Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 124 (2017) 142–150

Procedia Computer Science

www.elsevier.com/locate/procedia

4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia

Wiwik Anggraeni[a,*], Rafika Nurmasari[a], Edwin Riksakomara[a], Febriliyan Samopa[a], Radityo Prasetyanto Wibowo[a], Lulus Condro T.[b], Pujiadi[b]

[a]*Sepuluh Nopember Institute of Tehnology, Faculty of Information Technology, Department of Information Systems, Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia*

[b]*Malang Regency Health Service, Kepanjen Malang 65163, Indonesia*

Abstract

This paper adopted regression approach with Least Square and Natural Logarithmic transformation in response variables to predict the number of Dengue fever attacks in Malang Regency, Indonesia. The prediction involved weather factors. 8 models were prepared, and it was found that the weather factor was the most influential. Some tests, including hypothesis test, were adopted to identify the significance of the model found. The model using response variable with logarithmic natural transformation resulted better model compared to the ones without transformation. It was also supported by the average MAPE of the model that was less than 10%. Therefore, it was identified that the regression approach will work well if both dependent and independent variables have relatively similar variances so that the variability of the dependent variables can be well explained by the independent variable.

Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: Regression; Ordinary Least Square; Prediction; Dengue Fever; Natural Logarithm

1. Introduction

Dengue Fever is one of common tropical and sub-tropical diseases, specificically found in urban and sub-urban areas. According WHO (2015), Dengue Fever became one of the greatest child killers in some Asian and Latin America countries [1]. The website: www.healthmap.org presents displays Dengue Fever incident mapping found all

* Corresponding author. Tel.: +6-231-599-9944; fax: +6-231-596-4965.
E-mail address: wiwik@is.its.ac.id

1877-0509
Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017
10.1016/j.procs.2017.12.140

over Indonesia. At least 21 regencies/cities in East Java Province were under Dengue Fever Extraordinary Incident in January 2015. The total patients increased by 155.3% compared to the one in the same month in 2014 [2].

Considering the facts that such a great number of Dengue Fever incidents in Indonesia tends to present during the rainy season, the importance of utilizing weather variable for predicting the number of Dengue Fever attacks is certainly significant. WHO (2015) stated that Dengue Fever widely spreads all over tropical area, with local risk variations affected by rainfall, temperature and rapid but no well-planned urbanization [1]. Such a condition is supported by factors triggering the development of offsprings causing Dengue Fever [3]. In other countries, weather also proves to strongly affect the number of Dengue Fever incidents [4, 5, 6]. In addition, researches on relationship between weather and human health are still in progresses as it can be helpful to early detect, predict and prevent the negative effects to the human health [7]. This research predicted the number of Dengue Fever incidents in Malang Regency based on weather variable from 2009 through 2014. The method adopted was OLS (Ordinary Least Square) regression. It is reliable for predicting dependent variables having quantitative independent variables [8]. In addition, this method predicts well when the scope of the research comprehensively covers an area at large, instead of in a smaller area [9]. It is remarkable that this statistical method is capable of processing pieces of data qualifying certain criteria, and enables the researchers to set more than regression models to identify the most suitable compositions of variables in a model [9, 10, 11]. Regression approach also has advantages in general offered by a statistical method, i.e.: enabling descriptions exactly, conclusion drawing to an extend where such conclusion is correct, summary of more reliable and shorter research findings, general conclusion drawing, and predictions [9]. Regression is also a machine learning approach capable of correctly predicting output variables based on input variables [12]. Regression approach can estimate the relationship between output variable and a group of input variables by learning automatically based on a number of samples [13]. Some research methodology literatures adopting regression approach are logistic regression [14], Kriging [15], linear regression [16, 17], ALAMO [18, 19]. Besides, there is an approach called Ordinary Least Square (OLS). It is suitable to combine with regression [20] and it will be better if all variables have been normally distributed and are largely possible to have linear relationships, and the variances of the variables involved in the model are the same, balanced or slight different [9, 21]. In this research, the facts about data available were not normally distributed and the variances of the variables involved in the model were very different. Accordingly, it called for a modification so that the OLS Regression could result accurate predictions. The modification was affected to some hypothesis models and the results would be compared to the ones without modification.

Researches on Dengue Fever incident predictions have been completed. However, this concerned about prediction of Dengue Fever incidents, instead of number of Dengue Fever incidents [9]. In addition, the variables involved were metrological [22], entomological index and Dengue Fever dynamics. The researches on prediction of number of Dengue Fever patients have been conducted, yet the method adopted in those researches was Seasonal Autoregressive Moving Average (SARIMA)[11]. It was more difficult for SARIMA to define the strength of influence of independent variables to dependent variables. Modified Support Vector Machine had also used to forecast the dengue fever outbreak in Taiwan [23]. There have been previous studies that predicted the number of dengue fever incidents, but the method used was Fuzzy Inference System and the variables involved was the population only [25]. Therefore, this research adopted OLS Regression to predict the number of Dengue Fever Incidents in the area of Malang Regency that typically has different characteristics in the weather compared to the ones in the other areas all over East Java Province, Indonesia. The result of prediction of Dengue Fever incidents in Malang Regency is supposed to support the relevant health facilities to early detect the Dengue Fever endemic in that area, so that they can prepare required medicines, medical professionals, and other aspects that can curb the losses and damages due to the Dengue Fever attacks.

Specifically the objectives of the research were i) modeling the effect and relationship between the number of Dengue Fever incidents and temperature and predicting the number of Dengue Fever patients (thereafter called number of Dengue Fever incidents) based on variable of temperature, ii) modeling the effect and relationship between the number of Dengue Fever incidents and humidity and predicting the number of Dengue Fever patients based on variable of humidity, iii) modeling the effect and relationship between the number of Dengue Fever incidents and rainfall and predicting the number of Dengue Fever patients based on variable of rainfall, iv) modeling the effect and relationship between the number of Dengue Fever incidents and weather conditions comprising temperature, humidity and rainfall