

Innovative Research on Engineering Cost Control Based on Big Data Analysis Method

Tianyu Zhang*

College of Investment and Construction Management, Dongbei University of Finance and Economics, Dalian, China

*Corresponding Author: esther37011@gmail.com

Abstract. Traditional cost control methods in the field of engineering management often rely on experience judgment and limited data, which is challenge to satisfy the needs of increasingly complex engineering projects. This paper mainly explores the use of big data analysis methods in the field of cost control of engineering projects. First, data collection is carried out in combination with Web Crawler, and then data cleaning, data storage, data display, etc. Then, among the many available data, a project is selected as a typical case for Earned Value Management (EVM). Through data analysis of three basic parameters and four evaluation indicators, it is concluded that the project may have the risk of overspending and delay. The results show that the effective processing of the collected data not only improves the quality and availability of the data, but also lays a firm foundation for the data analysis of engineering cost control. At the same time, EVM analysis outcomes have important early warning significance for the management team of the project, which can help timely adjust strategies and avoid potential economic losses. This research combines the current big data boom and effectively integrates it into engineering management related fields. The field's research horizon is enhanced, as well as new tools and methods are provided to the practice community. As big data technology continues to develop and improve, its application prospects in engineering management are broad, and it is expected to play a greater role in the future project cost control.

Keywords: Big data analysis method; Earned Value Management; Risk warning.

1. Introduction

With the advent of the era of big data, data has become a key resource to promote social development and innovation [1]. In the field of engineering construction, cost control is the core of project management, which is directly related to the return on project investment and the competitiveness of enterprises [2]. Therefore, the innovative combination of big data analysis methods and engineering cost control has practical significance for meeting the needs of future projects. At the same time, it also has important theoretical significance in improving the accuracy of cost forecasting, optimizing the allocation of resources and reducing risks.

At present, engineering cost control is faced with many challenges [3]. First of all, the scale and complexity of engineering projects are increasing, and traditional cost control methods are difficult to adapt to the rapidly changing market demand and technological progress. Secondly, the phenomenon of information island is widespread, which leads to low efficiency of data sharing and utilization. In addition, the increase of uncertain factors in the process of cost control, such as material price fluctuations, labor cost changes, increases the difficulty of cost forecasting and management. Therefore, how to use big data technology to improve the accuracy and adaptability of cost control has become an urgent problem in the field of engineering management.

The purpose of this study is to explore the application of big data technology in engineering cost control, in order to realize the innovation of it. Firstly, a large amount of data in engineering projects will be integrated and analyzed. Secondly, big data technology will be used to help improve the accuracy of cost data, and finally, in the cost control process of a project, Earned Value Method will be used to predict risks and other aspects. Through these steps, this study is expected to provide new perspectives and solutions in the field of engineering cost control.



2. Theoretical Basis

Big data refers to the collection of data that cannot be captured, managed and processed with conventional software tools within a certain time frame, and it represents a massive, high-growth and diversified information asset [4]. These datasets require new processing patterns to enable greater decision-making, insight, and process optimization.

2.1. Characteristics of Big Data

Big data typically has a 5V characteristic, as shown in Table 1 [5].

Table 1. 5V Characteristics of Big Data

Characteristic	Concept	Effect	Influence
Volume	Size of data	Provides rich data resources	It promoted the development of storage technology and data processing architecture
Velocity	The rate at which data is generated and flowed	Ability to respond quickly to market changes and user behavior	It promotes the development of stream processing technology and improves the timeliness of decision making
Variety	Diversity of types and sources of data	Adapts to different types of data and provides a more comprehensive perspective	It promoted the development of data integration and Extract, Transform, Load (ETL) technology
Value	The amount of information that is meaningful and useful for decision making	Increase efficiency, reduce costs, and create new business opportunities and competitive advantages	It has promoted the development of data mining, machine learning and artificial intelligence technologies
Veracity	Quality and reliability of data	Ensure reliability and avoid bad decisions	The importance of data cleaning, validation and quality management is emphasized.

The processing process of big data mainly includes the following steps, as shown in Fig 1.

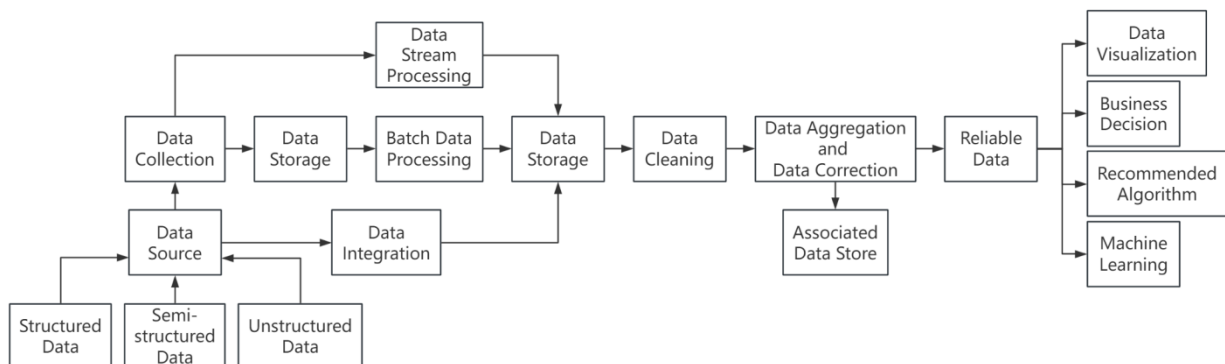


Figure 1. Data Processing Flow

In the process of big data processing, the first step is data collection, which involves obtaining data from different data sources, including structured data, semi-structured data and unstructured data. These data sources may include social media, transaction records, sensor data, etc. Part of the data obtained from the data sources can also be data integrated and directly stored. In addition, part of the data collected can be entered into the data stream processing, and the rest of the data collected will then enter the data storage stage. Here the data is stored in a suitable storage system for subsequent processing and analysis. This is followed by data batch processing and then data storage [5]. The next stage may involve initial processing of the data, which is cleaned to further ensure its accuracy and trustworthiness. In this step, recommendation algorithms may be used to help identify and correct problems in the data to ensure the quality and reliability of the data. Then through systematic data aggregation and correction, data from different sources and types are merged and unified to form a complete and more consistent dataset. The processed data is stored as reliable data, which is the basis for subsequent analysis and mining. The last is to let the data play a different role, there is data visualization, through the analysis of the results through the form of graphs or charts, so that non-technical users can understand and use the data. There is also processed and analyzed data to support business decisions, and data-driven insights to help organizations make more informed decisions. There is also the flow of data between the various stages, constantly being processed, analyzed, and utilized to support organizational goals and strategies, and so on.

2.2. The Benefit Evaluation Method of Building Engineering

The cost of housing construction project mainly includes the following aspects: (1) Direct cost: including labor cost, material cost, mechanical cost and other costs directly used for project construction. (2) Indirect costs: including management fees, financial fees, taxes and other costs that are not directly included in the project entity. (3) Risk cost: Additional costs that may be caused by uncertainties, such as design changes, construction delays.

The benefit evaluation of building engineering usually includes two aspects: economic benefit and social benefit [6]. Economic benefits are mainly measured by financial indicators, such as return on investment, net present value. Social benefits include non-financial indicators such as improving the living environment of residents and enhancing the image of the city.

The existing cost-benefit analysis methods of housing construction projects have the following limitations:

- (1) Data collection difficulties: it is difficult for traditional methods to collect all relevant data comprehensively and accurately.
- (2) Low processing efficiency: Faced with massive data, traditional methods are inefficient and difficult to analyze in real time.
- (3) Low prediction accuracy: Due to incomplete data or limited analysis methods, the accuracy of cost prediction and benefit evaluation is low.
- (4) Insufficient decision support: It is difficult for traditional methods to provide comprehensive and in-depth decision support information.

2.3. Application on Big Data Technology in the Field of Building Engineering

At present, the application of big data technology in the field of housing construction engineering is more and more extensive, and remarkable results have been achieved [7].

(1) Cost forecasting and control

Through the collection and analysis of historical project data, market data, etc., the cost forecast model is constructed. In-depth analysis of historical project data and market trends enables the construction of more accurate cost forecasting models. Accurate prediction and control of housing construction project cost can help project managers adjust resource allocation in time and reduce cost

risks [6]. This considers not only the cost of materials, labor and overhead, but also the impact of time value and market fluctuations. Project managers can use these models to predict future cost changes, adjust budget and resource allocation strategies in a timely manner, and effectively reduce the risk of cost overruns. In addition, by continuously monitoring cost data, managers are able to quickly identify cost deviations and take preventive or corrective actions to ensure that project costs are kept within predetermined limits.

(2) Engineering optimization

The application of big data technology is not limited to cost control, it can also significantly improve the efficiency and safety of the construction process [1]. The big data technology for real-time monitoring and analysis of construction progress, quality, safety, etc. is used to find potential problems and solve them in time. This helps to improve construction efficiency and quality and reduce construction risks. By collecting various data from the construction process in real time, such as project progress, material usage, equipment status and worker safety records, advanced analytical tools can be used to monitor construction status. Once the construction progress is found to be behind or there is a safety hazard, the system will automatically issue an early warning, prompting the project team to respond quickly and take measures to solve the problem. This intelligent construction management helps reduce construction delays, improve project quality, and ensure the safety of the construction site.

(3) Materials procurement and management

When it comes to material procurement and management, big data analysis can help companies deeply understand the dynamics of the supply chain and predict material demand and price fluctuations. Through big data analysis of supply chain information, optimize material procurement planning and management processes, reduce procurement costs and improve the efficiency of material use, and by analyzing historical procurement data and market information. Enterprises can optimize inventory levels and reduce inventory costs, while ensuring the timeliness and cost effectiveness of material supply [3]. In addition, through the evaluation of supplier performance, enterprises can choose the best supplier partners to further improve the efficiency and quality of material procurement.

(4) Market analysis and decision support

The complexity of the real estate market and the construction industry requires developers and construction companies to make decisions based on accurate market information. By analyzing big data of the real estate market and construction industry, it provides market insight and decision support information to developers and construction companies. This helps companies more accurately target markets, develop growth strategies, and improve return on investment. Big data analytics provides in-depth market insights, including consumer behavior, competitive environment, and macroeconomic trends. This information can help enterprises better understand market demand, predict market trends, and formulate effective market entry strategies and investment plans accordingly. In this way, enterprises can improve market response speed and optimize resource allocation, thereby gaining an advantage in a highly competitive market.

In short, the application of big data technology in the field of building engineering provides project managers with more comprehensive and in-depth data support and analysis capabilities, which helps to achieve the optimization of cost effectiveness, and provides in-depth insight and decision support in many aspects of construction management, material procurement and market analysis, thus promoting the innovation and development of the entire industry.

3. Research Methods

Earned Value is a cost and schedule control model designed to effectively measure the amount of work planned to be done and the amount of work already done [6]. In fact, Earned Value Management (EVM) is a mature management system based on the joint monitoring of cost and schedule, and

evaluates project performance and progress by comparing the actual work done with the original budget. EVM is a complete and effective project monitoring index and method. In fact, it is a method to analyze the difference between goal implementation and goal expectation, so it is often called deviation analysis [8].

3.1. Three Basic Parameters of EVM

(1) Budgeted Cost of Work Scheduled (BCWS)

The budgeted cost of the work (or part of the work) that should be done at a given time according to the schedule. In general, BCWS should remain unchanged during the implementation of the project unless the contract is changed. The project work budget is a benchmark to measure the progress and cost of the project.

$$BCWS = PW \times EUP. \quad (1)$$

where PW indicates planned workload and EUP indicates estimated unit price.

(2) Budgeted Cost of Work Performed (BCWP)

Refers to the total amount of money required for the work (or part of the work) that has been completed at a given time, based on the approved budget. Since it is based on this value that the employer pays the corresponding fee for the workload completed by the contractor, that is, the amount of money that the contractor gets (earned), so it is called Earned Value. It reflects the actual progress of the project that meets the quality and the performance of the completed work, and reflects the transformation of capital investment into the project results.

$$BCWP = CW \times EUP. \quad (2)$$

where CW indicates completed workload.

(3) Actual Cost of Work Performed (ACWP)

The total amount of money spent on the work or part of the work done up to a certain point in time. In numerical terms, ACWP is equal to the sum of the product of the completed works and the actual unit price of the project.

$$ACWP = CW \times AUP. \quad (3)$$

where AUP indicates actual unit price.

3.2. Four Evaluation Indicators of EVM

Based on these three basic parameters, four evaluation indexes of EVM can be determined, which are all functions of time [9].

(1) Cost Variance (CV)

$$CV = BCWP - ACWP. \quad (4)$$

When $CV < 0$, it means that the current cost control effect is not good, and the project operation exceeds the budgeted cost, that is, the actual cost of the project exceeds the established value.

When $CV > 0$, it means that the actual cost is lower than the budget and the project operation is economical, that is, the actual cost of the project is saved and the actual cost does not exceed the budget cost.

When $CV=0$, it means that the actual expense is equal to the budget.

(2) Schedule Variance (SV)

$$SV = BCWP - BCWS. \quad (5)$$

When $SV<0$, it means that the actual progress is less than the planned progress. It indicates that the progress is delayed, that is, the actual progress is behind the planned progress.

When $SV>0$, it means that the actual progress exceeds the planned progress. It indicates that the progress is ahead of schedule, that is, the actual progress is faster than the planned progress.

When $SV=0$, it means that the actual progress is consistent with the planned progress.

(3) Cost Performance Index (CPI)

$$CPI = BCWP / ACWP. \quad (6)$$

When $CPI<1$, it means that the actual cost of the project is higher than the budgeted cost, and the project is at risk of loss.

When $CPI>1$, it means that the actual cost of the project is lower than the budgeted cost, and the project has the possibility of profit.

When $CPI=1$, it means that the actual cost of the project is consistent with the budgeted cost, and the project is performing normally.

(4) Schedule Performance Index (SPI)

$$SPI = BCWP / BCWS. \quad (7)$$

When $SPI<1$, it means that the progress is delayed, and the actual progress of the project is slower than the planned progress.

When $SPI>1$, it means that the progress is ahead of schedule, and the actual progress of the project is faster than the planned progress.

When $SPI=1$, it means that the actual project progress is consistent with the planned progress.

4. Results and Discussion

4.1. Data Processing

Data acquisition is the first step in big data analytics processing and usually involves collecting data from a variety of data sources and storing it into a unified data platform. Data acquisition may involve crawler technology, API interface call, etc. The data collected is often incomplete, duplicate, or contains errors. Therefore, the data needs to be cleaned before data analysis to ensure the quality and accuracy of the data. Data cleaning usually involves removing duplicate data, filling in missing values, handling outliers, and so on [7]. After data cleaning is complete, the data is saved to a proper data storage engine for subsequent data analysis and display. Common data storage engines include relational databases, NoSQL databases, and data warehouses.

4.2. Data Analysis

Data acquisition is the first step in big data analytics processing and usually involves collecting data from a variety of data sources and storing it into a unified data platform. Data acquisition may involve crawler technology, API interface call, etc. The data collected is often incomplete, duplicate, or

contains errors. Therefore, the data needs to be cleaned before data analysis to ensure the quality and accuracy of the data. Data cleaning usually involves removing duplicate data, filling in missing values, handling outliers, and so on. After data cleaning is complete, save the data to a proper data storage engine for subsequent data analysis and display. Common data storage engines include relational databases, NoSQL databases, and data warehouses [10].

4.2.1. Work breakdown structure.

Work Breakdown Structure is the process of breaking down a project into tasks according to certain principles, breaking down tasks into tasks, and finally assigning these tasks to each person's daily activities until they can no longer be broken down, as shown in Table 2.

Table 2. Work Breakdown Structure

Number	Part Project	Subdivisional Work	Item Coding
1	1. Foundation Engineering	Earth-Rock Works	1.1
2		Foundation Pit Support Works	1.2
3		Foundation Pit Dewatering Engineering	1.3
4		Pile Foundation Project	1.4
5	2. Main Structure Engineering	Concrete Work	2.1
6		Steel Structure Work	2.2
7		Formwork	2.3
8		Masonry Work	2.4
9	3. Decoration Work	Floor, Roof Decoration Project	3.1
10		Waterproof Engineering	3.2
11		Insulation Works	3.3
12	4. Mechanical and Electrical Installation Engineering	Strong and Weak Electricity Engineering	4.1
13		Water Supply and Sewerage Works	4.2
14		Ventilation Duct Works	4.3
15		Civil Air Defense Engineering	4.4
.....

4.2.2. Cost calculation.

The actual project lasts 360 days, and the planned working period of the project up to this node is shown in Table 3.

Table 3. Planned Operation Duration

Number	Item Coding	Subdivisional Work	Planned days (a)	Planned days up to the current time (b)	Period proportion of planned construction days (c=b/a) /%
1	1.1	Earth-Rock Works	389	270	69.41%
2	1.2	Foundation Pit Support Works	501	500	99.80%
3	1.3	Foundation Pit Dewatering Engineering	482	433	89.83%
4	1.4	Pile Foundation Project	493	457	92.70%
5	2.1	Concrete Work	355	151	42.54%
6	2.2	Steel Structure Work	342	165	48.25%
7	2.3	Formwork	333	160	48.05%
8	2.4	Masonry Work	298	103	34.56%
9	3.1	Floor, Roof Decoration Project	302	23	7.62%
10	3.2	Waterproof Engineering	120	10	8.33%
11	3.3	Insulation Works	341	31	9.09%
12	4.1	Strong and Weak Electricity Engineering	866	122	14.09%
13	4.2	Water Supply and Sewerage Works	890	96	10.79%
14	4.3	Ventilation Duct Works	755	10	1.32%
15	4.4	Civil Air Defense Engineering	712	0	0.00%
				

(1) BCWS is shown in Table 4. As can be seen in Table 4, the BCWS of the current time node is 612,622,200 yuan.

Table 4. The Part Projects BCWS

Number	Item Coding	Subdivisional Work	Percentage of Planned Completion	BCWS/ ten thousand yuan
1	1.1	Earth-Rock Works	69.41%	3154.26
2	1.2	Foundation Pit Support Works	99.80%	8839.50
3	1.3	Foundation Pit Dewatering Engineering	89.83%	131.62
4	1.4	Pile Foundation Project	92.70%	27149.24
5	2.1	Concrete Work	42.54%	7142.30
6	2.2	Steel Structure Work	48.25%	7668.41
7	2.3	Formwork	48.05%	5122.35
8	2.4	Masonry Work	34.56%	413.20
9	3.1	Floor, Roof Decoration Project	7.62%	410.30
10	3.2	Waterproof Engineering	8.33%	101.68
11	3.3	Insulation Works	9.09%	413.39
12	4.1	Strong and Weak Electricity Engineering	14.09%	505.77
13	4.2	Water Supply and Sewerage Works	10.79%	198.30
14	4.3	Ventilation Duct Works	1.32%	11.88
15	4.4	Civil Air Defense Engineering	0.00%	0.00
		Total		61262.20

(2) BCWP is shown in Table 5. The BCWP of the current time node is 434,471,400 yuan.

Table 5. The Part Projects BCWP

Number	Item Coding	Subdivisional Work	Physical Percent Complete	BCWP/ ten thousand yuan
1	1.1	Earth-Rock Works	60.55%	2751.63
2	1.2	Foundation Pit Support Works	59.73%	5290.41
3	1.3	Foundation Pit Dewatering Engineering	55.69%	81.60
4	1.4	Pile Foundation Project	57.41%	16813.78
5	2.1	Concrete Work	33.20%	5574.15
6	2.2	Steel Structure Work	40.93%	6505.04
7	2.3	Formwork	45.26%	4824.92
8	2.4	Masonry Work	17.88%	213.77
9	3.1	Floor, Roof Decoration Project	4.35%	234.23
10	3.2	Waterproof Engineering	2.61%	31.86
11	3.3	Insulation Works	9.01%	409.75
12	4.1	Strong and Weak Electricity Engineering	13.89%	498.59
13	4.2	Water Supply and Sewerage Works	11.35%	208.59
14	4.3	Ventilation Duct Works	0.98%	8.82
15	4.4	Civil Air Defense Engineering	0.00%	0.00
		Total		43447.14

(3) ACWP is shown in Table 6. The ACWP of the current time node is 446,511,500 yuan.

Table 6. The Part Projects ACWP

Number	Item Coding	Subdivisional Work	Physical Percent Complete	ACWP/ ten thousand yuan
1	1.1	Earth-Rock Works	60.55%	2965.13
2	1.2	Foundation Pit Support Works	59.73%	6004.87
3	1.3	Foundation Pit Dewatering Engineering	55.69%	103.50
4	1.4	Pile Foundation Project	57.41%	15408.65
5	2.1	Concrete Work	33.20%	7061.42
6	2.2	Steel Structure Work	40.93%	6341.12
7	2.3	Formwork	45.26%	5010.36
8	2.4	Masonry Work	17.88%	397.02
9	3.1	Floor, Roof Decoration Project	4.35%	215.92
10	3.2	Waterproof Engineering	2.61%	37.87
11	3.3	Insulation Works	9.01%	381.61
12	4.1	Strong and Weak Electricity Engineering	13.89%	527.97
13	4.2	Water Supply and Sewerage Works	11.35%	188.34
14	4.3	Ventilation Duct Works	0.98%	7.37
15	4.4	Civil Air Defense Engineering	0.00%	0.00
		Total		44651.15

Based on the above data analysis, the current nodes BCWS, BCWP and ACWP are as follows:

- (1) The current BCWS of the node is 612,622,200 yuan.
- (2) The current BCWP of the node is 434,471,400 yuan.
- (3) The current ACWP of the node is 446,511,500 yuan.

4.2.3. Result of data.

Based on the above relevant data of the project, its correlation analysis was carried out according to the calculation principle of EVM, and evaluation indicators were calculated. The results are shown in Table 7.

Table 7. Evaluation Indicator Result

Name	Result	Comparison
CV	-12,040,100	<0
SV	-17,815,060	<0
CPI	0.97	<1
SPI	0.71	<1

To sum up, up to now, $CV < 0$, $CPI < 1$. The project has overspent and the risk of loss. Meanwhile, $SV < 0$, $SPI < 1$. The project has the risk of delay. This has an important early warning effect for the

project management team, helping them to take measures in advance to avoid potential financial losses.

By combining big data technology with project management, this study provides data-driven decision support for project management and enhances the scientific and foresight of decision making. Moreover, through the data analysis of EVM, this study provides a new analysis method for engineering cost control, which helps to predict and manage project cost more accurately. Through in-depth analysis of typical cases, this study can identify project risks in advance and provide a basis for project management teams to adjust strategies in time, which not only enriches the research content in the field of engineering management, but also extends the application of big data technology to the field of cost control, providing a new direction for subsequent research.

5. Conclusion

This paper collects and collates engineering data through big data analysis method and analyzes a project by EVM, and draws the following conclusions:

- (1) This paper confirms the effectiveness of big data analysis methods in cost control. Using big data crawler technology for data collection, as well as data cleaning, storage and presentation, the quality and availability of data is improved, laying a solid foundation for in-depth analysis.
- (2) The EVM analysis was carried out on a project, and the research revealed the possible risk of overspending and delay in the project. This finding has important warning significance for project management teams, which can help them adjust their strategies in time to avoid potential economic losses.
- (3) This paper provides a new research method for the field of engineering management, which is helpful to promote the innovation of research methods in this field. In addition, this paper complements traditional cost control methods that rely on empirical judgment and limited data, demonstrating the potential of big data technology in addressing the demands of increasingly complex engineering projects.
- (4) Although this paper takes a project as a case for in-depth analysis, the analysis results of a single case may not fully represent the general situation of all engineering projects. Project cost control is a multi-factor and multi-stage complex process, and this study may not cover all the factors that affect the cost. Moreover, the application of big data technology in project management is still in the exploration stage, and there may be problems such as technical implementation difficulty, data security and privacy protection. Future research can be extended to more project cases to verify the applicability and effectiveness of big data technology in different types and sizes of engineering projects. It is also possible to further explore more in-depth data analysis methods and how to optimize the data collection and analysis process to improve overall efficiency.

References

- [1] X.S. Li, D.F. Liu, J.W. Sun, Z.X. Zhu, Challenges of Industrial Engineering in Big Data Environment and Its new Directions on Extension Intelligence. *Procedia Computer Science* 214 (2022) 1561-1567.
- [2] Y.F. Li, Construction cost control and management, *China Tendering* 02 (2024) 96-98.
- [3] C.H. Feng, Factors affecting construction cost and cost control management methods, *Theoretical Research in Urban Construction* 20 (2023) 45-47.
- [4] S. Shi, W. Zheng, X.P. Xiao, Big data analysis methods based on artificial intelligence, *Digital Technology & Application* 41.02 (2023) 110-112.
- [5] D. Hammann, Big Data and Machine Learning in Cost Estimation: An automotive case study, *International Journal of Production Economics* 269 (2024) 109137.
- [6] J.B. Cheng, Analysis of cost management and control strategy of housing construction engineering, *Real Estate World* 23 (2023) 106-108.
- [7] D. Wang, On data security protection obligations of data processors, *Contemporary Law Review* 37.02 (2023) 40-49.

- [8] P. Ballesteros-Pérez, E. Sanz-Ablanedo, Daniel Mora-Melià, M^a.C. González-Cruz, J.L. Fuentes-Bargues, E. Pellicer, Earned Schedule min-max: Two new EVM metrics for monitoring and controlling projects, *Automation in Construction* 103 (2019) 279-290.
- [9] T. Hu, Software Engineering Classification Model and Algorithm Based on Big Data Technology, *Procedia Computer Science* 228 (2023) 119-128.
- [10] X.C. Tang, T. Lan, H. Zhong, D.F. Li, Q. Miao, An evolutionary game model for indirect data sharing in manufacturing big data consortium, *Expert Systems with Applications* 255 (2024) 124807.