ISSN: 2508-7894 © 2018 KAIA. http://www.kjai.or.kr
Doi: http://dx.doi.org/10.24225/kjai.2018.6.1.16

Development of Prediction Model for Diabetes Using Machine Learning

[1] Duck-Jin Kim, [2] Zhixuan Quan

[1] First Author Dept. of Food Service, Eulji University, Korea.
[2] Corresponding Author Dept. of Alternative Medical, Gwangju Women's University, Korea,
3700qzx@gmail.com

Received: Dec 23, 2017. Revised: Jan 02, 2018. Accepted: Jan 15, 2018.

Abstract

The development of modern information technology has increased the amount of big data about patients' information and diseases. In this study, we developed a prediction model of diabetes using the health examination data provided by the public data portal in 2016. In addition, we graphically visualized diabetes incidence by sex, age, residence area, and income level. As a result, the incidence of diabetes was different in each residence area and income level, and the probability of accurately predicting male and female was about 65%. In addition, it can be confirmed that the influence of X on male and Y on female is highly to affect diabetes. This predictive model can be used to predict the high-risk patients and low-risk patients of diabetes and to alarm the serious patients, thereby dramatically improving the re-admission rate. Ultimately it will be possible to contribute to improve public health and reduce chronic disease management cost by continuous target selection and management.

Keywords: Diabetes, Prediction Model, Machine Learning.

1. Introduction

1.1. Statistics of diabetes

According to data from the National Statistical Office, the death rate due to diabetes in 2015 was announced as ranking 6th in the Korean peninsula death cause. Diabetes is a high-risk disease due to complications, but due to the development of medicine prevention and treatment of diabetes is progressing steadily. Compared with the 2005 statistics, the number of diabetes in Korea in 2015 is decreased by one level, and it is considered to be treated. But based on the OECD (Organization for Economic Cooperation and Development) in 2012, the rate of death from diabetes in Korea was 32.3, which is higher than 9.5 persons from the OECD average of 22.8 people (Park, 2010).

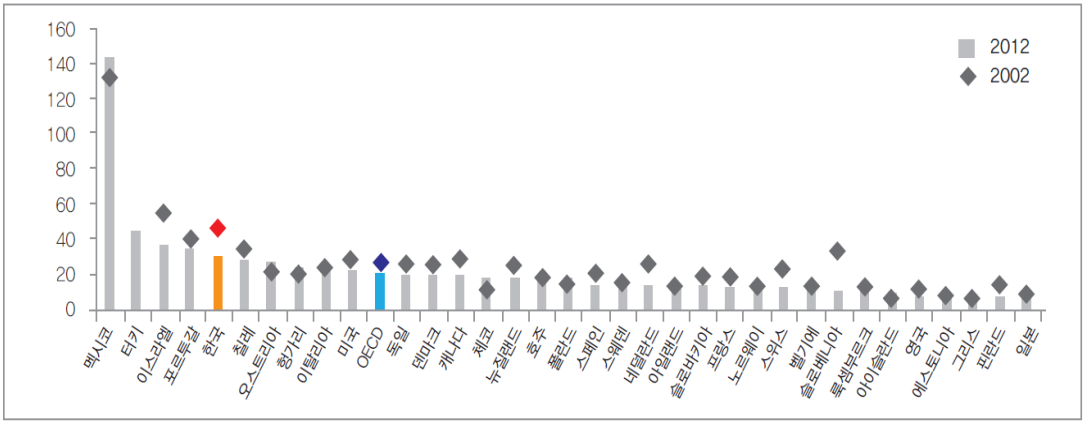

Figure 1. Diabetes Statistics by Country

According to the National Health Nutrition Survey of the 4th Phase (KNHANES) conducted from 2007 to 2009 (KNHANES), Korean adult diabetes over 30 years old, the prevalence rate is 9.9%, and it is estimated that there are about 2.8 million patients. In addition, 32% of patients in 2005 and 28% of patients in 2007-2009 were not diagnosed with diabetes. The number of patients with fasting glucose is also rapidly increasing from 17.4% in 2005 to 20.4% in 2007-2009. In the 2010 survey, it is estimated that the prevalence of diabetes is 10.1% in adults over 30 years old, there are about 3.2 million people with diabetes. Diabetes mellitus is usually accompanied by complications (Jung, 2016). Suddenly insufficient insulin in the body can lead to acute complications. Acute complications require immediate treatment with diabetic ketoacidosis and hyperglycemic hyperosmolar syndrome and are fatal if not properly treated. In extreme cases, it can lead to loss of consciousness and death. Chronic complications include microvascular complications such as retinopathy, nephropathy, neuropathy, and complications of macrovascular disease include coronary artery disease, peripheral arterial disease, and cerebrovascular disease.

1.2. Necessity of development of model for prediction of Korean diabetes

In this respect, it is necessary to develop a continuous management program and artificial intelligence program for the ultimate health promotion of the Korean people, reduction of mortality due to diabetes, and prevention and prediction of diseases. So far, many attempts have been made to screen out patients with diseases such as questionnaires and surveys, but most have been developed and tested for white Caucasian races, and only a handful of diabetes risk index models are available for Asians. It is essential to develop a risk prediction model because the risk index developed for a specific population or race is subject to great restrictions for use by other races. Also, many development models require additional blood test numerical results and complicated mathematical calculations, so it is difficult for ordinary people to access and use them. Therefore, the authors developed a self-measurement model for predicting diabetes in Korean adults and examined them, and evaluated the usefulness of the Korean model through comparison with other risk index models.

2. Development of Prediction Model for Diabetes

2.1. Analysis model

The analysis model was used the data provided by the data portal, we extracted only the diabetic code based on the main disease code and the injury code, and examined the precision of variables to be measured via people's age, residence, gender, day of treatment, number of days visited, insurance money.

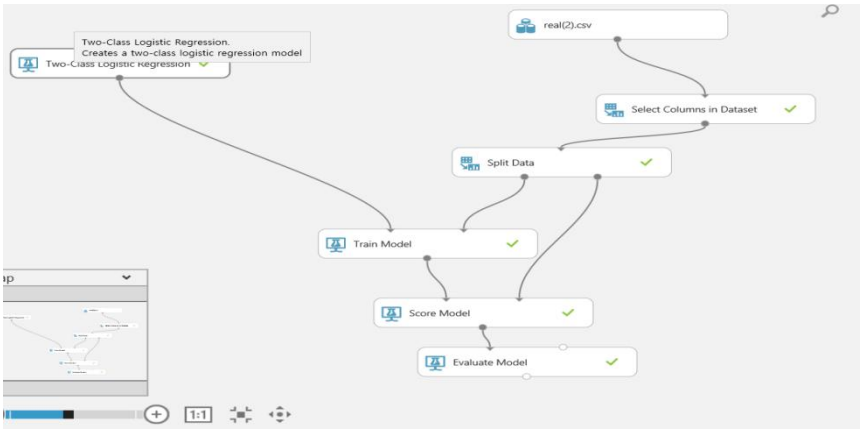

Figure 2. Learning with Two –Class Logistic Regression

2.2. Evaluation result

It was found that the probability of accurately predicting gender is 65% on average when variables of development of predictive model are taken as sex, and the probability of accurately is 85% when taken as residence.

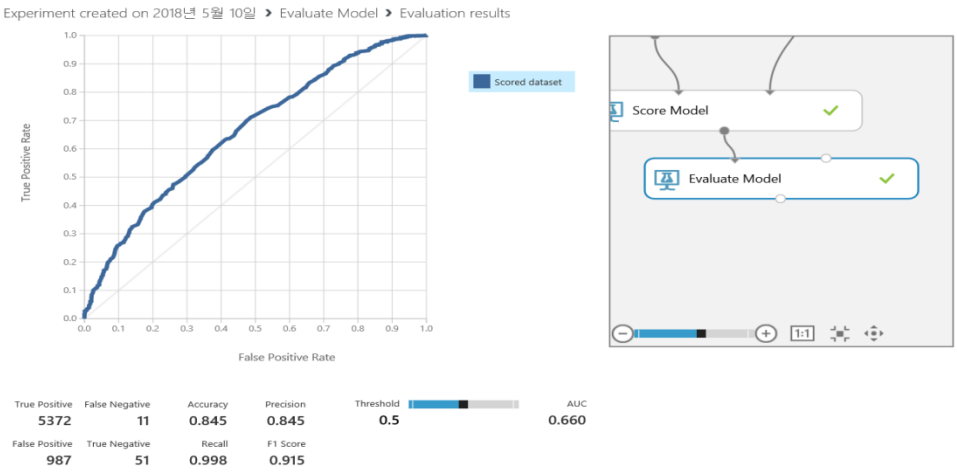

Figure 3. Logistic Regression ROC Curve

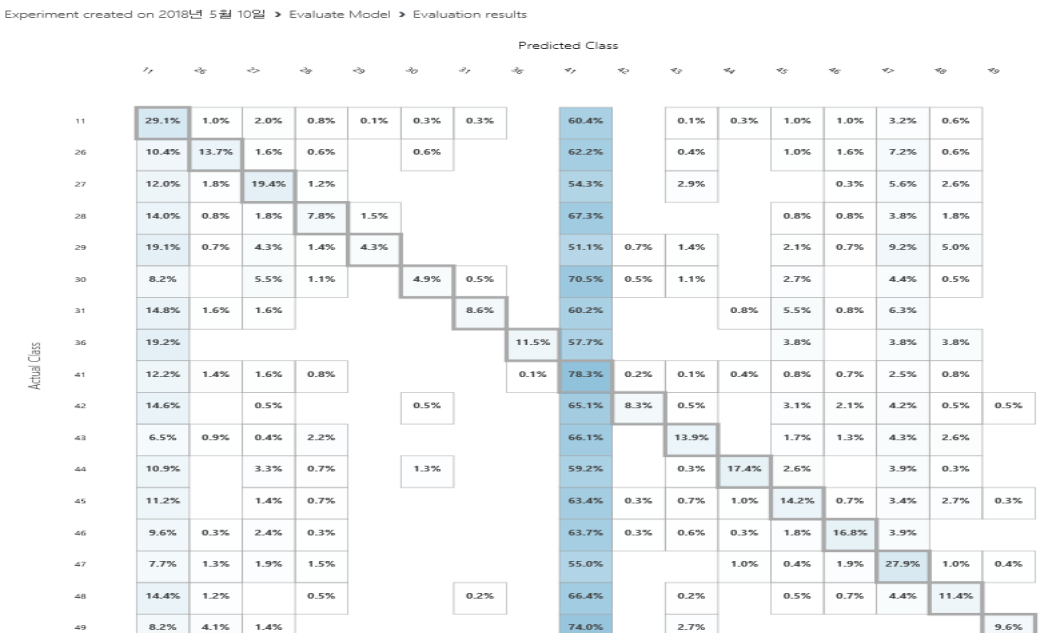

Figure 4. Predicted Value of Cites Using Multiclass Logistic Regression

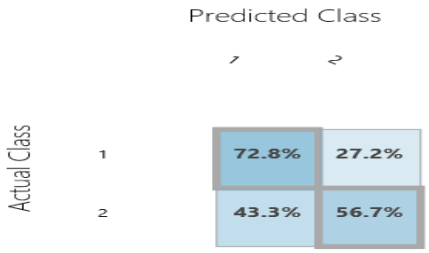

Figure 5. Predicted Value of Gender Using Multiclass Logistic Regression

3. Conclusion

Various variables can predict the risk factors of diabetes by using the above machine learning. As a result, it was found that men had a 15.1% higher risk of diabetes than women, and two - class logistic regression was more accurate than multiclass logistic regression when urban set as variable.

Although the results of previous studies have been obtained to some extent, if we study data with more reliable variables that affect diabetes and data with variables that directly or indirectly affect diabetes and conduct research in this way, It can be said that it has an important meaning in terms of building infrastructure of information technology that can develop into active healthcare business such as provision of customized health information, efficient selection of management subjects, and provision of management service.

References

Park, I. S., Han, J. T., Kang, S. B., & Ji, J. H. (2010). Developing the predictive model for stomach cancer using data mining. *Journal of the Korean Data and Information Science Society*, *21*(6), 1253-1261.

Jung, S. W. (2016). *Application of Big Data for Healthcare*. Seoul, Korea: Jeong sum.