Journal of Economic Theory and Business Management
ISSN 3006-4953 (Print) | ISSN 3006-4961 (Online) | Vol. 1, No. 3, 2024

Bank Credit Risk Early Warning Model Based on Machine Learning Decision Trees

GUO, Lingfeng [1*] LI, Zihan [2] QIAN, Kun [3] DING, Weike [4] CHEN, Zhou [5]

[1] Trine University, USA
[2] Northeastern University, USA
[3] Engineering School of Information and Digital Technologies, France
[4] University of Illinois at Urbana-Champaign, USA
[5] Zhejiang University, China

**** GUO, Lingfeng is the corresponding author, E-mail: glf9871@gmail.com***

Abstract: The study explores the application of the C5.0 decision tree algorithm to improve bank credit risk management. By transforming risk identification from subjective judgment to objective analysis, risk measurement from qualitative to quantitative, and risk control from static to dynamic, banks can enhance their credit risk management practices. Using data from the Center for Machine Learning and Intelligent Systems, we constructed a C5.0 decision tree model to predict high-risk bank loans. The model's performance was evaluated through various metrics, including a confusion matrix, revealing an error rate of 14.9%. The study demonstrates that decision tree models, by leveraging key features such as checking and savings balances, can significantly enhance the accuracy and efficiency of bank credit risk assessments.

Keywords: Credit Risk Management, Decision Tree Model, Loan Default Prediction, Machine Learning In Finance.

DOI: https://doi.org/10.5281/zenodo.11627011

1 Introduction

Bank credit risk management has initially realized the transformation from subjective to objective, from qualitative to quantitative, and from static to dynamic. Credit risk management involves credit approval, post-loan management, risk mitigation, etc., which runs through every link of credit, and its transformation has been embedded in the whole credit process. [1] First of all, risk identification has been transformed from subjective to objective. Traditional credit risk identification mainly relies on the experience and judgment of credit managers, which has intense subjectivity. With the accumulation of historical data and the application of mathematical-statistical analysis, the evaluation of borrowers' credit status gradually turns to an objective analysis based on data, and the objectivity of risk identification is significantly enhanced.

Secondly, the risk measurement has been transformed from qualitative to quantitative. The risk identification model based on experience judgment can only evaluate the risk qualitatively. [2] The emergence of credit risk quantitative management models, such as CreditMetrics, CreditRisk+, and KMV, enables credit risk to be measured objectively and scientifically. Finally, risk management and control have initially realized the transformation from static to dynamic. In the traditional risk management model, the borrower's information acquisition channel is limited, and the update cycle is long, resulting in the borrower's credit risk changes not being effectively measured. [3] Risk management is mainly static management. With the advancement of the digital transformation of economy and society and the progress of risk quantification technology, the ability of banks to obtain and process all kinds of data required for borrower risk assessment and control has been significantly enhanced, and dynamic monitoring and management of borrower credit risk has been realized.

Based on the decision tree model in machine learning, this study proposes a bank credit risk early warning model, aiming to improve further the early warning ability of banks in credit risk management and achieve more accurate and efficient risk management.

2 Related Work

2.1 Bank credit risk management

Commercial banks need to rely on credit risk mitigation tools to transfer or reduce credit risk due to the rigid requirement of capital and interest protection at the debt end and the regulatory requirement of maintaining economic and financial stability. [4-6]Combined with the development of China's financial system, it can be predicted that in the long term, collateral will still be the main starting point for China's commercial banks to ease the credit risk. As the real estate regulation continues and the government credit debt is broken, the future may lead to the lack of real estate, land, and other collateral and the contraction of government credit, which will have a more significant impact on the existing credit creation model and credit risk management strategy of banks, forcing them to expand the list of collateral and explore new collateral. At the same time, the continuous maturity and in-depth application of digital technology will also play an essential role in the innovation of banks' management models, such as credit risk identification and collateral monitoring.

The means of credit risk mitigation are expanding from a visible grip to an invisible grip. According to the "Commercial Bank Mortgage Management Guidelines", commercial banks should at least divide the mortgage into financial collateral, real estate, accounts receivable, and other collateral categories. Due to the straightforward valuation, clear ownership, effortless control, and disposal of financial collateral and real estate, the pledge mode is relatively mature, and the pledge rate is also high. [7] Due to the difficulties in valuation, unclear ownership, and disposal of accounts receivable and other pledges, the pledge mode still needs to be explored. With the transformation and upgrading of traditional industries and the increase of the proportion of new economy industries, the proportion of tangible assets such as real estate and land in the enterprise asset structure may decrease, and the proportion of equity assets such as intellectual property rights, asset use rights, accounts receivable, and goods rights may increase. [8] In serving customers, especially new economy customers, commercial banks have to expand the range of collateral, which will be transferred from physical, tangible assets such as real estate and land to equity intangible assets. Correspondingly, the grasp of the bank's credit risk mitigation is expanding from the visible grasp to the invisible grasp.

Further, the borrower's credit value and data assets can be regarded as new intangible assets. Taking a civil servant credit loan as an example, although it is nominally a credit loan and does not require collateral, it can be regarded as mortgaged by the credit value of civil servants, such as reputation and discount of future income. Therefore, the basic logic of risk management is the same as that of mortgage loans, and risk management is carried out based on the intangible asset of the borrower's credit value rather than completely unsecured and unsecured. Digital transformation provides the possibility to broaden the scope of collateral. [9-12] Based on massively big data, on the one hand, it can alleviate the information asymmetry between the two sides of credit and increase the credit of borrowers; on the other hand, it can help evaluate the credit value of borrowers and form data assets. From relying on real tangible assets such as real estate and land to relying on intangible assets such as borrowers' equity assets, credit value, and data assets, it is conducive to promoting credit model innovation.

Digital technology helps improve credit risk management capabilities. With the help of digital technology, the level of risk management at various stages, such as pre-loan identification, credit decision, collateral control, and disposal preservation, can be improved. Improve the accuracy of asset valuation. [13] Given the delay in estimating asset value based on financial accounting data, the application of digital technology to collect and analyze credit data, behavioral data, government data, blocklist, public data, and other data, combined with traditional asset value assessment methods, can achieve a more real-time and accurate assessment of the borrower's asset value. Then, the bank's pre-loan risk identification ability and credit management level will be enhanced. In addition, applying digital technology can also help improve the bank's post-loan management. The application of knowledge graphs and other advanced technologies to strengthen the monitoring of the flow of credit funds, monitor the actual use of loans is untrue, misappropriated, flowing into high-risk areas and other violations, strengthen post-loan management, and improve risk prevention and control level.

2.2 Application scenarios of bank credit risk management

Fintech is an emerging business model based on technological innovations such as big data, cloud computing, artificial intelligence, and blockchain, which is fully applied in banking, securities, insurance, funds, consumer finance, and financial regulation. New technologies have promoted the rapid development of the financial industry, derived many application scenarios, and built a financial technology ecosystem. [14-17] Commercial banks closely follow the development context, rely on accelerating the promotion of digital transformation, continue strengthening the ability to support the real economy, improve the availability, convenience, and precision of financial services, and enable high-quality development with financial technology.

1. Business operation innovation

In terms of business channels, the traditional asset-heavy business model with offline outlets as the core has changed, and the online-to-offline all-channel coordination model reflects the deep integration of banking and fintech. Offline channels use intelligent devices such as unmanned or smart banks to improve efficiency and reduce comprehensive operating costs. [18] A retail credit business online enriches the business ecology to achieve low-cost customers.

In terms of product design, banks provide customers with online services such as deposits and loans through online banking and mobile banking, deepen customer insights with big data and other technologies, provide differentiated products according to customer needs, use application programming interface[19-21] (API) and other technologies to integrate financial products into consumer scenarios, and then use artificial intelligence and other technologies to track and adjust customers in real-time continuously—banking products as a whole show four trends of online, personalized, scene and intelligent.

2. Data capacity building

Banks have accumulated massive amounts of data during the operation. In the era of the digital economy, data, as a new production factor, is an important asset and core competitiveness of financial institutions. In digital transformation, data capability is an essential driving force for banks, and ensuring security compliance and high-quality usage is the foundation of big data. The bank implements data governance in four aspects: data standardization, security control, quality control, and regulatory submission. On this basis, the bank's front, middle, and back office business departments should be connected to build a data platform to support data capacity building. Through intelligent risk control, clever marketing, innovative operation, and other scenarios that land big data applications, banks' digital transformation accelerates.

2.3 Decision Tree

A decision tree is a significant technique for classification and prediction. It deduces the classification rules of the decision tree from a set of irregular cases and constructs them in a top-down, recursive manner. The attribute values are compared in the inner nodes of the decision tree, and the conclusions are finally reached at the leaf nodes by branching down from the node according to the judgment of different attributes. [22-24] Thus, from the root node to the leaf node corresponds to a reasonable rule, and the whole tree corresponds to a set of expression rules. Decision trees play an essential role in data analysis, and they can be used for classification analysis and prediction. One of its most significant advantages is that the learning process does not require the user to have much background knowledge; as long as the training case can be expressed in attributes and conclusions, the algorithm can be used to learn effectively.

In bank credit risk management, decision tree technology can significantly improve risk assessment and early warning accuracy and efficiency. [25] By collecting a large amount of historical data on borrowers, banks can use a decision tree model to analyze and mine the potential patterns and correlations and then build a model for credit risk prediction. The model can identify the key factors affecting the borrower's credit risk and compare the key attributes at each decision tree node to achieve accurate risk classification and assessment. Banks can better conduct credit approval, post-loan management, and risk mitigation, reduce lousy debt rates, and improve overall risk management. At the same time, the transparency and interpretability of the decision tree model enable bank managers to clearly understand the logic behind each decision and enhance the trust and operability of the model in practical application.

Combining machine learning and decision tree technology, banks can not only achieve dynamic monitoring of credit risk of existing borrowers but also predict potential risks in the future. This predictive capability enables banks to take preventive measures in advance, optimize resource allocation, reduce operating costs, and improve the security and stability of financial services [26-27]. For example, by analyzing the borrower's past credit history, repayment behavior, economic status, and other data, the decision tree model can accurately predict the borrower's future default risk, thus helping the bank be more cautious in its credit decision. In addition, the decision tree model can be continuously updated and optimized. As new data is added, the model will learn and improve itself to ensure the accuracy and timeliness of risk prediction. This improves the efficiency of bank credit risk management and wins more initiative for banks in the highly competitive financial market.

3 Methodology

In the financial field, the risk assessment of bank loans has always been an important research topic. Accurate identification of high-risk loans can help banks reduce bad debt losses, optimize the loan approval process, and improve the efficiency of capital utilization. With the development of data science and machine learning technology, using algorithms to analyze loan data has become an effective method. This experiment aims to identify high-risk bank loans by applying the C5.0 decision tree method.

3.1 Experimental design

This experiment uses data from the Center for Machine Learning and Intelligent Systems. The data set contains bank loan applicants' personal data, loan history, financial status, and more. By analyzing and modeling these data, we can identify potentially high-risk loan applications and help banks make more accurate risk assessments.

The primary purpose of this experiment is to construct a classification model using the C5.0 decision tree method to predict the risk level of bank loans. Specific objectives include:

Data preprocessing and feature engineering to ensure data quality and model performance.

1. Build and train the C5.0 decision tree model.

2. Evaluate the model's performance and interpret and analyze the results.

3. Propose possible improvement methods to improve the accuracy and reliability of the model.

4. Through this experiment, we want to demonstrate the application value of the C5.0 decision tree in financial risk assessment and provide a reference for research in related fields.

3.2 Data Characteristics

Before the risk assessment of the bank loan data, we first conduct a detailed exploration and analysis of the data characteristics. The following are the descriptions and statistical results of the main characteristic variables in the dataset.

(1) checking_balance

The distribution of checking account balances is as follows:

Table 1. Checking Account Balance

Checking Balance	Frequency
< 0 DM	274
> 200 DM	63
1 - 200 DM	269
Unknown	394

From the results, it can be seen that out of 1,000 observations, 274 observations have a checking account balance of less than DM 0, 63 observations have a checking account balance of more than DM 200, 269 observations have a checking account balance of between DM 1 and DM 200, and 394 observations have an unknown checking account balance.

(2) savings_balance

The distribution of savings account balances is as follow

Savings Balance	Frequency
< 100 DM	603
> 1000 DM	48
100 - 500 DM	103
500 - 1000 DM	63
Unknown	183

From the results, it can be seen that 603 observations have savings account balances of less than 100 marks, 48 observations are greater than 1,000 marks, 103 observations are between 100 and 500 marks, 63 observations are between 500 and 1,000 marks, and 183 observations have savings account balances unknown.

(3) months_loan_duration

The statistics of loan terms are as follow

Statistic	Value
Minimum (Min.)	4
1st Quartile	12
Median	18
Mean	20.9
3rd Quartile	24
Maximum (Max.)	72

The results showed that the shortest loan term was 4 months, the longest was 72 months, the median loan term was 18 months, and the average was about 20.9 months.

(4) Loan amount

The statistical description is as follow

Statistic	Value
Minimum (Min.)	250
1st Quartile	1366
Median	2320
Mean	3271
3rd Quartile	3972
Maximum (Max.)	18424

The results showed that loan amounts ranged from 250 DM to 18,424 DM, with a median loan amount of 2,320 DM and an average of about 3,271 DM.

Through the analysis of the above data characteristics, we can preliminarily understand the basic distribution of each variable in the data set.

1. In terms of checking account balances (checking_balance), the data distribution is relatively balanced, with a few accounts with balances exceeding 200 marks and most with low or unknown balances. Similarly, in terms of savings account balances (savings_balance), the vast majority of accounts have balances below 100 marks, with only a few having higher or unknown balances. These characteristic variables show the status of the borrower's bank account at the time of applying for the loan and may be closely related to its loan risk.

2. The distribution of loan terms (months_loan_duration) shows that most loan terms are concentrated between 12 and 24 months, with a maximum loan term of 72 months. The amount of loans is widely distributed, with a minimum amount of 250 marks and a maximum amount of 18,424 marks, showing the diversity and difference of loans applied for.

3. In terms of default, the data shows that among 1000 observations, 700 observations have no default records, while 300 observations have default records, with a default rate of 30%. This data reflects the reality of loan defaults and provides key labeling information for model construction.

This information provides an important basis for the subsequent construction and analysis of the C5.0 decision tree model. Understanding the distribution of each feature variable can help us to better select and deal with features and improve the prediction accuracy of the model. Next, based on these characteristic variables, we will use the C5.0 decision tree method to build a classification model to identify high-risk bank loans and improve the risk management level of banks.

3.3 Experimental Decision Tree Overview Effect

In this experiment, we trained a decision tree model using the C5.0 algorithm to identify high-risk bank loans. The resulting decision tree consists of 45 decision nodes.

Below is a detailed analysis of the top three decision branches and their impacts:

- Checking Balance Decision Node
- Decision Condition: If the checking balance is greater than 200 DM or unknown.
- Classification Result: Classified as unlikely to default.
- Data Distribution: 422 cases meet this decision condition, with 57 cases misclassified as unlikely to default.

Further Classification by Checking the Balance

- Decision Condition: Otherwise, if the checking balance is less than 0 DM or between 0-200 DM.
- Classification Result: Further classification is required to determine the likelihood of default.

Credit History and Savings Balance Decision Node

Decision Condition: Perfect or very good credit history.

- Savings balance less than 100 DM: Classified as very likely to default.
- Savings balance greater than 1000 DM or between 500-1000 DM: Classified as unlikely to default.
- These decision nodes illustrate the key features the model relies on for classification. Particularly, checking and savings balances play a crucial role in assessing loan default risk.

Confusion Matrix and Classification Performance

According to the output confusion matrix, out of 900 training cases, 766 cases were correctly classified, and 134 cases were misclassified, resulting in a classifier error rate of 14.9%. The specific distribution of classification errors is as follows:

- False Positives (Type I Error): 44 instances where the actual value is "no" (not defaulted), but the model incorrectly classified them as "yes" (defaulted).
- False Negatives (Type II Error): 90 instances where the actual value is "yes" (defaulted), but the model incorrectly classified them as "no" (not defaulted).

4 Conclusion

The experimental results show that the C5.0 decision tree model effectively utilizes checking and savings balances, along with credit history, to classify loan default risk. While the model achieves a reasonably low error rate of 14.9%, there is still room for improvement, particularly in reducing false negative and false positive rates. Further optimization and feature engineering could enhance the model's accuracy and reliability in predicting high-risk loans.

The results indicate that the C5.0 decision tree model effectively identifies high-risk bank loans by analyzing critical features like checking and savings balances and credit history. The model's ability to classify loans with an error rate of 14.9% underscores its potential to improve risk management practices in banks. With 766 out of 900 training cases correctly classified, the model exhibits robust predictive capabilities. Furthermore, the decision tree's transparency and interpretability make it a valuable tool for bank managers, providing clear insights into the logic behind risk assessments. This approach not only facilitates dynamic monitoring of existing borrowers' credit risk but also allows for proactive measures to mitigate potential future risks. Therefore, integrating decision tree models into bank credit risk management can optimize resource allocation, reduce operating costs, and enhance the security and stability of financial services.

Acknowledgments

This article sincerely thanks Wang, Xiangxiang, et al made outstanding contributions in the article [1] "Short-Term Passenger Flow Prediction for Urban Rail Transit Based on Machine Learning". Their research not only provides new ideas and methods for short-term passenger flow prediction of urban rail transit but also opens up new possibilities for the application of machine learning in this field. Through their in-depth exploration of data analysis and model construction, we have a deeper understanding of the law and characteristics of passenger flow changes in urban rail transit systems. It is especially worth mentioning that the passenger flow forecasting method based on machine learning proposed by them is not only innovative in theory, but also shows good forecasting effects and application prospects in practice. Their research results not only promote the development of urban rail transit but also provide valuable experience and inspiration for our research and practice in artificial intelligence financial risk management and other fields.

In addition, I am so grateful to Zhan and the outstanding research by Xiaoan et al in [2] “Aspect category sentiment analysis based on multiple attention mechanisms and pre-trained models.” Their work has brought new ideas and methods to the field of sentiment analysis, especially the application of multiple attention mechanisms and pre-trained models, providing valuable contributions to the further development of the field. These results not only help companies better understand user needs and market feedback but also point the way to progress in the field of natural language processing. Thanks again for their hard work and research.

Funding

Not applicable.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author Contributions

Not applicable.

About the Authors

GUO, Lingfeng

Business Analytics, Trine University, AZ, USA.

LI, Zihan

Computer Science, Northeastern University, San Jose, CA, USA.

QIAN, Kun

Business Intelligence, Engineering School of Information and Digital Technologies, Villejuif, France.

DING, Weike

Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA.

CHEN, Zhou

Software Engineering, Zhejiang University, Hangzhou, China.

References

[1] Wang, Xiangxiang, et al. "Short-Term Passenger Flow Prediction for Urban Rail Transit Based on Machine Learning." Journal of Computer Technology and Applied Mathematics 1.1 (2024): 63-69.

[2] Zhan, Xiaoan, Chenxi Shi, Kangming Xu, Lianwei Li, and Haotian Zheng. "Aspect category sentiment analysis based on multiple attention mechanisms and pre-trained models." Applied and Computational Engineering 71 (2024): 21-26.

[3] He, Zheng, et al. "Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering."

[4] He, Zheng, et al. "Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering."

[5] Zhou, Y., Zhan, T., Wu, Y., Song, B., & Shi, C. RNA Secondary Structure Prediction Using Transformer-Based Deep Learning Models.

[6] Liu, B., Cai, G., Ling, Z., Qian, J., & Zhang, Q. Precise Positioning and Prediction System for Autonomous Driving Based on Generative Artificial Intelligence.

[7] He, Z., Shen, X., Zhou, Y., & Wang, Y. Application of K-means clustering based on artificial intelligence in gene statistics of biological information engineering.

[8] Cui, Z., Lin, L., Zong, Y., Chen, Y., & Wang, S. Precision Gene Editing Using Deep Learning: A Case Study of the CRISPR-Cas9 Editor.

[9] Wang, B., He, Y., Shui, Z., Xin, Q., & Lei, H. Predictive Optimization of DDoS Attack Mitigation in Distributed Systems using Machine Learning.

[10] Wang, Y., Zhu, M., Yuan, J., Wang, G., & Zhou, H. (2024). The intelligent prediction and assessment of financial information risk in the cloud computing model. arXiv preprint arXiv:2404.09322.

[11] Wang, X., Tian, J., Qi, Y., Li, H., & Feng, Y. (2024). Short-Term Passenger Flow Prediction for Urban Rail Transit Based on Machine Learning. Journal of Computer Technology and Applied Mathematics, 1(1), 63-69.

[12] Ding, W., Tan, H., Zhou, H., Li, Z., & Fan, C. Immediate Traffic Flow Monitoring and Management Based on Multimodal Data in Cloud Computing.

[13] Feng, Y., Li, H., Wang, X., Tian, J., & Qi, Y. (2024). Application of Machine Learning Decision Tree Algorithm Based on Big Data in Intelligent Procurement.

[14] Yu, D., Xie, Y., An, W., Li, Z., & Yao, Y. (2023, December). Joint Coordinate Regression and Association For Multi-Person Pose Estimation, A Pure

Neural Network Approach. In Proceedings of the 5th ACM International Conference on Multimedia in Asia (pp. 1-8).

[15] Ni, C., Zhang, C., Lu, W., Wang, H., & Wu, J. (2024). Enabling Intelligent Decision Making and Optimization in Enterprises through Data Pipelines.

[16] Bao, Q., Wei, K., Xu, J., & Jiang, W. (2024). Application of Deep Learning in Financial Credit Card Fraud Detection. Journal of Economic Theory and Business Management, 1(2), 51-57.

[17] Fan, Chao, et al. "Integrating Artificial Intelligence with SLAM Technology for Robotic Navigation and Localization in Unknown Environments."

[18] Bai, X., Jiang, W., & Xu, J. (2024). Development Trends in AI-Based Financial Risk Monitoring Technologies. Journal of Economic Theory and Business Management, 1(2), 58-63.

[19] Bao, Wenqing, et al. "The Challenges and Opportunities of Financial Technology Innovation to Bank Financing Business and Risk Management." Financial Engineering and Risk Management 7.2 (2024): 82-88.

[20] Qi, Y., Feng, Y., Tian, J., Wang, X., & Li, H. (2024). Application of AI-based Data Analysis and Processing Technology in Process Industry. Journal of Computer Technology and Applied Mathematics, 1(1), 54-62.

[21] Chen, B., Basak, S., Chen, P., Zhang, C., Perry, K., Tian, S., ... & Jin, R. (2022). Structure and conformational dynamics of Clostridioides difficile toxin A. Life Science Alliance, 5(6).

[22] Cao, J.*, Ku, D., Du, J., Ng, V., Wang, Y., & Dong, W. "A Structurally Enhanced, Ergonomically and Human-Computer Interaction Improved Intelligent Seat’s System," Designs, vol. 1, no. 2, 2017, p. 11, doi: 10.3390/designs1020011.

[23] Xu, J., Wu, B., Huang, J., Gong, Y., Zhang, Y., & Liu, B. (2024). Practical Applications of Advanced Cloud Services and Generative AI Systems in Medical Image Analysis. arXiv preprint arXiv:2403.17549.

[24] 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.

[25] 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.

[26] Huang, J., Zhang, Y., Xu, J., Wu, B., Liu, B., & Gong, Y. Implementation of Seamless Assistance with Google Assistant Leveraging Cloud Computing.

[27] Tian, J., Qi, Y., Li, H., Feng, Y., & Wang, X. (2024). Deep Learning Algorithms Based on Computer Vision Technology and Large-Scale Image Data. Journal of Computer Technology and Applied Mathematics, 1(1), 109-115.