Research Article

An Innovation and Entrepreneurship Management System for Universities Based on Cluster Analysis Theory

Weijie Zhao and Liang Fang

College of Innovation and Entrepreneurship, Pingdingshan University, Pingdingshan 467000, China

Correspondence should be addressed to Weijie Zhao; 2135@pdsu.edu.cn

Received 26 February 2022; Revised 31 March 2022; Accepted 4 April 2022; Published 9 May 2022

Academic Editor: Wen Zeng

Copyright © 2022 Weijie Zhao and Liang Fang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of the market economy and changes in the demand for talents in the market, the cultivation of innovation and entrepreneurship (IAE) among university students has been emphasized, and how to improve the quality of IAE education and promote the overall development of university students has become a key concern of university teachers. Therefore, the design and implementation of the IAE project management system should be explored in the study, starting from the IAE education of university students. The thesis addresses the current situation and characteristics of IAE development of university students and develops a WEB system applicable to the project management of IAE for university students at this stage. Cluster analysis is a mathematical statistical method for grouping and categorising similar data. The design of a classification system based on university innovation and entrepreneurship data proposes a new definition of outlier index for outliers detected in real time, combined with the concept of CBLOF(t). On this basis, the system also implements the monitoring function for abnormal data, thus serving as system optimisation and fault warning. The system implements the main functions such as login, registration, project release, project declaration, and management.

1. Introduction

To date, the majority of universities do not have a system for the declaration and management of innovative projects [1] for the students, and the declaration of projects mainly relies on manual registration and approval by special personnel, which is inefficient and not easy to manage. Therefore, it is urgent to develop an efficient and stable project declaration system [2]. The development of an innovative project declaration system will gradually move from the original single and simple to diversified and intelligent. The future innovation project declaration system for university students will be an efficient and stable information management system built on the basis of the Internet and supported by a database [3].

An analysis of the IAE project practice reveals that it plays a key role in cultivating the innovation, entrepreneurship, and practical ability of university students [4] and that after China proposed to implement IAE education for university students in the field of higher education, many schools have made corresponding practical exploration in this regard [5], creating idealized conditions for the cultivation of IAE ability of university students. However, it must be clearly noted that in order to give play to the important role of innovative and entrepreneurial educational practices of students [6], universities must promote the improvement of the technical level of educational practices with the help of scientific management. It is therefore essential to analyse the project management aspects of student IAE to ensure that scientific management can free student managers from the tedium of management, improve management efficiency and level, and strive to provide diversified services to the student community [7–9].

The IAE management system can create certain convenience for the management, and the analysis of the management system shows that the IAE project management platform system for students is mainly an information management system for university students to participate in IAE projects [10]. The system is designed with B/S architecture and MVC model, which can provide support for students
to participate in IAE project management and implement information audit and management of relevant information, which has strong practical significance for the management of university students’ IAE projects at present [11]. Therefore, it is necessary to strengthen the exploration of the management platform system in view of the basic situation of the management of the IAE projects of college students in China, so as to the platform system and promote the stable development of the IAE practice activities of college students.

The focus of university students’ IAE project management is to analyse the current problems on the basis of a comprehensive understanding of the project management characteristics and then to clarify the problems that exist in it. The work is then combined with different business requirements to complete UML modelling and determine scientific and reasonable measures for authority control [12]. In general, the comprehensive promotion of university students’ IAE project management can enhance the adaptability of university students to society, cultivate their entrepreneurial ability, and cultivate high-level entrepreneurial development talents for the country [13].

A systematic study of the construction of the IAE system for students reveals that the B/S architecture is introduced in the design and planning of the system, which can meet the needs of project management and promote the realization of diversified functions of the system [14]. With the support of the system, assessment experts and relevant instructors can have a comprehensive grasp of the project’s operation and provide effective guidance, as well as facilitate students’ access to their own IAE practices [15]. In planning the management platform system, it is necessary to promote the convenience of operation, but also to have the stability of functions, to ensure the accuracy and safety of data, so in the design and planning of the system, it must be analysed from multiple perspectives, to improve the management system, to develop the information backup and statistical analysis functions with the support of modern tools, and to ensure that it can be combined with the participation of students in IAE. The system should be designed and planned from multiple perspectives, and the management system should be improved [16], and the information backup and statistical analysis functions should be developed with the support of modern tools, so as to ensure the effective education and management of university students in the context of their participation in IAE education, to provide a corresponding platform for university students to participate in IAE education and practice, and to ensure that the basic work can be done well and the comprehensive development of university students’ IAE education can be improved. In this way, reform and innovation of IAE education for university students can be carried out [17], and the comprehensive work level will be improved comprehensively, which will play a good role in promoting the implementation of IAE education for university students. Based on the theory of cluster analysis, this paper adopts the weighted K-means algorithm based on Euclidean distance, establishes a clustering model by analyzing a large amount of normal historical data, and then, according to the actual operation, successively proposes the definition of the outlier index of parameter points within the cluster and the definition of the outlier index of abnormal parameter points outside the cluster and proposes a university innovation and entrepreneurship project management system based on this model.

2. Analysis of Basic System Requirements

According to the characteristics of the management of entrepreneurship for the Department of Innovation [18, 19], there are four roles involved in the system. The main functions are shown in Figure 1.

The system adopts an object-oriented approach [20], abandoning the traditional approach the development, instead of making full use database design tables based on the Java code corresponding to the entities written and the mapping files of the entity objects. This approach reduces the need to write cumbersome SQL statements to build the tables directly. The use of reverse engineering to generate the corresponding database tables from the entity objects’ JavaBean and their corresponding mapping files (xml files) [21] reduces data redundancy.

The supervisor provides guidance to the student leader and analyses the student leader’s project in the context of the actual situation and provides appropriate guidance. In general, the supervisor has the function of reviewing the project declaration, the midterm check, and the closing of the project. The EMS is based on the real needs of the development and management of student IAE projects. The system is used to ensure that students and teachers are involved in the process of each project, to guarantee the effective implementation of student entrepreneurship projects, to give full play to the important role of student IAE projects, and to promote the overall improvement of the quality of personnel training.

2.1. Entity Object Design. Based on a comprehensive analysis of the entity objects required by the system and their relationships, eight entity objects were identified, these are the administrator entity, the project application entity, the project applicant entity, the project reviewer entity, the new project entity, the old project entity, the project issuer entity, and the project executor entity.

In the proposed management system, there are 26 database tables, for the project declaration, and project evaluation modules are the core of the system, and there is a close relationship between project declaration and project evaluation. In the study, a variety of different database tables are discussed, and in this study, the identity information table and the project information table are selected for specific analysis. The administrator is able to keep up-to-date information about the student innovation and entrepreneurship projects in the management system. In general, their main job responsibility is to follow the realistic needs of student entrepreneurship, the output of the declaration plan and the evaluation plan, the information on the projects, etc., on the platform, thus ensuring that a scientific evaluation plan is developed. It is the focus centre of the whole system and can exchange information on the front page module.
The relationship between the various entities is shown in Figure 2.

### 3. Cluster Analysis Theory

The so-called cluster analysis algorithm works on the principle that for all vectors in an $m$-dimensional space $R$, the algorithm clusters each vector and achieves a category of vectors by minimising the distance between each vector and the centre of the cluster [23, 24]. That is, the greater the intraclass correlation, the better, and the smaller the interclass correlation, the better. As an unguided learning problem, the purpose of clustering is to obtain some intrinsic data pattern by dividing the original set of objects into similar groups or clusters.

#### 3.1. Introduction to Clustering Algorithms

There are many kinds of clustering algorithms, such as hierarchy-based, division-based, and density-based algorithms [25]. In this paper, the most common and reliable division-based K-means algorithm is used for analysis. The K-means algorithm divides $n$ vectors $x_i (i = 1, 2, \ldots, n)$ into $k$ classes $G_i (i = 1, 2, \ldots, k)$ and finds the cluster centres of each class so that the objective function of the nonsimilarity (or distance) index is minimized. When the metric between the vector $x_j$ and the corresponding cluster centre $C_i$ in the $i$-th class $G_i$ is chosen as the Euclidean distance, the objective function
can be defined as

\[ J = \sum_{k=1}^{i} J_k = \sum_{k \in G_i} x_i - c_i^2. \]  \hspace{1cm} (1)

Here, \( J_i = \sum_{x \in G_i} x - c_i^2 \) is the objective function \( G_i \) within the class, the apparent \( J \) size of which depends on the \( G_i \) shape of the cluster centre \( c_i \), and the \( J \) smaller the cluster, the better the clustering. 

The basic idea of the K-means algorithm is as follows.

1. First, \( k \) vectors are randomly selected as the centres of each class 
2. Let \( U \) be a 2-dimensional subordination matrix of \( c \times n \). If the \( j \)rd vector \( x_j \) belongs to class \( i \), then the element \( u_{ij} \) in matrix \( U \) is 1; otherwise, it is 0; i.e.,

\[ u_{ij} = \begin{cases} 1 & \text{For each } k \neq j \cap \| x_j - c_j \| \leq \| x_j - c_k \| \\ 0 & \text{others} \end{cases} \]  \hspace{1cm} (2)

3. Based \( u_{ij} \) on the value of the calculated objective function, stop if it is less than a threshold or if the difference between two consecutive times is less than a threshold
4. \( u_{ij} \) Calculate the centre \( G_i \) of each cluster according to \( C_i x_i = \sum_{x \in G_i} x_i / |G_i| \), \( |G_i| \) being the number of elements in the cluster. Then, return to (2)

The system entity object to database table relationship analysis and implementation entity object to relational table implementation process is shown in Figure 3.

3.2. Weighted Clustering Algorithms. The K-means algorithm has its limitations in that it takes into account all factors in the clustering analysis and assumes that these factors have an equal effect on distance. Relying equally on similarity measures for all attributes can be misleading \([26, 27]\); one effective way of overcoming this dimensionality trap can be achieved by assigning weights to each attribute, so that different attributes can achieve optimal results in the clustering process based on their own weights. In terms of the Euclidean space from a data perspective, adding weights is equivalent to lengthening the axes corresponding to the attributes of interest and shortening the axes corresponding to the attributes with smaller weights.

To this end, the objective function should be \( \| x_j - c \|_2 \) replaced by \( \sum_{k} (w_j x_{ij} - c_{ij})^2 \) an objective function that \( J \) also depends on the weights \( w_j \), the size of which can be determined by experienced staff using expert scoring methods to determine the importance of the equipment parameters for production optimisation and fault warning, thus solving the “dimensional trap” problem.

3.3. Concepts and Definitions of Outlier Indices. To date, there is no formal and generally accepted definition of an outlier. More recent findings are given in the relevant literature \([28, 29]\). To first give an example to illustrate this basic idea, consider the two-dimensional data set in Figure 1, where it can be found that there are four cluster centres, of which \( C1 \) and \( C3 \) are outliers. Since outliers are defined as data that are not inside any large cluster, here clusters \( C2 \) and \( C4 \) are to have the vast majority of the data and \( C1 \) and \( C3 \) are not part of clusters \( C2 \) and \( C4 \), so they are outliers. In order to show the physical characteristics of outlier data, an index is designed for each identified outlier data object, called CBLOF\( (t) \), some concepts are given below, and the absolute value of \( C \) is used to indicate the number of data points in \( C \).

Definition 1. Let \( A_1, \ldots, A_m \) denote a series of data points whose value domain is \( D_1, \ldots, D_m \) such that the result of clustering the database \( D \) is denoted \( C = \{ C_1, C_2, \ldots, C_k \} \); here, \( C_i \cap C_j \neq \emptyset \) and \( C_1 \cup C_2 \cup \cdots \cup C_k = D \), by the number of clusters \( k \).

Definition 2. Suppose \( C = \{ C_1, C_2, \ldots, C_k \} \) is the order of the clusters: \( |C_1| \geq |C_2| \geq \cdots \geq |C_k| \), given two numerical

| Serial number | Central value | Radius | Percent |
|---------------|---------------|-------|--------|
| Cluster 1     | 0.52,0.63     | 0.32  | 0.62   |
| Cluster 2     | 0.17,0.21     | 0.06  | 0.04   |
| Cluster 3     | 0.21,0.56     | 0.17  | 0.21   |
| Cluster 4     | 0.32,0.41     | 0.23  | 0.13   |

Figure 3: Generating data tables using control reversal reverse engineering.
parameters: \( \alpha \) and \( \beta \); we let \( b \) be the boundary of the size cluster, then
\[
\left( |C_1| + |C_2| + \cdots + |C_b| \right) \geq |D| \cdot \alpha,
\]
\[
|C_b|/|C_{b+1}| \geq \beta.
\]

The first part of equation (3) takes into account that the vast majority of data points in a data set are not outliers, so that a cluster that contains the vast majority of data points is a large cluster. The second part of equation (3) takes into account that the main difference between large and small clusters is the number of data points within the cluster, e.g., set \( \beta \) to 5, and then, any cluster in LC is 5 times larger than in SC [30].

Definition 3. Putting the above analysis together, assuming that \( C = \{C_1, C_2, \cdots, C_k\} \) is the order of clustering: \( |C_1| \geq |C_2| \geq \cdots \geq |C_k| \), and consistent in the definitions of \( p_{ij}(t) \), \( \alpha \), \( \beta \), \( b \), LC, and SC and setting the number of clusters in LC to \( n \), the resulting outlier index for any data point \( t \) is
\[
CBLOF(t) = |C_i| \cdot \min \left( \text{distance}(t, C_j) \right),
\]
where \( t \in C_i, C_i \in SC, C_j \in LC, j = 1 \) to \( b \).

According to Definition 3, the outlier index of a data point is determined by the size of the cluster in which it is located and the distance of this data point from the edge of its nearest large cluster.

4. System Design and Implementation

The process of using clustering models to guide production optimisation has been implemented in the previous work. The basic process is as follows: the clustering model is trained with a large amount of historical data of normal operation, and then, the corresponding actual data of the field operation is collected and put into the clusters of aggregation; if it is not included in any of the clusters, it means that it is an abnormality, and an alarm should be prompted to the staff to check the equipment, which serves as an early warning function. The detected anomalies belonging to any of the clusters mean that the equipment is operating normally and can be further explored in order to optimise the equipment.

4.1. Raw Data Preprocessing. Since the units of the parameters used in the design of the system vary considerably between orders of magnitude, the data must first be processed to make them dimensionless, and the standardised metric used in this paper is
\[
x_d = \frac{z_d - h_j}{S_j},
\]
where \( z_d \) is an actual parameter, \( h_j \) is the mean \( z_d \) of \( n \), and \( S_j \) is the mean standard deviation.

4.2. Analysis of the Off-Kernel Indices of the Parameter Points within Clusters. Using IAE at the Shanghai Jiaotong University as the application context, five years of relevant actual production data were collected, the data were processed with standardised metric values and stored in a database, and the corresponding parameters were trained to form a clustering model using a weighted K-mean algorithm. The similarity of the actual parameters within the clusters to the cluster centres was then analysed.

The similarity between the other parameters within the cluster and the central parameter is calculated using the correlation coefficient method and the Euclidean distance, respectively; the Euclidean distance has been described earlier, and the following equations are used for the correlation...
Figure 5: Ratio of the number of parameter points to the total number of clusters at different degrees of similarity.

Table 2: Relationship between anomalous parameter points and clustering models.

| Exceptional parameter points | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|------------------------------|-----------|-----------|-----------|-----------|
| 1                            | 0.65      | 0.51      | 0.75      | 0.58      |
| 2                            | 0.62      | 0.49      | 0.69      | 0.61      |
| 3                            | 0.59      | 0.41      | 0.71      | 0.56      |
| 4                            | 0.63      | 0.45      | 0.72      | 0.52      |
A pie chart of the radii and percentages of the four clusters is given in Figure 4. Then, the ratio of the number of equipment parameters to the number of parameters in the whole cluster was obtained for different correlation coefficients, and the ratio of the number of parameters in this point to the number of parameters in the whole cluster was obtained for different ratios of the distance from the parameter point to the centre of the cluster to the maximum radius, and the ratio was obtained in Figure 2(b). From Figure 2(a), it can be seen that the correlation with the central parameter is greater than 0.8 in more than 90% of the cases, while the proportion below 0.6 is less than 3%. Similar results are obtained for the other equipment parameters; the same analysis is also reflected in Figure 2(b). Examining these four clusters, the data points within 20% of the maximum radius each already account for more than 50% of all data points. This indicates that the equipment has been operating in a performance optimised condition for a long time and that once the equipment parameters have deviated from the optimised condition, the staff immediately call back to bring them back to the optimised condition, so that a large proportion of the parameters are close to the centre of the cluster and a small proportion are far from the centre. The above results show that the magnitude of the dispersion of these data is related not only to the similarity of this point to the central point but also to the proportion of the points occupied, such that the kernel dispersion index of the parameter points within the cluster can be defined as

$$SCAF(t) = \frac{r(t, C_j) \sqrt{p(C_j)}}{(1 - d(t, C_j))/r_j \sqrt{p(C_j)}}$$

where $p$ represents the number of device parameter points at point $t$.

The curve of where the data points are located and the error detection parameters in the system at this time is shown in Figure 5. Cluster 1 represents a normal and good operating condition of the system; cluster 2 is an overadjusted state with a low error rate and a good system; clusters 3 and 4 are the result of the system not being retuned in time, and clusters 3 and 4 also show that the marginal values (points with large dispersion) of the clusters are of poor error quality.

With the above analysis, after training the clustering model with a large amount of historical data of normal operation, when monitoring the operating conditions of the equipment in real time, when a real-time parameter point is within a cluster, this model can not only indicate in which cluster it is but also know the off-kernel index of the parameter point.
4.3. Analysis of Outlier Indices for Anomalous Parameter Points Outside Clusters. When the monitored real-time parameter point was outside the cluster, it was an anomaly. In order to be able to analyse the extent of the outliers of the anomaly, the outlier index $CBLOF(t)$ was first calculated using equation (3), and the propane content of the dry gas at this point was checked. As shown in Table 2, the outlier index 1 for the outlier point was larger than for the outlier point 2, but the propane content was not high at this time. The analysis shows that if outlier a is closer to a cluster with fewer parameter points and further away from a cluster with more and outlier b is closer to a cluster with more parameter points and further away from a cluster with fewer, $CBLOF(b) > CBLOF(a)$ is calculated, but in reality, outlier a may be more outlier than outlier b. The outlier index of an outlier is related not only to the distance of this outlier from the nearest cluster but also to the distance between this cluster and other clusters and the number of clusters within it. Also taking into account the weighted nature of clustering, a new definition of outliers for anomalous parameter points is obtained.

$$NCF(t) = \sum_{k} \sqrt{\frac{|c_j|}{N}} \min \{\text{distance}(t, c_j)\}. \quad (9)$$

When a parameter point is not in a cluster, a new outlier index can be used to analyse it, which gives insight into the system operation and provides a basis for monitoring the system operation in an optimised state.

Point pairs of anomalous parameters for four of the clustering models are shown in Figure 6.

Two sets of experiments were conducted using cluster analysis theory to cluster 20,000 data points from four clusters, and the detailed data distribution of cluster accuracy for the four clusters is given in Table 3. The comparative effect of the clustering accuracy data for the four clusters in the two sets of experiments is given in Figure 7.

5. Conclusion

The design and implementation of the IAE project management system is mainly based on the real demand for the development and management of university students’ IAE projects, to achieve coordinated management with the power of the system, thus ensuring that the teacher and student groups can participate in each project process, reducing the management workload and improving the comprehensive effect of management. Cluster analysis itself is a method of grouping data based on a certain similarity measure, and this paper adopts the weighted K-means algorithm based on the Euclidean distance to establish a clustering model by analysing a large amount of normal historical data and then, according to the actual operation, successively proposes the definition of the off-kernel index of parameter points within clusters and the definition of anomalous parameter points.
outside clusters. The definition of the outlier index of the parameter points within the cluster and the definition of the outlier index of the abnormal parameter points outside the cluster were then proposed according to the actual operation. The management system proposed in this paper can monitor the historical data of the system running in an optimised state, thus enabling the effective implementation of the innovation and entrepreneurship project, but due to the many uncontrollable factors in the actual operation, further development of the system is necessary to make the innovation and entrepreneurship project management system practically useful.

**Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**References**

[1] C. Jiang, W. Liu, and J. Zhang, “Risk assessment of generation and transmission systems considering wind power penetration,” *Transactions of China Electrotechnical Society*, vol. 29, pp. 260–270, 2014.

[2] M. Abdelrahem and R. Kennel, “Fault-ride through strategy for permanent-magnet synchronous generators in variable-speed wind turbines,” *Energies*, vol. 9, no. 12, pp. 1066–1070, 2016.

[3] M. Cao, Y. Qiu, Y. Feng, H. Wang, and D. Li, “Study of wind turbine fault diagnosis based on unscented Kalman filter and SCADA data,” *Energies*, vol. 9, no. 10, p. 847, 2016.

[4] S. J. Watson, B. J. Xiang, W. Yang, P. J. Tavner, and C. J. Crabtree, “Condition monitoring of the power output of wind turbine generators using wavelets,” *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 715–721, 2010.

[5] Z. Radoslav, B. Walter, and B. Tomasz, “Diagnostics of bearings in presence of strong operating conditions non- stationarity—a procedure of load-dependent features processing with application to wind turbine bearings,” *Mechanical Systems and Signal Processing*, vol. 46, no. 1, pp. 16–27, 2014.

[6] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, “Condition monitoring and fault detection of wind turbines and related algorithms: a review,” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 1–39, 2009.

[7] X. Y. Zhang, X. P. Zhang, and B. P. Su, “Vibrant fault diagnosis for hydro-turbine generating unit using minmax kernel K-means clustering algorithm,” *Power System Protection and Control*, vol. 43, no. 5, pp. 27–34, 2015.

[8] Z. Johanyák and O. Papp, “A hybrid algorithm for parameter tuning in fuzzy model identification,” *Acta Polytechnica Hungarica*, vol. 9, no. 6, pp. 153–165, 2012.

[9] R. Precup, R. David, E. Petriu, S. Preitl, and M. B. Rădac, “Novel adaptive charged system search algorithm for optimal tuning of fuzzy controllers,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 1168–1175, 2014.

[10] M. Kiran and O. Findik, “A directed artificial bee colony algorithm,” *Applied Soft Computing*, vol. 26, pp. 454–462, 2015.

[11] I. P. Solos, I. X. Tassopoulos, and G. N. Beliannis, “Optimizing shift scheduling for tank trucks using an effective stochastic variable neighbourhood approach,” *International Journal of Artificial Intelligence*, vol. 14, pp. 1–26, 2016.

[12] Q. Zhou, W. Sun, Y. Zhang, H. J. Ren, C. X. Sun, and J. Y. Deng, “A new method to obtain load density based on improved ANFIS,” *Power System Protection and Control*, vol. 39, pp. 29–34, 2011.

[13] O. Dragomir, F. Dragomir, and V. Stefan, “Adaptive neuro-fuzzy inference systems as a strategy for predicting and controlling the energy produced from renewable sources,” *Energies*, vol. 8, no. 11, pp. 13047–13061, 2015.

[14] A. Ahmed, S. Jun, M. Senior, and J. Yan, “Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning,” *IEEE Transactions on Learning Technologies*, vol. 5, no. 3, pp. 226–237, 2012.

[15] B. Chen, P. Matthews, and P. Tavner, “Automated on-line fault prognosis for wind turbine pitch systems using supervisory control and data acquisition,” *IET Renewable Power Generation*, vol. 9, no. 5, pp. 503–513, 2015.

[16] E. Lapira, D. D. Brisset, H. Ardakani, and J. Lee, “Wind turbine performance assessment using multi-regime modeling approach,” *Renewable Energy*, vol. 45, pp. 86–95, 2012.

[17] X. Zheng, M. Li, J. Wang, H. Ren, and Y. Fu, “Operational conditions classification of offshore wind turbines based on kernel principal analysis optimized by PSO,” *Power System Protection and Control*, vol. 44, pp. 28–35, 2016.

[18] Y. Li and R. Fang, “Reliability assessment for wind turbine based on weighted degree of improved grey incidence,” *Power Syst Prot Control*, vol. 43, pp. 12–69, 2015.

[19] Z. Yin, B. Han, and S. Xie, “An Improved Grey Correlation Algorithm and Its Application for Diesel Fault Prediction,” in *2015 Sixth International Conference on Intelligent Control and Information Processing (ICICIP)*, pp. 412–416, Wuhan, China, 2015.

[20] Y. Kong and M. Jing, “Research of the classification method based on confusion matrices and ensemble learning,” *Computer Engineering & Science*, vol. 34, no. 6, pp. 111–117, 2012.

[21] M. Bilal, M. Ullah, and H. Ullah, “Chemometric data analysis with autoencoder neural network,” *Electronic Imaging*, vol. 31, no. 1, pp. 679–1–6, 2019.

[22] S. Chatterjee and A. S. Hadi, “Estimation of the underlying structure of systematic risk with the use of principal component analysis and factor analysis,” *Contaduría y Administración*, vol. 59, no. 3, pp. 197–234, 2014.

[23] R. L. de Guevara Cortés, S. T. Porras, and E. M. Moreno, “Extraction of the underlying structure of systematic risk from non-Gaussian multivariate financial time series using independent component analysis: evidence from the Mexican stock exchange,” *Computación y Sistemas*, vol. 22, no. 4, 2018.

[24] R. Ladrón de Guevara Cortés, S. Torra Porras, and E. Monte Moreno, “Neural networks principal component analysis for estimating the generative multifactor model of returns under a statistical approach to the arbitrage pricing theory. Evidence from the Mexican stock exchange,” *Computación y Sistemas*, vol. 23, no. 2, 2019.

[25] K. Imaeo, K. Minami, K. Ito, and K. Nakagawa, “Deep portfolio optimization via distributional prediction of residual factors,” 2020, arXiv:2012.07245.
[26] A. Stacy, T. Perls, M. Nygaard, P. Sebastiani, and Q. Xiang, “Discovering modality of cognitive function using clustering analysis,” *Innovation in Aging*, vol. 5, Supplement 1, pp. 583-584, 2021.

[27] C. Iorio, G. Frasso, A. D’Ambrosio, and R. Siciliano, “Parsimonious time series clustering using p-splines,” *Expert Systems with Applications*, vol. 52, pp. 26–38, 2016.

[28] C. Iorio, G. Frasso, A. D’Ambrosio, and R. Siciliano, “A P-spline based clustering approach for portfolio selection,” *Expert Systems with Applications*, vol. 95, pp. 88–103, 2018.

[29] J. H. Kim, “Multicollinearity and misleading statistical results,” *Korean Journal of Anesthesiology*, vol. 72, no. 6, pp. 558–569, 2019.

[30] D. P. Kingma and M. Welling, “An introduction to variational autoencoders,” *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.