2024 World Journal of Innovation and Modern Technology, Vol.7, Issue 2 (Apr)
ISSN 2682-5910
DOI: 10.53469/wjimt.2024.07(02).14

Application of Machine Learning Decision Tree Algorithm Based on Big Data in Intelligent Procurement

Yuan Feng[1,*], Hanzhe Li[2], Xiangxiang Wang[3], Jingxiao Tian[4], Yaqian Qi[5]

[1]Interdisciplinary Data Science, Duke University North Carolina USA
[2]Computer Engineering, New York University , New York USA
[3]Computer Science, University of Texas at Arlington, Arlington, Texas, USA
[4]Electrical and Computer Engineering, San Diego State University, San Diego, USA
[5]Quantitative Methods and Modeling, Baruch Collegue, New York, USA
**Corresponding author: beckyfeng1219@gmail.com*

Abstract: ***With the massive data explosion, the era of big data has arrived. How to analyze big data efficiently has become an important topic. Machine learning, which uses artificial intelligence, is one of the most commonly used methods of data analysis. Traditional machine learning algorithms are often designed for offline batch training, but this approach is difficult to apply to large and growing data sets in big data environments. This paper takes the procurement business of manufacturing enterprises as the research object, and on the basis of comprehensive analysis of the characteristics of procurement business and data characteristics, determines the construction idea of intelligent procurement management framework, and builds an intelligent procurement management framework based on machine learning under big data. Based on this, this paper uses decision tree machine learning algorithm to transform the traditional procurement system. Make it more suitable for the current big data environment.***

Keywords: Procurement system; Big data; Machine learning; Decision tree

1. INTRODUCTION

With the rapid development of big data, artificial intelligence and other information technologies and the changes of digitalization in existing industries, the amount of data has increased rapidly, and massive data has appeared all over the world. [1] The era of big data is characterized by explosive growth and rapid organization of data. In purchasing management, the use of artificial intelligence and machine learning to help enterprises digital transformation has increasingly become the research focus of experts and scholars. Machine learning is the most intelligent and widely used core field of artificial intelligence. It is also the most important part of enterprises in the field of artificial intelligence. [2-3] Many scholars have studied the key technologies of machine learning, and applied clustering and classification algorithms to organize data, effectively extracting hidden valuable information from a large number of data, thus giving decision-making ability. [4] In the context of big data, improving the efficiency of procurement management through digital innovation, improving the efficiency of procurement management and strengthening risk control are the keys to improving the core competitiveness of enterprises.

Guo Min (2021) pointed out that in the digital age, the core of digital procurement is to create data value by using data analysis and mining methods. The goal of data analysis is not data, but business, that is, the core problem of digital procurement is to emphasize how to turn data problems into business problems. Business characteristics must be explained by effective data collection and the design of appropriate characteristics [5]. In order to solve the technical problems in the actual procurement system, this paper adopts the decision tree algorithm of machine learning, and establishes an intelligent procurement system based on the decision tree algorithm of machine learning, which can quickly classify the procurement data and speed up the retrieval of material information. [6]

2. RELATED WORK

2.1 Big data

Current user interactions with digital interfaces often face many challenges, and one of McKinsey's main definitions of Big Data in its report, Big Data: The Next Frontier for Innovation, Competition, and Productivity, is as follows: The collection of massive amounts of data that are captured, stored, managed, and analyzed far beyond the capabilities of traditional database software tools. [7]

Big Data permeates every field and is changing the way we live, work and do business. [8] Big data even has 5V characteristics, namely Volume, Variety, Velocity, and data Value, which are the most common and common characteristics. [9-11] In practice, big data systems include the following aspects :(1) data collection; (2) Data storage; (3) Data calculation; (4) Data mining and analysis; (5) Data application; (6) Data security.

In the manufacturing industry, industrial big data is used to improve the manufacturing level, including product fault diagnosis and prediction, process analysis, production process improvement, production process energy consumption optimization, industrial supply chain analysis and optimization, production planning and scheduling.

In the enterprise intelligent procurement management system, we can use big data to build enterprise intelligent data warehouse, organize all procurement data sets, and use artificial intelligence for planning and analysis to achieve the lowest cost of business procurement.

2.2 Machine learning

The core of machine learning is to make machines imitate people and own the way people think. The mechanism of machine learning can be summarized in two stages. First, a series of training data is given, and then a learning system model is used to predict and test the data set. [13-14] The goal is to find a specific model hidden in the data and use it to infer new data.

Supervised learning is the most widely used and abundant machine learning method at present. [15] For example, common algorithms such as support vector machines (SVM), logistic regression, and decision trees are all supervised learning. This paper uses decision tree method to demonstrate the application of intelligent purchasing system. As one of the cornerstones in the field of machine learning, decision tree algorithm has won a wide range of applications due to its intuitiveness and easy interpretation. Today, with the increasing depth and complexity of technology, decision tree algorithm still maintains its unique charm. [16] It can be combined with emerging machine learning techniques such as deep learning and reinforcement learning to create more powerful and adaptable models. For example, the predictive performance of decision trees has been significantly improved through random forests or lifting methods in ensemble learning, while retaining the interpretability of the model. The structure of decision trees makes them a powerful tool for understanding data properties and making predictive decisions, especially in scenarios where quick decision making and explanation of the decision process are required. This is particularly important in the intelligent procurement system.

2.3 Purchasing intelligent management

Intelligent management is artificial intelligence and management science, knowledge engineering and system engineering, computer technology and communication technology, Software engineering, information engineering and other multi-disciplines, multi-technology combination, mutual penetration into a new technology, a new discipline. [17] Intelligent management is gradually playing a greater role in business management. The intelligent management concept is applied to the procurement business, that is, relying on big data, artificial intelligence and other technical means to achieve intelligent procurement management, reduce the cost and risk of enterprise procurement, and develop more reasonable procurement plans. [18] Intelligent procurement management reflects the people-oriented, but also reflects the material (data) as the support basis, only according to the enterprise's own intelligent structure, tailored to their own procurement management mode and method, in order to play the value of intelligent procurement management. [19]

To solve the problem that it is difficult to obtain effective information from massive procurement data in enterprise procurement management under big data, machine learning decision tree can solve this problem well. In order to meet the requirements of data analysis in procurement management, a data warehouse for data procurement management is designed. Firstly, the data is sorted and graded through artificial intelligence, and then the procurement intelligent management model is designed through the machine learning decision tree algorithm to further mine and analyze the procurement data and make the accuracy judgment. To find out more valuable information to assist management decision-making, make purchasing management possible.

2.4 Decision tree algorithm

Decision tree algorithm is a typical representative, which has the advantages of high classification accuracy and simple model generation. [20] The C4.5 decision tree inherits the advantages of ID3 and is optimized and improved. It can discretize continuous eigenvalues and generate classification rules that are easy to understand and have high prediction accuracy. It is the most commonly used decision tree algorithm. [21]C4.5 decision tree, based on information theory, selects information entropy and information gain as the criteria for data classification. [22-23]C4.5 decision tree inherits the advantages of ID3 and is optimized and improved. It can discretize continuous eigenvalues and generate classification rules that are easy to understand and have high prediction accuracy. It is the most commonly used decision tree algorithm. [21]C4.5 decision tree, based on information theory, selects information entropy and information gain as the criteria for data classification.

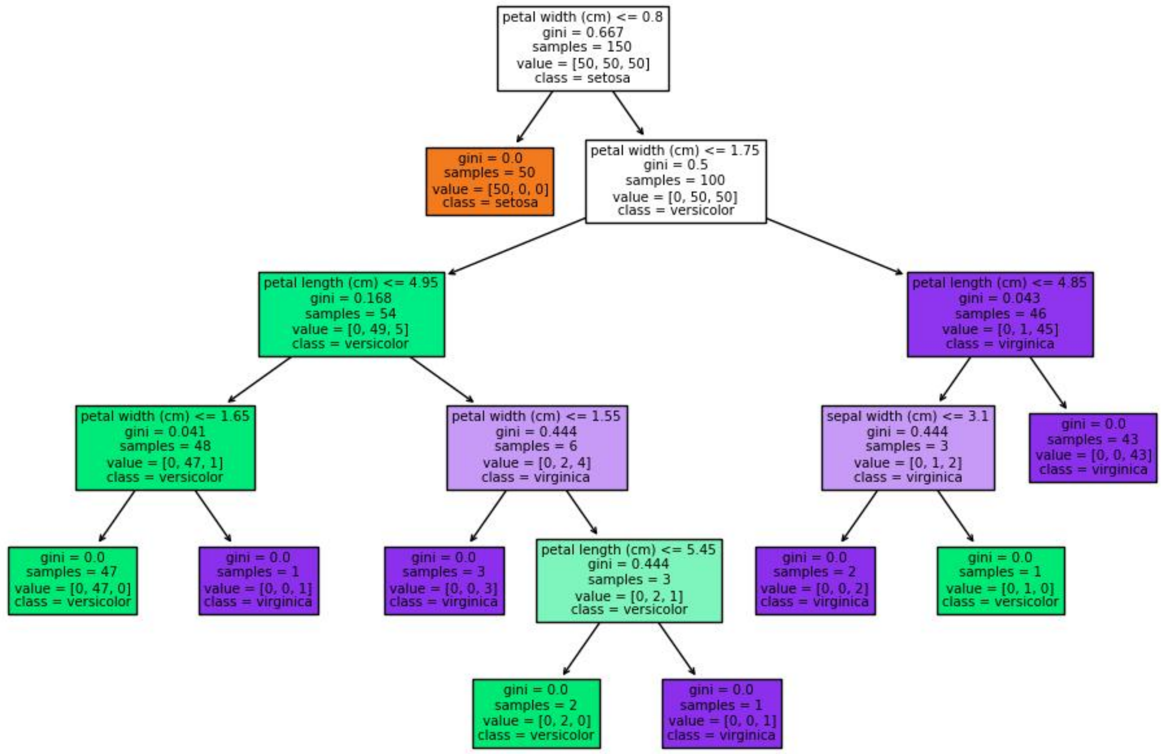

Figure1: Decision tree example diagram

The data warehouse is established in the intelligent procurement management of enterprises, and the procurement data is mined and analyzed based on the support of decision tree algorithm, so as to solve the problem that the procurement cost is too high due to too much information and too many influencing factors in the traditional procurement system, which is not conducive to the development of enterprises. [22-23] This paper mainly studies the application of decision tree algorithm in intelligent purchasing system. Classification is a supervised learning algorithm that classifies data objects into predefined classes based on known attribute values. Classification prediction is the most widely used method in data mining, and decision tree algorithm is its typical representative, which has the advantages of high classification accuracy and simple model generation.

3. APPLICATION OF MACHINE LEARNING DECISION TREE ALGORITHM IN INTELLIGENT PURCHASING

In the era of big data, the use of data has become a major issue, and it is difficult for decision makers in procurement operations to rely on their eyes and experience. [24] Machine learning algorithms mine valuable information hidden in massive data. This paper starts with the procurement business and draws a conclusion.

Based on the decision tree algorithm, the intelligent management needs of each business topic are analyzed, the corresponding features are selected, and the intelligent management model is built through the corresponding machine learning algorithm to achieve intelligent management.

This chapter takes C enterprise as an example for practice. The enterprise is a machinery manufacturing enterprise, and its procurement business is usually based on the procurement of raw materials, partly involving the procurement of activity materials, employee welfare, research and development materials, office supplies, etc. C enterprises have a wide variety of raw materials, a total of 383 different types of different types, a large amount of procurement, an average monthly intervention procurement of more than 1000 times.

3.1 Experimental model

Decision tree algorithm. In this experiment, we use the decision tree algorithm model to simulate and analyze the intelligent procurement system of C enterprise. Since C company has thousands of purchases every month, it is difficult for the company to use traditional methods to accurately grasp and analyze the procurement cost control situation. [25]Based on this, the corresponding attributes of procurement cost control are extracted, and then C4.5 decision tree is used to learn the classification rules of control levels, so as to realize the forecast of procurement cost control and timely reflect the situation of procurement cost control. Then, the correlation analysis of the calibrated procurement cost control data is carried out, and the management of procurement cost is finally realized.

(1) Create a decision tree classifier instance

```
clf = DecisionTreeClassifier(criterion='entropy', random_state=42)
```
(1)

(2) Training model

```
clf.fit(X_train, y_train)
```
(2)

(3)Make predictions on the test set

```
y_pred = clf.predict(X_test)
```
(3)

(4) Evaluation model

```
accuracy = accuracy_score(y_test, y_pred)
print(f'Model accuracy: {accuracy:.2f}')
```
(4)

3.2 Experimental process

(1) Data feature extraction and data preprocessing
A total of 1212 items of procurement cost control data were extracted from the procurement cost fact table in the procurement management data warehouse of enterprise C, including procurement materials, actual procurement costs, planned costs, differences between plans and realities, suppliers, purchasers, and levels. In the procurement cost control level, 1 represents procurement cost savings and 2 represents consistency. 3 indicates a slight overspend and 4 indicates a significant overspend

(2) Establish the model

Import procurement cost control data, use decision tree method to train the classification forecasting model of procurement cost data. [26]The classification model needs to use two types of data, namely the test set and the training set, so it also needs to divide the procurement cost control data. Combining decision tree and decision rule, the analysis conclusion is: when the difference between planned cost and actual cost, the actual purchase cost control level can be calculated, which can be divided into one level, two level, three level and four level. As shown in the following table:

Table 1: Procurement cost data division table

Decision Rule	Procurement Cost Control Level	Confidence Level
If difference > -0.3 (131)	Level 1	98.5%
If -0.3 (692) < difference <= -361.03 (559)	Level 2	98.6%
If difference <= -1389.95 (37) AND actual cost <= 19,230.8 (6)	Level 3	83.3%
If difference <= -1389.95 (37) AND actual cost > 19,230.8 (6)	Level 4	83.9%
If -1389.95 (96) < difference <= -484.43 (68)	Level 3	85.3%
If -1389.95 (96) < difference AND difference > -484.43 (28)	Level 2	57.1%

(3) Network training:

Add analysis nodes to the decision tree classification model to analyze the accuracy of procurement cost control classification results. Based on the analysis results of the coincidence matrix output, it is found that in the procurement cost control decision tree classification prediction model, the cost control level and procurement control level of the training data set are shown in the following table:

Table 2: Training data set accuracy table

Procurement Cost Control Level	Training Data (Actual)	Training Data (Misjudged)	Test Data (Actual)	Test Data (Misjudged)
Level 1	131	2	62	1
Level 2	587	20	284	8
Level 3	74	11	27	5
Level 4	31	5	16	1

3.3 Experimental algorithm

The procurement cost data collected by the enterprise was pre-processed before going online, and a total of 1212 items of special diagnosis data were extracted. It is imported into the decision tree algorithm, the preprocessed data is converted into numerical values, and the feature set and label are defined. By using the classification model, the procurement cost control data is divided into two types, namely the test set and the training set, and the generated decision tree algorithm and classification rules are used to partition the procurement cost control data. [27]Through the training data, we can add analysis nodes to the decision tree classification model to analyze the accuracy of the classification results of procurement cost control, and finally use the coincidence matrix to output the analysis results.

Finally, 7 procurement cost saving data sets can be obtained by generating association rules and analyzing association relationships, and procurement management suggestions are finally obtained by analyzing their relationships. Enterprises with a strong association with savings can further increase their procurement efforts. For enterprises that have a strong association with overspending, the intensity of cooperation with them can be reduced and strict.

Review the purchase price; For the procurement personnel with a strong correlation with overspending, the relevant personnel can be interviewed to investigate the causes of cost overspending, and linked to the performance of the procurement personnel, and finally, in the procurement approval, the procurement price needs to be more strictly controlled.

3.4 Practical Conclusions

It is found in the correlation analysis of procurement cost control that enterprises with strong correlation with savings can further increase their procurement efforts. For enterprises that have a strong association with overspending, the intensity of cooperation with them can be reduced and strict

Review the purchase price; [28]For the procurement personnel with a strong correlation with overspending, the relevant personnel can be interviewed to investigate the causes of cost overspending, and linked to the performance of the procurement personnel, and finally, in the procurement approval, the procurement price needs to be more strictly controlled.

Compared with the traditional procurement system that relies on manpower, the intelligent procurement system using machine learning decision tree has higher accuracy and more accurate actual data, greatly reducing errors and reducing enterprise costs. Standardized high-precision algorithms and models and standardized forecast data can better manage enterprises, show the causes of cost overruns more clearly, and be more conducive to the procurement choice of enterprises. [29]However, the traditional procurement often requires a lot of manpower and material costs and it is difficult to get the best procurement choice, and the application of machine learning in manufacturing intelligent procurement technology solves these problems well.

To sum up, the application of machine learning in intelligent procurement systems reduces costs for enterprises, and it also has the advantages of greater accuracy, efficiency, standardized reporting, and personalized interpretation.

4. CONCLUSION

In conclusion,starting from the basic knowledge of decision tree, gradually in-depth algorithm optimization and research progress, and finally demonstrates its application in practical problems through a practical case. A purchasing cost intelligent management model based on decision tree -Apriori algorithm is constructed. First of all, through the analysis of material procurement attributes, design a unified procurement cost control attributes, and then use C4.5 decision tree algorithm to achieve the classification of procurement cost control levels. Secondly, the [30]Apriori algorithm is used to output the association rules in the purchase cost control data. Can put forward targeted decision suggestions for purchasing cost management. Procurement cost control is an important way for enterprises to reduce costs and increase efficiency, provide quantitative data for assessing employees' work performance, and provide more accurate decision support for procurement cost management. In the future, with the further integration of information technology and management science, business management

The update and iteration of theory and management environment, and the design of intelligent management characteristics of procurement can have a deeper study.

Better meet the needs of procurement management. Through a high degree of combination with artificial intelligence and the use of deep learning algorithms to achieve independent learning, can better improve the self-learning ability of intelligent procurement management, in order to improve the level of procurement management and intelligent effect.

With the explosion of big data in the era of big data, the accumulation and availability of all kinds of data continue to increase, as well as the development of computers, the use of powerful data mining capabilities based on machine learning to overcome traditional manual procurement methods, and the choice of artificial intelligence can greatly empower enterprises. Artificial intelligence can help enterprises more accurately purchase means of production and select supply chains, greatly improving the efficiency of enterprises.

Driven by both business and technology, enterprises have an increasing demand for intelligent management of procurement business, and are eager to improve the efficiency of procurement management, reduce management costs, and promote the digital transformation of enterprises through digital and intelligent means. Therefore, how to make full use of big data, artificial intelligence and other technologies to empower procurement management is of great significance to the digital and intelligent construction of enterprises.

ACKNOWLEDGEMENT

We would Special thanks to Zeng Qiang, Q. Cheng, M. Tian and other authors, Their article "Enhancing High-Frequency Trading Strategies with Edge Computing and Deep Learning" provides important reference and inspiration for this research. Their work has provided valuable support and guidance to our research, especially in integrating intelligent applications and development with machine learning. Through their research, we have a

deeper understanding of the potential and application prospects of machine learning, which provides a strong theoretical foundation and practical guidance for our work.

REFERENCES

[1] “Unveiling the Future Navigating Next-Generation AI Frontiers and Innovations in Application”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 147-56, https://doi.org/10.62051/ijcsit.v1n1.20.

[2] Q. Cheng, M. Tian, L. Yang, J. Zheng, and D. Xin, “Enhancing High-Frequency Trading Strategies with Edge Computing and Deep Learning”, Journal of Industrial Engineering & Applied Science, vol. 2, no. 1, pp. 32–38, Feb. 2024.

[3] Ji, Huan, et al. "Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising." Academic Journal of Science and Technology 9.2 (2024): 215-220.

[4] Liu, Beichang, et al. "Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System." International Journal of Computer Science and Information Technology 1.1 (2023): 139-146.

[5] Wang, Y., Bao, Q., Wang, J., Su, G., & Xu, X. (2024). Cloud Computing for Large-Scale Resource Computation and Storage in Machine Learning. Journal of Theory and Practice of Engineering Science, 4(03).

[6] Xu, J., Jiang, Y., Yuan, B., Li, S., & Song, T. (2023, November). Automated Scoring of Clinical Patient Notes using Advanced NLP and Pseudo Labeling. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 384-388). IEEE.

[7] Xu, Z., Yuan, J., Yu, L., Wang, G., & Zhu, M. (2024). Machine Learning-Based Traffic Flow Prediction and Intelligent Traffic Management. International Journal of Computer Science and Information Technology, 2(1), 18-27.

[8] Wang, Yixu, et al. "Exploring New Frontiers of Deep Learning in Legal Practice: A Case Study of Large Language Models." International Journal of Computer Science and Information Technology 1.1 (2023): 131-138.

[9] Duan, Shiheng, et al. "THE INNOVATIVE MODEL OF ARTIFICIAL INTELLIGENCE COMPUTER EDUCATION UNDER THE BACKGROUND OF EDUCATIONAL INNOVATION." The 2nd International scientific and practical conference “Innovations in education: prospects and challenges of today”(January 16-19, 2024) Sofia, Bulgaria. International Science Group. 2024. 389 p.. 2024.

[10] Wu, Jiang, et al. "Case Study of Next-Generation Artificial Intelligence in Medical Image Diagnosis Based on Cloud Computing." Journal of Theory and Practice of Engineering Science 4.02 (2024): 66-73.

[11] Qian, Wenpin, et al. "NEXT-GENERATION ARTIFICIAL INTELLIGENCE INNOVATIVE APPLICATIONS OF LARGE LANGUAGE MODELS AND NEW METHODS." OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS (2024): 262.

[12] Che, C., Lin, Q., Zhao, X., Huang, J., & Yu, L. (2023, September). Enhancing Multimodal Understanding with CLIP-Based Image-to-Text Transformation. In Proceedings of the 2023 6th International Conference on Big Data Technologies (pp. 414-418).

[13] Xu, Z., Gong, Y., Zhou, Y., Bao, Q., & Qian, W. (2024). Enhancing Kubernetes Automated Scheduling with Deep Learning and Reinforcement Techniques for Large-Scale Cloud Computing Optimization. arXiv preprint arXiv:2403.07905.

[14] Gong, Y., Huang, J., Liu, B., Xu, J., Wu, B., & Zhang, Y. (2024). Dynamic Resource Allocation for Virtual Machine Migration Optimization using Machine Learning. arXiv preprint arXiv:2403.13619.

[15] Zhang, Y., Liu, B., Gong, Y., Huang, J., Xu, J., & Wan, W. (2024). Application of Machine Learning Optimization in Cloud Computing Resource Scheduling and Management. arXiv preprint arXiv:2402.17216.Wang, Yong, et al. "Construction and application of artificial intelligence crowdsourcing map based on multi-track GPS data." arXiv preprint arXiv:2402.15796 (2024).

[16] Xu, X., Xu, Z., Ling, Z., Jin, Z., & Du, S. (2024). Comprehensive Implementation of TextCNN for Enhanced Collaboration between Natural Language Processing and System Recommendation. arXiv preprint arXiv:2403.09718.

[17] Song, B., Xu, Y., & Wu, Y. (2024). ViTCN: Vision Transformer Contrastive Network For Reasoning. arXiv preprint arXiv:2403.09962.

[18] Gong, Yulu, et al. "RESEARCH ON A MULTILEVEL PRACTICAL TEACHING SYSTEM FOR THE COURSE'DIGITAL IMAGE PROCESSING." OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS (2024): 272.

[19] Du, S., Qian, W., Zhang, Y., Shen, Z., & Zhu, M. (2024, February). IMPROVING SCIENCE QUESTION RANKING WITH MODEL AND RETRIEVAL-AUGMENTED GENERATION. In The 6th International

scientific and practical conference “Old and new technologies of learning development in modern condition s”(February 13-16, 2024) Berlin, Germany. International Science Group. 2024. 345 p. (p. 252).
[20] K. Xu, X. Wang, Z. Hu and Z. Zhang, "3D Face Recognition Based on Twin Neural Network Combining D eep Map and Texture," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1665-1668, doi: 10.1109/ICCT46805.2019.8947113.
[21] Xu, X., Niu, H., Ji, H., Li, H., & Wang, J. (2024). AI Empowered of Advancements in Microbial and Tumo r Cell Image Labeling for Enhanced Medical Insights. Journal of Theory and Practice of Engineering Scien ce, 4(03), 21-27.
[22] Shi, Peng, Yulin Cui, Kangming Xu, Mingmei Zhang, and Lianhong Ding. 2019. "Data Consistency Theor y and Case Study for Scientific Big Data" Information 10, no. 4: 137. https://doi.org/10.3390/info10040137 .
[23] Predict and Optimize Financial Services Risk Using AI-driven Technology. (2024). Academic Journal of S cience and Technology, 10(1), 299-304. https://doi.org/10.54097/6zrqef25
[24] Zhenghua Hu, Xianmei Wang, Kangming Xu, and Pu Dong. 2020. Real-time Target Tracking Based on PC ANet-CSK Algorithm. In Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence (CSAI '19). Association for Computing Machinery, New York, NY, USA, 343–346. https://doi.org/10.1145/3374587.3374607.
[25] Medication Recommendation System Based on Natural Language Processing for Patient Emotion Analysis. (2024). Academic Journal of Science and Technology, 10(1), 62-68. https://doi.org/10.54097/v160aa61
[26] Li, X., Zheng, H., Chen, J., Zong, Y., & Yu, L. (2024). User Interaction Interface Design and Innovation Ba sed on Artificial Intelligence Technology. Journal of Theory and Practice of Engineering Science, 4(03), 1-8.
[27] Li, H., Wang, X., Feng, Y., Qi, Y., & Tian, J. (2024). Integration Methods and Advantages of Machine Lea rning with Cloud Data Warehouses. International Journal of Computer Science and Information Technolog y, 2(1), 348-358. https://doi.org/10.62051/ijcsit.v2n1.36
[28] Biology-based AI Predicts T-cell Receptor Antigen Binding Specificity. (2024). Academic Journal of Scien ce and Technology, 10(1), 23-27. https://doi.org/10.54097/wy28c490
[29] Song, T., Li, X., Wang, B., & Han, L. (2024). Research on Intelligent Application Design Based on Artifici al Intelligence and Adaptive Interface.
[30] Chen, Y., Wang, S., Lin, L., Cui, Z., & Zong, Y. (2024). Computer Vision and Deep Learning Transformin g Image Recognition and Beyond. International Journal of Computer Science and Information Technology, 2(1), 45-51.