individual, Group, Legal entity, Nadzir <=3, Nadzir >3, High education, Low education, productive, not productive}.

Transaction data were collected from 159 records in the dataset. The rules were defined based on 8-itemsets representing 7 independent attributes values plus 1 attribute value of prediction result by Naïve Bayes.

By defining the minimum support = 1, all items were frequent (F1) as the results of 1-itemsets. Table 7 shows the effect of 1-itemsets.

Table 7. 1-itemsets

Attribute	Itemsets	Number of Transaction	Support
A1	Certified	127	13%
	Not certified	32	4%
A2	Near the main road	95	10%
	Near Public place	32	3%
	In the housing area	32	4%
A3	Very large	48	5%
	Large	64	6%
	Medium	16	2%
	Small	32	4%
A4	Very densely	48	5%
	Densely	48	6%
	Less densely	32	3%
	Not densely	32	3%
A5	Individual	80	6%
	Group	48	2%
	Legal entity	32	3%
A6	Nadzir >=3	64	2%
	Nadzir<3	95	4%
A7	High education	64	6%
	Middle education	70	4%
	Low education	25	10%
A8	Productive	95	10%
	Not Productive	64	6%
Total		1275	100%

The next step was searching for 2-itemsets using itemsets of an attribute of A8 (productive and not productive) as the key to finding confidence values of 2-itemsets of attributes A1-A7. Table 8 shows 2-itemsets defined based on characteristics A1-A7 using itemsets of attribute A8 as the key.

The confidence values of 2-itemsets based on attributes were used to measure a record's information to reclassify results of classification by Naïve Bayes with a deviation of posterior value below the threshold 2-itemsets were summed up to find the mean value of the confidence. The mean value was then tested by a minimum confidence value to define whether Naïve Bayes' prediction was right or wrong.

Table 8. 2-itemsets based on attributes

Attrib.	2-Itemsets
A1	{Certified, productive}, {certified, not productive}, {not certified, productive}, {not certified, not productive}
A2	{Near main road, productive}, {near the main road, not productive}, {near public place, productive}, {near public place, not productive}, {in the housing area, productive}, {in the housing area, not productive}
…	…
A7	{High education, productive}, {high education, not productive}, {low education, productive}, {low education, not productive}

The calculation of support and confidence value for 2-itemsets of each attribute was separately conducted. Table 9 shows examples of 2-itemsets calculation.

Table 9. 2-itemsets of Attribute A1 and A7

Attr.	2-Itemsets		X∪Y	Support (X)	Confidence
	X	Y			
A1	Certified	Productive	37	46	80.4%
		Not Productive	0	13	19.6%
	Not Certified	Productive	9	46	0.0%
		Not Productive	13	13	100.0%
A2	Near main road	Productive	8	14	57.1%
		Not Productive	6	14	42.9%
	Near public place	Productive	24	34	70.6%
		Not Productive	10	34	29.4%
	In housing area	Productive	6	11	54.5%
		Not Productive	5	11	45.5%
…	…	…	…	…	…
A7	High educatio	Productive	17	23	73.9%
		Not Productive	6	23	26.1%
	Middle education	Productive	20	36	55.6%
		Not Productive	16	36	44.4%
	Low education	Productive	5	11	45.5%
		Not Productive	6	11	54.5%

Rules were defined based on the mean value of 2-itemsets of attributes A1-A7 with the minimum confidence set at 70%. The algorithm works by testing the prediction by Naïve Bayes. For an example, given information as follows:
A1: Certified; A2: In the housing area; A3: Medium; A4: Densely; A5: Group; A6: >=3 A7: Low Education

The information above was predicted as not productive by Naïve Bayes. The proposed apriori algorithm tested this prediction by finding the mean value of attributes A1-A7 confidence using the prediction result by Naïve Bayes. Table 10 shows the calculation of confidence means the value, which was at 51.1%. Since the mean value was lower than the minimum confidence value, which was 70%, the apriori algorithm convinced that the prediction result by Naïve Bayes was accurate.

Table 10. Works of proposed apriori algorithm

Attribute	Itemsets	Confidence
A1	Certified →Not Productive	100,00%
A2	In the house area →Not Productive	42,0%
A3	Medium →Not Productive	46,0%
A4	Densely →Not Productive	44,0%
A5	Group → Not Productive	51,0%
A6	>=3 → Not Productive	38,0%
A7	Low edu. → Not Productive	36,0%
Mean		51.1%

The rules were implemented to reclassify the classification results by Naïve Bayes with a deviation of posterior value below the threshold value. Table 11 shows the prediction results by the Apriori algorithm on training data and testing data. No., Id, Naïve Bayes Prediction, confidence mean value, T/F stands for true or false, reclassification by apriori, and marks of asterisks denote the wrong prediction by Naïve Bayes or the apriori.

Based on reclassification results of training data, rules defined by the apriori algorithm can keep all the right predictions by Naïve Bayes and correct three of four wrong predictions, which was Land with Id. 036, 039, and 041. Another wrong prediction, Id. 032 with productive class, was judged as accurate by apriori. While testing data, the apriori algorithm could also keep all the right predictions and correct one of two wrong predictions: Land with Id. 118.

Table 11. Reclassification of training data by Apriori Algorithm

Data	No.	Id.	Naïve Bayes Prediction	Apriori confidence mean value	T/F	Reclassification
Training	1	022	Productive	70.4%	T	Productive
	2	025	Not Productive	44.7%	T	Not Productive
	3	027	Productive	72.4%	T	Productive
	4	032	Productive*	57.5%	T	Productive*
	5	033	Not Productive	56%	T	Not Productive
	6	036	Productive*	68.6%	F	Not Productive
	7	039	Productive*	67.6%	F	Not Productive
	8	041	Not Productive*	74.6%	F	Productive
	…	…	…	…	…	…
Testing	1	118	Not Productive*	62.4%	F	Productive
	2	154	Not Productive	61.7%	T	Not Productive
	3	155	Productive*	70.8%	T	Productive*
	4	158	Not Productive	59.8%	T	Not Productive
	…	…	…	…	…	…

The accuracy of reclassification using the apriori algorithm was measured using a confusion matrix. Based on total classification results by Naïve Bayes, which was then reclassified by a modified apriori algorithm, results of classification for training data achieved accuracy at 98%. For testing, data achieved accuracy at 93% (Table 12).

Table 12. Accuracy of reclassification by the Apriori algorithm

Data	True Positive	False Positive	True Negative	False Negative	Accuracy
Training	80	1	45	1	0.98
Testing	22	1	9	1	0.93

4   Conclusion

The proposed method by implementing Naïve Bayes and Apriori algorithm to build a classifier can improve the accuracy of classification conducted by Naïve Bayes only. The Apriori algorithm's task was to reclassify the classification by Naïve Bayes assumed as confusion based on a threshold value. By identifying frequent items of lands attributes values and items of prediction results by Naïve Bayes, the modified apriori can keep right predictions and correct wrong predictions by Naïve Bayes. The accuracy value in the training process increased from 91% using Naïve Bayes alone to 98% by the Apriori algorithm, and i84% increased to 93% in testing.

It is interesting to improve this proposed method to build a decision support system of allocation types of business prospects for lands empowerment for the following works.

References

1   M. De-Arteaga, W. Herlands, D. B. Neill, and A. Dubrawski, "Machine Learning for the Developing World," *ACM Trans. Manag. Inf. Syst.*, vol. 9, no. 2, p. 9:1-9:14, Aug. 2018, doi: 10.1145/3210548.

2   T. Cagala, U. Glogowsky, J. Rincke, and A. Strittmatter, "Optimal Targeting in Fundraising: A Machine Learning Approach," *ArXiv210310251 Cs Econ Stat*, Mar. 2021, Accessed: Mar. 30, 2021. [Online]. Available: http://arxiv.org/abs/2103.10251

3   D. Helbing and S. Balietti, "From social data mining to forecasting socio-economic crises," *Eur. Phys. J. Spec. Top.*, vol. 195, no. 1, p. 3, May 2011, doi: 10.1140/epjst/e2011-01401-8.

4   J. T. d Souza, A. C. d Francisco, C. M. Piekarski, G. F. d Prado, and L. G. d Oliveira, "Data mining and machine learning in the context of sustainable evaluation: a literature review," *IEEE Lat. Am. Trans.*, vol. 17, no. 03, pp. 372–382, Mar. 2019, doi: 10.1109/TLA.2019.8863307.

5   D. P. W. Direktorat Jenderal Bimbingan Masyarakat Islam, *Panduan Pemberdayaan Tanah Wakaf Produktif Strategis di Indonesia*. Jakarta: Departemen Agama RI, 2009. Accessed: Mar. 30, 2021. [Online]. Available: http://simbi.kemenag.go.id/eliterasi/katalog-buku/panduan-pemberdayaan-tanah-wakaf-produktif-strategis-di-indonesia-tahun-2009

6   A. Rosadi, D. Effendi, and B. Busro, "The Development of Waqf Management Throught Waqf Act in Indonesia (Note on Republic of Indonesia Act Number 41 of 2004 regarding Waqf)," *MADANIA J. Kaji. Keislam.*, vol. 22, no. 1, pp. 1–18, 2018.

7   : "Sistem Informasi Wakaf :." http://siwak.kemenag.go.id/ (accessed Dec. 08, 2021).

8   A. Fahmi, E. Sugiarto, A. Winarno, S. Sumpeno, and M. H. Purnomo, "Waqf Lands Assets Classification Based On Productive Value For Business Development Using Naïve Bayes," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Nov. 2018, pp. 622–626. doi: 10.1109/ISRITI.2018.8864489.

9   A. Rikalovic, I. Cosic, R. D. Labati, and V. Piuri, "A Comprehensive Method for Industrial Site Selection: The Macro-Location Analysis," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2971–2980, Dec. 2017, doi: 10.1109/JSYST.2015.2444471.

10 G. Turhan, M. Akalın, and C. Zehir, "Literature Review on Selection Criteria of Store Location Based on Performance Measures," *Procedia - Soc. Behav. Sci.*, vol. 99, pp. 391–402, Nov. 2013, doi: 10.1016/j.sbspro.2013.10.507.

11 S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowl.-Based Syst.*, vol. 192, p. 105361, Mar. 2020, doi: 10.1016/j.knosys.2019.105361.

12 H.-B. Wang and Y.-J. Gao, "Research on parallelization of Apriori algorithm in association rule mining," *Procedia Comput. Sci.*, vol. 183, pp. 641–647, Jan. 2021, doi: 10.1016/j.procs.2021.02.109.

13 D. B. F. Sihombing and A. Sugiantoro, "A Comparative Study of Waqf Land for Commercial Use in Indonesia," vol. 11, no. 5, p. 4, 2016.

14 N. B. Nor and M. O. Mohammed, "Categorization of waqf lands and their management using Islamic Investment Models: The case of the State of Selangor, Malaysia," in *International Conference on Waqf Laws & Management: reality and Prospects, IIUM*, 2009, pp. 20–22.

15 M. S. Shabbir, "Classification and prioritization of waqf lands: a Selangor case," *Int. J. Islam. Middle East. Finance Manag.*, 2018.

16 V. R. Vedula and S. Thatavarti, "Binary association rule mining using bayesian network," in *Proceedings of 2011 International Conference on Information and Network Technology (ICINT 2011)*, 2011, vol. 4, pp. 171–176.

17 S. Xiao, Y. Hu, J. Han, R. Zhou, and J. Wen, "Bayesian Networks-based Association Rules and Knowledge Reuse in Maintenance Decision-Making of Industrial Product-Service Systems," *Procedia CIRP*, vol. 47, pp. 198–203, Jan. 2016, doi: 10.1016/j.procir.2016.03.046.

18   G. D'Angelo, S. Rampone, and F. Palmieri, "Developing a trust model for pervasive computing based on Apriori association rules learning and Bayesian classification," *Soft Comput.*, vol. 21, no. 21, pp. 6297–6315, 2017.

19   S. M. Kamruzzaman and C. M. Rahman, "Text categorization using association rule and naive Bayes classifier," *ArXiv Prepr. ArXiv10094994*, 2010.

20   A. R. Kulkarni, V. Tokekar, and P. Kulkarni, "Identifying context of text documents using Naïve Bayes classification and Apriori association rule mining," in *2012 CSI Sixth International Conference on Software Engineering (CONSEG)*, Sep. 2012, pp. 1–4. doi: 10.1109/CONSEG.2012.6349477.

21   T. Yang, K. Qian, D. C.-T. Lo, Y. Xie, Y. Shi, and L. Tao, "Improve the Prediction Accuracy of Naïve Bayes Classifier with Association Rule Mining," in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, Apr. 2016, pp. 129–133. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.38.

22   I. Ahmed, D. Guan, and T. C. Chung, "SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset," *Int. J. Mach. Learn. Comput.*, vol. 4, no. 2, pp. 183–187, Apr. 2014, doi: 10.7763/IJMLC.2014.V4.409.

23   K. J. Patel and K. J. Sarvakar, "Web Page Classification Using Data Mining," vol. 2, no. 7, p. 8, 2013.

24   A. Purohit, D. Atre, P. Jaswani, and P. Asawara, "Text Classification in Data Mining," vol. 5, no. 6, p. 6, 2015.

25   A. Dhanne, G. Kale, and P. Kulkarni, "Context based document searching using Apriori Association for Mobile Devices," *undefined*, 2014, Accessed: May 03, 2021. [Online]. Available: /paper/Context-based-document-searching-using-Apriori-for-Dhanne-Kale/70330f6f1e6a28fd400a648ed51e940a2632de7b

26   S. Sofiandi, "Towards Reformulation of Waqf: An Indonesians Discourse," *Asia-Pac. J. Relig. Soc.*, vol. 3, no. 2, Art. no. 2, Dec. 2019.

27   C. C. Aggarwal, *Data mining: the textbook*. Springer, 2015.

28   L. Hanguang and N. Yu, "Intrusion Detection Technology Research Based on Apriori Algorithm," *Phys. Procedia*, vol. 24, pp. 1615–1620, Jan. 2012, doi: 10.1016/j.phpro.2012.02.238.

29   M. G. Ingle and N. Y. Suryavanshi, "Association Rule Mining using Improved Apriori Algorithm," *Int. J. Comput. Appl.*, vol. 112, no. 4, pp. 37–42, Feb. 2015.

30   S. A. Abaya, "Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation," vol. 3, no. 7, pp. 1–4, 2012.

31   X. Liu and H. Liu, "An Improved Apriori Algorithm for Association Rules," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 11, Art. no. 11, Nov. 2013.