Classification of elderly based on attitude for health promotive in Deli Serdang using machine learning approach

Masdalifa Pasaribu¹,²,*, Albiner Siagian¹*, Rahayu Lubis¹, Dewi Ramadani², Rika Subarniati Triyoga³, Etti Sudaryati¹

¹Faculty of Public Health, Universitas Sumatera Utara, Medan, North Sumatera-20222, Indonesia
²Sekolah Tinggi Ilmu Kesehatan Rumah Sakit Haji, Deli Serdang, Indonesia
³Faculty of Public Health, Universitas Airlangga, Surbaya, Indonesia

Corresponding author: * masdalifa.hajismut@gmail.com

Abstract. Health promotion is the process of enabling people to increase control over and to improve their health. Health promotion is a global issue. In the most city in Indonesia, health is improving. In this case, we collected the elderly data from Deli Serdang, North Sumatera, Indonesia. The aim of this study to classify the elderly dataset using Support Vector Machine and Logistic Regression as a machine learning approach. The result of this study indicated the SVM and Logistic Regression is available to train and test the dataset. The dominant sample in this study is female. The evaluation of the elderly dataset using two models of machine learning indicated there is a positive attitude with the age for health promotion.

1. Introduction
The situation and the main problem faced in Indonesia show undergoing a major process of social, and economical change. The global issued about health promotion makes a different situation in Indonesia. Indonesia has a large population in the world. The demographics composition of Indonesia shown great economic potential and a large population with more islands. The public health problem in a big population is more related to health promotion. The health promotion based on the [1] need not only to be built into all the policies but also if utilized efficiently will lead to positive health outcomes. Indonesia with many islands and the city is a priority for health promotion. The health promotion for the elderly is very important. In Indonesia, health promotion for many older adults is motivated to avoid processing negative information. Health promotion is meant for the entire population [2], [3], if a specific group within a population is singled out as the recipient of health promotion interventions, it’s because of valid reason, such us epidemiological concern or preferences in social policy [4]. The elderly has long been deserted as the addressee of health promotion activities. In this study, we focused on collected and observed the dataset from Deli Serdang, North Sumatera, Indonesia. The demographic of the Deli Serdang showed that the productive age (15-64 years) was greater than the non-productive age (children and elderly). The local government of Deli Serdang had been given information about the demographic bonus for a teenager in 2016 [5]. The location Deli Serdang unclose to Medan shows the population
increase per year. Based on the histories of the population in Deli Serdang, we try to collect and observe
the data in the field, for recoded the dataset of the elderly from more than 54 years. The elderly data will
be pre-processing and processing first by classify using some model in statistic approach or machine
learning. The main aim of this study to identification and classification the elderly to compare the
attitude observed in Deli Serdang, North Sumatera, Indonesia. Classification of the elderly dataset by
training the data and testing the data using a machine learning approach. We run the data by using the
Support Vector Machine and Logistic Regression.

Machine learning is a perspective on modeling data with focused model skill, and close to the statistical
learning as a perspective on modeling data. To train and test the data, in this case, we use the SVM as a
supervised learning model and logistic regression that analyze data. The classification of SVM is a
supervised machine learning algorithm that can be used for the regression problem. And the
classification of logistic regression based on the concept of probability. The classification of the elderly
for health promotion is very urgent for training and testing the data, it will be better to evaluate the data
by using some model like SVM and logistic regression.

2. Data and Methods
2.1 Data
The dataset of the elderly had been collect in Deli Serdang, North Sumatera from 2018 years. There is
more than 200 sample in this study was recorded. From the dataset of the elderly, the female is dominant
than male. There are 242 females and 40 males in these samples. The percentage of the dataset can show
in Figure. 1.

![Elderly Dataset 2018 in Deli Serdang, North Sumatera](image)

**Figure. 1** Elderly dataset 2018 in Deli Serdang, North Sumatera, Indonesia

The health promotion in Deli Serdang had been studied [6]. There is a positive attitude with the sample
with the predisposing factor associated with clean and healthy life behavior. The map of this study show
in Figure.2. Deli Serdang is located in North Sumatera and the capital of the district is Lubuk Pakam.
The research area is located approximately 30 km east of Medan.
Figure. 2 The research area for collected the elderly dataset in Deli Serdang, North Sumatera

2.2 Methods

The model to classify the elderly dataset by using SVM and Logistic Regression. SVM is one of the machine learning approaches. Other methods such as Learning Vector Quantization also can be applied[20]. Logistic regression is a supervised classification algorithm, in a classification problem, the target variable or output y can take only discrete values for a given set of features X [7], [8]. The model of logistic regression using the sigmoid function as shown in Eq. 1

\[ g(z) = \frac{1}{1 + e^{-z}} \] (1)

The logistic regression means binary which having binary target variables, but there can be two more categories of target variables that can be predicted by it [9]. SVM is a discriminative classifier formally representation of different classes in hyperplane in multidimensional space. SVM is a supervised machine learning algorithm that can be used for both regression challenges and classification [10]–[13]. The SVM mostly used in classification problems, in this case, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate [14]–[17]. SVM classification had been studied for multidiscipline field [18], [19].

3. Result and Discussions

The result of the analysis of the classification of the elderly dataset for health promotion with 282 sample data in Deli Serdang can show in Figure. 3 and Figure. 4. The visualizing of SVM result in Figure. 3 based on selected the age (elderly) with the score of attitude remark in this research in Deli Serdang, North Sumatera. In the SVM 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 visualizing of SVM in Figure 3 shows the result of the elderly dataset 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 SVM result in Figure 3 (a) shows the separation boundary of the two classifiers. This creates a way to classify the vector or the features of the elderly dataset. In this study, there are observable blue dots in the red region and red dots in the green region. The SVM result in Figure 3 (b) show the results of the testing elderly dataset, which could minimize the correct prediction. The green dots in the green region and red dots in the red region could be separated unless an non-linear boundary was used since the hyperplane is linear.

Figure 3 Visualizing SVM result for elderly dataset: a. Training set, b. Testing set

Figure 4 Visualizing Logistic Regression result: a. Training elderly dataset, b. Testing elderly dataset, c. Probability values

Figure 4 (a) is a logistic regression result of the elderly dataset, the graph shows the clear correlation between the dependent and independent variables. It is obvious that as Age (elderly 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. (b) shows the testing result using logistic regression. The graph looks very similar to the last, but with fewer observations. We can classify the age (elderly dataset) and attitude score increase will lead to a higher probability of clicking the advertisement. Figure. 4 (c) is probability value for classify the elderly dataset and create a prediction based on specified cut of value.

4. Conclusions
In this study, we present the elderly dataset from collected in Deli Serdang, 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 elderly dataset. The SVM model shows 90 percent accuracy and the logistics regression show 93 percent accuracy. Both of the SVM and logistics regression models show good performance for train and testing the dataset of the elderly in health-promoting.

Acknowledgment
We are grateful to STIKES RS. Haji and Universitas Sumatera Utara. We would like thank to Kabupaten Deli Serdang.

References
[1] Kumar S and Preetha G S, 2012 Health promotion: An effective tool for global health, Indian J. Community Med., 37(1), pp. 5–12.
[2] Sujudi, 1997 Health promotion towards the 21st century - Indonesian policy for the future, Health Promot. Int., 12(4), pp. 279–282.
[3] Delhi, N 2001 Health promotion in Indonesia, [Online]. Available: https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148. [Accessed: 09-Mar-2020].
[4] Golinowska S, Groot W, Baji P, and Pavlova M, 2016 Health promotion targeting older people, BMC Health Serv. Res., 16, no. Suppl 5, pp. 4–6.
[5] Santosa H, Mutiara E, and Juanita J, 2017 An Understanding Demographic Bonus and Its Implication among Teenagers in Deli Serdang District, vol. 1, no. PHICO 2016, pp. 404–408.
[6] Simbolon P and Sianipar C M, 2018 Predisposing Factors Associated with Health Behavior in Deli Serdang, North Sumatera, J. Heal. Promot. Behav., vol. 3, no. 2, pp. 130–138, 2018.
[7] Introduction to Logistic Regression - Towards Data Science. [Online]. Available: https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148. [Accessed: 09-Mar-2020].
[8] Understanding Logistic Regression - Tutorialspoint.dev. [Online]. Available: https://tutorialspoint.dev/language/python/understanding-logistic-regression. [Accessed: 09-Mar-2020].
[9] Hutabarat L T, H. A. S, M. Sinambela M, and Limbong T, 2019 Classification of Student ’ S Air Traffic Control Skill Using, 3(2), pp. 166–169.
[10] Hsu C W and Lin C J, 2002 A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks, 13(2), pp. 415–425.
[11] Smola A J and Schölkopf B, 2004 A tutorial on support vector regression, Statistics and Computing, vol. 14, no. 3, pp. 199–222.
[12] Meyer D, Leisch F, and Hornik K, 2003 The support vector machine under test, Neurocomputing, 55, no. 1–2, pp. 169–186.
[13] Support-vector machine - Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Support-vector_machine. [Accessed: 09-Mar-2020].
[14] Decoste D and Schölkopf B, 2002 Training invariant support vector machines, Mach. Learn., vol. 46, no. 1–3, pp. 161–190.
[15] Boser B E, Guyon I M, and Vapnik V N, 1992 Training algorithm for optimal margin classifiers,
in *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pp. 144–152.

[16] Wenzel F, Galy-Fajou T, Deutsch M, and Kloft M, 2017 Bayesian Nonlinear Support Vector Machines for Big Data, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10534 LNAI, pp. 307–322.

[17] Cortes C and Vapnik V, 1995 Support-vector networks, *Mach. Learn.*, 20(3), pp. 273–297.

[18] Ritonga A P, Adithya A R, Agustina I, and Limbong T, 2019 Classification of Labor Using Support Vector Machine in North Sumatera, vol. 04, pp. 240–243.

[19] Anwar D, Sembiring K, Tarigan K, Sinambela M, Abubar A R, and Hasibuan D, 2020 Machine learning approach for turbulence forecasting using support vector machine, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1

[20] Rahmat R F, Pulungan A F, Faza S, Budiarto R 2017 Image Classification on Ribbed Smoked Sheet using Learning Vector Quantization *Journal of Physics: Conference Series* 801(1).