

## Personality Classification Experiment by Applying k-Means Clustering

<https://doi.org/10.3991/ijet.v15i16.15049>

Assem Talasbek, Azamat Serek  
Suleyman Demirel University, Kaskelen, Kazakhstan

Meirambek Zhaparov  
Paragon International University, Phnom Penh, Cambodia

Seong Moo-Yoo <sup>(✉)</sup>  
University of Alabama in Huntsville, Alabama, USA  
yoos@uah.edu

Yong Kab Kim, Geun-Ho Jeong  
Wonkwang University, Iksan, Korea

**Abstract**—This paper describes personality classification experiment by applying k-means clustering machine learning algorithms. Several previous studies have been attempted to predict personality types of human beings automatically by using various machine learning algorithms. However, only few of them have obtained good accuracy results. To classify a person into personality types, we used Jungian Type Inventory. Our method consists of three parts: data collection, data preparation, and hyper-parameter tuning. Our testing results showed that the k-means model has 107 inertia value, which is a good number for an unsupervised learning model as an interim result. With the result, we divided the data into 16 clusters, which can be considered as personality types. We continue this research with analysis of large data to be collected in the future.

**Keywords**—Personality types, machine learning, Jungian Type Inventory, k-means clustering on personality test.

### 1 Introduction

Personality is a combination of a person's characteristics and attitudes in dealing with different social situations as in kindergarten, school, university, family, working team, etc. [1]. Humans are addicted to biases and prejudices that might affect their judgmental accuracy. Personality can be taken as assessment in various fields such as selection of staff, choice of profession, relationship and health counseling. There is a great effect of personality on learning capabilities of humans. For instance, in learning performance we may see significant differences between persons who belong to extroverts and the ones belonging to introverts [2]. One of the main reasons why

students drop their studies in universities is a poor academic performance (AP), but personality also affects AP at the same level as intellectual abilities, self-esteem, motivation, etc. [3].

Some studies show that personality can be taken as an effective measurement in predicting academic performance, especially at the university level [4]. Studies [5, 6, 7] based on the Big Five Model [8] describe the consecutive relationship between conscientiousness and AP. To describe personality the proposed model uses five factors such as openness, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN). However, in different work directions of relationship between personality traits such as agreeableness and AP are irreconcilable and different. For instance, some studies [9, 10] validate positive relationship between agreeableness and AP but another [11] claims negative relationship between them.

Personality has a great effect from the perspective of a career. Caldwell and Burger [12] show that in case of career difficulties extraversion of a person facilitates the way of searching and gathering information about possible career path. Because extroverts are individuals who seek support from society [13]. However, people with high scores of neuroticism suffer from career indecision because they become more vigilant in job search [14].

Attempts to analyze personality manually led to several issues. Identification of personality type of human beings manually requires a lot of time when it needs to analyze many people. Although personality, its types and variation among individuals is the study of personality psychology [15] which is also focused on human nature, psychological similarities between people, cooperation with computer science is necessary to be able to efficiently analyze large amounts of data.

To classify people into personality types psychologists have developed various types of questionnaires. Some popular types of them are NEO-Personality-Inventory Revised [16], Big-Five Inventory [17], and Ten Item Personality Inventory [18]. In this study, questions based on Jung [19] theory were used.

NEO-Personality-Inventory Revised consists of 240 questions and is intended for adults (from 18 years) without mental pathologies [16]. The questionnaire was developed within the concept of the Big Five personality factors and each of the five "big" factors, in turn, is divided into six aspects. For instance, neuroticism consists of such aspects as anxiety, hostility, depression, self-consciousness, impulsiveness, and vulnerability to stress. Extraversion is divided into gregariousness, assertiveness, activity, excitement, seeking, and positive emotion. Openness is the mix of fantasy, aesthetics, feelings, actions, ideas, and values. Agreeableness is divided into trust, straightforwardness, altruism, compliance, modesty, and tender mindedness. Conscientiousness is separated into competency, order, dutifulness, achievement, striving, self-discipline, and deliberation.

One of the latest methods of testing the Big Five Model is Big-Five Inventory (BFI). Unlike other Big Five tests, BFI uses fragments of sentences as a stimulus material, which, as planned, should increase the reliability of the test [17]. The most well-known brief questionnaire for the diagnosis of the Big Five is Ten Item Personality Inventory (TIPI), which consists of, as the name suggests, 10 questions and is successfully and repeatedly adapted in many countries in Europe, Asia and

South America [18]. The field of computer science such as machine learning provides various methods to collect, process and classify large amounts of data in a systematic fashion.

This paper is organized as follows. Section II provides literature review, Section III and IV describe methodology and inner mechanisms of k-means clustering. Section V provides results of the experiment. Section VI concludes this paper.

## 2 Literature Review

Since H. J. Eysenck outlined features that describe types of personality [20], various studies have been conducted. In [21] authors collected text and image data from Instagram and Twitter to predict personality type. Pratama and Sarno [22] applied algorithms such as Naive Bayes Classifier [23], K-Nearest Neighbors (KNN) [24] and Support Vector Machine (SVM) [25] to classify personality based on text from Twitter. Testing results showed that Naive Bayes outperformed the other methods [22]. In [26] to extract the common data between modalities and mapping it to the personality type, an unsupervised learning algorithm such as Heterogeneity Entropy Neural Network (HENN) was proposed. According to Goldberg’s research [20] psychological models which dependent on trait are more effective measurement for aspects in subject of life. Rushton, et al. [27] focused on members of Florida League of Teachers. Here, they used Myers-Briggs Type Indicator (MBTI) and conclude that ENFP type teachers are more likely to be top teachers.

MBTI is one of the popular methods to classify personality type based on Jung’s theory [19]. His theory itself based on four dichotomies (pairs of opposite signs), extraversion (E) - introversion (I), sensing (S) - intuiting (N), thinking (T) - feeling (F), and judgment (J) - perceiving (P). Combination of these four dichotomies reveals sixteen types of personality [28]. Table 1 shows MBTI personality types.

**Table 1.** Combination of dichotomies [8]

ISTJ	INTJ	ESTJ	ENTJ
ISTP	INTP	ESTP	ENTP
ISFJ	INFJ	ESFJ	ENFJ
ISFP	INFP	ESFP	ENFP

- The extraversion (E) - introversion (I) is considered as the superior direction of MBTI types. Extroverts concentrate on ideas and events in the outside world, while introverts are more concentrated on internal thoughts. If extroverts like to share their ideas with the outside world, introverts will think about it very carefully before they say or do something [28].
- The sensing (S) – intuiting (N) direction is about how a person receives information from the external world. While the intuitives clearly see the possibilities and think globally noticing the consequences and the relationship between events, the sensors can pay attention to details, marking the nuances [29].

- The thinking (T) – feeling (F) direction describes the way of making decisions. Thinkers come to a decision logical and rational, not relying on emotions and sensations. Feelers make decisions by following their hearts and tend to be more emotional, soft heart [2].
- The judging (J) - perceiving (P) direction reveals how people orient themselves in the world. Judges prefer to organize and make plans to achieve goals, while perceivers solve problems and make choices only when it is necessary [28].

In the study about teachers and their MBTI types [30] Lawrence conducted that the combination of Extraversion, Sensing, Feeling and Judging (ESFJ) is more frequently common and preferable rather than combination of other dichotomies. It should be noted that factors such as a level of a teacher (middle or high school, etc.) and the frequency rate in each group are not specified. As a sample, personality types of 5366 teachers were recorded. Fairhurst and Fairhurst [31] compile that ENFP type teachers more inspire students to solve problems that they did not faced before and learn new things through imagination and creativity. Kamal and Radhakrishnan [32] outlined that consideration of student’s learning style by personality traits may affect the success of e-learning courses. E-learning can provide learning materials in various formats as an example audio-streaming, presentations, video tutorials, group work, etc. For extroverts, chat discussions are preferable while for introverts, individual locus control should be provided.

Komisin and Guinn [33] took writing samples and MBTI from 40 graduate students to identify personality type of participants using document classification. In this experiment Naive Bayes Classifier [23], SVM [24] and analysis software Linguistic Inquiry and Word Count (LIWC) were used to predict personality type based on word-choice of participants. To select parameters one-leave-one-out cross validation was used. A comparison of the results of applying Naive Bayes Classifier and SVM revealed that Naive Bayes Classifier outperformed SVM in prediction of Sensing-Intuition, Thinking-Feeling traits.

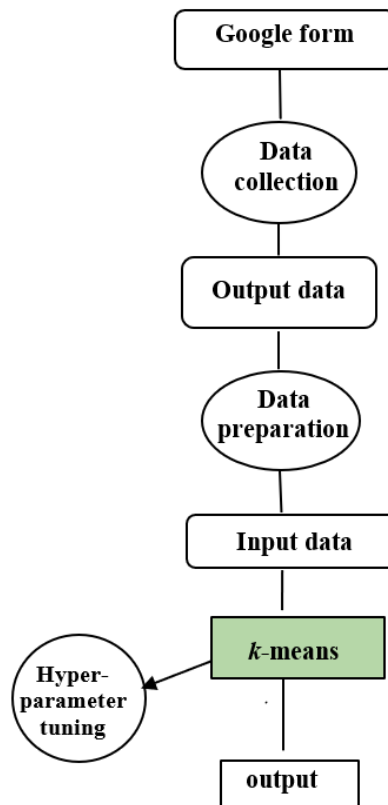
Mairesse and Walker [34] in their work attempted to predict personality type using pieces of conversation records. As a base for identification type of personality Big Five Model was used. They conducted that informal format of dialog has high rates of correlation between linguistic factors and personality traits. Results show that “neuroticism” is more complex trait to analyze rather than other traits while “extroversion” is a more accurate predictor. They conclude that proposed hypothesis of personality prediction through linguistic characteristics is proved to be applicable in various fields.

### **3 Our Methodology**

Figure 1 illustrates the general representation of this research. Our method consists of three parts: data collection, data preparation, and hyper-parameter tuning. To develop a prototype of a classification system, we created a Google form. Here, questions were based on four dichotomies discussed in Section II. For each dichotomy there are allocated five questions resulting in overall number of 40 questions.

Participants of our survey had to answer to each question “Yes” or “No”. Using Google’s script, we automated the process of personality calculation and made some simulations using k-means clustering. But collected data did not give us exact output because some participants found it difficult to answer some questions and, as the result, for more than 35% of the respondents we could not determine the exact type of personality. For example, one person could relate to an extrovert and introvert 50% to 50% or several types at once.

As the next stage to resolve this problem, we decided to use in Likert scale from 1 to 5 instead of “Yes” or “No” answers, where 1 is the maximum disagreement and 5 is the maximum agreement. After changing the type of answers, we took as a sample 105 bachelor-degree students of the Computer Science Department of Suleyman Demirel University.



**Fig. 1.** Personality Classification by applying K-means clustering.

As the second experiment we launched the second form which consists of 167 questions adopted to the population of post-Soviet states. Participants of these surveys had to answer each question by selecting preferable options for specific situations and by selecting an adjective that describes or suits them. In this experiment we took as a

sample 410 bachelor-degree of Computer Science Department of Suleyman Demirel University.

After collection of all answers to make a prediction whether a person who passed the test belong to what group, i.e., to classify persons by given data into 16 types, we needed to use machine learning. During creation process, we needed to specify the number of clusters to choose. We decided to choose it based on inertia property of the model, which is the sum of squared distances of samples to their closest cluster center. The less is inertia property, the better is our work. Thus, we started from three clusters, and our inertia was 700. After tuning it a little bit and coming to 16 numbers of clustered, we reduced our inertia parameter to 107. After the creation and fitting process (training our k-means model), we tested it on some arbitrary input data to verify whether it works or not. Then, we found it worked. After passing such data, it determines the index of the cluster it belongs to.

Given that we did not have some target variable for our training data, we had to use unsupervised learning. Unsupervised learning allows algorithms to act on given data without any guidance [35]. Out of all unsupervised learning algorithms, we decided to choose the most optimal one - k-means clustering algorithm [36]. After identification of k number of clusters, we allocate each data point to the nearest cluster. After importing data, making some changes to it (e.g., deleting unnecessary columns from our data frame to pass to the model of the algorithm), we created a model of k-means algorithm.

To simplify the process of machine learning part in order to apply k-means algorithm on our dataset, we used BigML.com [37] website, which automates the process of a variety of data processing, analysis parts, machine learning, supervised and unsupervised learning parts. That really makes the process of working with data flow smoothly. In terms of preprocessing, the values of categorical variables have been transformed from textual representation into numeric one, so that k-means algorithm will be able to work with it. As to hyper-parameter tuning, we used the same infrastructure BigML.com to find the most optimum inertia value.

Inertia value is one of the most important parameters of k-means, of which the definition is how far the data points are from their centroids. That can be manifested itself as a standard deviation regarding the mean notion. So, the corollary of that is the less is the inertia, the better is our k-means model in terms of its cluster division.

#### **4 k-means Clustering Algorithm**

The k-means method is the most popular clustering method. It refers to unsupervised learning part in machine learning field. The most well-known algorithm utilizes an iterative refinement technique. By cause of pervasiveness it is regularly called the k-means algorithm. The algorithm continues by switching back and forth between two stages [35].

*Assignment step:* Assign every observation (sample, example, feature) to the cluster with mean (average of a set of values), which has the slightest squared

Euclidean distance which shows how close or far away two observations from each other [36]. We calculated the distance as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

where  $x$  and  $y$  are samples in  $n$ -dimensional feature space.

*Update step:* Calculation of values of new means is made of clusters that will become the centroids of observations in the new clusters. If the assignments do not change, the algorithm converges.

Generally, we define randomly  $k$  clusters in the plane, then we calculate distances of each data point to these  $k$  clusters and assign the observation to the closest centroid, after that we move centroids to the mean of assigned values to them. We repeat this process until convergence, when the values of centroids do not change after iteration.

Using this algorithm, we cannot be sure that the most optimal result will be found. We can stop the algorithm from converging using other distance functions as Manhattan distance functions for instance. There are other  $k$ -means transformations, especially the spherical  $k$ -means and  $k$ -medoids. They allow the use of other remote measures, based on the following distance calculation:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

where  $x$  and  $y$  are data points and  $n$  are the number of dimensions. Manhattan distance is also called L1 norm or L1 distance in machine learning which are used for regularization problems.

#### 4.1 Initialization methods

The most used methods for initialization are Forgy and Random Partition [36]. The Forgy method from the dataset inadvertently selects  $k$  observations to use them as a source of funds. The Random Partition method also randomly assigns a cluster, but after that it goes immediately to the update phase. The Forgy method tends to spread the initial means out, but Random Partition works quite differently. It puts them all close to the center of the dataset.

#### 4.2 Demonstration

- a) As indicated in the Figure 2,  $k$  initial "means" is randomly selected. In this case,  $k$  equals to 3
- b)  $k$  clusters are generated by matching each observation with the closest mean value
- c) The new average value is set by the centroid of each of the clusters  $k$
- d) Steps b and c are repeated until convergence is achieved

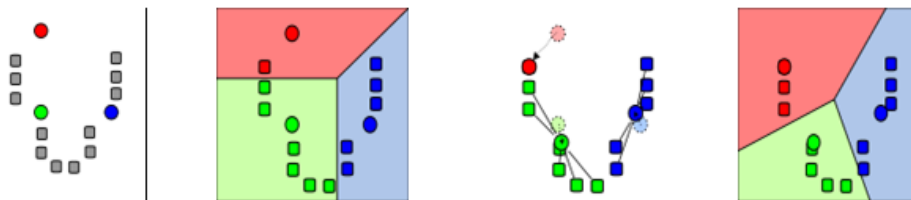


Fig. 2. Demonstration of K- means clustering [19]

Clustering form is one of the major limitations of k-means algorithm. The concept is based on spherical clusters which are divided in a similar way that average significance would be like the middle of the cluster. When clusters have the same volume so that the destination in the closest cluster center will be the right destination. For instance, in the process of applying k-means clustering to data set such as Iris flower data [38], the result is not able to divide three Iris types which are in information set. Iris flower is a data set for the classification problem, by the example of which Ronald Fisher in 1936 demonstrated the work of the discriminant analysis method developed by him [39]. In other words, this is classic set, which is used to illustrate the work of various statistical algorithms.

If we consider the case when  $k = 2$ , two explicit clusters will be evaluated; one of them holds two types of cluster, as with  $k = 3$ , only one with two clusters will be separated into two even sections.  $k = 2$  is the most suitable for this dataset, against the fact that dataset has three classes. As in cases of other clustering algorithms, k-means clustering result assumes that data satisfy criteria.

## 5 Discussion and Results

The main three distinctive advantages of “k-means”, which makes it productive and effective for the provided task, are often become great disadvantages in overall scheme of things.

- Euclidean distance [40] is applied exactly as a metric, and dispersion is applied exactly as a measure of scattering.
- The number of clusters  $k$  is considered an input parameter; an incorrect choice of  $k$  can give non-positive results. For this reason, it is substantial to make diagnostic check to determine the number of clusters.
- Convergence to a local minimum can cause a contradiction.

We made two surveys. 106 participants responded to the first survey, and 410 participants responded to the second survey. The results of both surveys are listed in Table 2 and Table 3. Both surveys covered the questions according to Carl Jung's personality theory through the metrics as extrovert/introvert, sensing/intuiting, thinking/feeling, and judging/perception. As the result, it ended up with 16 features. In order to differentiate between different personalities, given that we did not have a



label about person what personality they have, we decided to use unsupervised learning. In this branch of machine learning we selected k-means algorithm, whose basic idea is to divide the data into specified number of clusters using so-called centroids, which is the alternative of mean of a cluster.

**Table 2.** Cluster instances based on collected data from survey 1

Cluster name	Instances	%	Mean distance
Cluster 00	30	28.6	0.133
Cluster 01	5	4.72	0.140
Cluster 02	9	8.49	0.144
Cluster 03	4	3.77	0.147
Cluster 04	2	1.89	0.098
Cluster 05	3	2.83	0.100
Cluster 06	5	4.72	0.111
Cluster 07	1	0.94	0
Cluster 08	17	16.04	0.140
Cluster 09	2	1.89	0.103
Cluster 10	2	1.89	0.103
Cluster 11	3	2.83	0.118
Cluster 12	3	2.83	0.113
Cluster 13	16	15.09	0.138
Cluster 14	1	0.94	0
Cluster 15	3	2.83	0.098
Total	106	100	0.164

Table 2 lists data distribution results using k-means clustering algorithm. According to the results of our research above 28 percent (30 survey participants) belong to cluster 00. Clusters 07 and 14 placed the farthest from all and have the smallest amount which is equal to 0.94 percent of total. Since we have only one person in each cluster like 07 and 14, the distance means of this clusters are equal to 0. Mean distance in the table shows the mean distance of cluster to centroid of the cluster.

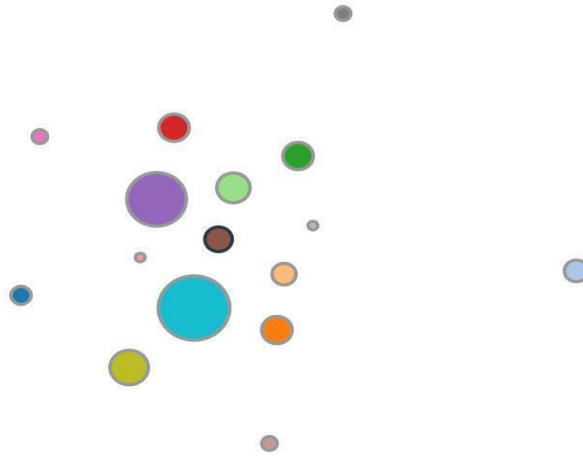
In Google form we asked questions like “Do you communicate openly without censoring?” because it shows the level of motivation that belongs to extraversion and introversion traits. To see the way of acquiring information, questions that belong to sensing and intuiting traits were used, for instance, “You show evidence (e.g., facts, details, examples, etc.)” Questions related to approaching life that belong to judging and perceiving traits and questions that present the way of making decisions that belongs to thinking and feeling traits were also asked to participants such as “You acknowledge the time for creativity.” and “You demonstrate empathy by showing areas of agreement first.”

**Table 3.** Cluster instances based on collected data from survey 2

Cluster name	Instances	%	Mean distance
Cluster 00	52	12.7	0.04301
Cluster 01	58	14.15	0.04391
Cluster 02	57	13.9	0.03853
Cluster 03	40	9.7	0.04316
Cluster 04	24	5.8	0.04438
Cluster 05	25	6.10	0.04648
Cluster 06	12	2.9	0.04217
Cluster 07	26	6.34	0.04275
Cluster 08	21	5.12	0.04602
Cluster 09	9	2.19	0.0449
Cluster 10	11	2.68	0.04596
Cluster 11	10	2.44	0.04501
Cluster 12	10	2.44	0.04566
Cluster 13	16	3.90	0.04677
Cluster 14	8	1.95	0.04398
Cluster 15	31	7.56	0.04584
Total	410	100	0.04924

Table 3 lists the result of a questionnaire which consists of 167 questions. Here, we asked participants to describe themselves by choosing proper adjectives and asked questions like “What will you do if ...?”. Questions were asked in Russian language. As shown in the table, 58 participants belong to cluster 01. Furthermore, we noticed that people tend to participate in such a survey if questions are asked in the language that they use in daily life. Also, we asked them to provide us some text about themselves in Kazakh (native) language and send us their image and Instagram accounts for our future work to make an experiment using such a kind of data.

Figure 3 and Figure 4 illustrate 16 clusters which were obtained by applying k-means clustering on data from the surveys. Figure 3 shows the result from survey 1, and Figure 4 shows that from survey 2 (Russian language).



**Fig. 3.** The result of applying k-means clustering on data from survey 1.



**Fig. 4.** The result of applying k-means clustering on data from survey 2.

After applying k-means clustering on given data and classifying it into 16 clusters shown in Figure 3, it is noticeable that in general, people who have passed the test are like each other, because clusters are close to each other. But there are small clusters that are too far separated from the most. The main problem and peculiarity of unsupervised learning algorithms is what it gives us to what kind of cluster a person belongs to without the description of personality type. Therefore, to assign exact personality type to each person, the work of professional psychologists with both the people themselves and their results is needed. Through application of grid tuning and coming up with the decreased number of clusters in the parameters of the algorithm, we improved our model and the inertia has been dropped from 700 to 107. After applying k-means algorithm, we decided to execute association analysis in order to

see how variables are related to each other in terms of what factors contribute to its values interchangeably.

From Table 4 we can see that if you are pretty aware that “how you communicate” is as important as “what you are communicating”, then it will lead that you are in a good level of taking care about how the idea will affect people and what people’s reaction would be, and vice versa. Also, we found that if you are brief and concise, then you are a priory often calm and reasonable, and vice versa.

**Table 4.** Association analysis

Antecedent	Consequence
You aware that how you communicate is as important as what you’re communicating > 4	You take care about how the idea will affect people and what people’s reaction would be > 4
You take care about how the idea will affect people and what people’s reaction would be > 4	You aware that how you communicate is as important as what you’re communicating > 4
You are brief and concise > 4	You are often calm and reasonable > 4
You are often calm and reasonable >4	You are brief and concise > 4

## 6 Conclusion

This paper describes personality classification by applying k-means clustering. We provided explanations of advantages and disadvantages of k-means clustering. Given that we do not have some target variable for our training data, we had to use unsupervised learning. The k-means method has been utilized for this purpose. After fitting the model on the training data, it divided the data into 16 clusters, with an overall inertia of 107, which is a good value for unsupervised learning model. These 16 clusters can be considered as 16 types of personality. Furthermore, we executed association analysis in terms of antecedents and consequences to see how variables are related to each other, what factors contribute to the change of values of features.

Our future work will be continued in analysis of text data and image data that we collected from this experiment.

Remark: The short version of this paper was presented in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), Fukuoka, Japan, Feb. 2020.

## 7 References

- [1] David C. Funder. Personality. Annual Review of Psychology, 52(1):197–221, 2001.
- [2] Jieun Kim, Ahreum Lee, Hokyoung Ryu. Personality and its effects on learning performance: Design guidelines for an adaptive e-learning system based on user model In: 2013 Elsevier. International Journal of Industrial Ergonomics. 2013 p. 1–12. <https://doi.org/10.1016/j.ergon.2013.03.001>

- [3] Sawsen Lakhali, Serge Sévigny, Éric Frenette Personality and student performance on evaluation methods used in business administration courses. *Educational Assessment, Evaluation and Accountability*. May 2015, Volume 27, Issue 2, pp 171–199 Springer. <https://doi.org/10.1007/s11092-014-9200-7>
- [4] Chamorro-Premuzic, T., & Furnham, A. (2003a). Personality traits and academic examination performance. *European Journal of Personality*, 17(3), 237–250. <https://doi.org/10.1002/per.473>
- [5] Bauer, K. W., & Liang, Q. (2003). The effect of personality and precollege characteristics on first-year activities and academic performance. *Journal of College Student Development*, 44(3), 277–290. <https://doi.org/10.1353/csd.2003.0023>
- Conard, M. A. (2006). Aptitude is not enough: how personality and behaviour predict academic performance. *Journal of Research in Personality*, 40(3), 339–346. <https://doi.org/10.1016/j.jrp.2004.10.003>
- De Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, 10(5), 405–425. [https://doi.org/10.1002/\(sici\)1099-0984\(199612\)10:5<405::aid-per255>3.0.co;2-m](https://doi.org/10.1002/(sici)1099-0984(199612)10:5<405::aid-per255>3.0.co;2-m)
- [8] R. R. McCrae. Cross-cultural research on the five-factor model of personality. *Online Readings in Psychology and Culture*, 4(4), 2002. <https://doi.org/10.9707/2307-0919.1038>
- [9] Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: the role of personality, intelligence and application. *Personality and Individual Differences*, 34(7), 1225–1243. [https://doi.org/10.1016/s0191-8869\(02\)00111-3](https://doi.org/10.1016/s0191-8869(02)00111-3)
- [10] Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, 70(2), 177–206. <https://doi.org/10.1111/1467-6494.05002>
- Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behaviour. *Journal of Personality and Social Psychology*, 74(2), 538–556. <https://doi.org/10.1037/0022-3514.74.2.538>
- [12] Caldwell, D. F., & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology*, 51, 119–136. <https://doi.org/10.1111/j.1744-6570.1998.tb00718.x>
- Caldwell, D. F., & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology*, 51, 119–136. <https://doi.org/10.1111/j.1744-6570.1998.tb00718.x>
- [14] Tokar, D., Fischer, A., & Subich, L. M. (1998). Personality and vocational behavior: A selective review of the literature, 1993–1997. *Journal of Vocational Behavior*, 53, 147–179. <https://doi.org/10.1006/jvbe.1998.1660>
- [15] Friedman, Howard; Schustack, Miriam (2016). *Personality: Classic theories and modern research*. USA: Pearson. ISBN 978-0-205-99793-0.
- [16] Paul T. Costa and Robert R. McCrae. Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI). Psychological Assessment Resources, Odessa, Fla. P.O. Box 998, Odessa 33556, 1992. <https://doi.org/10.1037/t03907-000>
- [17] Oliver P John, Eileen M Donahue, and Robert L Kentle. The big five inventory—versions 4a and 54. University of California, Berkeley, Institute of Personality and Social Research, 1991.

- [18] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504 – 528, 2003. [https://doi.org/10.1016/s0092-6566\(03\)00046-1](https://doi.org/10.1016/s0092-6566(03)00046-1)
- [19] Charles C. Healy *Journal of Career Development* In: June 2000 Springer, Volume 26, p.295–308.
- [20] H. J. Eysenck, *Dimensions of Personality*, 1947.
- [21] Skowron M, Tkalčič M, Ferwerda B, Schedl M. “Fusing social media cues: personality prediction from twitter and instagram”. In: *Proceedings of the 25th International Conference Companion on World Wide*. <https://doi.org/10.1145/2872518.2889368>
- [22] Pratama BY, Sarno R. “Personality classification based on Twitter text using Naive Bayes, KNN and SVM.” In: *2015 International Conference on Data and Software Engineering (ICoDSE)*. IEEE; 2015. p. 170–4. <https://doi.org/10.1109/icodse.2015.7436992>
- [23] simple Bayesian classifier under zero-one loss". *Machine Learning*. **29**: 103–137.
- Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression". *The American Statistician*. **46** (3): 175–185. <https://doi.org/10.1080/00031305.1992.10475879>.
- [25] Ben-Hur, Asa; Horn, David; Siegelmann, Hava; and Vapnik, Vladimir N.; "Support vector clustering"; (2001); *Journal of Machine Learning Research*, 2: 125–137
- Xianyu H, Xu M, Wu Z, Cai L. Heterogeneity-Entropy Based Unsupervised Feature Learning For Personality Prediction With Cross-Media Data. In: *2016 IEEE International Conference on Multimedia and Expo (ICME)*. 2016. p.1–6. <https://doi.org/10.1109/icme.2016.7552980>
- Stephen Rushton, Jackson Morgan, Michael Richard. Teacher’s Myers-Briggs personality profiles: Identifying effective teacher personality traits *An International Journal of Research and Studies Teaching and Teacher Education* 23 (2007) p. 432–441. <https://doi.org/10.1016/j.tate.2006.12.011>
- [28] V. Rodri’guez Montequi’n, J. M. Mesa Fernández, J. Villanueva Balsera, A. Garcí’a Nieto. Using MBTI for the success assessment of engineering teams in project-based learning In: *2013 Springer. International Journal of Technology and Design Education Vol 23 2013* p.1127–1146. <https://doi.org/10.1007/s10798-012-9229-1>
- [29] Charles C. Healy “Interpreting the Myers-Briggs Type Indicator to Help Clients in Understanding Their Strong Interest Inventory” In: *2000 Springer Journal of Career Development Vol 26 2000* p 295–308. <https://doi.org/10.1177/089484530002600405>
- [30] Lawrence, G. (1979). *Peoples types and tiger stripes: A practical guide to learning styles*. Gainesville, FL: Center for Applications of Psychological Type.
- [31] Fairhurst, A. M., & Fairhurst, L. L. (1995). *Effective teaching effective learning: Making the personality connection in your classroom* (1st ed). Palo Alto, CA: Davis-Black.
- [32] Anju Kamal & Souhreda Radhakrishnan. Individual learning preferences based on personality traits in an E-learning scenario. January 2019, Volume 24, Issue 1, pp 407–435. <https://doi.org/10.1007/s10639-018-9777-4>
- [33] Mike Komisin and Curry Guinn. Identifying Personality Types Using Document Classification Methods. *Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference*.
- [34] F. Mairesse, M. and Walker, M., 2006. Automatic Recognition of Personality in Conversation. In: *Association for Computational Linguistics ed., Proceedings of the*

- Human Language Technology Conference of the North American New York: Association for Computational Linguistics. Automatic Recognition of Personality in Conversation, pp. 85. <https://doi.org/10.3115/1614049.1614071>
- [35] T. Kohonen, O. Simula, “Engineering Applications of the SelfOrganizing Map”, Proceeding of the IEEE, Vol. 84, No. 10, 1996, pp.1354 – 1384.
- Gregory A. Wilkin and Xiuzhen Huang “K-Means Clustering Algorithms: Implementation and Comparison” In: 2007 Second International Multisymposium on Computer and Computational Sciences. In IEEE pp: 133-136 ISBN: 0-7695-3039-7. <https://doi.org/10.1109/imsccs.2007.51>
- [37] BigML <https://bigml.com/> date of access 21.11.2018.
- [38] Bezdek, J.C. and Keller, J.M. and Krishnapuram, R. and Kuncheva, L.I. and Pal, N.R. (1999). "Will the real iris data please stand up?". IEEE Transactions on Fuzzy Systems. 7(3). <https://doi.org/10.1109/91.771092>
- [39] Fisher, R.A. (1936). “The Use of Multiple Measurements in Taxonomic Problems” (PDF). Annals of Eugenics. 7: 179—188.
- [40] Anton, Howard (1994), Elementary Linear Algebra (7th ed.), John Wiley & Sons, pp. 170–171, ISBN 978-0-471-58742-2

## 8 Authors

**Assem Talasbek** is a Senior Lecturer of Computer Science department at Suleyman Demirel University (SDU), Kaskelen, Kazakhstan and a PhD student at SDU. She earned a BS degree in Computer Science at Suleyman Demirel University, Kaskelen, Kazakhstan, in 2015, MSc degree in Computer Science at M.Auezov South Kazakhstan State University, Shymkent, Kazakhstan in 2017. She is an Oracle Academy instructor at Suleyman Demirel University. Her research interests include IT education, Machine Learning.

**Azamat Serek** is an instructor of Computer Science department at Suleyman Demirel University (SDU), Kaskelen, Kazakhstan and a MSc student at SDU. He earned a BS degree in Computer Science at Suleyman Demirel University, Kaskelen, Kazakhstan, in 2018. His research interests include IT education, Machine Learning, Big Data.

**Meirambek Zhaparov** is a Vice Rector of student Affairs, dean of Information Communication Technologies faculty at Paragon International University, Phnom Penh, Cambodia. He was an Assistant Professor of Computer Science and Dean of Engineering and Natural Sciences Faculty at Suleyman Demirel University, Kaskelen, Kazakhstan in 2018. He earned a BS degree in Computer Science at Suleyman Demirel University in 2003, MS degree in Computer Science at D. Serikbayev East Kazakhstan State University, Ust-Kamenogorsk, Kazakhstan in 2006, and PhD degree in Computer Science at Suleyman Demirel University in 2013. His research interests include microlearning and IT education.

**Seong Moo-Yoo** is an Associate Professor of Electrical and Computer Engineering at the University of Alabama in Huntsville (UAH), Huntsville, Alabama, USA. Before joining UAH, he was an Assistant Professor at Columbus State University,

Columbus, Georgia, USA. He earned a BA degree in Economics at Seoul National University, Seoul, Korea, and MS and PhD degrees in Computer Science at the University of Texas at Arlington, USA, in 1989 and 1995, respectively. His research interests include computer network security, malware detection, and wireless networks such as mobile ad hoc networks, vehicular networks, sensor networks, and mesh networks. He has authored/co-authored over 100 scientific articles in refereed journals and international conferences. He is a life member of IEEE and a senior member of ACM.

**Yong-Kab Kim** received his Ph.D degree in Electrical & Computer Engineering in North Carolina State University. He is currently in Professor at School of Electrical Information Communication Engineering, Wonkwang University, Korea. His research interests are Remote sensing for Visible Communication, Optical Fiber Sensing, Intelligent Control, and Advanced Computational Intelligence.

**Geun-Ho Jeong** received his BS degree and MS degree in Electrical Engineering in Wonkwang University. His research interests are Electric Machinery, Electronic Circuit, and Light source of LED.

Article submitted 2020-04-12. Resubmitted 2020-05-14. Final acceptance 2020-05-14. Final version published as submitted by the authors.