Maternal mortality classification for health promotive in Dairi using machine learning approach

Henry Manik1,2,*, M.Fidel G.Siregar1*, R.Kintoko Rochadi1, Etti Sudaryati1, Ida Yustina1, Rika Subarniati Triyoga3

1Faculty of Public Health, Universitas Sumatera Utara, Medan, North Sumatera-20222, Indonesia
2Rumah Sakit Umum Daerah Sidikalang, Dairi, Indonesia
3Faculty of Public Health, Universitas Airlangga, Surabaya, Indonesia

Corresponding author: * henry_m2368@yahoo.com

Abstract. Reducing maternal mortality rate is a key concern of health promotion in developing countries or city face. The investigated and survey for maternal mortality had been done in Dairy City. There are 149 samples got from the survey directly in this area for 2017. In this study, we use a machine learning approach to train and test the data of maternal mortality. The aim of this study to classification maternal mortality in health promotion for reducing the maternal mortality rate in Dairi. The result of this study indicated the decision tree and Naïve Bayes are available to train and test the dataset. The accuracy of the decision tree of maternal mortality is 100 % and the Naïve Bayes model indicates 97.37 % of maternal mortality.

1. Introduction
Maternal mortality is unacceptably high in the word. Based on the word, health organization report in 2019, about 295 000 women died during and following pregnancy and childbirth in 2017 [1], [2]. The situation and the main problem faced in Indonesia show the maternal mortality rate is higher than those of neighboring countries in ASEAN. [3] report the risk of maternal death per five million is 1 in 65 mothers, compared to 1 in 1100 mothers in Thailand, or they estimate that 20000 women die from complications of pregnancy per five million birth each year. Based on this situation, Indonesia tries to solve this issue by reducing maternal mortality using education and health promotion action. To solve this issue need to mapping and survey of maternal in each city in Indonesia. The represent of maternal mortality show the risk associated with each pregnancy. Maternal mortality has become important in this case, a measure of human and social development in Dairi, North Sumatera, Indonesia. It is particularly revealing of women’s overall status, access to health care, and the responsiveness of the health care system to their needs. In this study, we collect the maternal mortality dataset from Dairi which 149 samples. Dairi regency is the area on the northwest shore of Lake Toba in North Sumatera [4], [5].
Machine learning is a perspective on modeling data with focused maternal mortality on modeling data and evolving research fields that are gaining popularity in many areas including health care, health promotion, and education [7–9]. One major health issue is maternal mortality, that needs to be worried about. Women are the pillars of any family and society but maternal death during pregnancy is a great loss to baby, family, society, and country in Dairi, Indonesia. The goal of this study to classify the maternal mortality observed in Dairi, North Sumatera, Indonesia. Classification of the maternal mortality dataset by training the data and testing the data using a machine learning approach. We run the data by using the Discussion Tree and Naïve Bayes model.

2. Data and Methods

2.1 Data

The dataset of maternal mortality had been collected in Dairi, North Sumatera from 2017 years. There is more than 149 sample in this study was recorded. From the dataset of maternal mortality is female as age productive. The health promotion in Dairi had been studied [8]. There is a positive attitude with the sample with the predisposing factor correlated with clean and healthy life behavior. Dairi is located in North Sumatera and the capital of the district is Sidikalang. The research area is located approximately 148.3 km of Medan [5].

2.2 Methods

The model to classify the maternal mortality dataset by using Decision Tree and Naïve Bayes. A decision tree is the most influential and popular tool for prediction and classification. A decision tree is a flowchart like a tree structure, where each internal node denotes a test on an attribute, each branch signifies outcomes of the test and each leaf node holds a class label. A tree can be “learned” by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions [10], [11].

Naive Bayes algorithms are naturally used in sentiment analysis, spam filtering, recommendation systems. They are fast and easy to use but their biggest drawback is that the constraint of predictors to be independent. In the real of life case, the predictors are dependent, this hinders the performance of the classifier [12]–[14]. A Naïve Bayes classifier is a probabilistic machine learning model that’s used for the classification task. The crux of the classifier is based on the Bayes theorem. The Bayes theorem can show in eq. 1.

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

From Bayes theorem, we can discovery the probability of A happening, given that B has ensued. Here, B is the evidence and A is the hypothesis. The supposition made here is that the predictors/features are independent. That is the presence of one particular feature does not affect the other. Hence it is called naïve.

3. Result and Discussions

The result of the analysis of the classification of the maternal mortality dataset for health promotion with 149 sample data in Dairi can show in Figure. 1 and Figure. 2. The visualizing of the Decision Tree result in Figure. 1 based on selected the age (maternal mortality) with the score of attitude remark in this research in Dairi, North Sumatera. In the Decision Tree model, we train the dataset to one or the other of two categories, and assigns new examples to one category or other, making it a non-probabilistic binary linear classifier. The confusion matrix result use Decision Tree show the classification rate or
accuracy is 100%, where the true positive from maternal mortality dataset is nine, the true negative is twenty-nine, false negative and false positive is nil.

**Figure. 1** Visual of decision Tree in Training Set of Maternal

The visualizing of the Decision Tree in Figure. 1 shows the result of the maternal mortality data set since in the result a hyper-plane has been found in the training set result and verified to be the best one in the set result. The decision tree result in Figure. 1 shows the separation boundary of the two classifiers. This creates a way to classify the vector or the features of the maternal mortality dataset. In this study, there are observable green dots in the yellow region and red dots in the blue region.

**Figure. 2** Visualizing of decision Tree result for testing maternal dataset
The decision result in Figure 2 shows the results of the testing of the maternal mortality dataset, which could minimize the correct prediction. The green dots in the yellow region and red dots in the blue region could be separated unless a non-linear boundary was used since the hyperplane is linear.

**Figure. 3** Visualizing of Naïve Bayes result for training maternal dataset

Figure 3 is a Naïve Bayes result of the training maternal mortality dataset, the graph shows the clear correlation between the dependent and independent variables. It is obvious that as Age (maternal mortality dataset) and Attitude score increase, each individual has a higher likelihood of being green dots. There will almost always be some degree of error - or at least there should be, otherwise, our model is probably guilty of overfitting.

**Figure. 4** Visualizing of Naïve Bayes result for testing maternal dataset
Figure 4. shows the testing result using Naïve Bayes. The graph looks very similar to the last, but with fewer observations. We can classify the age (maternal mortality dataset) and attitude score increase will lead to a higher probability of clicking the advertisement. The accuracy value of Naïve Bayes is 97.36 % when the confusion matrix indicated the true positive in maternal mortality is 8, the true negative is 29, false positive is 0 and false negative is 1.

4. Conclusions
In this study, we present the maternal mortality dataset from collected in Dairi, North Sumatra, Indonesia. Based on the result, indicated the training and testing dataset with age and attitude scores in pre-processing indicated there are positive attitudes. The classification using two models in this study is better to identify the maternal mortality dataset. The classification accuracy of the decision tree is 100 % and the Naïve Bayes Model is 97.37 %. The Decision Tree and Naïve Bayes models show good performance for train and testing the dataset of the maternal mortality in health-promoting.

References
[1] "WHO | Maternal mortality." [Online]. Available: https://www.who.int/gho/maternal_health/mortality/maternal_mortality_text/en/. [Accessed: 21-Mar-2020].
[2] "WHO | Maternal mortality," WHO, 2019.
[3] "UNFPA Indonesia | Disparity of Access & Quality Review of Maternal Mortality in Five Region in Indonesia." [Online]. Available: https://indonesia.unfpa.org/id/node/11920. [Accessed: 21-Mar-2020].
[4] Demografi. Government of Dairi Regency.
[5] "Dairi Regency - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Dairi_Regency. [Accessed: 21-Mar-2020].
[6] Purba J, Bahri S, Khair H, Darnila E, and Sinambela M, 2020 Performance of TB health center service officers based on compensation effect, supervision and competence using machine learning technique IOP Conf. Ser. Mater. Sci. Eng., vol. 725(1)
[7] Hutasabarat L T, Sunardy, Sysvia T, Dwiyanto, and Sinambela M, 2020 Machine learning approach to air traffic control skill based on mastery theory of aerodrome control procedures, self-concept and practice drills IOP Conf. Ser. Mater. Sci. Eng. 725(1)
[8] Hutasabarat L T, H. A. S, Sinambela M, and Limbong T, 2019 Classification of Student’s Air Traffic Control Skill Using 3(2), pp. 166–169.
[9] Tahmassebi A, Gandomi A H, Schulte M H J, Goudriaan A E, Foo S Y, and Meyer-Baese A, 2018 Optimized Naive-Bayes and Decision Tree Approaches for fMRI Smoking Cessation Classification Complexity.
[10] Rahmat R F, Nasution F R, Seniman, Syahputra M F, Sitompul O S 2018 Implementation of Bayesian model averaging on the weather data forecasting applications utilizing open weather map IOP Conf. Ser. Mater. Sci. Eng., 309(1), 012054.
[11] Safri Y F, Arifudin R, and Muslim M A, 2018 "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," Sci. J. Informatics, 5(1)
[12] Aravinthan K and Vanitha D M, 2016 A Novel Method For Prediction Of Heart Disease Using Naive Bayes," Int. J. Adv. Res. Trends Eng. Technol., 3(2), pp. 56–60.
[13] Kour P et al 2019 Classification of Maternal Healthcare Data using Naive Bayes Int. J. Comput. Sci. Eng., 7(3), pp. 388–394