March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

Deep Learning : Artificial Neural Network Retail Stores VIP Customers Prediction Study

Jong-chan Kim[1], Jae-Chul Noh[2], Sung-jun Kim[*3]
[1]Research Scholar, Innovalue Partners, 06724, SouthKorea
[2]Professor, The Department of Social Welfare, Hoseo University, 06724, SouthKorea
[*3]Professor, The Department of Big Data Industry Security, Namseoul University Graduate School, 06724, SouthKorea.
0001jmp@gmail.com[1], noh-jc@hanmail.net[2], mvstar@hanmail.net[*3]

Article Info
Volume 83
Page Number: 4257 - 4263
Publication Issue:
March - April 2020

Abstract
Background/Objective : As recent studies have begun to shed a light on an artificial neural network deep learning based on machine learning techniques, attempts to apply this technology have been increasing in various industries such as production, manufacturing, quality control, marketing, social networks and so on.
Methods/Statistical : In this study, this study analyzes sales data generated from retail stores in Gyeonggi-do, Korea (4 stores), and customer data (approximately 13,000 people) using the stores. In this study, the neural network algorithm was used to show better performance than the existing machine learning techniques. It also shows how to utilize retailer data in multiple data mining techniques using unsupervised learning.
Statistical analysis/Findings : As a result of the analysis, cluster analysis was divided into four clusters, and the model showed the best performance of the neural network model. Of the four clusters, a cluster was created that represented the characteristics of VIP customers and that cluster was modeled as a target. The group's customers were found to be mainly using the store's mobile alarm service and delivery service.
Improvements/Applications : The analysis makes it easy to identify your differentiated marketing and outlook.

Article History
Article Received: *24 July 2019*
Revised: *12 September 2019*
Accepted: *15 February 2020*
Publication: *26 March 2020*

Keywords: *Artificial Neural Network, Business problem solving methodology, K-means clustering, retail store, Customer relational management data, Data mining.*

1. Introduction

Customer Relational Management (CRM) data derived from marketing is being utilized in various areas. In a field of E-Commerce, the most data is recorded and stored on the server. Retailers and large retailers are paying attention to marketing using customer characteristics based on sales trends by combining customer purchase data with basic information. Normally, customers' purchasing history data were sampled in traditional statistics and marketing was conducted by identifying customer purchasing trends through population estimation methods. However, many studies have applied of data mining techniques more often, customers' characteristics can be learned from computers based on a whole group of data rather than estimates using samples. We can find new business insights by applying not traditional statistics methods but machine learning methods. In this study, the main techniques of data mining, supervised learning and unsupervised learning, are used to establish strategies and provide guidelines on how to apply marketing

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

practice of CRM data. In case of unsupervised learning, integrated customer data is clustered through cluster analysis, and classification models are designed to create classification models for VIP customers by utilizing neural network models that are currently in the spotlight. Chapter 2 will introduce the preceding study. Chapter 3 will introduce a methodology and data mining techniques in this study. In Chapter 4, results were derived in detail. Chapter 5 highlights the effects of used data mining techniques.

2. Methods

2.1 Business Model Design

Using ICAIS business problem solving methodology, this study aims to build a neural network model that predicts VIP customers based on customer data. The ICAIS model is divided into a total of five stages. It is an enterprise problem-solving methodology in which problems are defined, data is collected, analyzed, and improved through the ICAIS(Identify problem, Collect Data, Analyze Data, Improve, and Systemize). This is a model that has been improved based on the 6 sigma problem-solving methodology, which has been used primarily in the existing manufacturing and production sectors. Specifically, in the process of 'Identify problem', the problem is defined and data related with the problem is understood. One of the key aspects of data analysis is to identify prior information. Pre-survey gets information from media articles or external data. Prior to the data collection phase, if data is already collected, the problem can be figured out from the company's own data. After figuring out the problem, the goal is set as the Key Performance Indicator (KPI).

We prepare data to achieve KPI in the 'Collect Data' step. Data can be collected from both internal data and external data. In this study, we will analyze VIP customers prediction by collecting internal data and refining it.

In a stage of 'Analyze Data', data analysis is performed by applying machine learning algorithm based on refined data. Here, data analysis refers to the entire process of finding

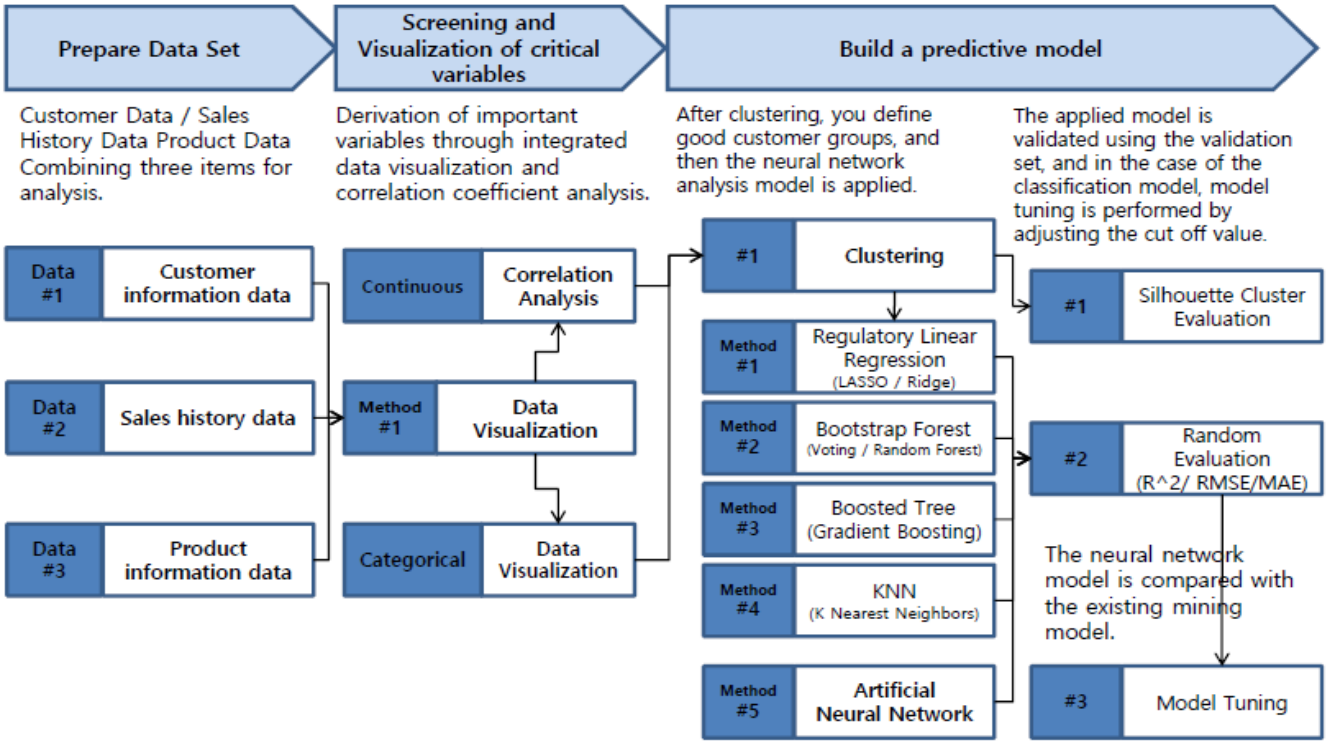

Figure 1. 'Analyze Data' Step : In a step of 'Build a predictive model'.

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

insights from data with comprehensive concepts including visualization and modeling analysis. [Figure 1] is a schematic of the 'Analyze Data' step in ICAIS business problem solving methodology. we choose a suitable analyzing model compared with ANN(Artificial Neural Network) algorithm and the other machine learning algorithms.

In a process of 'Improve' and 'Systemize', we apply the previous analyzed results. For example, a website-based platform allows users to analyze data that is collected anytime, anywhere, and to discover Insight. The most important part is to check the previous results can be applied in practice.

2.2 Data Mining

Data mining is trained to distinguish between relationships and correlations between data. In general, data is complex, with hundreds of thousands of records and hundreds of independent variables. The advantage that data mining has over traditional statistical models is that the relationship between independent and dependent variables is nonlinear and performs well on data where the specific rules of the relationship between data are unknown. [1]

Data mining is largely divided into supervised learning, unsupervised learning, and reinforcement learning. Supervised learning specifies a label that we want to predict (or classify) and makes computer algorithms learn the label. Unsupervised learning uses relationships among several features to measure data similarity or to cluster data. In this study, we use unsupervised learning to cluster the data.

2.2.1 K-Means Clustering

K-means cluster (MacQueen, 1967) is a commonly used method for dividing more than approximately 10,000 data into k groups on their own. Select k initial cluster centers and proceed as follows. [2]

1. Select an initial partition with K clusters; repeat steps 2 and 3 until cluster membership stabilizes.

2. Generate a new partition by assigning each pattern to its closest cluster center.

3. Compute new cluster centers. [3]

Cluster analysis is unsupervised learning, and the user must decide how many clusters to analyze initially. The number of clusters to be determined initially is usually referred to edge analysis and silhouette analysis. In this study, we used silhouette analysis to determine the number of clusters.

Silhouette analysis is a good analysis to use beforehand when determining the number of clusters. Calculated by Euclidean-based distance ratios, useful for finding distinct clusters. This analysis uses the average close distance between the data and is known to be best analyzed geometrically in spherical data distribution. To construct a silhouette, we need things: subgroups (created with some clustering) and the average of all distances between objects. We calculate a specific value s (i) for j as below and then graph this value. [4] The silhouette score is calculated as follows, which is an indicator for determining how close the clusters are and how far apart the clusters are.

$$S(j) = \frac{m(j) - n(j)}{\max\{m(j), n(j)\}}, \qquad \text{if } |C_j| > 1 \qquad (1)$$

2.2.2 Artificial Neural Network Classification

Neural networks have been developed to virtualize the human nervous system for machine learning by computing learning models with algorithms similar to human brain neurons.. [5]

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

There are many types of algorithms in the Artificial neural network. The CNN algorithm, which is used for image recognition, and the cyclic neural network algorithm (RNN), which is used when processing time series data or natural language data, are derived. In particular, image data analysis has made great strides, and nowadays algorithms such as GAN and LSTM are widely used. In neural network, if "Deep Neural Network" is classified and described, representatively, CNN model is distinguished from ANN. The CNN model includes layers related to pooling and convolution, so you need to use multiple layers. When multiple layers occur like this, it can be described as "Deep". This characteristic also changes the number of hidden layers. ANN contains 2 ~ 3 hidden layers, but CNN can have more than 100 hidden layers.[6] In this study, we use 'Feed forward neural networks algorithm' [7] because it uses structured data.

Feed forward neural network is a type of Artificial Neural Network(ANN) that computes data from the first layer and then sends the output using the weight of each node to the next layer. The concept of Multiple Layer Perceptron (MLP) emerged in this process. [8]. Another kind of learning is the backpropagation algorithm. This is the most used Multiple Layer Perceptron(MLP). This updates the weight between layers to minimize errors. This model works very powerfully for new data learning. However, the processing speed is slow and there is a risk of local overfitting [9]. It is very critical here to determine the number of layers and the number of nodes and the active function between them in the hidden layer. The parameters of these models are very important for determining performance. It is important to identify nodes, layers, and active functions according to data type. Through trial and error, finding the model that best fits your data is a critical capability. [10]

3. Research Method

In this study, customer data and sales data of four retail stores operated by membership system in Yongin, Gyeonggi-do, Korea, were combined and analyzed. The data is all structured data, and the data preprocessing is done in Python. In the case of sales data, a total of 706,320 data were obtained, and purchase trends were calculated based on customer ID and merged with customer data.

3.1 Analysis Object

The analysis was conducted for 13,039 customers using the Yongin retail store. These are customers who have visited four stores each for about one year (January 2018-December 2018) and have made a purchase. The collected customer data was combined with the sales data and preprocessed. The results are shown in the table below [Table1]. These data were extracted from each store's data base, received as csv files, and preprocessed using Python's Pandas library.

3.2 Analysis Tool

The analysis tool used the Python programming language, which is an open source that recently gained popularity for data analysis. This programming language gives a lot of flexibility to developers or analysts. There are many packages for doing other tasks. Python is relatively simpler to use and easier to access than other languages.[11] Python provides several libraries to make data analysis easier: the Pandas library for preprocessing data, the Matplot for visualization, the Seaborn library, and the Sklearn library for data mining. Sklearn makes it easy for users to handle many of the functions related to data mining, including supervised and unsupervised learning.

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

Table 1. Customer data collected at 4 stores in Yongin, Gyeonggi-do, Korea

Column name	Type	Explanation
ID	ID	Customer's unique ID value.
State	Categorical	Customer's status (new sign up, withdrawal, escalation)
Address	Categorical	Customer's home address
Gender	Categorical	Customer's gender
Age	Continuous	Customer's age
Total Visits	Continuous	Total visits (1 year)
Total Purchase Amount	Continuous	The total amount the customer purchased in the store
Total Purchase Quantity	Continuous	The total quantity the customer purchased in the store
Purchase Amount For One Visit	Continuous	Purchase Amount For One Visit
Visit per Week	Continuous	The number of times a customer visited a store each week.
Shipping Service	Categorical	Whether the customer applied for shipping.
Mobile Alarm	Categorical	Whether the customer will receive an alarm.

3.3 Analysis Methods

The analysis is based on Python's Skearn library. First, before the cluster analysis, silhouette analysis was performed based on customer data to determine the appropriate number of clusters.

After that, clustering was performed using K-means Clustering algorithm. After clustering, the purchasing characteristics of each cluster were identified by using a visualization technique, and the cluster with the highest sales was selected and defined as an excellent cluster. The superior cluster was set as the positive of the target variable for classification analysis.

In the classification analysis, binary classification analysis was performed using the superior cluster as the positive residual cluster and the multiple neural network model was used. The layer was composed of two layers, and the analysis was performed by putting nine active functions for each layer. (Three Gaussian, Linear, and Sigmoid functions were added for each layer.) The diagram of the model is shown below. 9 active functions are included in each of 2 layers [Figure 2].

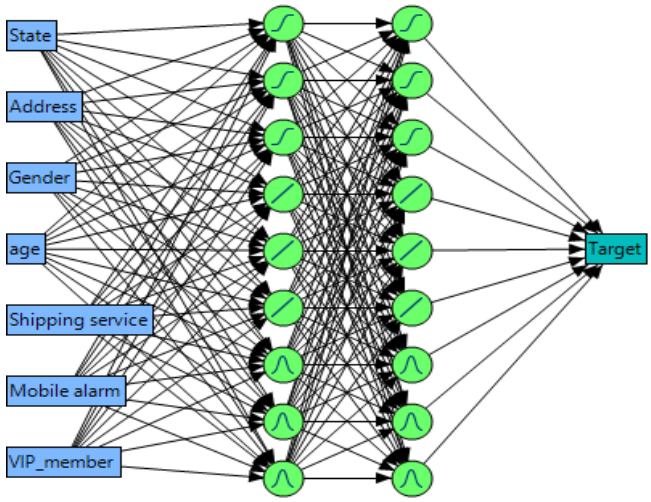

Figure 2. Artificial Neural Network (ANN) algorithm used in the study.

4. Result And Discussion

As a result of the cluster analysis, we analyze the data dividing into four clusters. For the cluster analysis, the results of the silhouette analysis are shown below [Figure 3]. The Silhouette score was higher when we created five clusters. However, when we created five clusters, the data were out of balance, so that we chose four cluster.

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

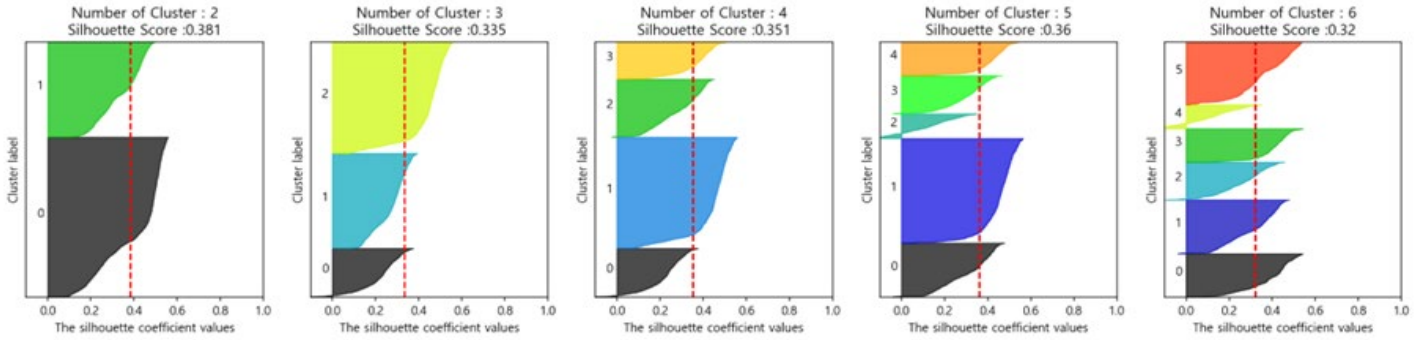

Figure 3. As a result of the silhouette analysis.

As a result of the cluster analysis results, we calculate each characteristics of clusters as belows [Table2]. Among the clusters, we define Group 3 as the Target group. We reduced all data columns to two dimensions using PCA in order to visualize the cluster analysis results [Figure 4].

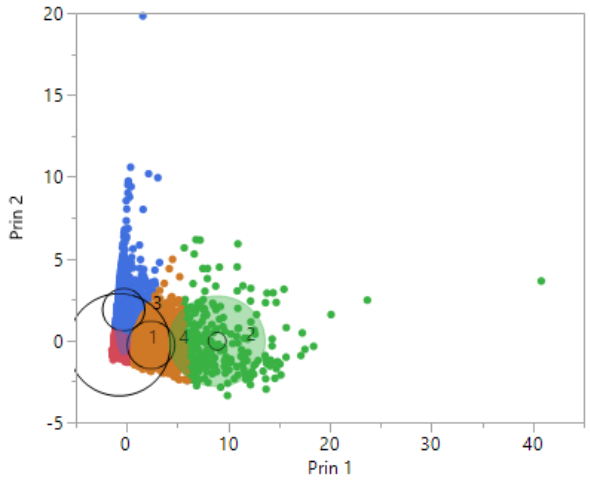

Figure 4. A total of four clusters were formed using K-Means clustering.

Based on the Target Group, we create several classification models and use four algorithms. Among the algorithms, the artificial neural network algorithm is highly recommended [Table3].

Table 2. According to the results of K-Means Clustering, we calculate Purchase Amount For One Visit and Visit per Week.

	Purchase Amount For One Visit (Unit : Won)	Visit per Week
Group 1	40370	0.17
Group 2	96665	2.75
Group 3	141432	0.14
Group 4	62170	1.14

Table 3. Classification Analysis Results by using several classification models.

Model	**RMSE (Log Like Hood)**	**MAE (BIC)**	**R Square**
Logistic Regression	1793 (Log Like Hood)	3671.67 (BIC)	0.68
Random Forest	0.139	0.069	0.80
Gradient Boosting	0.091	0.028	0.91
Multiple Neural Network	0.02	0.001	0.96

March - April 2020
ISSN: 0193-4120 Page No. 4257 - 4263

5. Conclusion

As a result of the analysis, we derived that the multi-neural network model is highly recommended. In addition, using Feature Importance and Profiler the factors that affect the model at most are 'age' and 'mobile alarm'. According to the analyzed results, the company should enforce differentiated marketing related to ‘age’. Futhermore, the company should increase the number of customers who receive news from stores immediately through mobile alarms. In addition, when the company finds the new customers, the derived model can be used to calculate the probability of changing the position from the new customers to VIP customers.

This study might be useful in retail stores where the customer's personal information is secured. We expect to find business insights not only visualizing the data but also calculating the probability of the number of new customers who willing to become VIP customers and manage the VIP customers.

Reference

[1] FRANCIS, Louis. The basics of neural networks demystified. *Contingencies*, 2001, 11.12: 56-61.

[2] Krishna, K., and M. Narasimha Murty. "Genetic K-means algorithm." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29.3 (1999): 433-439.

[3] Jain, A. K. Data clustering: 50 years beyond K-means. Pattern recognition letters, 2010, 31(8), 651-666.

[4] ROUSSEEUW, Peter J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 1987, 20: 53-65.

[5] AGGARWAL, Charu C. Neural networks and deep learning. *Springer*, 2018, 10: 978-3.

[6] KŁOSOWSKI, Grzegorz; RYMARCZYK, Tomasz. Using neural networks and deep learning algorithms in electrical impedance tomography. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, 2017, 7.

[7] Bebis, G. and M. Georgiopoulos, Feed-forward neuralnetworks. IEEE Potentials, 1994. 13(4): p. 27-31.

[8] Haykin, S.S., et al., Neural networks and learning machines. Vol. 3. 2009: Pearson Upper Saddle River.

[9] Rumelhart, D.E., G.E. Hinton, and R.J. Williams, Learning representations by back-propagating errors. nature, 1986.323(6088): p. 533.

[10] SARITAS, Mucahid Mustafa; YASAR, Ali. Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification. International Journal of Intelligent Systems and Applications in Engineering, 2019, 7.2: 88-91.

[11] LAYTON, Robert. *Learning Data Mining with Python*. Packt Publishing Ltd, 2017. p. 1. .