2018 2nd International Conference on Electrical Engineering and Informatics (ICon EEI 2018), Batam - Indonesia, 16th-17th October 2018

Using Bayesian Network for Determining The Recipient of Zakat in BAZNAS Pekanbaru

Akbarizan
Faculty of Sharia and Law, Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Subrantas Km. 15, Pekanbaru 28293, Indonesia
akbarizan@uin-suska.ac.id

Rahmad Kurniawan
Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Subrantas Km. 15, Pekanbaru 28293, Indonesia
rahmadkurniawan@uin-suska.ac.id

Mohd Zakree Ahmad Nazri
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
zakree@ukm.edu.my

Siti Norul Huda Sheikh Abdullah
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
snhsabdullah@ukm.edu.my

Sri Murhayati
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Subrantas Km. 15, Pekanbaru 28293, Indonesia
sri.murhayati@uin-suska.ac.id

Nurcahaya
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Subrantas Km. 15, Pekanbaru 28293, Indonesia
nurcahaya@uin-suska.ac.id

***Abstract*—The National Amil-Zakat Agency (Baznas) in Pekanbaru has the function to collect and distribute zakat in Pekanbaru city. Baznas Pekanbaru should be able to determine *Mustahik* properly. *Mustahik* is a person eligible to receive zakat. The Baznas committee interviews and observes every *Mustahik* candidates to decide whom could be receive the zakat. Current *Mustahik* determination process could lead to be subjective assessment, due to large number of zakat recipient applicants and the complexity of rules in determining a *Mustahik*. Therefore, this study utilize artificial intelligence in determining *Mustahik*. The Bayesian Network method is appropriate to apply as an inference engine. Based on the experimental results, we found that Bayesian network produces a good accuracy 93.24% and effective to use in data set have an uneven class distribution. In addition, based on experiments by setting an alpha estimator's values, at 0.6 to 1.0 can increase the accuracy of a Bayesian Network to 95.95%.**

Keywords—bayesian network, baznas pekanbaru, mustahik, zakat

I. INTRODUCTION

Pekanbaru is the capital of Riau province, Indonesia and a major economic center on the eastern part of Sumatra Island. Based on Indonesia Central Agency on Statistics, in 2017, the number of Muslim in Riau were 4.8 million people. The good economy in Riau has an impact on enthusiastic citizens to give zakat. Zakat is a Muslim duty. It is obligatory for Muslims who are able to give it. Every Muslim who has a surplus of wealth must pay Zakat [1].

Zakat is the third pillar of the Islamic pillars. *Muzakki* is a Muslim who has qualified and obliged in financial to give zakat. Meanwhile, *Mustahik* is a people entitled to receive zakat. The main tasks of Baznas in distributing zakat is to prioritize the distribution of zakat based on applicant's accurate data. Some of the programs of giving zakat are for productive enterprises, scholarship and training, construction of facilities and infrastructure. Baznas in Pekanbaru must be immediately channeled the zakat to the *Mustahik* according to the priority scale that has been compiled in the program. Islam has regulated the *Mustahik* in Quran, verse 60 of Surah *At-Taubah.*

Indonesia Central Agency on Statistics reported the number of poor people (population below the Poverty Line) in Riau. September 2017 amounted to 496.39 thousand people (7.41%). When compared with the poor population in September 2016 which amounted to 501.59 thousand people (7.67%), the poor in Riau decreased by 5.20 thousand people. Zakat has the potential to reduce the social gap between *Muzakki* and *Mustahik*. The number of *Muzakki* in Riau province more than *Mustahik*. The ratio of *Mustahik* and *Muzakki* reaches 0.56. Particularly, 100 *Muzakki* can bear 56 *Mustahik*.

Procedurally, *Muzakki* make the payment of zakat in Baznas office. Baznas receives zakat from *Muzakki* and distributes to *Mustahik*. Baznas determinates a *Mustahik* candidate manually, i.e. interviewing candidates and doing field observation. Feelings, emotions, sentiments are related with atmospheres and moods that can lead to miss-judgment and affecting the zakat distribution. Meanwhile, artificial intelligence can assist the repetitive and monotonous works.

Classification technique in artificial intelligence think faster than humans think and can performs multi-tasking. Classification parameters is different with humans, it can be adjusted. Their speed and time are calculation based parameters only. On other hands, artificial intelligence do not need frequent rests. The humans can used AI for long hours and can continuously perform the works without getting bored or tired. One of the popular classification methods is Bayesian Network.

978-1-5386-6000-3/18/$31.00 ©2018 IEEE

We considered the Bayesian Network to be more popular since Bayesian Network inference is consistent to handle uncertainty cases. Bayesian networks can facilitate learning about the causal relationship between the variables [2]. Bayesian Network is easy to be converted into tools for decision support such as to assist the management of natural resources [3]. Hence, we chose the Bayesian Network method as inference engine that proved reliable for uncertainty problem.

Bayesian Networks obviously show the relationship between different components of the system using the graphics of Bayesian. Therefore, Bayesian Network is comfortable for researchers from different backgrounds to understand the concept of Bayesian Network [4]. Bayesian inference can use readily updated for new knowledge and Bayesian networks can produce a good prediction accuracy [5]. Therefore, this study uses Bayesian Network as a method of classification in determining *Mustahik*.

II. RELATED WORKS

Lately, artificial intelligence is widely used in the field of Islamic science. Sulaiman *et al* presented an expert system using Islamic approaches [6]. They have proposed a knowledge-based approach in expert system for Hajj Pilgrims. They planned a dynamic knowledge-based approach to diagnose possible problems and solutions from the expert. They have implemented expert system using Forward Chaining. Users can obtain enquire any questions related to Hajj ritual.

The zakat has also adopted the artificial intelligence. Afaf-Al-Riaymi *et al* have used expert system to help people calculate the amount of each type of Zakat [1]. This study help *Muzakki* to determine if they are required to pay Zakat, determine the unique conditions, and the amount of each type of Zakat they have to give every year. They have developed the system based on a rules based expert system shell.

Harry *et al* have used the artificial intelligence especially Weight Product Method to determine *Mustahik* [7]. This decision support system calculated and sorted *Mustahik* using Weighted Product method with monthly income criteria, number of dependents, home ownership, ownership of motor vehicles and presence or absence of responsibility.

Kurniawan *et al* have developed a prototype expert system using Bayesian Network for early diagnosing and educating the peoples in social ills cases [8, 9]. It has proved its capability for educating the peoples in social ills cases based on Quran, Hadith and advices of experts. The implementation of Bayesian Network shows a promising result.

According to related works, the zakat studies merely focus on calculation of zakat and they used a rule-based method, which is traditional inference of the expert system. A traditional inference could be possible to conclude the same classes despite having different symptoms [10]. Another related work regarding zakat is a decision support system for determining *Mustahik*, but it has not explained the accuracy of the method. The Weighted Product method provides decision with rankings. Therefore, we considering this research use statistical methods that is more intelligent for inference engine. In contrast to the traditional inference, the Bayesian Network method is a combination of probability theory and graph theory so it is more appropriate in giving decision. This study aims to assist the BAZNAS to determining recipient's candidate of zakat (*Mustahik*) more quickly, fairly and transparently.

III. MATERIALS AND METHODS

The method used is according to data mining process, starting from the data collection and preprocessing. Data obtained from Baznas Pekanbaru in 2017, counted 1000 with 14 attributes (*sequence number, districts, IC number, date, name, address, phone number, age, marital status, dependents, occupation, program, income and decision*).

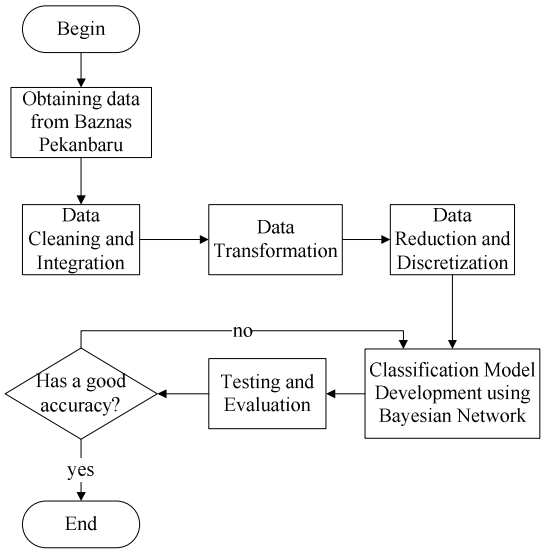

Fig. 1. Methodology of Bayesian Network classification technique

Fig.1 shows six steps to conducting the Bayesian Network classification technique. We perform data pre-processing after data collection to obtain clean and useful of the data. Data pre-processing greatly determines the classification results. Data pre-processing includes data cleaning, data transformation, data reduction and discretization, classification, testing and evaluation. Discretization of data is very important in data pre-processing, because we use the Bayesian Network as classification.

A. Data Preprocessing

The data preprocessing technique includes data cleaning, data integration, data transformation and data reduction. This technique uses to make sure that the data is suitable for analysis. In the cleaning data concept includes filling missing values, identifying of outliers, and handling of noise data, correcting of inconsistent data and solving the redundancy after data integration. We were conducted data integration to combine data from multi-sources. We have obtained some data from different sources and formats. We have to integrate of schemes, identifying of entity, detecting and solving of data values conflicts.

We did not find missing value in this data. However, we have removed attributes that have no effect on classification results such as *sequence number, districts, program, IC number, date, address and phone number*.

The final steps in data pre-processing is data transformation and data reduction. We have transformed data into appropriate forms. We also applied data reduction techniques and discretization to reduce the presentation of the data set. The discretization involves reduction number of values of a continuous attribute by dividing the range of attribute intervals such as age and income. We have categorized age to be:

1) *Adult (aged 20 to 30 years)*
2) *Adult (aged 31 to 59 years)*
3) *Adult (aged >60 years)*

Based on the regulations applicable to Baznas Pekanbaru, We have categorized the income as follows:

1) *Income I = Rp. 0 – 500,000*
2) *Income II = Rp. 500,001 – 1,000,000*
3) *Income III = Rp. 1,000,001 – 1,500,000*

Meanwhile, for the class, automatically determined by the machine using Bayesian Network algorithm. We have specified three types of the classes.

1) *Zakat I = Rp. 1,000,000 – 1,500,000*
2) *Zakat II = Rp. 1,500,001 – 2,000,000*
3) *Zakat III = Rp. 2,000,001 – 2,500,000*

B. Classification Model

The Bayesian network is also known as Belief Network, Bayesian Belief Network, Bayes Nets or referred to as the Probabilistic Network. The Bayesian network is a method based on the Bayes theorem created by Thomas Bayes in 1763. The Bayesian Network method has become very popular over the last decade as used for a variety of intelligent applications such as machines learning, text processing, bioinformatics processing, medical diagnostics, weather forecasts, and other intelligent system applications.

Bayesian Network modeling techniques have several features that make this method useful in many data analysis and management issues. The Bayesian network can also show good prediction accuracy even with a small data. The Bayesian Network Method is also useful for combining a variety of knowledge well.

Let *U= {age, status, dependents, occupation, income}* be a set of variables. A Bayesian network *B* over a set of variables *U* is a network structure, which is a Directed Acyclic Graph (DAG) over *U* and a set of probability tables $B_P = \{p(u|pa(u))|u \in U\}$ where $p(u)$ is the set of parents of u in structure [11]. A Bayesian network represents a probability distribution $P(U) = \Pi_{u \in U} p\left(u|pa(u)\right)$. The classifier has learned from a zakat data set consisting of samples over (*attributes, zakat classes*). The learning task consists of finding an appropriate Bayesian network given a data set *D* over *U*.

Fig.2 shows the structure of Bayesian Network for each node contains probability.

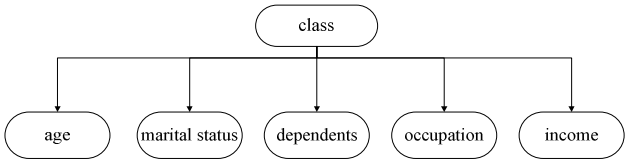

Fig. 2. The simple structure of Bayesian Network to determining *Mustahik* in Baznas Pekanbaru

C. Testing and Evaluation

Every experiment was performed using the dataset split factor (training data : testing data) from 10 to 90 for classification modeling. These nine different models are obtained from each experiment. The parameters to evaluate the performance of each classification modeling as follow.

a. Accuracy $= \frac{TN+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$

b. Precision $= \frac{TP}{TP+FP}$

c. Recall $= \frac{TP}{TP+FN}$

d. F-Measure $= 2. \frac{Precision\ .\ Recall}{Precision + Recall}$

Where:

Condition Positive (P): the number of real positive cases in the data

Condition Negative (N): the number of real negative cases in the data

True Positive (TP): These refer to the positive cases in the data and correctly labeled by the classifier. Let TP be the number of true positives.

True Negative (TN): These are the negative cases were correctly labeled by the classifier. Let TN be the number of true negatives.

False Positive (FP): These are the negative cases were incorrectly labeled as positive

False Negative (FN): These are the positive cases were mislabeled as negative

Suppose there are three classes on zakat data set C1, C2, and C3, so that:

a. "TP of C1" is all C1 instances are classified as C1.

b. "TN of C1" is all non-C1 instances are not classified as C1.

c. "FP of C1" is all non-C1 instances are classified as C1.

d. "FN of C1" is all C1 instances are not classified as C1.

IV. Results and Discussion

We were conducted the data preprocessing and zakat data set is ready to classify using the Bayesian network algorithm. We used a Weka version 3.8.2 as a tool for machine learning which has capability in classifying data using Bayesian network. All Bayes network algorithms

implemented in Weka assume all variables are discrete finite variables.

We have set the split percentage from the data training 10% to 90%. We are made some arrangements in Weka classification options e.g. estimator and search algorithm. We have used Simple Estimator for estimating the conditional probability tables of a Bayes network once the structure has been learned. Estimates probabilities directly from data.

We settled alpha is 0.5 as default value in Weka. We have used Alpha for estimating the probability tables and can interpreted as the initial count on each value. In Weka classification, there was featured search algorithm to assist the classification tasks. This Bayesian Network learning algorithm uses a hill climbing algorithm restricted by an order on the variables. Table I shows the result of the measurement of the Bayesian classification process.

TABLE I. CLASSIFICATION ACCURACY OF BAYESIAN NETWORK

Exp	(Training) (%)	Accuracy	Precision	Recall	F-Measure
1.	10	94.8571	0.930	0.949	0.939
2.	20	93.8907	0.936	0.939	0.938
3.	30	95.2206	0.938	0.952	0.944
4.	40	94.8498	0.930	0.948	0.938
5.	50	94.3299	0.899	0.943	0.921
6.	60	92.9487	0.876	0.929	0.902
7.	70	92.3077	0.923	0.923	0.960
8	80	88.4615	0.826	0.885	0.855
9.	90	92.3077	0.912	0.923	0.916
Average		93.24	0.91	0.93	0.92

Based on the Table I, we have found that Bayesian network produces a good average accuracy 93.24%. Bayesian network is accurate in zakat data set, even has an average of precision 0.91. In average of recall 0.93, so Bayesian network has predicted positive observations to the all observations in actual class is yes. F-Measures is weighted average of precision and recall. Bayesian network has earned good F-Measures 0.92. Although accuracy is the reference in assessing the correctness of an algorithm, but F-Measures has verified Bayesian network can use if data set have an uneven class distribution.

Fig.3 shows the best accuracy in the third experiment, which is 30% for training data allocation with accuracy 95.22%. We have continued experiment by looking at the alpha estimator's relationship among accuracy. The training data allocation used is 30%. Table II shows the relationship between alpha and accuracy.

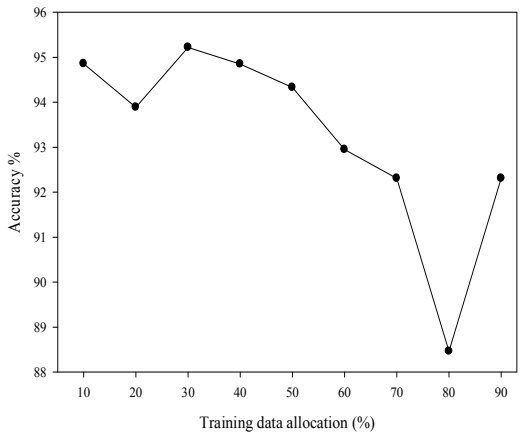

Fig. 3. The relationship of accuracy and training data allocation

Based on Table II, we obtain a correlation of accuracy with alpha estimator's determination. Although Weka by default has set the alpha value is 0.5, but the changes of alpha has been able to affect to accuracy.

TABLE II. THE RELATIONSHIP OF ALPHA ESTIMATOR'S

Exp	Alpha	Accuracy (%)
1.	0.1	94.4853
2.	0.2	95.2206
3.	0.3	95.2206
4.	0.4	95.2206
5.	0.5	95.2206
6.	0.6	95.9559
7.	0.7	95.9559
8	0.8	95.9559
9.	0.9	95.9559
10.	1.0	95.9559

Based on the experiment by changing the alpha value from 0.1 to 1.0, it has changed of accuracy. Experiments with setting alpha value 0.6 can increase the accuracy of a Bayesian Network. The alpha setting from 0.6 to 1.0 produces the same accuracy 95.95%.

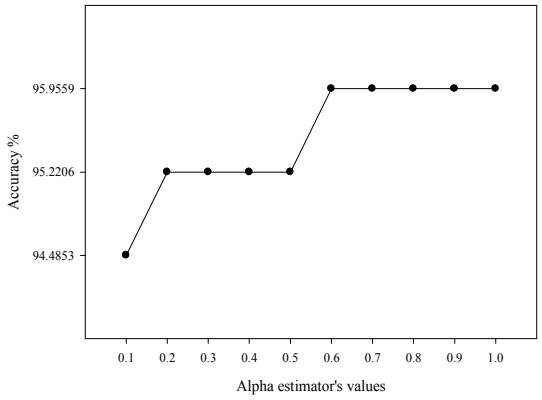

Fig. 4. The relationship of alpha estimator's values and accuracy

Fig. 5 shows the precision and recall have earned a high score in the third experiment, which is 30% for training data allocation. Meanwhile on F-Measure have earned the highest score in the seventh experiment with 70% training data allocation. Fig. 5 shows the relationship among precision, recall and F-Measure according to training data allocation.

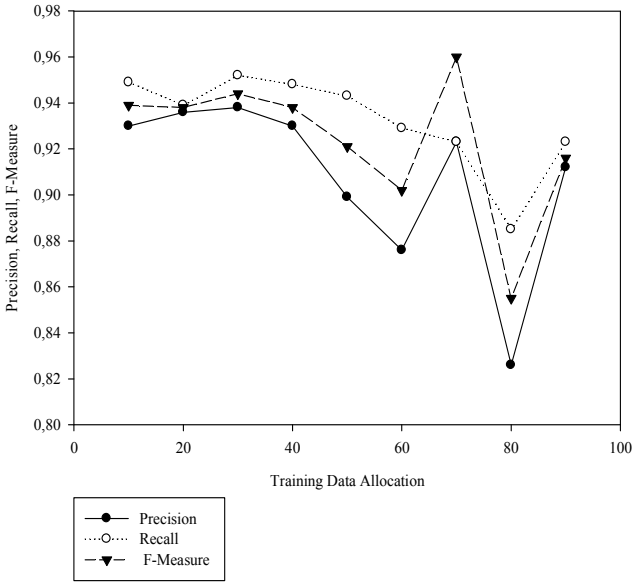

Fig. 5. The relationship of precision, recall, F-Measure and training data allocation

The final phases is evaluate the effectiveness of the system. The developed system has to produce the desired outcome in accordance with its objectives with high accuracy without any systemic error. We compared the outcomes of *Mustahik* determination system with expert.

We have conducted final test to validate the system output directly to the experts. We have selected, assessed and tested 15 cases with experts. We have supplied test set with the 15 new applicant's data. The 15 new applicant's data to predict needs to have the same structure that data used to learn the model. We set the value of the class attribute is *"?"* for all instances (question marks represent missing values in Weka). The following table compares diagnostic results by expert systems and the human expert.

TABLE III. THE SYSTEM RESULTS

Cases	Determination by the System	Human Expert
1	Zakat 2	Zakat 2
2	Zakat 1	Zakat 1
3	Zakat 1	Zakat 1
4	Zakat 3	Zakat 3
5	Zakat 1	Zakat 1
6	Zakat 2	Zakat 2
7	Zakat 3	Zakat 3
8	Zakat 1	Zakat 1
9	Zakat 3	Zakat 3
10	Zakat 1	Zakat 1
11	Zakat 1	Zakat 1
12	Zakat 2	Zakat 2
13	Zakat 3	Zakat 3
14	Zakat 1	Zakat 1
15	Zakat 2	Zakat 2

The 15 cases have similar results with human expert. The *Mustahik* determination system results is according to expect.

V. CONCLUSION AND FUTURE WORK

We have successfully classified whom eligible to receive zakat (*Mustahik*). *Mustahik* determination is very important to distribute the zakat correctly. We have selected the Bayesian Network as a method to classify of *Mustahik*. Bayesian network produces promising classification average accuracy 93.24%. F-Measures has shown the capability of Bayesian network to cover an imbalanced class distribution.

Based on experiments by setting an alpha estimator's values, there is a change of accuracy. An alpha value setting in 0.6 to 1.0 can increase the accuracy of a Bayesian Network to 95.95%.

A machine-learning algorithm works based on sufficient and accurate of training data. In this study, we have used 1000 data of zakat recipients. We have used Bayesian Network as an inference engine in determining who is eligible to receive zakat. The classification data mining has many algorithms and each algorithm has its own advantages. The accuracy generated is also depending on the instance. Therefore, in future research it can be apply another algorithm in machine learning and collecting more training data accurately to produce the highest accuracy.

ACKNOWLEDGMENT

We would like to thank Badan Amil Zakat Nasional (BAZNAS) in Pekanbaru, Riau, Indonesia by provided precious data on zakat.

REFERENCES

[1] A. Al-Riyami, A. A.-H. K. Al-Amri, and K. A. Al-Busaidi, "Zakat Expert System," *Vol. One,* p. 31, 2014.

[2] L. Uusitalo, "Advantages and challenges of Bayesian networks in environmental modelling," *Ecological modelling,* vol. 203, pp. 312-318, 2007.

[3] B. G. Marcot, R. S. Holthausen, M. G. Raphael, M. M. Rowland, and M. J. Wisdom, "Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement," *Forest ecology and management,* vol. 153, pp. 29-42, 2001.

[4] F. M. de Santa Olalla, A. Domínguez, A. Artigao, C. Fabeiro, and J. Ortega, "Integrated water resources management of the hydrogeological unit "Eastern Mancha" using Bayesian belief networks," *Agricultural Water Management,* vol. 77, pp. 21-36, 2005.

[5] J. L. Ticehurst, R. A. Letcher, and D. Rissik, "Integration modelling and decision support: A case study of the Coastal Lake Assessment and Management (CLAM) Tool," *Mathematics and Computers in Simulation,* vol. 78, pp. 435-449, 2008.

[6] S. Sulaiman, H. Mohamed, M. R. M. Arshad, and U. K. Yusof, "Hajj-QAES: A knowledge-based expert system to support hajj pilgrims in decision making," in *Computer Technology and Development, 2009. ICCTD'09. International Conference on*, 2009, pp. 442-446.

[7] H. O. Maulana, I. M. Shofi, N. F. Rozy, and F. E. M. Agustin, "Application for determining mustahiq based on the priority using weight product method (case study: BAZIS DKI Jakarta)," in *Cyber*

and IT Service Management (CITSM), 2017 5th International Conference on, 2017, pp. 1-4.

[8] R. Kurniawan, A. M. Nur, R. Yendra, and A. Fudholi, "Prototype expert system using bayesian network for diagnose social illness," *Journal of Theoretical and Applied Information Technology,* vol. 93, p. 338, 2016.

[9] R. Kurniawan, Akbarizan, K. Jamal, A. Nur, M. Z. A. Nazri, And D. Kholilah, "Advise-Giving Expert Systems Based On Islamic Jurisprudence for Treating Drugs and Substance Abuse," *Journal of Theoretical & Applied Information Technology,* Vol. 96, 2018.

[10] R. Kurniawan, N. Yanti, and M. Z. Nazri, "Expert systems for self-diagnosing of eye diseases using Naïve Bayes," in *Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of*, 2014, pp. 113-116.

[11] R. R. Bouckaert, "Bayesian network classifiers in weka for version 3-5-7," *Artificial Intelligence Tools,* vol. 11, pp. 369-387, 2008.