

A study of creative human judgment through the application of machine learning algorithms and feature selection algorithms

Yong Jun Kim*, Jung Min Park**

*Tomato system, Seoul, Korea,
**Professor, Tech university of Korea

*yjkim615@gmail.com, jmpark@tukora.ac.kr

Abstract

In this study, there are many difficulties in defining and judging creative people because there is no systematic analysis method using accurate standards or numerical values. Analyze and judge whether In the previous study, A study on the application of rule success cases through machine learning algorithm extraction, a case study was conducted to help verify or confirm the psychological personality test and aptitude test. We proposed a solution to a research problem in psychology using machine learning algorithms, Data Mining's Cross Industry Standard Process for Data Mining, and CRISP-DM, which were used in previous studies. After that, this study proposes a solution that helps to judge creative people by applying the feature selection algorithm. In this study, the accuracy was found by using seven feature selection algorithms, and by selecting the feature group classified by the feature selection algorithms, and the result of deriving the classification result with the highest feature obtained through the support vector machine algorithm was obtained.

Keywords: Big 5, Data mining, Machine Learning, Select Attributes, Case Study

1. Introduction

With the continuous development of the computer field, computers are taking over the role in many fields that require humans. Various studies are underway, such as a study on the relationship between stress and personality ^[1], a study on the relationship between interpersonal relationships and personality ^[2], and a study that personality is also related to learning and eating habits ^[3]. Therefore, for modern people, curiosity about personality is being applied to work and human relationships. And while many workplaces need someone who understands and implements new knowledge well, more and more companies want to create new knowledge creatively. As the American psychologist Guilford once said, "A person with creative thinking is a person who can come up with novel and new answers or ideas." We defined these people as creative people. Although the number of companies with the slogan of cultivating creative talents is increasing, there is no systematic classification and analysis method through accurate standards or numerical values for the classification of

Manuscript Received: March. 18, 2022 / Revised: March. 22, 2022 / Accepted: March. 25, 2022

Corresponding Author: jmpark@tukorea.ac.kr

Tel:*** - **** - ****

Professor, Tech university of Korea

creative talents. A study on the application of regular success stories through machine learning algorithm extraction using CRISP-DM^[4], an industry standard process for data mining, with case studies helpful in verifying or confirming psychological personality tests and aptitude tests. A solution to a research problem in psychology was proposed using machine learning algorithms and data mining. It finds similar studies through related studies and draws conclusions after showing results through experiments. In this study, we intend to propose a solution that helps to judge creative people by applying a feature selection algorithm.

2. Related Research & Techniques

2.1 Related Research

2.1.1 A study on the application of rule success cases through machine learning algorithm extraction.

In previous studies, we created a set for applying data artificial intelligence and data mining techniques based on questionnaires obtained through actual surveys, and performed clustering using the K-means algorithm, an unsupervised learning algorithm that is not specialized knowledge. Through analysis, similar groups were classified, and rules were created using the random forest algorithm, a supervised learning algorithm, to find out the characteristics of similar groups. By creating a conceptual model, a survey was conducted with graduate students, and evaluation and validation through case studies did the result was a mapping between undergraduate and graduate students^[5].

2.1.2 Automatic Classification of Advertising Restaurant Blogs Using Machine Learning Techniques

In the study of Jae-Young Jang, Byung-Jun Lee, Se-Jin Cho, Ha Ha-Hye, and Kyu-Hong Lee, advertisement blogs were collected among restaurant blogs and common features were analyzed, and classification accuracy was achieved using an automatic classification algorithm using a naive Bayes classification algorithm and a neural network classification algorithm. was tested and the optimal algorithm and feature combination was searched^[6].

2.1.3. Personality classification based on Twitter text using Naive Bayes, KNN and SVM.

In Pratama's study^[7], the responses and posts of questionnaires were analyzed in social media to analyze and classify human personality. He constructed text data sets through data collection processes from Twitter and Facebook. He used naive Bayes, K-nearest neighbor algorithm, and support vector machine, among which naive Bayes classification showed the best results.

2.2 Techniques

2.2.1 Big Five Personality Traits

Big Five Personality Traits are the five major factors or dimensions of personality traits established through empirical research and research in psychology. There are five factors: neuroticism, extraversion, affinity, conscientiousness, and openness to experience. The model compiled by Costa & McCrae has been validated in various countries. This is the most widely accepted personality theory in modern psychology. Numerous studies have shown that the personality five factor theory is not only related to personal happiness, physical and mental health, religiousness, and identity, but also to various relational consequences among family, friends, and lovers, job choice, job satisfaction, performance, social participation, criminal behavior, it has been found to predict factors such as political position well. A test based on this theory is the NEO-PI-R personality test. This Big-5 model has reliability and validity in various data.

3. Description and application of the proposed model

Existing research suggests solutions to research problems in psychology using artificial intelligence algorithms and Data Mining's Cross Industry Standard Process for Data Mining, CRISP-DM in the field of psychology. First, in the business understanding stage, based on the introduction of the existing thesis, through understanding and analysis of the Big 5 personality analysis in the Department of Psychology, problems that can be solved using data mining tools are defined, and the purpose of the study is classified. Second, in the data understanding stage, data collected through surveys are extracted and stored as Big 5 analysis and classification criteria. In the third data preparation stage, the data extracted from the data understanding stage is newly constructed and built into a data set for use by data mining tools. In the fourth modeling phase, the data mining tool is used to apply and refine the data set using the clustering algorithm, and the source code is examined and applied to the data mining tool to use the rules. In the fifth evaluation phase, the K- Analyze and evaluate using source code extracted random forest rules to determine the similarity and characteristics of clustered groups using the means algorithm. In the final development phase, we discuss how to use group analysis and rules to transform data sets more systematically and continuously improve, test, and validate the algorithms used.

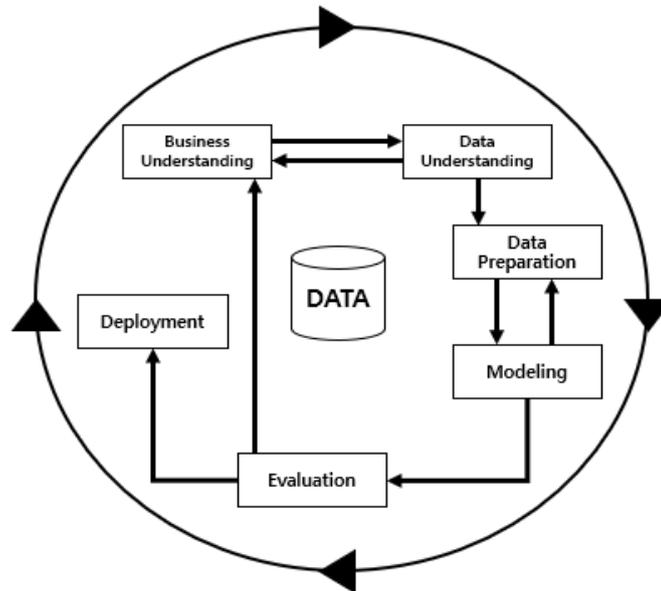


Figure 1. Cross Industry Standard Process for Data Mining, CRISP-DM

The feature selection algorithms used in this paper were Ant Algorithms, Best First Algorithm, Cuckoo Search Algorithms, Elephant Search Algorithms, Genetic Algorithms, Greedy Step Wise Algorithms, and Ranker Algorithms. As a classification method to verify the results derived through the feature selection algorithm, the machine learning algorithms, Naive-Bayes Algorithms, Multilayer Perceptron Algorithms (MLP), Support Vector Machine Algorithms (SVM), and Decision Tree Algorithms were used for verification.

4. Results of the Experiment

First, we analyze the data set using WEKA, observe what characteristics of creative people are, and select attributes that are relatively related to class labels and have high priority. As part of the creation of a new data set, only high-priority attributes are selected and refined into a new data set. As a result of data set analysis, which is the first step in data mining, the property "Is original, comes up with new ideas" in the Big 5

personality trait data set is defined as a class label. Next, each data is listed for this class label, and only the major attributes are selected from the remaining 43 attributes. In this process, 10 to 20 attributes are extracted using the 7 feature selection algorithms described above, respectively, and then a new data set is created individually. Next, the new data set created in this way is applied to the three algorithms described above and verified through 10-fold cross validation.

	Naïve Bayes	Multi-Layer Perceptron	Support Vector Machine	Decision Tree
Ant	66%	63%	69%	61%
BestFirst	68%	65%	69%	62%
Cuckoo	69%	66%	71%	61%
Elephant	66%	64%	70%	59%
Genetic	65%	63%	68%	60%
Greedy	68%	65%	69%	62%
Ranker	62%	64%	70%	56%

Table 1. Feature Selection Algorithm Accuracy

Table 1 shows the accuracy of the results of the seven feature selection algorithms based on the class label, where “Is Original, comes up with new idea” in the data set derived from the Big-5 survey is the standard for creative people.

Multiples	Property	Multiples	property
10 (100%)	1 1	0 (0%)	23 23
0 (0%)	2 2	0 (0%)	24 24
4 (40%)	3 3	10 (100%)	25 25
0 (0%)	4 4	10 (100%)	26 26
0 (0%)	5 5	0 (0%)	27 27
0 (0%)	6 6	0 (0%)	28 28
0 (0%)	7 7	0 (0%)	29 29
0 (0%)	8 8	5 (50%)	30 30
10 (100%)	9 9	0 (0%)	31 31
5 (50%)	10 10	0 (0%)	32 32
7 (70%)	11 11	10 (100%)	33 33
0 (0%)	12 12	0 (0%)	34 34
0 (0%)	13 13	0 (0%)	35 35
0 (0%)	14 14	0 (0%)	36 36
10 (100%)	15 15	0 (0%)	37 37
0 (0%)	16 16	0 (0%)	38 38
0 (0%)	17 17	0 (0%)	39 39
0 (0%)	18 18	9 (90%)	40 40
0 (0%)	19 19	0 (0%)	41 41
10 (100%)	20 20	0 (0%)	42 42
0 (0%)	21 21	0 (0%)	43 43
0 (0%)	22 22	0 (0%)	44 44

Table 2. Cuckoo Algorithm results

As a result of performing the four algorithms, it can be seen that the values of MLP and Decision-Tree showed the lowest accuracy, and SVM showed the highest accuracy. In other words, it can be determined that

the algorithm is more accurate than other algorithms. Input the values of the questionnaire made by hand with a total of 44 questions, and SVM and Cuckoo Algorithm are applied to show the selection results for the feature group. In Table 2, the shaded areas mean the characteristics related to the creative person's class label, and the percentage shows how related they are. Since 10-fold cross-validation was used in the verification process, it shows that the final feature group was created by selecting up to 10 to at least 3 related features out of 10 cases.

Question	Contents	Big 5
1	I like to talk.	E
3	I try to be meticulous in anything I do.	C
9	I am comfortable with others and handle stress well.	N
10	I am a curious person.	O
11	I am a person full of energy.	E
15	I am a person who thinks deeply.	O
20	I am a person with a lively imagination.	O
25	I am a person who likes to invent.	O
26	I am an active person.	R
30	I am a person who values artistic and aesthetic experiences.	O
33	I am an efficient person.	C
40	I like to share my thoughts with others.	O

Table 3. Big 5 results with relevant personality

Finally, Table 3 provides explanations of the items that are highly related to creative people derived from the cuckoo algorithm. We have classified a total of 12, and you can see what similar personalities people with creative personalities have. And it can be judged that a creative person is a person who corresponds to O in the Big 5.

When applied to Big 5, when the items classified by applying to Big 5 Inventory are summarized, the results shown on the far right of Table 3 indicate it. Among the Big 5, OCEAN has the most O's, which means creative people represent Openness to Experience. More specifically, it relates to diversity due to individual psychology and experience, and features such as intelligence, imagination, breaking of stereotypes, interest in aesthetics and desire for diversity, and characteristics related to dignity. As a result, through this study, judgment on the characteristics of people with creative personalities and similar personalities. And the judgment and classification through the algorithm were successful.

5. Conclusion

This study refines the data set for applying artificial intelligence and data mining techniques based on the questionnaire obtained through the actual survey and the data set directly collected and uses a machine learning algorithm and feature extraction algorithm based on objective evaluation criteria to classify and compare. In Kim Wan-seop's study^[8], an in-depth analysis technique applied with data mining was presented, and the C5.0 decision tree classification algorithm supported by the data mining tool called Clementine of SPSS was used as the classification method. Analysis of questionnaire data was performed. As a result, the correlation between grades and other items could be analyzed hierarchically. In my thesis, I experimented to make more accurate

judgments using WEKA, a data mining tool, and machine learning algorithms. Designate a class label with "I'm original and like to suggest new things" as an attribute that indicates that I'm a creative person, The remaining properties were set as factors and analyzed, and the accuracy was verified using the feature selection algorithms Ant Algorithm, Best-First Algorithm, Cuckoo Search Algorithm, Elephant Algorithm, Genetic Algorithm, Greedy Step Wise Algorithm, and Ranker Algorithm, And, by using three machine learning algorithms: Naive Bayes, Multi-Layer Perceptron, and Support Vector Machine, the features produced through the support vector machine yielded the highest accuracy and classification results. The reason why it performs better than other classifiers is that the data set of this study contains more data that are related to each other than various data extracted from Survey as attributes. It can be judged that these characteristics affected the performance ranking. If it is a small data set that is diverse or unrelated, naturally, it would have performed better when classified with the basic and fast Naive Bayes, if it was a larger data set than it is now, which is diverse and unrelated, I think that the classifier using the Multi-Layer Perceptron would have the best performance.

For future research, we plan to develop and analyze a system that can be connected, such as applications to which this study can be applied, used for classification and analysis using the Big 5 personality trait technique. In addition, it is expected that various results can be analyzed using not only the machine learning algorithms applied in this study, but also other algorithms such as clustering techniques.

References

- [1] Guilford, Joy Paul. "Three faces of intellect." *American psychologist* 14.8 (1959): 469.
- [2] Provost, Foster, and Tom Fawcett. "Data Science for Business: what you need to know about data mining and data-analytic thinking." O'Reilly media inc. 2013
- [3] Wang, Rui, et al. "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones." *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, 2014.
- [4] Shearer, Colin. "The CRISP-DM model: the new blueprint for data mining." *Journal of data warehousing* 5.4 (2000): 13-22.
- [5] Yong-Jun Kim, "A study on Big-5 based Personality Analysis through Analysis and Comparison of Machine Learning Algorithm." *The Journal of the Institute of Internet, Broadcasting and Communication (IIBC)*, Vol.19, No.4, pp.169-174, August 2019. DOI: <https://doi.org/10.7236/IIBC.2019.19.4.169>.
- [6] Jae-Young Chang, Byung-Jun Lee, Se-Jin Cho, Da-Hye Han, Kyu-Hong Lee, "Automatic Classification of Advertising Restaurant Blogs Using Machine Learning Techniques", *The Journal of the Institute of Internet, Broadcasting and Communication (IIBC)*, Vol.16, No.2, pp.55-62, April,20-16. DOI: <http://dx.doi.org/10.7236/IIBC.2016.16.2.55>.
- [7] Pratama, Bayu Yudha, and Riyanarto Sarno. "Personality classification based on Twitter text using Naive Bayes, KNN and SVM." *Data and Software Engineering (ICoDSE), 2015 International Conference on*. IEEE, 2015. [6] Han, K., A Study of Acetic Acid Formation in *Escherichia coli* Fermentation, Ph.D. Thesis. University of California, Irvine, CA, USA., 2010.
- [8] Wanseop Kim, Soowon Lee. "An In-depth Survey Analysis Applying Data Mining Techniques." *Korean Society for Engineering Education & Technol* 9.4 71-82. 2006.
- [9] The Institute of Internet, Broadcasting and Communication, Submission of manuscript. <http://www.iibc.kr>.