Classification of treatment tuberculosis history based on machine learning techniques

by Ermatita Ermatita

PAPER · OPEN ACCESS

Classification of treatment tuberculosis history based on machine learning techniques

To cite this article: Eva Darnila *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **725** 012102

View the article online for updates and enhancements.

Classification of treatment tuberculosis history based on machine learning techniques

Eva Darnila[1,*], Mutamminul Ula[1], Mauliza[2], Iwan Pahendra[3], Ermatita[4]

[1]Department of Informatics, Universitas Malikussaleh, Lhokseumawe, Indonesia
[2]Department of Informatics System, Universitas Malikussaleh, Lhokseumawe, Indonesia
[3]Department of Electrical Engineering, Universitas Sriwijaya, Palembang, Indonesia
[4]Department of Informatics, Universitas Sriwijaya, Palembang, Indonesia

*eva.darnila@unimal.ac.id

Abstract. An important step in data Tuberculosis analysis is data exploration and representation. Tuberculosis treatment is crucial to protect the patients and it can lead to death in untreated in countries with low income. In this case, we use the machine learning technique by using Decision Tree and Random Forest for classification the tuberculosis to analysis and represented based on the treatment history. We use Tuberculosis dataset which employed from Province Aceh, Indonesia. The result indicated the performance of the designed arrangement was successful and could be used in Tuberculosis treatment analysis based on the histories in Aceh Utara and Lhoksumawe.

1. Introduction

The global enemy of TB in the word is Indonesia. In general case, the TB patient had not been notified in Indonesia. Based on the [1] Indonesia is a new active of TB case. In this case, we tried to use the TB histories in recent years in Province of Aceh, Indonesia. The healthcare has attracted much attention, which is looking for more data analytics in healthcare to relieve medical problems in medical staff age, population, people living alone, and quality of life. Data mining and forecasting play a vital role in modern social and medical fields [2]. Tuberculosis remains a major global health problem despite recent and continues progress in prevention and treatment[3].

In general, the data analysis techniques will be widely used in disease surveillance, decision-making, health management, and other fields, which focus on current intelligent medical care [4]. According to the above discussion, this paper seeks to use a machine learning technique by using Decision Tree and Random Forest to classification and represented the Tuberculosis dataset in Aceh Province based on treatment history.

2. Data and Method
2.1 Data

The Tuberculosis data were retained from Dinas Kesehatan Kabupaten Aceh Utara, Indonesia. We use tuberculosis which records of Rumah Sakit Cut Meutia, Lhoksemuawe, Indonesia for two years 2016 – 2017. The raw data at this paper is the standard metadata exchange formats of Excel.

Distribution the Tuberculosis data in Province of Aceh show in Table 1. In this table, Lhoksumawe and Aceh Utara are the wide areas which distribution of tuberculosis based on histories.

**Table 1**.Distribution Tuberculosis Data in Province of Aceh, Indonesia based on object area.

No	Regency	City
1.	Lhoksumawe	Banda Sakti, Muara Satu, Blang Mangat, Muara Dua
2.	Aceh Besar	Kuta Malaka
3.	Aceh Selatan	Meukek

2.2 Method

The Decision Tree and Random Forest analysis are a general, predictive modelling that have applications spanning a number of different areas [5]–[8]. Decision tree and Random Forest are not only constructed via an algorithmic approach that identifies way to split a data set based on different condition, but also are a non-parametric supervised learning methods use for both classification and regression task. The presented tuberculosis data for two years in 2016-2017 by using Machine learning technique Classification Model [9]–[11]. We computed tuberculosis treatment at data and observatories on python. It provides read/write support the most relevant tuberculosis and formats data [12]. We use the Machine Learning by using Decision Tree and Random Forest [13]–[16] to classification the Tuberculosis time series by Classification and to know the distribution of Tuberculosis in the regency of Province of Aceh, Indonesia.

3. Result and Discussions

Figure 1(a) and 1(b) shows the result of Tuberculosis data from machine learning using the Decision Tree on which respectively plotted in classification. The hierarchical tuberculosis prevalence rate dataset, applied in the simulation, the distribution of Tuberculosis treatment based on histories show that the new case (1) on green color is more than gave distribution (0) on red color.
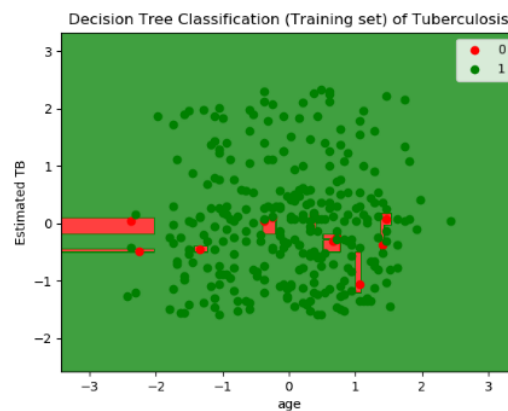
**Figure 1.** Decision Tree Classification Training Set of TB

In figure 1, the decision tree classification result as the training set of TB. The training set had been visual from making the confusion of TB histories. By using machine learning in decision tree models needs to be tested in the real TB to measure how robust its predictions are. TB histories data in the figure 1 was labeled where there is new case (1).

Figure 2. Decision Tree Classification Test Set Result

The result of decision tree classification as test set show in figure 2. The test set indicated there are sample data of TB used to provide an unbiased evaluation of final model fit on the training data set (red colors).

The term of the machine learning model on decision tree represent entire sample of TB devided into more homogeneous set. The splitting was process of deviding a node into more sub-nodes of TB and making strategics split heavily affects a tree's accuracy.
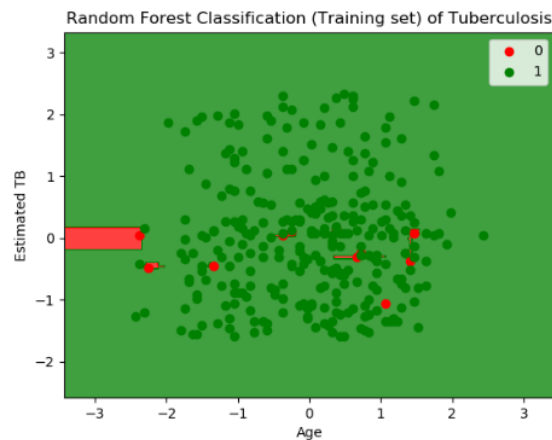
Figure 3. Random Forest Classification for Training set Result

Random forest classifiation for training set result in Figure 3 are an esemble learning for classification the TB data. The training set of TB show there are many new case show in table 1.
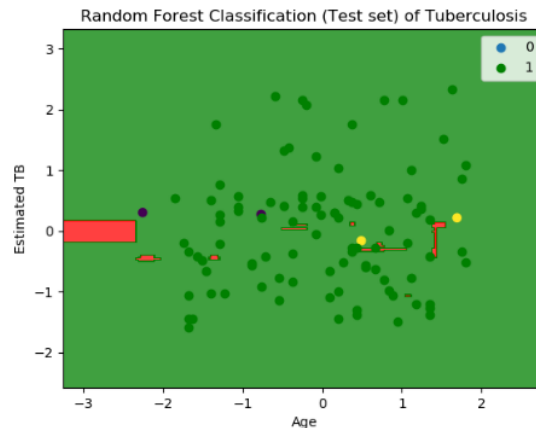

Figure 4.Random Classification as Test Set Result of TB

The random forest in figure 4 is a TB Test set result and idictaed the bias in red color. In this case, random forest and decision tree are difference, where the random forest is an ensemble clasifier which uses many decision tree model to predict the TB histories. A decision tree built using the whole consider all the TB features.

4. Conclusions

Compare with the previous studies, the classification accuracy obtained by using Decision Tree and Random Forest was better to analysis and presented the Tuberculosis treatment based on time series histories in Province of Aceh, Indonesia. The result indicated the performance of the designed system and not only was successful and could be used in Tuberculosis treatment analysis based on the histories in Aceh Utara and Lhoksuemawe has increased but also it's simple to understand the TB data and make some good interpretations.

References

[1]   D. Collins, F. Hafidz, and D. Mustikawati, "The economic burden of tuberculosis in Indonesia," *Int. J. Tuberc. Lung Dis.*, vol. 21, no. 9, pp. 1041–1048, 2017.

[2]   J. Wang, C. Wang, and W. Zhang, "Data Analysis and Forecasting of Tuberculosis Prevalence Rates for Smart Healthcare Based on a Novel Combination Model," *Appl. Sci.*, vol. 8, no. 9, p. 1693, 2018.

[3]   A. Siroka, *2017 Global Tuberculosis Report*, no. April. 2018.

[4]   Geneva: World Health Organization., "Communicable disease surveillance and response systems: Alertresponse, Epidemicpandemic," pp. 1–90, 2006.

[5]   B. M. Pereira *et al.*, "The role of point-of-care ultrasound in intra-abdominal hypertension management.," *Anaesthesiol. Intensive Ther.*, vol. 49, no. 5, pp. 373–381, 2012.

[6]   G. Louppe, "Understanding Random Forests: From Theory to Practice," no. July, 2014.

[7]   B. N. Patel, "Efficient Classification of Data Using Decision Tree," *Bonfring Int. J. Data Min.*, vol. 2, no. 1, pp. 06-12, 2012.

[8]   J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random forests and decision trees," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 272–278, 2012.

[9]   A. Yahiaoui, O. Er, and N. Yumusak, "A new method of automatic recognition for tuberculosis disease diagnosis using support vector machines," *Biomed. Res.*, vol. 28, no. 9, pp. 4208–4212, 2017.

[10]   F. H. O. Alfadhli, A. A. Mand, M. S. Sayeed, K. S. Sim, and M. Al-Shabi, "Classification of tuberculosis with SURF spatial pyramid features," *Proceeding 2017 Int. Conf. Robot. Autom. Sci. ICORAS 2017*, vol. 2018–March, no. November, pp. 1–5, 2018.

[11]   A. Gomez, "Evaluation of Machine Learning Frameworks on Finis Terrae II," *arXiv1801.04546 [cs]*, no. TECHNICAL REPORT CESGA-2017-004, p. 24, 2017.

[12]   A. Ira, A. Simbolon, and M. Pujiastuti, "Machine Learning for Handoffs Classification Based on Effective Communication History," vol. 3, no. 2, pp. 265–267, 2019.

[13]   T. B. Meier *et al.*, "Support vector machine classification and characterization of age-related reorganization of functional brain networks," *Neuroimage*, vol. 60, no. 1, pp. 601–613, 2012.

[14]   A. Sasirekha and P. G. Kumar, "Support Vector Machine For Classification of Heartbeat Time Series Data," no. 10, pp. 38–41, 2013.

[15]   K. S. Park *et al.*, "Hierarchical support vector machine based heartbeat classification using higher order statistics and hermite basis function," *Comput. Cardiol.*, vol. 35, pp. 229–232, 2008.

[16]   I. Mandal, "SVM-PSO b ased F eature S election f or I mproving M edical D iagnosis R eliability u sing M achine L earning E nsembles," pp. 267–276, 2012.

Classification of treatment tuberculosis history based on machine learning techniques

1 Eva Darnila, Mutamminul Ula, Mauliza, Iwan Pahendra, Ermatita. "Classification of treatment tuberculosis history based on machine learning techniques", IOP Conference Series: Materials Science and Engineering, 2020
Publication
71%

2 Submitted to Universitas Sumatera Utara
Student Paper
6%

3 Murali Krishna Puttagunta, S. Ravi. "Detection of Tuberculosis based on Deep Learning based methods", Journal of Physics: Conference Series, 2021
Publication
2%

4 Submitted to Universitas Diponegoro
Student Paper
2%

5 iopscience.iop.org
Internet Source
2%

6 www.researchgate.net
Internet Source
2%

7	repository.uki.ac.id	1 %
	Internet Source	

8	www.coursehero.com	1 %
	Internet Source	

Exclude quotes	On		Exclude matches	< 1%
Exclude bibliography	On			