Proceedings of CONF-CDS 2025 Symposium: Data Visualization Methods for Evaluation
DOI: 10.54254/2755-2721/2025.PO24507

Research on the Impact of Digitization on Economic Growth Based on Principal Component Analysis

Jiale Tian

Huazhong University of Science and Technology, Wuhan, China
3295182429@qq.com

Abstract. In the era of the digital economy, digital technologies have deeply integrated into various sectors of the economic system. Analyzing the impact of digitization on economic development is of guiding significance for deepening economic reform. Existing studies often rely on single indicators to measure the level of digitization. Such methods are limited in their ability to comprehensively capture the systemic characteristics of digitization. Principal Component Analysis (PCA) provides an effective tool for addressing the issues of information redundancy and multicollinearity in multi-dimensional indicators. By performing dimensionality reduction through linear transformation, PCA can extract key common factors from digitization indicators and construct a comprehensive index, thereby providing a more systematic portrayal of the overall development level of digitization and its structural differences. To validate the effectiveness of PCA in this regard, this study performs principal component analysis on data from 130 regions and conducts comparisons and analysis for regions in Asia and some other areas. The conclusions derived from PCA are generally consistent with the actual conditions of the regions, demonstrating the efficacy of PCA in addressing these issues.

Keywords: Principal Component Analysis, Digital Economy Analysis, Dimensionality Reduction of Economic Structural Features

1. Introduction

In the era of the digital economy, digital technologies have deeply integrated into various sectors of the economic system, becoming a key driver of global economic growth. From the intelligent upgrading of infrastructure to the automation transformation of production processes, from the platform innovation of business models to the digital reshaping of consumer behaviors, digitization is reshaping the underlying logic of economic development by improving production efficiency, optimizing resource allocation, and expanding market boundaries. However, the mechanisms by which digitization affects economic growth are highly complex: its impact is not only reflected in the direct effects of technological penetration but also through indirect pathways such as industrial collaboration, knowledge spillovers, and institutional innovation. How to scientifically quantify the multidimensional characteristics of digitization and precisely identify its contribution to economic growth has become a central concern for both academics and policymakers.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.PO24507

This study uses the PCA method to analyze the factors influencing income for residents around the world and proposes methods to improve income levels. Traditional research often uses single indicators (such as internet penetration rate, R&D investment, or the scale of digital infrastructure) to measure the level of digitization. However, such methods are unable to fully capture the systemic characteristics of digitization. For example, the impact of digital technologies on the economy often depends on the collaborative effects of technological applications, data accumulation, human capital adaptation, and institutional environments. In this context, Principal Component Analysis (PCA) offers an effective tool for addressing issues of information redundancy and multicollinearity in multi-dimensional indicators. By performing dimensionality reduction, PCA can extract key common factors from digitization indicators and construct a comprehensive index, thereby offering a more systematic depiction of the overall level of digitization and its structural differences.

The article is structured into four main sections. The first section introduces the topic. The second section reviews some of the existing research. The third section discusses the analysis of the factors influencing economic growth based on PCA. The fourth section summarizes the contributions of the study.

2. Analysis of existing research

PCA, as a dimensionality reduction technique, reduces data redundancy by extracting principal components and has demonstrated its efficiency across various fields of study. When combined with other models, it can further enhance the model's performance. In the study by Wang Xiaohua and Chen Linfan on a financial early warning model for the new energy industry based on PCA-SVM [1], PCA was used to reduce 36 indicators in the new energy industry study to 10 principal components, which accounted for a cumulative contribution rate of 70.719%. These 10 indicators were then used to construct an SVM model, which achieved a prediction accuracy of over 85%. In the research by Jiang Yue and Qiao Yuyang on the evaluation of logistics enterprises' operational performance based on the PCA-DEA model [2], the DEA model was combined with PCA, and 6 principal components were extracted to replace 21 original indicators, effectively evaluating the comprehensive efficiency of logistics enterprises. They found that 14 enterprises achieved a technically efficient status. A similar approach was used in the study on the comprehensive efficiency evaluation of population, energy, economy, and environment in various regions of China based on PCA-DEA [3]. In the study by Zhang Junxiong and Wei Rong on the impact factors of low-carbon city construction based on PCA [4], PCA was used to extract 8 principal components from 26 indicators (with a cumulative variance contribution rate of 79.27%), identifying key factors such as the environment, energy, and policy. In the study by Gao Ting and others on the spatiotemporal characteristics and driving forces of cultivated land to construction land transformation in Shaanxi Province based on PCA-MGWR [5], PCA and MGWR were combined, selecting 22 factors in total from natural location, socio-economic, and agricultural aspects, and extracting 11 principal components for the MGWR model.

Among these methods, combining machine learning with PCA is considered an ideal approach. In the new energy industry study [1], the PCA-SVM model, using 10 principal components as input, achieved a prediction accuracy of over 85%, significantly outperforming traditional linear models (e.g., the Z-score model). In the study by Tang Shuyi on predicting total health expenditure in China using the PCA-BP neural network model [6], PCA was used to reduce 15 indicators to 3, and a BP neural network was then constructed to predict the total expenditure. In Ma Ruoting’s research on the PCA-GAT-based quantitative investment model [7], the PCA-GAT model reduced 68 factors to 13 and used graph neural networks for stock selection. The model’s performance surpassed that of

DOI: 10.54254/2755-2721/2025.PO24507

traditional models. The advantage of the SVM learning method lies in solving nonlinear separability issues through kernel functions, while also avoiding the heavy dependence of neural networks on large sample sizes. Compared to PCA-DEA (logistics performance evaluation [2]) and PCA-MGWR (cultivated land transformation analysis), PCA-SVM is more suitable for classification problems, whereas the latter are often used for efficiency evaluation or spatiotemporal driving force analysis.

Using PCA for analysis alone can also yield good results. In the study on a factor model for the Chinese stock market based on principal component analysis [8], through Fama-Macbeth regression, 25 investment portfolios were obtained. PCA was used to extract 3 principal components, revealing significant differences in the influencing factors between the Chinese and U.S. stock markets. In Zhang Tian's study on analyzing and forecasting the factors influencing residents' income based on PCA and regression combined models, PCA was used to extract 2 principal components from 9 influencing factors, offering feasible methods for improving residents' income [9].

Although PCA combined models have shown significant results, there are still many challenges. Some studies (e.g., financial early warning [1]) suffer from classification errors due to small sample sizes (15 companies) and require expansion to cross-industry and cross-period data to verify model robustness. Most studies rely on static data and do not adequately consider the time series characteristics. For instance, while the PCA-BP neural network model in health expenditure prediction [5] retains time features, it needs further integration with time series models like LSTM. Additionally, PCA dimensionality reduction may result in the loss of economic meaning of some indicators, requiring the use of interpretability tools such as SHAP values to strengthen the connection between model results and policy formulation.

3. Research content and methods

3.1. Summary

In order to explore the impact of digitalization on economic growth, this study performs principal component analysis (PCA) on the dataset "Database International Panel Data Analysis of the Effect of Digitalization on Economic Growth" collected from the Kaggle dataset website. The dataset contains data for 130 regions, with annual data collected from 2000 to 2020. Each year's data consists of 12 indicators. PCA was performed on the data for each country, with the goal of ensuring that the cumulative variance contribution rate of the extracted principal components reached 95%. The results show that for typical regions, the cumulative variance contribution rate of the first two principal components exceeds 95%, successfully reducing the data from 12 dimensions to 2. The paper then analyzes the factors affecting economic growth based on the PCA results and the actual conditions of each region, effectively demonstrating the good performance of PCA in data analysis.

3.2. Theory

PCA (Principal Component Analysis) is a dimensionality reduction method that represents data with fewer features while retaining the most important characteristic information (data compression; PCA is a form of lossy compression). The main goal of PCA is to find a set of new orthogonal bases $\{u_1, u_2, \cdots, u_k\}$ (reducing from n dimensions to k dimensions) that maximize the variance of the data after projection. If the variance of the data projected onto each orthogonal basis is maximized, the projection distance on the plane formed by the orthogonal basis will also be maximized.

For a base u_j, the projection distance of data point x_i onto this base is $x_i^T \bullet u_j$ (data is already centered), so the variance of all data projected onto this base is:

DOI: 10.54254/2755-2721/2025.PO24507

$$J_j = \frac{1}{m} \sum_{i=1}^{m} \left(x_i^T \bullet u_j - x_c^T \bullet u_j \right)^2 \tag{1}$$

Where m is the number of samples. Since the data has been centered, $\mathrm{x_c^T} = 0$, so:

$$J_j = \frac{1}{m} \sum_{i=1}^{m} \left(x_i^T \bullet u_j \right)^2 = u_j^T \bullet \frac{1}{m} \sum_{i=1}^{m} \left(x_i x_i^T \right) \bullet u_j = u_j^T \bullet \frac{1}{m} X X^T \bullet u_j \tag{2}$$

Where $\mathrm{X} = [\mathrm{x_1},\ \mathrm{x_2}, \cdots, \mathrm{x_m}]$, and $\mathrm{S} = \frac{1}{\mathrm{m}} \mathrm{X X^T}$. To maximize $\mathrm{J_j} = \mathrm{u_j^T} \bullet \mathrm{S} \bullet \mathrm{u_j}$, the goal is to find $\mathrm{u_j}$ such that:

Subject to:

$$max J_j = u_j^T \bullet S \bullet u_j \tag{3}$$

$$s.t. \quad u_j^T u_j = 1 \tag{4}$$

Constructing the Lagrange function:

$$F(u_j) = u_j^T \bullet S \bullet u_j + \lambda_j \left(1 - u_j^T u_j \right) \tag{5}$$

Taking the derivative:

$$\frac{\partial F}{\partial u_j} = 0 \qquad \text{yields} \qquad S u_j = \lambda_j u_j \tag{6}$$

Thus, when λ_j and $\mathrm{u_j}$ are the eigenvalues and eigenvectors of S, the maximum value is:

$$J_{j\,m} = u_j^T \lambda_j u_j = \lambda_j \tag{7}$$

Thus, for any orthogonal basis that satisfies the condition, the variance of the data after projection onto that basis corresponds to the eigenvalues of the matrix S . Therefore:

$\mathrm{J_{max}} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \lambda_{(\mathrm{j})}$, where λ are sorted in descending order.

Thus, the projection orthogonal basis consists of the eigenvectors corresponding to the top k largest eigenvalues of S.

Performing Singular Value Decomposition (SVD) on $\mathrm{S} = \frac{1}{\mathrm{m}} \mathrm{X X^T}$ yields the eigenvector set $\mathrm{U} = \{\mathrm{u_1}, \mathrm{u_2}, \cdots, \mathrm{u_k}\}$

Therefore, the relationship between the reduced-dimensional data and the original data is:

$$X_{n \times m}^{new} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_k^T \end{bmatrix}_{k \times n} \bullet X_{n \times m} \tag{8}$$

DOI: 10.54254/2755-2721/2025.PO24507

3.3. Data analysis

Terminology Definitions:

•FPS (Fixed Phone Subscriptions) refers to traditional landline phone services subscribed to by users through telecom operators, typically connected via physical lines (such as copper cables or fiber optics), commonly found in households, businesses, or institutions.

•MPS (Mobile Phone Subscriptions) refers to wireless communication services subscribed to by users through cellular network operators, enabling voice calls, SMS, and data usage through mobile devices.

•BS (Broadband Subscription) refers to high-speed internet access services subscribed to by users via either wired or wireless technologies, supporting high-capacity data transmission (e.g., video streaming, online gaming, remote working).

•IU (Internet Use) refers to activities where individuals, businesses, or institutions access information, services, or interact via the internet, encompassing a broad range of activities from basic browsing to more complex digital engagements.

•GFCF (Gross Fixed Capital Formation) refers to the total value of fixed assets purchased, transferred, or self-produced for use by resident units, excluding the sale or transfer of fixed assets, and is divided into tangible and intangible fixed capital formation.

•TO (Trade Openness) measures the extent to which a country or region participates in international trade, reflecting its economy's openness and reliance on cross-border goods, services, and capital flows.

•LCPI (Local Consumer Price Index) is a macroeconomic indicator that measures the price level of a basket of goods and services over time, reflecting the extent of inflation or deflation.

•LPOP (Local Population) refers to the number of people residing in a specific area.

This paper analyzes data from countries and cities in the Asian region and some other regions selected from the 130 countries in the dataset. The PCA results show that for each Asian region, the cumulative variance contribution rate of the first two principal components exceeds 95%, effectively reducing the dimensionality of the data.

Results for the Asian regions are as follows: For Singapore, the first principal component has the highest contributions from LPOP, labor, and RD, followed by GFCF, LCPI, and government consumption. The second principal component has the highest contributions from FPS, BBS, and GFCF, followed by MPS, IU, labor, and RD. For South Korea, the first principal component has the highest contributions from BBS, IU, and RD, followed by MPS, GFCF, labor, and LCPI. The second principal component has the highest contributions from FPS, BBS, and IU, followed by MPS, DDI, LCPI, and RD. For Japan, the first principal component has the highest contributions from BBS, IU, and RD, with GFCF contributing second. The second principal component has the highest contributions from MPS, BBS, and IU, followed by FPS. For Hong Kong, the first principal component has the highest contributions from DDI and government consumption, followed by IU, GFCF, and labor. The second principal component has the highest contributions from FPS and BBS, followed by MPS and IU.

Results for other regions are as follows: For Australia, the first principal component has the highest contributions from BBS, IU, and RD, followed by FPS, MPS, DDI, and GFCF. The second principal component has the highest contributions from MPS, BBS, and IU, followed by FPS. For New Zealand, the first principal component has the highest contributions from MPS and RD, followed by BBS, DDI, GFCF, and TO. The second principal component has the highest contribution from BBS, followed by FPS and MPS.

DOI: 10.54254/2755-2721/2025.PO24507

Table 1: The composition of the first two principal components in each region

area		FPS	MPS	BBS	IU	DDI	GFCF	TO	Labor	LCPI	LPOP	consum	RD
Singapore	PC1	0.076588	-0.00678	-0.045406	0.152462	-0.041612	-0.288541	-0.090606	-0.392396	0.205265	-0.67635	0.029583	0.470268
	PC2	-0.456354	0.18448	-0.475672	0.266876	-0.024954	0.591058	-0.128119	-0.222763	0.027907	0.077278	-0.002446	0.193533
South Korea	PC1	0.012013	0.243608	-0.430369	0.602296	-0.064961	-0.27552	0.001694	0.181881	0.239835	0.021371	-0.013004	0.470289
	PC2	0.386023	-0.255806	-0.557484	-0.536285	-0.277912	0.038358	-0.106712	-0.073473	0.22758	0.045719	-0.017995	0.193548
Australia	PC1	-0.148528	0.167455	0.550539	-0.548716	-0.17132	-0.281713	-0.065519	0.007412	-0.105248	-0.019697	-0.006546	-0.470264
	PC2	0.344316	-0.58548	0.505171	0.459566	-0.001999	-0.098863	-0.087099	0.118035	0.014868	0.055181	-0.015622	-0.193533
Japan	PC1	0.018915	0.122845	0.652774	-0.495094	-0.239543	-0.174609	0.063438	-0.000487	-0.006978	0.015163	-0.470263	-0.000006
	PC2	0.410833	0.805234	-0.236504	-0.002246	-0.011376	0.281081	0.024202	-0.0961	-0.01441	0.01603	-0.193531	-0.000003
New Zealand	PC1	0.075928	0.702071	0.219905	-0.043703	0.267153	0.254163	0.286935	-0.095559	-0.02998	-0.030521	-0.025501	0.470265
	PC2	-0.418024	-0.320758	0.749226	0.207202	-0.007075	-0.130213	0.128324	-0.024326	-0.211609	-0.021403	0.036671	0.193528
Hong Kong	PC1	0.002968	0.057995	-0.06511	0.226528	-0.772767	0.268133	0.120118	0.184391	0.04851	0.009301	0.470257	-0.000027
	PC2	0.568499	-0.337139	-0.579996	-0.380103	0.089277	0.115163	0.151195	-0.026831	0.001899	-0.00011	0.193536	-0.000011

The yellow highlight represents the highest contribution rate, while the bold font represents the second highest contribution rate.

In the data from the Asia region, the common characteristic is that RD typically has a high contribution rate in the first principal component, while FPS and BBS are the main influencing factors in the second principal component. The situation in Hong Kong is somewhat different. The main influencing factors in its first principal component are DDI and government consumption. This could be attributed to China's designation of Hong Kong as a Special Administrative Region (SAR), with the Chinese government having a unique development plan for Hong Kong. Every year, the government invests funds in building Hong Kong, which leads to government consumption becoming a key influencing factor.

The results for Australia show that BBS and IU are the main influencing factors, closely linked to its geographical location. Situated between the South Pacific and the Indian Ocean and surrounded by the sea, the internet naturally becomes one of the primary channels for communication and trade with the rest of the world. As a result, BBS and IU have had a significant impact on Australia's economic growth.

For New Zealand, the main influencing factors are MPS and RD. Despite being located in the southwestern Pacific like Australia, the main factors differ entirely from those in Australia. There are notable differences between Australia and New Zealand in terms of the digital economy, mainly reflected in policy orientation, industrial structure, international cooperation, and applications in specific sectors. New Zealand's digitalization process is characterized by policy leadership and a unique strategic framework. It joined the Digital Economy Partnership Agreement (DEPA) early on, emphasizing flexible cooperation and modular governance to promote innovations in areas such as digital identity and cross-border data flow. In contrast, Australia focuses more on industrial

DOI: 10.54254/2755-2721/2025.PO24507

digitalization and market-driven processes. However, its digitalization policies are decentralized, and there is a lack of unified privacy legislation at the federal level, relying primarily on corporate-driven innovation.

Table 2: The output results of Singapore and Hong Kong

Singapore	FPS	MPS	BBS	IU	DDI	GFCF	TO	Labor	LCPI	LPOP	consum	RD
PC1	0.076588	-0.00678	-0.045406	0.152462	-0.041612	-0.288541	-0.090606	-0.392396	0.205265	-0.67635	0.029583	0.470268
PC2	-0.456354	0.18448	-0.475672	0.266876	-0.024954	0.591058	-0.128119	-0.222763	0.027907	0.077278	-0.002446	0.193533
PC3	-0.129083	0.06917	-0.041785	0.140251	-0.163051	-0.317273	0.139152	0.406066	-0.093829	0.371799	-0.018142	0.705881
PC4	-0.243414	0.027692	0.571327	-0.438702	0.415421	0.287699	-0.061811	-0.141611	-0.034323	0.027055	0.007589	0.37406
PC5	-0.729877	0.285445	0.220425	0.181309	0.049489	-0.428441	0.023947	0.018758	0.071188	-0.095179	0.008641	-0.321253
PC6	-0.004486	-0.009531	-0.552482	-0.243773	0.727233	-0.300646	-0.037392	0.062465	0.083835	0.059805	0.008192	-0.000012
PC7	0.314214	0.935276	-0.010097	-0.15317	-0.033553	0.020373	0.032203	0.001749	0.001461	-0.018952	0.004723	-0.000003
PC8	-0.061558	-0.014463	-0.125051	-0.145961	-0.079206	-0.171731	0.42744	-0.535595	-0.665772	0.099285	0.024831	-0.000035
PC9	-0.18426	-0.043576	-0.202002	-0.650535	-0.488023	-0.181592	-0.403616	-0.132958	0.17023	0.135416	-0.033273	0.000021
PC10	-0.20159	-0.030807	-0.181492	-0.360015	-0.136587	0.216799	0.476009	0.46971	-0.057627	-0.528423	0.020976	0.000003
PC11	0.010612	-0.031485	0.014007	-0.059358	-0.050938	0.011258	0.609293	-0.283637	0.676796	0.277176	0.076735	-0.000061
PC12	0.000206	0.003153	0.003144	0.002758	0.018754	-0.000927	0.075835	-0.0398	0.037462	-0.019105	-0.995247	0.000693
Hong Kong	FPS	MPS	BBS	IU	DDI	GFCF	TO	Labor	LCPI	LPOP	consum	RD
PC1	0.002968	0.057995	-0.06511	0.226528	-0.772767	0.268133	0.120118	0.184391	0.04851	0.009301	0.470257	-0.000027
PC2	0.568499	-0.337139	-0.579996	-0.380103	0.089277	0.115163	0.151195	-0.026831	0.001899	-0.00011	0.193536	-0.000011
PC3	0.072594	-0.001781	0.231596	-0.142174	0.14393	-0.290366	-0.539667	-0.152561	-0.055393	-0.004406	0.705865	-0.000041
PC4	0.15024	-0.077025	0.388969	0.269767	0.414654	0.088773	0.651499	0.05306	0.01933	-0.003852	0.374099	-0.000022
PC5	0.681276	-0.21181	0.517083	0.104354	-0.278303	-0.072804	-0.160146	-0.019651	-0.027227	0.00036	-0.321259	0.000019
PC6	-0.008179	0.022546	-0.148506	0.132507	-0.256216	-0.597758	0.335211	-0.651122	0.02291	-0.000644	0.000011	0
PC7	0.414618	0.891185	-0.136637	0.06094	0.104442	0.003426	-0.002873	0.023613	0.003423	-0.000077	0.000001	0
PC8	-0.00276	0.000858	0.02506	0.161315	0.077551	0.664635	-0.161008	-0.699877	-0.099134	-0.002952	-0.000015	0
PC9	0.108919	-0.190398	-0.381221	0.802442	0.193466	-0.131747	-0.26557	0.145934	-0.126393	0.007522	0.000009	0
PC10	0.031214	-0.034679	-0.018629	0.09539	0.068251	0.030426	-0.112	-0.055765	0.983091	-0.025906	-0.000068	0
PC11	0	0	0	0	0	0	0	0	0	0	-0.000058	-1
PC12	0.000696	-0.000185	0.00523	-0.004774	0.009917	-0.000036	-0.002076	-0.007208	0.025541	0.999571	0.000312	0

4. Conclusion

In conclusion, this study employs the PCA method to analyze the impact of digitalization on economic growth, reducing the 12 independent variables (FPS, MPS, BBS, etc.) in the dataset to two principal components, with a cumulative variance contribution rate reaching 95%. The study compares the similarities and differences in influencing factors between different regions (such as Singapore and Hong Kong in Asia, and New Zealand and Australia from other regions). The accuracy of the PCA method in data analysis has been validated.

Although the PCA method has yielded good results, there are still many aspects of this study that could be improved due to time and resource limitations. In terms of dataset selection, due to data source constraints, this study used the publicly accessible Kaggle website for open-source datasets. However, the dataset is small and does not comprehensively account for factors from all perspectives (such as…). If researchers wish to delve deeper into this issue, further investigation in this area could be explored.

DOI: 10.54254/2755-2721/2025.PO24507

References

[1] Wang, X., & Chen, L. (2024). A study on the financial early warning model for the new energy industry based on PCA-SVM. Business Observation, 23, 52-55.
[2] Jiang, Y., & Qiao, Y. (2024). Performance evaluation of logistics companies based on the PCA-DEA model. Logistics Technology, 5, 44-47+54.
[3] Wu, J. (2014). A study on the evaluation of comprehensive efficiency in population, energy, economy, and environment across various regions of China based on PCA-DEA (Master's thesis, Jiangxi University of Finance and Economics). Master's thesis. Retrieved from https: //kns.cnki.net/kcms2/article/abstract?v=6h6U53PWxNTg73nESGJuJAbMOWV9UQeQhuy6p6C7lSAovGxENmMFkNoV2RC_jdFAmg84SaHXvmCsz9mflM0OPt0EYf2cR2Bx6TiwvVgilTGQC1nv4an9etIesjM6HFA_l7s0-L_f-rfOS44JpZYb50X0pIo2ByM9XNidJ2zm5zcVFDBYqeYYLvea7Q0IPTPJgyimJOs1Lfw=& uniplatform=NZKPT& language=CHS
[4] Zhang, J., & Wei, R. (2023). A study on the influencing factors of low-carbon city construction based on PCA algorithm. Science and Technology & Industry, 22, 67-72.
[5] Gao, T., Chai, H., Wang, J., & Zhang, X. (2023). Analysis of the spatial and temporal characteristics and driving forces of cultivated land-construction land transformation in Shaanxi Province based on PCA-MGWR. Henan Science and Technology, 23, 95-101.
[6] Tang, S. (2021). Forecasting of total health expenses in China based on the PCA-BP neural network model (Master's thesis, Beijing Jiaotong University). Master's thesis. Retrieved from https: //link.cnki.net/doi/10.26944/d.cnki.gbfju.2021.000952
[7] Ma, R. (2024). A study on quantitative investment models based on PCA-GAT (Master's thesis, Xi'an International Studies University). Master's thesis. Retrieved from https: //link.cnki.net/doi/10.27815/d.cnki.gxawd.2024.000996. doi: 10.27815/d.cnki.gxawd.2024.000996
[8] Peng, C. (2023). A study on the factor model of the Chinese stock market based on principal component analysis (Master's thesis, Shanghai University of Finance and Economics). Master's thesis. Retrieved from https: //link.cnki.net/doi/10.27296/d.cnki.gshcu.2023.002173. doi: 10.27296/d.cnki.gshcu.2023.002173
[9] Zhang, T. (2023). Analysis and prediction of factors influencing residents' income based on PCA and regression combination models: A case study of Tibet Autonomous Region. Commercial Exhibition Economics, 18, 18-21. doi: 10.19995/j.cnki.CN10-1617/F7.2023.18.018