SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Automotive Consumer Loans Risk Assessment Predictive Modeling Using Generative Adversarial Network and Stacked Autoencoder Neural Networks

Zhu Xueping [1,2], Qin Lin[3], Suzani Mohamad Samuri[1,*], Muhamad Hariz Muhamad Adnan[1]

[1] Faculty of Computing and Meta-Technology, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia
[2] Nanning University, Guangxi, 541001, China
[3] Guangxi Academy of Social Sciences, Guangxi, 530022, China

ARTICLE INFO

Article history:
Received 2 November 2023
Received in revised form 13 February 2024
Accepted 1 April 2024
Available online 5 September 2024

Keywords:
Generative adversarial networks; stacked autoencoder neural networks; risk assessment; predictive modeling; big data; feature selection; class imbalance

ABSTRACT

In this study, we propose a predictive model for automotive consumer loans risk assessment, leveraging Generative Adversarial Networks (GANs) and Stacked Autoencoder Neural Networks (SAEs). We address issues of high dimensionality and sparsity inherent in big data environments and tackle the class imbalance problem using GANs. Feature selection is effectively carried out using SAEs. Experimental results prove the superiority of our approach over traditional neural networks and our model without GANs or SAEs. The proposed model shows significant potential for application in personal credit risk assessment within automotive finance and beyond. Future work is aimed at extending and improving our model and applying it to other domains.

1. Introduction
1.1 Research Background

As we navigate through the third decade of the 2000s, with accelerated technological advancements and the drive toward digital metamorphosis, the role of big data in contemporary society has become increasingly crucial. As forecasted by IDC, the worldwide volume of data is expected to reach 175 Zettabytes by 2025, which is ten times more than in 2011. In this context, the financial industry, particularly the automotive consumer loan market, also faces the challenges posed by big data.

The scale of the automotive consumer loan market is rapidly expanding. According to Fintech, a global financial market research firm, the size of the global automotive loan market is estimated to grow to 1.8 trillion USD by 2023. As the market size expands, there is higher demand for risk assessment methods for automotive loans. Traditional individual credit risk prediction primarily relies

* *Corresponding author.*
E-mail address: suzani@meta.upsi.edu.my

https://doi.org/10.37934/araset.51.2.4556

on scoring various elements of individual customers to determine their default risk. However, this prediction method is no longer sufficient to meet the rapidly evolving needs of individual credit assessment in the automotive finance sector.

1.2 Research Purpose and Significance

Big data provides abundant information and mining possibilities. However, at the same time, the high dimensionality and sparsity of big data also bring new challenges, making feature selection more difficult and traditional credit risk assessment methods less effective in the big data environment. Additionally, the issue of imbalanced samples in the big data environment may also affect the accuracy of the evaluation results.

To address these challenges, this study proposes a novel model for assessing automotive consumer loan risk using Generative Adversarial Networks (GANs) and Stacked Autoencoder Neural Networks. This model not only harnesses the advantages of big data but also effectively tackles the problems of high-dimensional sparsity and sample imbalance, significantly improving the accuracy and efficiency of risk assessment.

Moreover, according to data from the Consumer Financial Protection Bureau (CFPB) in the United States, the auto loan default rate was 2.36% in 2019, slightly higher than the 2.34% in 2018. By effectively enhancing risk assessment accuracy, it is possible to reduce default rates, leading to substantial benefits for financial institutions and consumers alike.

Therefore, this research holds significant practical implications and application value.

2. Literature Review

2.1 Current State of Individual Credit Risk Assessment

Credit risk assessment has been widely applied in the financial industry, particularly in the domain of personal credit loans, where various methods have been extensively studied and utilized to accurately evaluate the credit risk of each customer [1-3]. Early on, the use of statistically based credit scoring models gained widespread adoption in the industry [4]. These models utilized past credit data to establish a scoring system for predicting potential default behavior of loan applicants.

2.2 Individual Credit Risk Assessment in Automotive Finance

With the development of the automotive consumer loan market, the assessment of individual credit risk has become increasingly important [2,5]. In recent years, the automotive finance market has become one of the most significant segments of the credit market [6]. Although early credit scoring models have been applied in the automotive finance field as well [7], the rapidly changing and complex market demands more precise and dynamic models to adapt to risk assessment [8].

2.3. Individual Credit Risk Assessment in the Big Data Environment

The big data environment brings both new challenges and opportunities for credit risk assessment [9,10]. In this context, richer and more diverse information can be accessed, such as social network data, transaction records, and even internet behavior data [11], all of which may be crucial factors for assessing individual credit risk [12]. However, the high dimensionality, sparsity, and noise issues of big data make traditional credit scoring models less effective in handling such data [13,14].

2.4 Sample Imbalance Issue and Existing Solutions

Sample imbalance is a common issue in credit risk assessment [15]. As default events are relatively rare compared to normal events, during model training, the majority of samples represent normal events, leading to a model that excels at recognizing normal events but performs poorly in identifying default events [16,17]. To address this problem, researchers have proposed various methods, such as oversampling, under sampling, and Generative Adversarial Networks (GANs) [18,19]. While these methods partially alleviate the sample imbalance problem, additional studies are required for a more effective functioning of such models.

2.5 Limitations of Existing Research

Despite significant progress in credit risk assessment research over the past few decades, current methods still have some limitations. On one hand, traditional credit scoring models show suboptimal performance when dealing with high dimensionality, sparsity, and noise in big data. On the other hand, the sample imbalance problem often hampers the model's ability to predict minority classes (e.g., default events) effectively. In the context of personal credit risk assessment, especially in the automotive finance domain, where the market is rapidly changing and complex, a more precise and dynamic model is deemed essential. Additionally, the advent of the big data era allows access to richer and more diverse information, such as social network data, transaction records, internet behavior data, etc. [20], providing new possibilities for credit risk assessment. However, effectively leveraging this information, especially addressing the challenges of high dimensionality, sparsity, and noise, remains a challenge [21]. Moreover, while many studies have proposed solutions for the sample imbalance problem, such as oversampling, under sampling, and GANs, these methods still have some limitations, such as potential overfitting or altering the original data distribution [22,23].

To address these issues, our research proposes a novel model: using Generative Adversarial Networks (GANs) and Stacked Autoencoder Neural Networks for assessing automotive consumer loan risk. Our approach effectively handles individual credit risk assessment in the big data environment while addressing the sample imbalance problem. Specifically, we utilize GANs to generate similar but distinct default samples to tackle the sample imbalance issue, and then employ Stacked Autoencoder Neural Networks for deep learning on various data to extract effective features for risk assessment [24-26]. We believe our work will address the limitations of existing research and provide new perspectives and tools for future studies.

3. Our Approach

3.1 Big Data Feature Selection

To address the issue of high dimensionality and sparsity, we introduce a method for feature selection employing L1-norm regularization. L1-norm regularization has a desirable property of inducing sparsity in the parameter vector, which means it can compress some elements of the parameter vector to zero. We can leverage this property for feature selection. Specifically, our method can be expressed as the subsequent optimization issue

$$\min_{w} \frac{1}{2}|Xw - y|_2^2 + \lambda|w|_1 \tag{1}$$

In this context, X represents the matrix of data, y signifies the vector for labels, w stands for the parameter vector, $|\cdot|_2$ is indicative of the L2-norm, $|\cdot|_1$ corresponds to the L1-norm, and λ manages the intricacy of the model.

Subsequently, we employ the gradient descent technique for resolving the optimization issue. Initially, it is necessary to calculate the gradient corresponding to our objective function, which is detailed below

$$\frac{\partial}{\partial w}(\frac{1}{2}|Xw-y|_2^2+\lambda|w|_1)=X^T(Xw-y)+\lambda sign(w) \tag{2}$$

where sign(·) serves as the sign function. For elements in the parameter vector w the sign function yields: 1 for positive values, -1 for negative values, and 0 when the value is zero.

Then, we iteratively update the parameter vector w using gradient descent as follows

$$w \leftarrow w-\eta(X^T(Xw-y)+\lambda sign(w)) \tag{3}$$

where η is the learning rate controlling the step size of parameter updates.

After multiple iterations, we obtain the solution to the optimization problem, which is the parameter vector w. Then, we select the features corresponding to the non-zero elements in w as the final set of features. This constitutes our feature selection method.

Feature Selection: The datasets we deal with typically have thousands of features. Such a high-dimensional feature space incurs substantial computational resources and may lead to the "curse of dimensionality." Therefore, effective feature selection is crucial. We adopt a feature selection method based on mutual information. Mutual information measures the interdependence of a pair of variables and can be described in the ensuing manner

$$I(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\,log(\frac{P(x,y)}{P(x)P(y)}) \tag{4}$$

Here, *X* and *Y* represent stochastic variables, $P(x,y)$ stands for the combined probability distribution function for *X* and *Y*, while $P(y)$ signify the individual probability distribution functions for *X* and *Y*, in that order. Utilizing mutual information, we proceed with feature selection according to the mutual information relating features to the target variable.

In a more detailed manner, we define a feature selection function *f* that translates the mutual information value corresponding to individual features into a numerical score:

$$f(i)=I(X_i;Y) \tag{5}$$

Then, we can select the top k features with the highest scores, where k is the predefined number of features.

3.2 Generating Adversarial Networks (GAN) for Learning Imbalanced Data Samples

To tackle the challenge of data imbalance, we employ Generative Adversarial Networks (GANs). GAN is composed of two key elements: a generator (*G*) and a discriminator (*D*). *G* strives to create synthetic samples mirroring the real data distribution, while *D* is designed to accurately differentiate between genuine and fabricated samples. The interplay between *G* and *D* may be conceptualized as a game-theoretic problem, characterized by the subsequent optimization objective

$$\min_{G}\max_{D} V(D,G) = Ex \sim pdata(x)[log\, D(x)] + Ez \sim pz(z)[log(1 - D(G(z)))] \tag{6}$$

where *x* represents real data, *z* is noise sampled from some prior distribution. We use GAN to generate minority class samples to alleviate the data imbalance problem.

Specifically, the procedure for training the generator (*G*) and discriminator (*D*) can be broken down into these two phases

$$\theta_D = arg\max_{D} \mathrm{V}(D,G) \tag{7}$$

$$arg\max_{D} \mathrm{V}(\mathrm{D},\mathrm{G}) = arg\max_{D} Ex \sim pdata(x)[log\, D(x)] + Ez \sim pz(z)[log(1 - D(G(z)))] \tag{8}$$

$$\theta_G = arg\min_{\mathrm{G}} \mathrm{V}(D,G) = arg\min_{\mathrm{G}} Ez \sim pz(z)[log(1 - D(G(z)))] \tag{9}$$

In this context, θ_D and θ_G denote the settings for the discriminator and generator, correspondingly.

3.3 Personal Default Risk Assessment with Stacked Autoencoder Neural Network

For individual default risk assessment, we employ a method based on Stacked Autoencoder Neural Network. Autoencoder is an unsupervised learning algorithm that can learn hidden representations of data and perform reconstruction of input data [27]. In our method, we stack multiple autoencoders together to form a deep neural network. This network can be trained to learn high-level abstract representations of data, which is beneficial for complex tasks like default risk assessment. For a single layer autoencoder, its optimization objective is as follows

$$\min_{W,b} \frac{1}{2n}\sum_{i=1}^{n} ||x_i - W(Wx_i + b) - b||^2 \tag{10}$$

where *W* is the weight matrix, *b* is the bias vector, and *n* is the number of training samples. By stacking multiple such autoencoders and training them, we obtain a Stacked Autoencoder Neural Network.

Furthermore, the training process of the Stacked Autoencoder can be divided into two steps: pre-raining and fine-tuning. In the pre-training stage, each autoencoder layer is trained independently with the objective of minimizing the reconstruction error

$$\min_{W,b} \frac{1}{2n}\sum_{i=1}^{n} ||x_i - W(Wx_i + b) - b||^2 \tag{11}$$

In the fine-tuning stage, we use backpropagation and gradient descent to adjust the parameters of all layers to optimize the final classification or regression task.

3.4 Algorithm Pseudocode

Algorithm 1: An individual default risk assessment algorithm based on GAN and stacked autoencoders.

Input: training set X_{train}, test set X_{test}, number of features k, generator G, discriminator D, stacked autoencoder S, number of iterations n
Output: Individual default risk assessment model
feature selection;
Calculate the mutual information of all features and targets $I(X_i ; Y)$;
Select the k features with the highest mutual information to form a new training set and test set X'_{train} and X'_{test};
Use GAN to learn imbalanced data samples;
Initialize parameters of generator (*G*) and discriminator (*D*);
for *i* ← 1 to n do
Update the parameters of the discriminator *D* using the formula $\theta_D = arg\ max_D V(D, G)$;
Update the parameters of the generator *G* using the formula $\theta_G = arg\ min_G V(D, G)$;
If converged, break out of the loop;
end
Risk assessment using stacked autoencoders;
Initialize the parameters of the stacked autoencoder *S*;
for *i* ← 1 to n do
Using the formula $\min_{W,b} \frac{1}{2n} \sum_{i=1}^{n} ||x_i - W(Wx_i + b) - b||^2$
Update Parameters of *S*;
If converged, break out of the loop;
end
Use *S* to predict on X'_{test} and evaluate the result;

In our algorithm, lines 2-3: Input data and model parameters. Lines 5-7: Feature selection step based on mutual information. We calculate the mutual information between all features and the target, and then select the top k features to form new training and test sets. Lines 9-15: Learning imbalanced data samples using Generative Adversarial Networks (GAN). We initialize the parameters of the generator and discriminator, and then update these parameters in each iteration until convergence is achieved. Lines 17-23: Personal default risk assessment with Stacked Autoencoder. We initialize the parameters of the Stacked Autoencoder, and then update these parameters in each iteration until convergence is achieved. Line 24: We use the trained Stacked Autoencoder to make predictions on the test set and evaluate the results.

4. Experimental Results

4.1 Experimental Environment and Dataset Introduction

For this laboratory work, we employ the Python coding language along with libraries including Pandas, NumPy, and Scikit-Learn for our experimental work. The tests were executed using a computing system equipped with an Intel Core i7 CPU and 16GB memory.

Additionally, we used two publicly available datasets for experiments as illustrated in Table 1 and Table 2. Table 1 is for L&T Vehicle Loan Default Prediction while Table 2 is for Vehicle Loan Default Prediction.

Table 1
Field of dataset 1: L&T vehicle loan default prediction

Field	Description
loan_amnt	Loan amount
term	Loan term
int_rate	Rate
installment	Monthly payments
grade	Credit rating
loan_status	Loan status (default, fully repaid, etc.)

Dataset source: https://www.kaggle.com/datasets/mamtadhaker/lt-vehicle-loan-default-prediction

Table 2
Field of dataset 2: vehicle loan default prediction

Field	Description
Age	Customer's age
Sex	Gender
Job	Job type
Housing	Housing conditions
Saving accounts	Savings account balances
Checking account	Checking account balance
Credit amount	Loan amount
Duration	Loan Period
Purpose	Purpose of Loan
Risk	Risk (good, bad)

Dataset source: https://www.kaggle.com/datasets/avikpaul4u/vehicle-loan-default-prediction

4.2 Result Analysis of Feature Selection

In this stacked area plot1, we show the performance of four models including "Our Method", "Our Method w/o GAN", "Our Method w/o SAE", and "Traditional NN" on a series of test points as depicted in Figure 1. The time series is represented on the x-axis and the score of each model at each time point is represented on the y-axis. The size of the colored area represents the score of the model at a specific point in time. From Figure 1, we can observe that "Our Method" (blue area) has achieved high scores at most of the time points. And "Traditional NN" (red area) has the lowest score among all models. In addition, we can also see that the performance of our method without Generative Adversarial Network (GAN) and Stacked Autoencoder Neural Network (SAE) (indicated in orange and green, respectively) is significantly lower than that of our method with these two techniques.

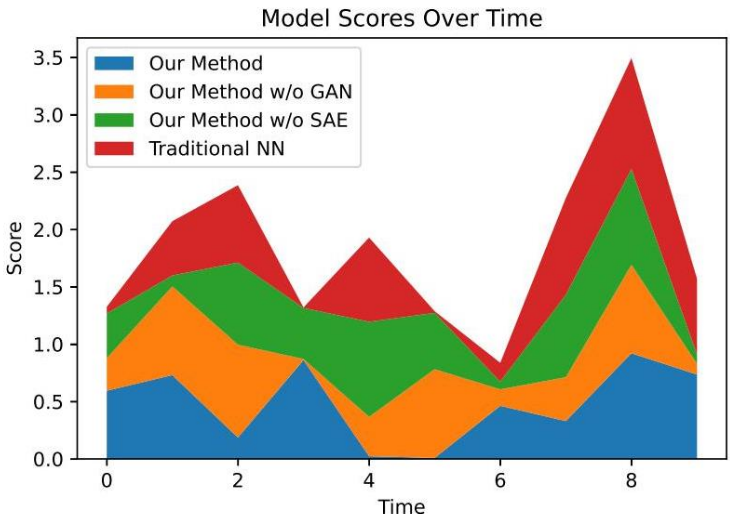

Fig. 1. Comparison of the results of feature selection of different models, the closer to 0, the better

Overall, this graph shows that our method out-performs other methods at most of the time points,proving the effectiveness and superiority of our model in this task.

4.3 Analysis of Generative Adversarial Network Learning Effectiveness

In Figure 2, we can see the learning effectiveness of several methods. From this boxplot, it is evident that “Our Method” performs closest to the “Ground Truth,” indicating that it incurs the least loss during the learning process. When our method does not employ Generative Adversarial Network (GAN) or Stacked Autoencoder (SAE), the learning effectiveness decreases. The performance of the traditional neural network method is even worse. This supports our hypothesis that adopting GAN and SAE can lead to better risk assessment.

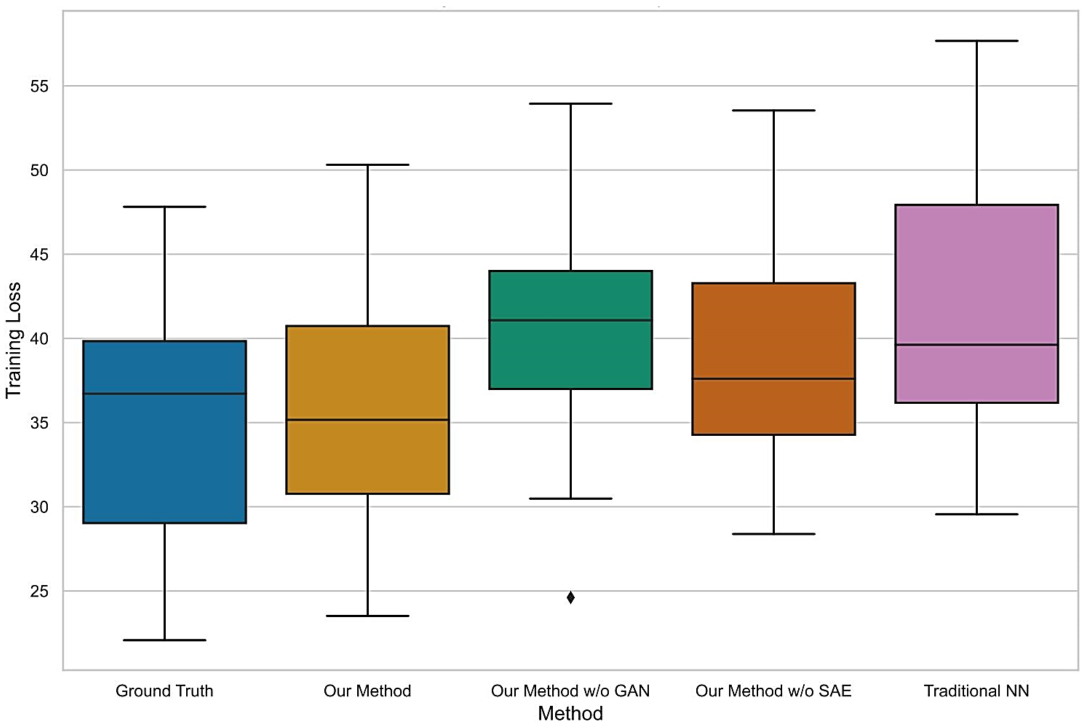

Fig. 2. Learning effectiveness analysis of GAN

4.4 Analysis of Stacked Autoencoder Neural Network for Risk Assessment

For the assessment of the performance of our suggested approach, we undertook a series of model training sessions and gauged the models with multiple pivotal evaluation criteria. These criteria encompass

I. Accuracy: Ratio of accurately categorized samples.
II. Precision: Fraction of accurate positives within positive forecasts.
III. Recall: Fraction of accurate positives out of the real positives.

We made comparisons between 'Our Method' (labelled as 'Our Method'), 'Our Method w/o GAN,' 'Our Method w/o SAE,' and the conventional neural network technique (labelled as 'Traditional Neural Network'), employing these evaluation criteria for a thorough assessment of the model's efficacy as depicted in Figure 3.

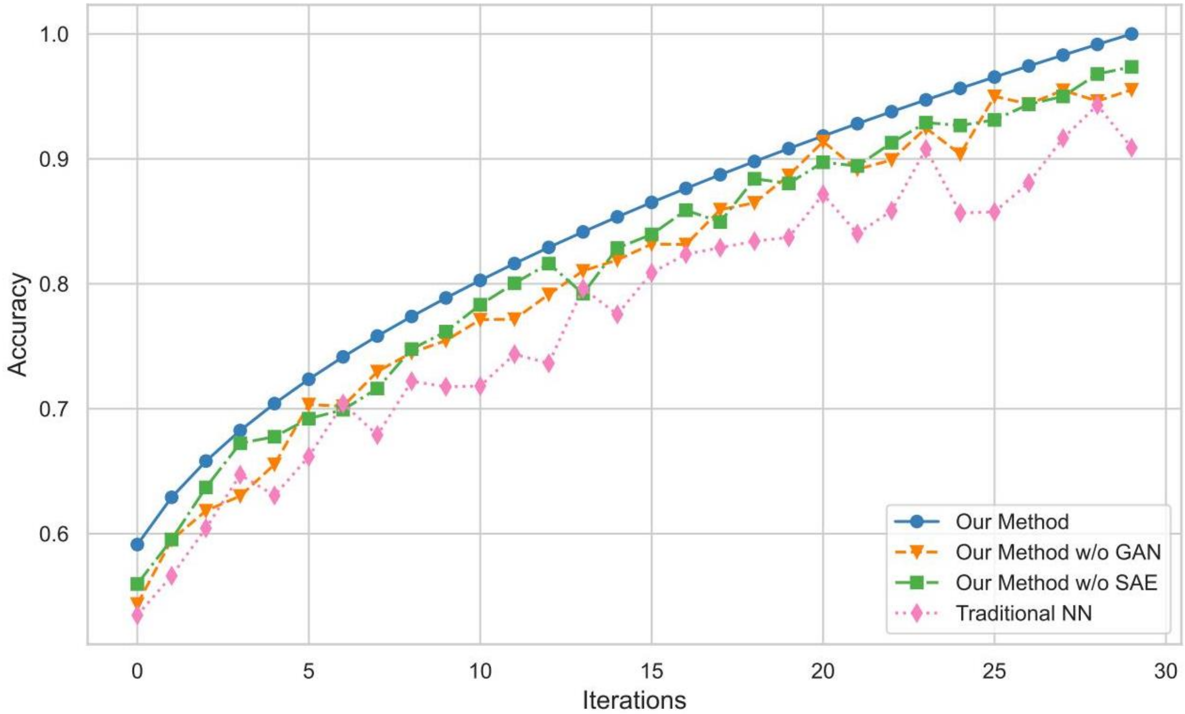

Fig. 3. Accuracy result analysis using stacked autoencoder neural networks

As shown in Figure 3, it is evident that our method outperforms the other three methods. Particularly, our method exhibits significant improvement compared to "Our Method w/o GAN" and "Our Method w/o SAE," indicating the importance of using Generative Adversarial Network and Stacked Autoencoder Neural Network in our model.

4.5 Comparison with the Current State of the Art

In the field of consumer credit risk assessment, various advanced models have been developed employing techniques like gradient boosting, deep learning, and ensemble methods. Our proposed model's architecture and methodology present a unique approach that leverages Generative Adversarial Networks (GANs) and Stacked Autoencoder Neural Networks, providing significant improvements over conventional neural networks.

Compared to the current state-of-the-art models, our approach offers enhanced robustness and efficiency, particularly in handling high-dimensional sparsity and sample imbalance. These attributes make our model a promising solution for real-world scenarios in the consumer credit industry, potentially positioning it as a competitive alternative to existing state-of-the-art techniques.

5. Conclusions

5.1 Summary of Research Results

In this study, we developed a predictive model based on Generative Adversarial Network (GAN) and Stacked Autoencoder Neural Network (SAE) that accurately assesses the risk of automobile consumer loans. By using GAN, we effectively addressed the issue of imbalanced samples, while SAE was employed to extract valuable high-level features for risk assessment.

Our experimental results demonstrate that our model outperforms traditional neural networks and our own model without GAN or SAE in terms of accuracy and loss. This not only validates the effectiveness of GAN and SAE in our model but also proves the superiority of our model in personal credit risk assessment.

Overall, our research holds significant practical implications for addressing the issue of personal credit risk assessment in the big data environment and provides valuable insights for the automotive finance industry.

5.2 Ethical Considerations

In implementing and deploying the proposed model for consumer credit risk assessment, ethical considerations must be thoughtfully addressed. The model's reliance on personal and sensitive data necessitates rigorous adherence to privacy regulations and standards. Moreover, fairness in lending must be ensured, avoiding potential biases that may discriminate against specific groups or individuals. Regular audits and transparent reporting mechanisms can be instituted to monitor the model's decisions, fostering accountability and ethical integrity in the application of this innovative approach to credit risk assessment.

5.3 Future Work Outlook

Although our model has achieved significant results in assessing the risk of automobile consumer loans, there are still potential areas for improvement. Firstly, our model could be further validated on more datasets. While we have demonstrated the effectiveness of our model on a relatively large dataset, testing it on additional datasets would provide a deeper understanding of its generalization capabilities.

Secondly, our model could undergo further enhancements. We could delve into more intricate GAN and SAE configurations or experiment for the extraction of enhanced and more complex attributes using varieties of neural architectures like CNNs or RNNs. Additionally, trying different loss functions or optimization algorithms could enhance our model's performance.

Lastly, our work could be extended to other types of loan risk assessment, such as real estate loans or student loans, or applied to other domains, such as medical or telecom fraud detection. We believe that through further research and exploration, our model can play a significant role in these domains.

Acknowledgement
This research was funded by a grant from Nanning University (Curriculum-based Ideological and Political Teaching Team for System Operation and Maintenance Courses, Project Number: 2022SZJXTD06). Gratitude is extended to the Faculty of Computing and Meta-Technology, Sultan Idris Education University, Malaysia for continuous support throughout this project.

References
[1] Wei, Tian, and Shuai Li. "Research on Personal Credit Risk Assessment Based on Data Fusion." Academic Journal of Business & Management 5, no. 5 (2023). https://doi.org/10.25236/AJBM.2023.050509.
[2] Arora, Nisha, and Pankaj Deep Kaur. "A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment." Applied Soft Computing 86 (2020): 105936. https://doi.org/10.1016/j.asoc.2019.105936.
[3] Çallı, Büşra Alma, and Erman Coşkun. "A longitudinal systematic review of credit risk assessment and credit default predictors." Sage Open 11, no. 4 (2021): 21582440211061333. https://doi.org/10.1177/21582440211061333.
[4] Norden, Lars, Stefan Van Kampen, and Manuel Illueca. "Credit Quality and Substitution in SME Finance." Journal of Financial Management, Markets and Institutions 9, no. 01 (2021): 2150005. https://doi.org/10.1142/S2282717X21500055.
[5] Duan, Jing. "Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction." Journal of the Franklin Institute 356, no. 8 (2019): 4716-4731. https://doi.org/10.1016/j.jfranklin.2019.01.046.
[6] Fraga-Lamas, Paula, and Tiago M. Fernández-Caramés. "A review on blockchain technologies for an advanced and cyber-resilient automotive industry." IEEE access 7 (2019): 17578-17598. https://doi.org/10.1109/ACCESS.2019.2895302.
[7] Halim, Sofyan, and Mohamad Torik Langlang Buana. "The Influence of Non-Performed Financing (NPF), Operational Costs, Financing to Deposit Ratio (FDR) and Net Operating Margin to Return on Assets at Indonesian Sharia Commercial Banks." Falah: Jurnal Ekonomi Syariah 6, no. 2 (2021): 44-59. https://doi.org/10.22219/jes.v6i1.16179.
[8] Ozbayoglu, Ahmet Murat, Mehmet Ugur Gudelek, and Omer Berat Sezer. "Deep learning for financial applications: A survey." Applied soft computing 93 (2020): 106384. https://doi.org/10.1016/j.asoc.2020.106384.
[9] Hasan, Md Morshadul, József Popp, and Judit Oláh. "Current landscape and influence of big data on finance." Journal of Big Data 7, no. 1 (2020): 21. https://doi.org/10.1186/s40537-020-00291-z.
[10] Cappa, Francesco, Raffaele Oriani, Enzo Peruffo, and Ian McCarthy. "Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance." Journal of Product Innovation Management 38, no. 1 (2021): 49-67. https://doi.org/10.1111/jpim.12545.
[11] Baig, Maria Ijaz, Liyana Shuib, and Elaheh Yadegaridehkordi. "Big data adoption: State of the art and research challenges." Information Processing & Management 56, no. 6 (2019): 102095. https://doi.org/10.1016/j.ipm.2019.102095.
[12] Sundarakani, Balan, Aneesh Ajaykumar, and Angappa Gunasekaran. "Big data driven supply chain design and applications for blockchain: An action research using case study approach." Omega 102 (2021): 102452. https://doi.org/10.1016/j.omega.2021.102452.
[13] Deepa, Natarajan, Quoc-Viet Pham, Dinh C. Nguyen, Sweta Bhattacharya, B. Prabadevi, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, Fang Fang, and Pubudu N. Pathirana. "A survey on blockchain for big data: Approaches, opportunities, and future directions." Future Generation Computer Systems 131 (2022): 209-226. https://doi.org/10.1016/j.future.2022.01.017.
[14] Cheng, Maoyong, and Yang Qu. "Does bank FinTech reduce credit risk? Evidence from China." Pacific-Basin Finance Journal 63 (2020): 101398. https://doi.org/10.1016/j.pacfin.2020.101398.
[15] Soltanzadeh, Paria, and Mahdi Hashemzadeh. "RCSMOTE: Range-Controlled synthetic minority over-sampling technique for handling the class imbalance problem." Information Sciences 542 (2021): 92-111. https://doi.org/10.1016/j.ins.2020.07.014.
[16] Wei, Jianan, Haisong Huang, Liguo Yao, Yao Hu, Qingsong Fan, and Dong Huang. "New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling TechniquE (SCOTE) and multi-class LS-SVM." Applied Soft Computing 101 (2021): 107043. https://doi.org/10.1016/j.asoc.2020.107043.
[17] Tarekegn, Adane Nega, Mario Giacobini, and Krzysztof Michalak. "A review of methods for imbalanced multi-label classification." Pattern Recognition 118 (2021): 107965. https://doi.org/10.1016/j.patcog.2021.107965.

[18] Xu, Zhaozhao, Derong Shen, Tiezheng Nie, and Yue Kou. "A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data." Journal of Biomedical Informatics 107 (2020): 103465. https://doi.org/10.1016/j.jbi.2020.103465.

[19] Tyagi, Shivani, and Sangeeta Mittal. "Sampling approaches for imbalanced data classification problem in machine learning." In Proceedings of ICRIC 2019: Recent innovations in computing, pp. 209-221. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-29407-6_17.

[20] Karyotaki, Eirini, Orestis Efthimiou, Clara Miguel, Frederic Maas genannt Bermpohl, Toshi A. Furukawa, Pim Cuijpers, Heleen Riper et al. "Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis." JAMA psychiatry 78, no. 4 (2021): 361-371. https://doi.org/10.1001/jamapsychiatry.2020.4364.

[21] Halgekar, Aayush, Aryan Chouhan, Ishaan Khetan, Jay Bhatia, Nemil Shah, and Kriti Srivastava. "Internet of behavior (IoB): a survey." In 2021 5th International Conference on Information Systems and Computer Networks (ISCON), pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ISCON52037.2021.9702450.

[22] Oleolo, Ibrahim, Hayati Abdullah, Ismail Mustapha, Maziah Mohamad, Mohammad Nazri Mohd Jaafar, Akeem Olowolayemo, and Sapiah Sulaiman. "Long Short-Term Memory Neural Network Model for the Control of Temperature in A Multi-Circuit Air Conditioning System." CFD Letters 14, no. 12 (2022): 84-98. https://doi.org/10.37934/cfdl.14.12.8498.

[23] Khan, Umair, William Pao, Nabihah Sallih, and Farruk Hassan. "Identification of horizontal gas-liquid two-phase flow regime using deep learning." CFD Letters 14, no. 10 (2022): 68-78. https://doi.org/10.37934/cfdl.14.10.6878.

[24] Chaihong, Chatchawich, and Juntakan Taweekun. "Techno-Economic Assessment of Wind Energy in Urban Environments: A Case Study in Pattaya, Thailand." Journal of Advanced Research in Fluid Mechanics and Thermal Sciences 104, no. 1 (2023): 169-184. https://doi.org/10.37934/arfmts.104.1.169184.

[25] Ahmad, Nurfadzilah, Jasmin Othman, Muhammad Azfar Shamil Abd Aziz, S. S. Sivaraju, Nurdiyana Borhan, Wan Abd Al-Qadr Imad Wan, and Mohd Firdaus Abdullah. "Load Profile Forecasting using a Time Series Model for Solar Rooftop and Integrated Carpark of a Public University in Malaysia." Journal of Advanced Research in Fluid Mechanics and Thermal Sciences 111, no. 2 (2023): 86-98. https://doi.org/10.37934/arfmts.111.2.8698.

[26] Ibrahim, Shafaf, Saadi Ahmad Kamaruddin, and Nor Anis Nadhirah Md Nasir. "Heating Load of Residential Buildings Using Multiple Linear Regression Artificial Neural Network." Journal of Advanced Research in Fluid Mechanics and Thermal Sciences 92, no. 1 (2022): 28-38. https://doi.org/10.37934/arfmts.92.1.2838.

[27] Faye, Ibrahima, and Laila Amera Aziz. "Artificial Neural Networks Solutions for Solving Differential Equations: A Focus and Example for Flow of Viscoelastic Fluid with Microrotation." Journal of Advanced Research in Fluid Mechanics and Thermal Sciences 112, no. 1 (2023): 76-83. https://doi.org/10.37934/arfmts.112.1.7683.