

# **A Study on Prediction of Business Status Based on Machine Learning<sup>1</sup>**

<sup>1</sup> Ki-Pyeong Kim, <sup>2</sup> Seo-Won Song

<sup>1</sup>, First Author Professor of Daejeon University, Korea, Email: [gpkim2970@hanmail.net](mailto:gpkim2970@hanmail.net)

<sup>2</sup>, Corresponding Author Master course student, Dept. of Medical IT Marketing, Eulji University, Seongnam, Korea, Email: [songst32@gmail.com](mailto:songst32@gmail.com)

Received: September 03, 2018. Revised: December 10, 2018. Accepted: December 13, 2018.

## **Abstract**

Korea has a high proportion of self-employment. Many of them start the food business since it does not require high-techs and it is possible to start the business relatively easily compared to many others in business categories. However, the closure rate of the business is also high due to excessive competition and market saturation. Cafés and restaurants are examples of food business where the business analysis is highly important. However, for most of the people who want to start their own business, it is difficult to conduct systematic business analysis such as trade area analysis or to find information for business analysis. Therefore, in this paper, we predicted business status with simple information using Microsoft Azure Machine Learning Studio program. Experimental results showed higher performance than the number of attributes, and it is expected that this artificial intelligence model will be helpful to those who are self-employed because it can easily predict the business status. The results showed that the overall accuracy was over 60 % and the performance was high compared to the number of attributes. If this model is used, those who prepare for self-employment who are not experts in the business analysis will be able to predict the business status of stores in Seoul with simple attributes.

**Keywords:** Food Hygiene Business, Business Analysis, Machine Learning, Prediction of Business Status, Neural Network

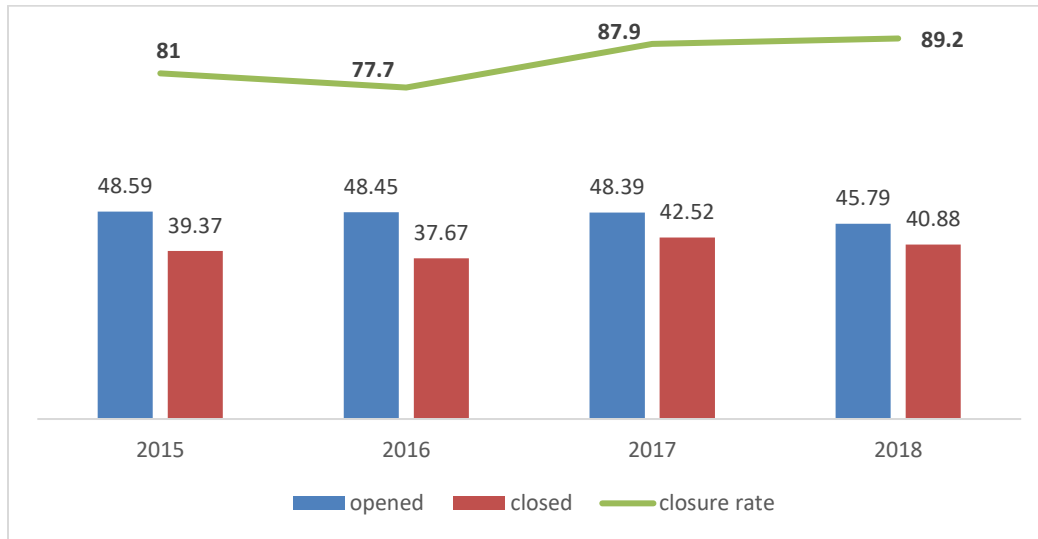
## **1. Introduction**

Korea has a high proportion of self-employment. According to data released by the Statistics Korea in 2017, the number of self-employed workers in Korea is about 5.63 million, accounting for 21% of the total employment, and occupies a large part of the national economy. However, the survival rate of the business is weak due to market saturation and excessive competition. In other words, the closure rate is high. Although it was decreased slightly in 2016, the closure rate was increased to 89.2% due to factors such as rise in the minimum wage.

Food hygiene is popular for self-employed workers because it has a lower entry barrier than other businesses because it can be started without special skills. Unemployed people choose to start a food hygiene business even they know that it is not easy to run the business industry because they have no particular alternative. In addition, consumer sentiment is shrinking due to labor costs, increase in rent, economic slowdown, and the operation of food hygiene is becoming more difficult as time goes by.

---

<sup>1</sup> This work was supported by the research grant of the KODISA Scholarship Foundation in 2018.



**Figure 1.** Closure rates of Self-employment in Korea

Food hygiene can be found around us. Typical examples include cafes and restaurants, and this type of business is the one which makes the business analysis very important. However, it is not only difficult for a start-up preparatory person to systematically conduct a business analysis, but also hard to find appropriate information. Therefore, in this paper, we applied the machine learning algorithm with the minimum data of the establishment to predict the operation status of the food hygiene store.

## 2. Theoretical background

### 2.1. Machine Learning

Machine learning is a part of artificial intelligence, focusing on representation and generalization as a field for developing algorithms and technologies that enable computers to learn. In addition, data can be analyzed using algorithms, learning through analysis, and make decisions or predictions based on learning. Machine learning is largely classified into supervised and unsupervised learning, and supervised learning includes classification algorithms, regression algorithms, and deep learning.

### 2.2. Deep Learning

Deep learning is simply a field of machine learning that teaches computers how people think. It can be defined as a part of machine learning that enables computers to learn on their own based on artificial neural networks, which are high-level abstractions that mimic the human brain. Neurons is a nerve cell in the brain, are unique cells that carry electrical signals, and the brain is a collection of neurons. These neurons are connected by one after the other, and these neurons are called neural networks, and what is artificially created is called the artificial neural network.

In other words, this artificial neural network means a network structure that is modeled after the human neural network. One neuron receives signals from other neurons and transmits signals to other neurons. It consists of the input layer, the hidden layer, and the output layer, and the processing takes place sequentially and the final result is

output. When three or more of these neural networks overlap, it is called a deep neural network (DNN), and the machine learning uses this is called deep learning.

### 3. Experiment

#### 3.1 Experiment environment

For this experiment, we used Microsoft’s Azure Machine Learning Studio and Python. MS Azure ML studio is a cloud-based service, with which users can access easily at any time and everywhere. It provides various algorithms such as classification and regression, and models can be easily created by drag and drop without complicated programming. Data preprocessing was performed using Python, and the machine learning algorithm was applied in MS Azure ML studio.

#### 3.2 Experiment data

The resting and eating house information data from the Ministry of Public Administration and Security in the Open Data Portal was used for the experiment. Table 1 shows the description of the experiment data. This data consists with store’s name, area, location, telephone number, business status, etc. Although the number of columns of the original data was more than 40, in this study, only four attributes were used to predict business status.

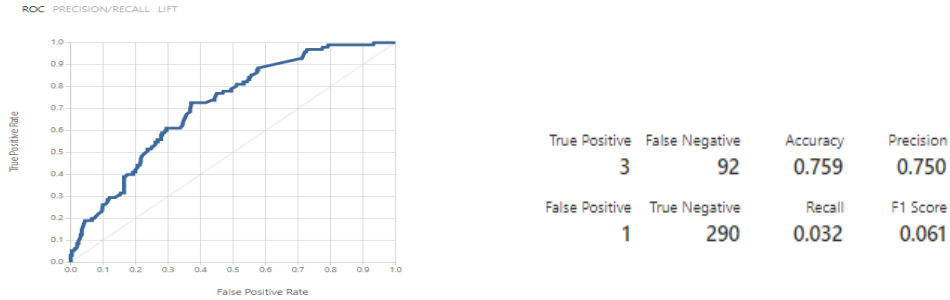
The Road Name Address was converted to "gu" and "road". Also, items with missing values were removed and experimented with a total of 56,157 data. If the number of attributes increases, the curse of dimensionality problem would occur, thus it is important to choose meaningful features. In this study, only four columns were used for predicting the business status and unnecessary attributes were removed. The business status column used for the label and the others were used making the predicting model. Table 1 shows the used data.

**Table 1.** Description of experimental data

Attributes	Description
Business Status	open / close
Type	Highway, bakery, tea room, tteok cafe, department store, ice cream, amusement park, traditional tea cafe, cafe, kids’ cafe, fast-food, convenience store, food-truck etc.
Location	All districts of Seoul city (gu)
Area	0 ~ 99.81

#### 3.3 Results

In this experiment, two-class neural network algorithm was applied to predict business status based on business type, location(gu) and location area data. The part of experiment result is shown in Figure .2



**Figure 2.** Experiment results

Figure 2 shows the result of the store’s business status in Gwangjin-gu, the accuracy was 75.9%, the precision was 75%. The performance was relatively high compared to the number of attributes.

#### 4. Conclusion

The closure rate of self-employment is increasing due to structural factors such as economic factors and intensifying competition. Korea has more self-employed workers than other countries. Related measures for the current self-employed or policies that can help prepare self-employment are needed. Among various sectors, food hygiene is a sector that attracts self-employed workers due to its relatively low entry barrier. Restaurants and convenience stores commonly found in the neighborhood belong to the food hygiene industry, and commercial and customer analysis are important. Before starting a business, it is necessary to collect, analyze, and prepare a lot of information, but it is not only difficult to do on its own but also challenging to obtain relevant information.

Therefore, in this study, an artificial intelligence algorithm was applied to easily obtainable data to create a model for predicting store business status. The food hygiene store data in Seoul, which is provided by the Open Data Portal, and user can predict the business status based on three types of information: gu, type of business, and area of the store. The results showed that the overall accuracy was over 60% and the performance was high compared to the number of attributes. If this model is used, those who prepare for self-employment who are not experts in the business analysis will be able to predict the business status of stores in Seoul with simple attributes. In addition, If the model is supplemented by adding details of each store, such as the number of employees hired and the annual operation, it is expected that higher accuracy can be obtained.

#### References

Azure Machine Learning Studio, “Azure Machine Learning Studio”, Retrieved September 01, 2018, from <https://studio.azureml.net/>

Kang, M.S., Kang, H.J., Yoo, K.B., Ihm, C.H., & Choi., E. S. (2018), *Getting started with Machine Learning using Azure Machine Learning studio*, seoul, Hanti media.

Kim, E.J. (2016). *Introduction to Artificial Intelligence, Machine Learning and Deep Learning*. Paju-si, Korea: Wikibook.

Kujira Hikouzkume. (2017), *Introduction to hands-on development of machine learning and deep learning using Python*, Paju-si, Korea: Wikibook.

National Tax Service, National Tax Information, Retrieved September 01, 2018, from [https://www.nts.go.kr/info/info\\_05.asp](https://www.nts.go.kr/info/info_05.asp)

Park, J.S. (2018, December), Income and Consumption of Self-Employed Households. Korean Social Trends 2018, Statistics Research Institute, 190-198.