International Journal of Internet, Broadcasting and Communication Vol.14 No.4 228-239 (2022)
http://dx.doi.org/10.7236/IJIBC.2022.14.4.228

IJIBC 22-4-27

A Study on Prediction of Linear Relations Between Variables According to Working Characteristics Using Correlation Analysis

Seung Jae Kim

Assistant Professor, Department of Convergence, HONAM University,
Koreacdma1234@hanmail.net

Abstract

Many countries around the world using ICT technologies have various technologies to keep pace with the 4th industrial revolution, and various algorithms and systems have been developed accordingly. Among them, many industries and researchers are investing in unmanned automation systems based on AI. At the time when new technology development and algorithms are developed, decision-making by big data analysis applied to AI systems must be equipped with more sophistication. We apply, Pearson's correlation analysis is applied to six independent variables to find out the job satisfaction that office workers feel according to their job characteristics. First, a correlation coefficient is obtained to find out the degree of correlation for each variable. Second, the presence or absence of correlation for each data is verified through hypothesis testing. Third, after visualization processing using the size of the correlation coefficient, the degree of correlation between data is investigated. Fourth, the degree of correlation between variables will be verified based on the correlation coefficient obtained through the experiment and the results of the hypothesis test

Keywords: Correlation Analysis, Correlation Coefficient, Machine Learning, Classification Analysis, Data Mining

1. Introduction

The 4th Industrial Revolution was discussed at the WER(World Economic Forum) in 2016, and so far, many places using ICT technologies around the world have possessed various technologies in line with the 4th Industrial Revolution. Based on the network, studies have been conducted to explore major technologies through centrality and keyword group analysis, and to improve educational performance by analyzing the learning trends related to self-regulated learning of adult learners using CNN technology in Korea [1,2]. In addition, various algorithms and systems have been developed. A study was conducted to apply the AHP (Analytic Hierarchy Process) to experts in the agricultural R&D field in Korea [3]. At this point, the decision-making by BDA(Big Data Analysis) applied to the AI system should be equipped with more and more sophistication. In the convergence of AI technology and IT technology, studies such as content reproduction system, household waste monitoring system, emotional analysis-based psychological counseling AI chatbot, etc. were conducted in Korea [4-6]. PCA(Principal component analysis) using BDA, technology marketing

Manuscript Received: October. 25, 2022 / Revised: November. 2, 2022 / Accepted: November. 5, 2022
Corresponding Author: cdma1234@hanmail.net
Tel: +82-062-940-5639
Assistant Professor, Department of Convergence, HONAM University, Korea

application, direction of change of fintech service model, APT attack precursor analysis were conducted and AR tourism recommendation system based on character and tourism preference in Korea [7-12]. In order to make a decision with high precision and reliability, accurate classification and accurate prediction of data must be accompanied, and whether there is a relationship between the data and check the correlation.

For sophisticated decision making, first, data classification belongs to DA(Data Mining) among ML(Machine Learning), and DA(Discriminant Analysis) and classification based on data By applying CA(Classification Analysis), effective data classification is possible. Research has been conducted on the learning process of ML, SR(Speech Recognition), CV(Computer Vision), and DLA(Designing Learning Algorithms) routinely used in commercial systems for a variety of other tasks in USA [13-15]. Various studies were conducted in DM, such as establishing a win-loss prediction model for Korean professional baseball using DT(Decision Trees) in Korea [16]. There are also studies suggesting that data collection and data preposition are very important in DM in Australia [17]. For LDA(Linear Discriminant Analysis), LASSO RA(Regression Analysis) was used to effectively select variables in a situation with a small number of variables, and a study to find out what characteristics adolescents fell into game addiction was conducted in Korea [18,19]. Second, if PA(Predictive analysis) using RA is performed based on data, predicted results for the future can be obtained from the past and present. However, even in PA, if the data is not properly classified, the value obtained by PA is also unreliable. In PA, a study was conducted to predict the probability of fire occurrence when weather conditions are given using a DT in Korea [20]. Third, if the original meaning of the data classification and prediction data mentioned above is changed due to the relationship between the data, the original meaning of the data will also change. In order to remove these insecure factors, CA(Correlation Analysis), which can measure the degree of correlation between data, should be performed before data analysis to determine the degree of correlation between data. In the CA, CA on the synchronization phenomenon of the global stock market, CA between two sets, and cluster standard CA for joint dimension reduction were studied in Korea and Netherlands and USA[21-23].

In this study, before proceeding with various analyzes using the analysis data, we try to find out whether it is possible to extract statistical values suitable for the purpose of analysis from the analysis data. In the research method, CA is applied to six independent variables to find out the job satisfaction that office workers feel according to their job characteristics. There are 327 data, and we examine how each variable relates to each other. First, we examine the overall data structure for each variable. Second, a CC(Correlation Coefficient) is obtained to find out the degree of correlation with each other for each of the six pieces of data. Third, the presence or absence of correlation for each data is verified through hypothesis testing. Fourth, the degree of correlation between data will be analyzed after visualization processing using the size of the CC. Fifth, the degree of correlation between variables will be verified based on the CC obtained through the experiment and the results of hypothesis testing. This is a pre-processing process that must be performed before analyzing data and extracting the results, and it is an analysis step to derive sophisticated and reliable BDA results.

2. Correlation Analysis

CA is an analysis technique for examining the degree of relationship between variables. Correlation refers to the relationship between variables, and the index indicating the extent to which a change in one variable affects other variables is called CC. Therefore, CA shows the degree of correlation between variables, not causality.

2.1 Pearson's CA Definition

There are two types of CA: Pearson CA and Spearman CA. Pearson and Spearman are determined according to the scale of the two variables participating in the analysis. As shown in Figure 1, When performing CA, the analysis model should be set. At this time, the analysis model changes depending on the characteristics of the data to be analyzed. The characteristics of data are information possessed by the data, meaning a unique value, and there are nominal, sequence, interval, and ratio scales. (Figure 1) shows the variable relationship of Pearson's CA.

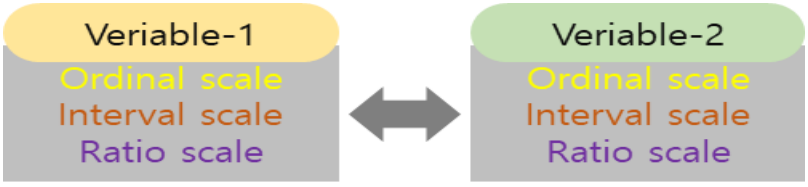

Figure 1. Variable relationship in Pearson's CA

In general, CA, which analyzes the relationship between variables rather than a causal relationship based on the temporal priority of two variables, requires that both the independent variable and the dependent variable be above the order scale.

2.2 Pearson's CA

Pearson's CA is most commonly used to determine the relationship between variables using continuous data. This method is suitable when the data have normality. In general, when two variables are above the interval scale, the degree of linear relationship is expressed using Pearson's CA. Figure 3 is Pearson's CC representing the relationship between variables, and shows the value and direction according to the degree of linearity. As shown in Figure 2, it shows the linear relationship of data according to the degree of Pearson's CC. The determination of the shape according to the linear relationship is determined depending on whether the relationship between the variables is in a positive (+) direction or a negative (-) direction based on 0. Also, the distribution of the linear relationship is determined by the size of the Pearson's CC.

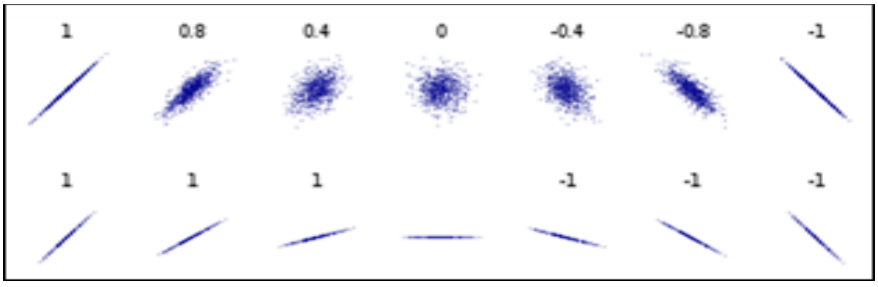

Figure 2. Pearson CC figure

2.3 Pearson CA Steps

In Pearson's CA stage, the 'CC range' indicates the degree of relationship between variables, 'covariance' to determine whether the relationship between two variables is linear, and 'significance test' for hypothesis testing.

The CC has a value from -1 to +1, and the meaning is as shown in Table 2. As shown in Table 2, the CC(r) has a value of -1 to +1($-1 \leq r \leq +1$) and the interpretation of the analyzed result is different depending on whether it is close to -1 or +1. The closer it is to -1, the more it is interpreted as a negative factor, and the closer it is to +1, the more it is interpreted as a positive factor. Also, depending on the degree of the CC, it can be

called a weak relationship, a clear relationship, or a strong relationship. If the CC is 0, it means that there is no correlation between the variables. (Figure 2) shows a linear plot drawn by the CC values in Table 1.

Table 1. Meaning of Pearson's CC

CC range	mean
if -1 < r < -0.7	strong negative linear relationship
if -0.7 < r < -0.3	distinct negative linear relationship
if -0.3 < r < -0.1	weak negative linear relationship
if -0.1 < r < +0.1	negligible linear relationship
if +0.1 < r < +0.3	weak positive linear relationship
if +0.3 < r < +0.7	distinct positive linear relationship
if +0.7 < r < +1	strong positive linear relationship

3. Experiments

3.1 Experimental method

In this study, among the CA methods, Pearson's CA method was used to investigate the degree of relationship for employee's job satisfaction. The data collection was obtained by conducting a survey, and the questionnaire was used for analysis by setting up 6 independent variables to improve the job satisfaction of office workers. The total number of data used for analysis is 327 data. As an experimental method, the tool used for analysis was coded with R program version 4.2.1 in Windows 10 environment and tested. Pearson's CA can be used only when the data type of two variables for which to check the linear relationship between variables is continuous data. As shown in Figure 3, this is the data for examining the correlation between variables for job satisfaction. Before proceeding with correlation analysis, using R programming technique, based on the data used in the experiment, the number of independent variables and the total number of data were investigated. If you show all the data, the image will get bigger, so I will show only a part of it. Figure 3 is a part of 327 data used for CA, and the file format is (*.csv) file.

	A	B	C	D	E	F
1	profession	accountat	accomplis	satisfactio	commitme	action
2	6.25	6	6.25	6	6	4
3	4.5	5.5	5.25	4.8	4	4
4	4.5	6.25	6	6.2	7	5
5	6.5	6.5	6	5.8	3.5	4
6	3.75	5	5	4.8	4.333333	3.333333
322	5	5.25	4.75	4	4.666667	4
323	4.25	4.25	4.75	3.6	5.666667	4
324	4.75	5.5	4.75	3.6	5.666667	4.333333
325	4.5	5.75	5	4	4.333333	3
326	4.5	6	4.5	5.4	5.333333	3
327	5	6.75	7	6	6.833333	4
328	4	4	4	3.8	3.333333	3

Figure 3. Job satisfaction data (327)

3.2 Data Analysis model

In order to analyze data using Pearson's CA, it is recommended to set up a data analysis model first and then analyze it. As shown in Figure 6, a data analysis model should be established. This is because if analysis is performed without a data analysis model, the form of data may be incorrectly used. In addition, it is necessary to determine whether the characteristics of the data to be set are categorical data or continuous data. This is because, depending on the characteristics of the data, Pearson CA or Spearman CA is determined. (Figure 4) shows the data analysis model of Pearson's CA. Variable 1 and Variable 2 are both continuous data.

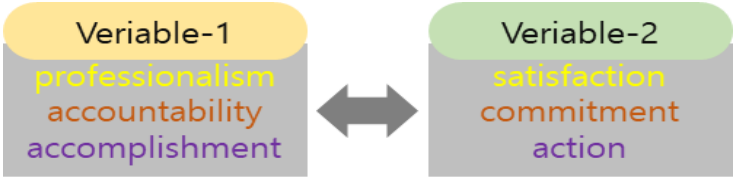

Figure 4. Data analysis model

The data analysis model procedure is as follows. First, the data characteristics for the variables should be continuous. Second, DS(Descriptive Statistics) are performed to find out the data characteristics according to each variable. Third, Covariance analysis is performed to determine the influence between variables, and correlation is identified through CC. Fourth, CA is performed on variables with continuous data characteristics. Fifth, the results of Pearson's CA are visualized according to CC, and the conclusions are drawn by plotting the results of hypothesis testing through significance tests between variables.

3.3 DS Analysis

DS is an analysis technique that enables quick diagnosis of data interpretation for all data, and calculates and shows representative values for each variable. Representative values include average, median, maximum, minimum, minimum, and number of data.

3.3.1 Data structure by DS

In each command, str shows the number of data used in the analysis and some of the values each variable has, showing that the structure of the data is 'data frame', and the total number of data is 327 with six variables. As shown in Figure 5, as a result of the R programming technique, each of the 6 variables consists of 327 data, but only the data located in the front of each variable is shown. The str(p) function is a command that displays

the structure in an easy-to-understand manner, and the structure can be identified by the str(p) function without checking the entire data. (Figure 5) shows the data structure by DS analysis.

```
> str(p)
'data.frame':   327 obs. of  6 variables:
 $ professionalism: num  6.25 4.5 4.5 6.5 3.75 5.5 4.25 4.75 4.75 7 ...
 $ accountability : num  6 5.5 6.25 6.5 5 6 5 4.25 6.75 7 ...
 $ accomplishment : num  6.25 5.25 6 6 5 5.5 4.75 4.5 6 7 ...
 $ satisfaction   : num  6 4.8 6.2 5.8 4.8 4 4.8 4.2 5.2 5.6 ...
 $ commitment     : num  6 4 7 3.5 4.33 ...
 $ action         : num  4 4 5 4 3.33 ...
```

Figure 5. Data structure

As shown in Figure 6, summary information about the entire data is displayed using the R programming technique. summary(p) shows the average value, median value, lower 25% value, upper 25% value, maximum value, and minimum value as summary information for each of the six variables. The mean of each of the six variables is professionalism (5.052), responsibility (5.468), achievement (5.342), job satisfaction (4.725), organizational commitment (4.910), and job satisfaction (3.856). It can be seen that the score scale for each item is on a 5-point scale only for job satisfaction, and all other variables are on a 7-point scale. (Figure 6) shows summary information by DS analysis.

```
> summary(p)
 professionalism accountability  accomplishment  satisfaction    commitment       action     
 Min.   :1.750   Min.   :2.750   Min.   :2.750   Min.   :2.000   Min.   :1.500   Min.   :2.000  
 1st Qu.:4.250   1st Qu.:5.000   1st Qu.:4.750   1st Qu.:4.000   1st Qu.:4.083   1st Qu.:3.667  
 Median :5.000   Median :5.500   Median :5.500   Median :4.800   Median :5.000   Median :4.000  
 Mean   :5.052   Mean   :5.468   Mean   :5.342   Mean   :4.725   Mean   :4.910   Mean   :3.856  
 3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:6.000   3rd Qu.:5.400   3rd Qu.:5.750   3rd Qu.:4.000  
 Max.   :7.000   Max.   :7.000   Max.   :7.000   Max.   :7.000   Max.   :7.000   Max.   :5.000  
```

Figure 6. Data summary

As shown in Figure 7, the data distribution of the front part and the back part is shown for the entire data using the R programming technique. The distribution of the entire data can be inferred by examining some data distributions for the front and rear parts of the entire data.In addition, the number of data to be investigated can be changed using the R program.

The head() function shows the number of data according to the setting from the top data based on each variable value. The number of data depends on the factor value of the set function, but if there is no value, five data are shown. The tail() function is the same function as the head() function, but the criteria for the data to be shown are applied from the bottom. (Figure 7) shows the upper and lower data by the head() and tail() functions.

```
> head(p)
  professionalism accountability accomplishment satisfaction commitment   action
1            6.25          6.00           6.25          6.0   6.000000 4.000000
2            4.50          5.50           5.25          4.8   4.000000 4.000000
3            4.50          6.25           6.00          6.2   7.000000 5.000000
4            6.50          6.50           6.00          5.8   3.500000 4.000000
5            3.75          5.00           5.00          4.8   4.333333 3.333333
6            5.50          6.00           5.50          4.0   4.333333 3.000000
> tail(p)
    professionalism accountability accomplishment satisfaction commitment   action
322            4.25          4.25           4.75          3.6   5.666667 4.000000
323            4.75          5.50           4.75          3.6   5.666667 4.333333
324            4.50          5.75           5.00          4.0   4.333333 3.000000
325            4.50          6.00           4.50          5.4   5.333333 3.000000
326            5.00          6.75           7.00          6.0   6.833333 4.000000
327            4.00          4.00           4.00          3.8   3.333333 3.000000
```

Figure 7. Result of head(), tail() function

3.4 CA Result

3.4.1 CC Extraction

CC are extracted to find out what kind of relationship the six variables to be subjected to CA have with each other. The degree of relationship between two variables can be defined based on the extracted CC. (Figure 8) shows that the CC is extracted.

As shown in Figure 8, CC were obtained for each variable using the R programming technique. What kind of relationship each variable has with other variables and, if so, how much relationship they have with each other, can be known through the CC. As a result of the analysis, the CC of all variables were above 0.30, indicating that there is a clear positive correlation. As such, it can be said that the correlation increases as the magnitude of the CC is greater than 0.70 or closer to +1. Since the diagonal information is each variable itself, we get 1.0.

```
> cor_p
               professionalism accountability accomplishment satisfaction commitment    action
professionalism       1.0000000      0.6597096      0.6208878    0.3745572  0.3366143 0.3950411
accountability        0.6597096      1.0000000      0.7276348    0.4676185  0.4103792 0.4991189
accomplishment        0.6208878      0.7276348      1.0000000    0.5194062  0.4432512 0.5061232
satisfaction          0.3745572      0.4676185      0.5194062    1.0000000  0.6573291 0.2908027
commitment            0.3366143      0.4103792      0.4432512    0.6573291  1.0000000 0.3048332
action                0.3950411      0.4991189      0.5061232    0.2908027  0.3048332 1.0000000
```

Figure 8. CC extraction

3.4.2 Probability of significance between variables

To find out which of the six independent variables used in the CA has an influence on job satisfaction, the significance probability values between the variables are extracted. Based on the extracted significance probability value, it is possible to evaluate whether there is an influence or not. (Figure 9) shows the extracted significance values between variables.

As shown in Figure 9, it can be seen whether each variable that can affect job satisfaction actually has an influence. Each variable is linked to job satisfaction (action) and is expressed by the p-value as a result of the analysis. From the top of Figure 14, they are 1.169e-13, 2.2e-16, 2.2e-16, 8.572e-08, and 1.848e-08.

In the analysis result of (Fig. 9), the significance probability p value between the variables was found to be less than 0.05. This proves that there is a significant correlation between job satisfaction and other variables.

The cor value representing the CC was also greater than 0.3, indicating a clear positive correlation.

```
> cor.test(action, professionalism)

        Pearson's product-moment correlation

data:  action and professionalism
t = 7.7522, df = 325, p-value = 1.169e-13
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2994110 0.4828133
sample estimates:
      cor
0.3950411
```

```
> cor.test(action, accountability)

        Pearson's product-moment correlation

data:  action and accountability
t = 10.384, df = 325, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4130185 0.5763761
sample estimates:
      cor
0.4991189
```

```
> cor.test(action, accomplishment)

        Pearson's product-moment correlation

data:  action and accomplishment
t = 10.579, df = 325, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4207617 0.5826009
sample estimates:
      cor
0.5061232
```

```
> cor.test(action, satisfaction)

        Pearson's product-moment correlation

data:  action and satisfaction
t = 5.4793, df = 325, p-value = 8.572e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1882826 0.3870536
sample estimates:
      cor
0.2908027
```

```
> cor.test(action, commitment)

        Pearson's product-moment correlation

data:  action and commitment
t = 5.7701, df = 325, p-value = 1.848e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2030891 0.4000650
sample estimates:
      cor
0.3048332
```

Figure 9. Extraction of significance values between variables

3.4.3 Visualization according to CC

Through CA, the correlation between two variables can be confirmed as numerical data, but it is difficult to visually confirm the extent of the relationship. In order to visually check the numerical data like this, it is easy to understand by using the visualization in the form of a graph. Visualization according to the CC can be expressed as a scatter plot using a point distribution and a distribution chart using the color density. At this time, it is necessary to carefully determine whether the linear relationship between the variables is linear in the clockwise direction or linear in the counterclockwise direction. If it is linear in a clockwise direction, it means a positive linear relationship, and if it is linear in a counterclockwise direction, it indicates a negative linear relationship. A positive linear relationship is interpreted as a positive aspect, and a negative linear relationship is interpreted as a negative aspect. (Figure 10) shows the correlation between two variables as a scatter plot. If you look at each scatterplot, you can see in which direction it is linear.

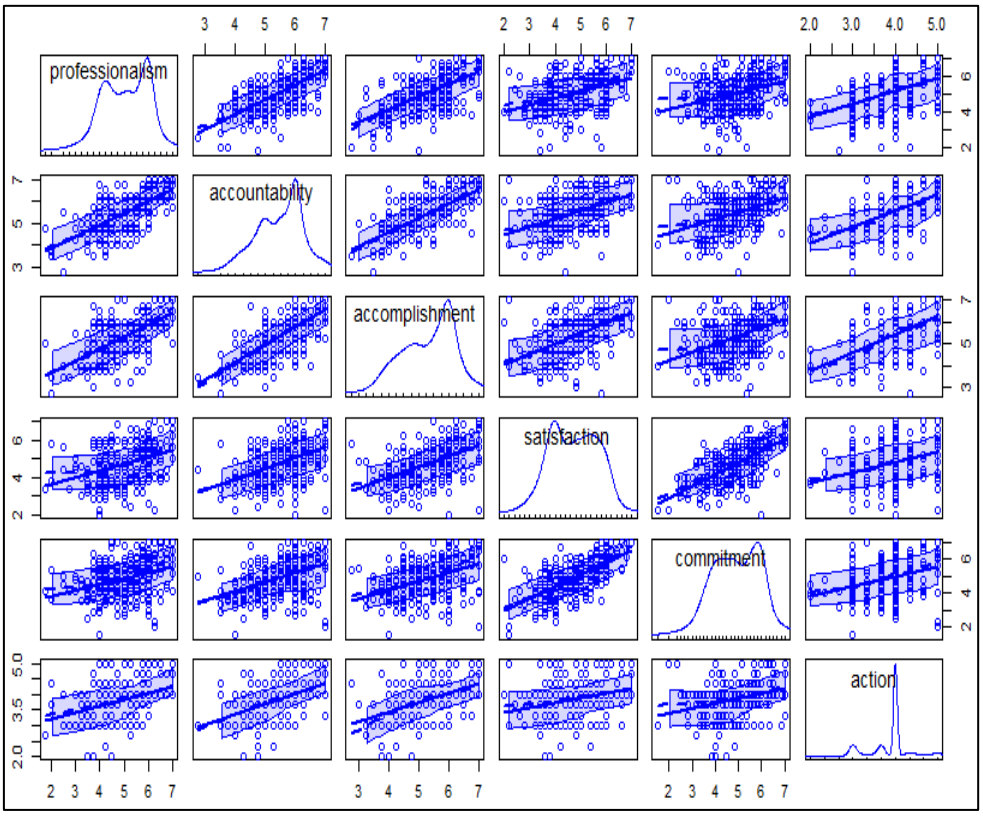

Figure 10. Scatter plot according to CC

As shown in Figure 10, the visualization shows a linear relationship between all variables. The degree of strength or weakness of a linear relationship can be checked depending on whether it is linear or not, so visually check how large the CC is. Each of the six variables showed a linear relationship using a scatterplot according to the magnitude of the CC obtained earlier. The smaller the size of the CC, the farther it is from the diagonal component (weaker influence), and the larger the size of the CC, the closer it is to the diagonal component (strong influence).

As shown in Figure 11, the linear relationship of all variables is expressed using the original distribution plot. Another visualization for examining the correlation between two variables is visualization using color intensity. (Figure 11) is a visualization that shows correlation using color density. Looking at the visualization results in (Figure 11), it indicates that the larger the circle size and the darker the color, the greater the correlation between the two variables. In the results, it can be seen through visualization that there is a strong correlation between professionalism and a sense of responsibility and achievement, and that there is a clear correlation between job satisfaction and organizational commitment.

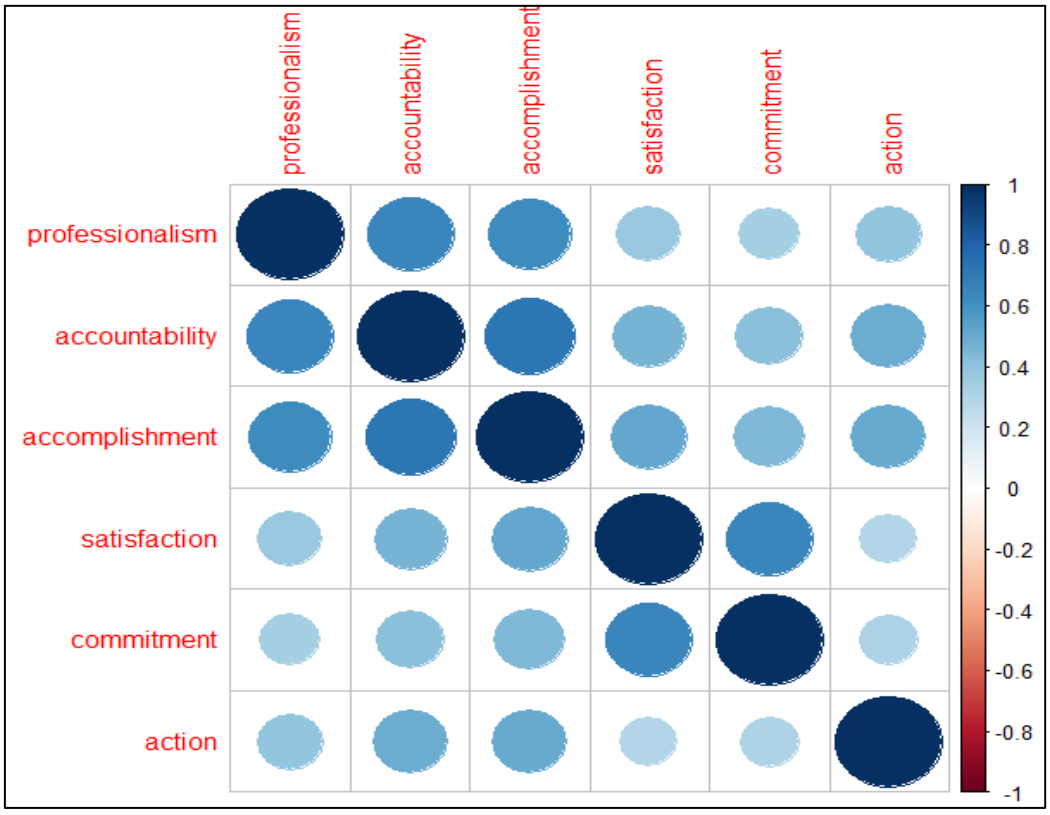

Figure 11. Distribution chart according to color density

As shown in Figure 21, CC for each variable were obtained through Pearson's CA, and what kind of influence each other had was shown using a chart. Looking at the results of Pearson's CA, the correlation between all variables showed a p-value of less than 0.05, so the null hypothesis that there is no correlation between each variable was rejected and the alternative hypothesis was adopted. Therefore, by adopting the alternative hypothesis that there is at least one correlation between each variable, it was found through the experiment that the six variables of job satisfaction had a relationship with each other and had an effect on job satisfaction. (Figure 12) is a diagram of the CC between two variables.

	professionalism	*accountability*	*accomplishment*	*satisfaction*	*commitment*	*action*
professionalism						
accountability	0.660					
accomplishment	0.621	0.728				
satisfaction	0.375	0.468	0.519			
commitment	0.337	0.410	0.443	0.657		
action	0.395	0.499	0.506	0.291	0.305	

Computed correlation used pearson-method with pairwise-deletion.

Figure 12. CC plotting

4. Conclusion

This study tries to find out whether it is possible to extract statistical values suitable for the purpose of

analysis from the analysis data before proceeding with various analyzes using the analysis data. We apply to as for the research method, 6 variables were set and 327 data were used to find out the job satisfaction of office workers. In order to find out what kind of relationship each independent variable has with each other's job satisfaction, we experimented with CA. As a result of the experiment, it was proved that the six independent variables showed a linear relationship with each other and had a clear correlation.

We are, among the six independent variables for job satisfaction, a high CC was found in professionalism, responsibility, and sense of achievement. Among them, the relationship between achievement and responsibility has the highest CC, indicating that there is a strong correlation. In the future, we plan to conduct additional CA on how the difference in job satisfaction occurs between men and women. In addition, we plan to conduct various satisfaction surveys using CA, and in addition, analyzes such as “Why did the difference occur?” and “What if there is a causal relationship?” will be continued using various analysis methods.

References

[1] S. Lee and M. Ko, “Exploring the Key Technologies on Next Production Innovation,” Journal of the Korea Convergence Society, vol. 9, no. 9, pp. 199–207, Sep 2018.
DOI: https://doi.org/10.15207/JKCS.2018.9.9.199

[2] Y. Jeong, E. Lee, and J. Do, “Development and evaluation of AI-based algorithm models for analysis of learning trends in adult learners,” Journal of The Korean Association of Information Education, vol. 25, no. 5. Korean Association of Information Education, pp. 813–824, Oct 2021.
DOI: https://doi.org/10.14352/jkaie.2021.25.5.813

[3] D. H. Lee, S. Kim, K. I. Jang, T. Sa, and D. Yoo, “Study on Agricultural Science Convergence R&D Agenda under the Fourth Industry Revolution,” The Journal of the Korea Contents Association, vol. 19, no. 7, pp. 323–334, 2019(Jul).
DOI: https://doi.org/10.5392/JKCA.2019.19.07.323

[4] S. Yang and Y. S. Lee, “Study on AI-based content reproduction system using movie contents,” Journal of Korea Multimedia Society, vol. 24, no. 2, pp. 336–343, Feb 2021.
DOI: https://doi.org/10.9717/kmms.2020.24.2.336

[5] S. H. Kim, Y. H. Kang, and D. H. Yoon, “Implementation of Monitoring System of the Living Waste based on Artificial Intelligence and IoT,” Journal of IKEEE, vol. 24, no. 1, pp. 302–310, Mar 2020 .
DOI: https://doi.org/10.7471/ikeee.2020.24.1.302

[6] S. H. An and O. R. Jeong, “A Study on the Psychological Counseling AI Chatbot System based on Sentiment Analysis,” Journal of Information Technology Services, vol. 20, no. 3, pp. 75–86, Jun 2021.
DOI: https://doi.org/ 10.9716/KITS.2021.20.3.075

[7] S. J. Lee, “Big Data Analysis Using Principal Component Analysis,” Journal of Korean Institute of Intelligent Systems, vol. 25, no. 6. Korean Institute of Intelligent Systems, pp. 592–599, Dec 2015.
DOI: https://doi.org/10.5391/JKIIS.2015.25.6.592

[8] C. N. Jun and I. W. Seo, “Analyzing the Bigdata for Practical Using into Technology Marketing : Focusing on the Potential Buyer Extraction,” Korean Strategic Marketing Association, Vol. 21, No. 2(58), pp. 181-203. Jun 2013.
UCI : G704-001657.2013.21.2.008

[9] M. S. Suh and D. H. Kim, “A Study on the Changing Direction of FinTech Service Model based on Big Data,” Global e-Business Association, Vol. 20, No. 2, pp. 195-213, Apr 2019.
DOI: https://doi.org/10.20462/TeBS.2019.4.20.2.195

[10] C. Choi and D. Park, “The Analysis of the APT Prelude by Big Data Analytics,” Journal of the Korea Institute of Information and Communication Engineering, vol. 20, no. 6, pp. 1129–1135, Jun 2016.
DOI: httpss://doi.org/10.6109/jkiice.2016.20.6.1129

[11] H. S. Lee, E. A. Kwak and D. S. Han, “A Study on Factors Affecting Avoidance of Based on Big Data AI Retargeting Advertising”, Korean Association for Advertising and Public Relations, Advertising Research (120),

pp. 80-111, Mar 2019.
DOI: http://dx.doi.org/10.16914/ar.2019.120.80

[12] In-Seon Kim, Chi-Seo Jeong, Tea-Won Jung, Jin-Kyu Kang and Kye-Dong Jung, "AR Tourism Recommendation System Based on Character-Based Tourism Preference Using Big Data", international Journal of Internet, Broadcasting and Communication. Vol. 13. NO. 1. pp. 61-68. 2021.
DOI: https://dx.doi.org/10.7236/IJIBC.2021.13.1.61

[13] T. M. Mitchell, "The discipline of machine learning," Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.
http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf

[14] Albertsson, Kim and et al, "Machine learning in high energy physics community white paper." Journal of Physics: Conference Series. Vol. 1085. No. 2. IOP Publishing, 2018.
DOI: https://doi.org/10.1088/1742-6596/1085/2/022008

[15] Mahesh and Batta, "Machine learning algorithms-a review," International Journal of Science and Research (IJSR), Vol. 9, Issus. 1. pp. 381-386. Jan 2020.
DOI: https://doi.org/10.21275/ART20203995

[16] Y. H. Oh, H. Kim, J. S. Yun and J. S. Lee, "Using Data Mining Techniques to Predict Win-Loss in Korean Professional Baseball Games," Journal of the Korean Institute of Industrial Engineers(KIIE), Vol. 40, No. 1, pp. 8-17, Feb 2014.
DOI: http://dx.doi.org/10.7232/JKIIE.2014.40.1.008

[17] Zhang, Shichao, Chengqi Zhang, and Qiang Yang, "Data preparation for data mining," Applied artificial intelligence 17. 5-6. pp. 375-381. 2003.
DOI: https://doi.org/10.1080/713827180

[18] J. W. Hwa and C. Y. Park, "Variable Selection in Linear Discriminant Analysis", Journal of The Korean Data Analysis Society(JKDAS), Vol.11, No.1 (B), pp. 381-389, Feb 2009.
UCI : G704-000930.2009.11.1.020

[19] J. H. Kwon and E. H. Lee, "Predicting Game Addiction in Adolescents: An Application of Discriminant Function Analysis", THE KOREAN JOURNAL OF HEALTH PSYCHOLOGY, The Korean Psychological Association, Vol. 10, NO. 1, pp. 95-112, Mar 2005.
https://kiss.kstudy.com/thesis/thesis-view.asp?key=2434958

[20] Y. J. Kim, J. W. Ryu, W. M. Song and M. W. Kim, "Fire Probability Prediction Based on Weather Information Using Decision Tree", Journal of KIISE, JOK: software and application", Vol. 40, No. 11, Nov 2013.
UCI: G704-E00398.2013.40.11.003

[21] Choi and Jeong-Il, "Synchronization Phenomenon and Correlation Analysis of Global Stock Market," The Journal of the Korea Contents Association, vol. 16, no. 1, pp. 699–707, Jan 2016.
DOI: https://doi.org/10.5392/JKCA.2016.16.01.699

[22] Weenink, David, "Canonical correlation analysis," Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam. Vol. 25. Amsterdam: University of Amsterdam, 2003.
http://graphics.stanford.edu/courses/cs233-18-spring/ReferencedPapers/CCA_Weenik.pdf

[23] Rasiwasia, Nikhil and et al, "Cluster canonical correlation analysis," Artificial intelligence and statistics. PMLR, 2014.
DOI: https://doi.org/10.7232/iems.2012.11.2.134