Measurement: Sensors 27 (2023) 100756

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Large-scale data-driven financial risk management & analysis using machine learning strategies

M. Senthil Murugan [a,*], Sree Kala T [b,**]

[a] Immaculate College for Women, Viriyur, Tamil Nadu, India
[b] Department of Computer Science, VISTAS, Pallavaram, Chennai, Tamil Nadu, India

ARTICLE INFO

Keywords:
Big data
Financial risk
Risk management
Risk analysis
Machine learning
Nearest neighbor
Logistic regression
XG boost
Clustering approaches

ABSTRACT

Recently the wave of financial crises that have shaken the economy and financial world have caused severe bank losses. Some researchers have focused on examining catastrophes to develop an early warning system to handle financial risks. Financial experts and academics are increasingly interested in developing big data financial risk prevention and control capabilities based on cutting-edge technologies like big data, machine learning (ML), and neural networks (NN), as well as accelerating the implementation of intelligent risk prevention and control platforms. This research analyzed and processed the large-scale datasets before training and evaluated using the three models – cluster based K-nearest neighbor (KNN), cluster based logistic regression (LR), and cluster based XG Boost for their ability to predict loan defaults and their occurrence of likelihood. The investor's wealth proportion measure of the proposed model ranges from 0.02 to 0.09. Applying the value-at-risk strategy, the optimal consumption stability not exceeded 5% of the total investment wealth. The simulation results of the proposed model obtained better results of large-scale data-driven financial risks over the state-of-the-art methods. In this article XG Boost, KNN are the machine learning are proposed for financial risk management with IOT deployement.

1. Introduction

The necessity of preventing and managing financial risks has become more crucial in recent years due to several factors, including rising macroeconomic pressure, more regulatory requirements, increased market competitiveness and increased criminal activity. Commercial banks are risk takers and risk managers in their capacity as financial intermediaries. Commercial banks' business environments are getting more complex and riskier as the financial system becomes more complex and global financial integration picks up speed. Commercial banks' ability to gain competitive advantages in this new environment depends on their ability to prevent and control risk intelligently. Extensive data risk prevention and control capabilities based on artificial intelligence (AI), biometrics, and big data have become hot topics for financial experts and researchers [1]. Internet of technologies (IoT) can help banks and other financial institutions get real-time data on their own and their client's assets, improving their algorithm's effectiveness for evaluating financial risks [2].

Several businesses had a significant boost in growth due to big data, which was made possible by advances in the field of information technology. Big data refers to data collection that a computer can quickly obtain, manage, store, and analyze. Big data has improved the efficiency of data processing, exchange, and storage. All of an organization's business, financial, and related data can be seen at the same time. It provides essential data for the early detection of internet credit troubles. Applying big data analytic tools to credit risk management enhances the precision and objectivity of risk estimation and early warning. The multiple discriminant and LR discriminant analysis methods efficiently analyze credit risk. The disadvantages are that it relies too heavily on past data, requires a significant amount of actual data as the premise, and has limited dynamic early warning capabilities [3].

The structure of financial management assessment is sketched in Fig. 1. In addition to using risk commentary models and writing commitments on the board, this is accurate. Risk assessment in land relations is usually left to the evaluator's skill and whims. The concrete quickly understood numerical estimations that often apply to Hageman land appraisals Risk is a crucial component of the business. It can be challenging to distinguish between poor appraisal and oversight because of

* Corresponding author.
** Corresponding author.
E-mail addresses: vasukisenthil33@gmail.com (M.S. Murugan), sreekalatm@gmail.com (S.K. T).

https://doi.org/10.1016/j.measen.2023.100756
Received 1 December 2022; Received in revised form 24 February 2023; Accepted 26 April 2023
Available online 28 April 2023

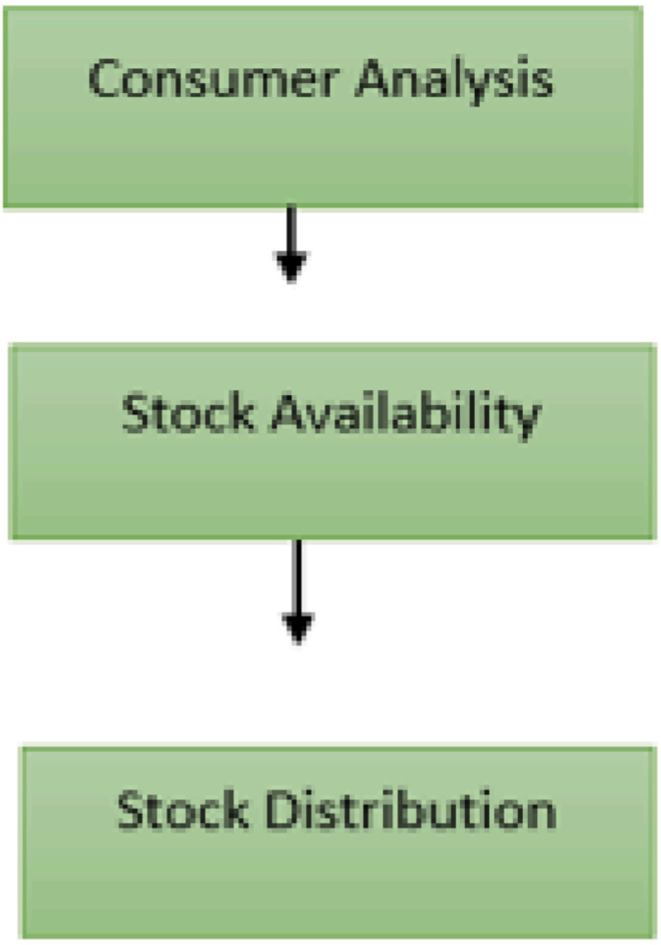

Fig. 1. Financial management assessment.

the danger's puzzling attitude, which can take many different shapes [4]. There are two motivations for conducting bank default research. Regulators can better regulate banks if they understand the variables leading to failure. A bank's bail-in, which its shareholders and bondholders fund, and its bail-out can be lowered. In this case, regulators can prevent a bank from failing if they uncover concerns early enough. A second factor to consider is that the failure of one bank can often lead to the downfall of other institutions in the financial system. Understanding the causes of a single bank failure can help understand the causes of systemic financial difficulties, whether microeconomic, characteristic, or macroeconomic imbalances cause them. The root causes of issues can be eliminated or isolated, preventing contagion effects when problems are discovered [5].

Both physical and financial ventures must consider risk. The rate of return and the level of risk are taken into account by both individual and institutional investors when making accurate and individual investing decisions. Financial risk tolerance plays a substantial role in selecting financial assets and allocating funds in the financial markets. As a result, this element is essential for investor portfolio optimization and personal financial planning. For financial service providers to offer the right services to their clients, it is crucial to ascertain a person's financial risk tolerance. The utmost uncertainty that can be endured when making a financial decision is referred to as financial risk tolerance [6]. Most financial organizations today have developed objective validity ranking systems based on empirical data and methods. Modern systems for evaluating and measuring applicants' credit rely on mechanized methods. Certain applicants' unique essential credit traits are given significant benefits. The client is deemed creditworthy if the sum of these benefits exceeds expectations. Otherwise, the client is presumptively uncreditworthy [7]. The model determines the borrower's credit score based on their past credit data, and the creditor then decides on whether or not to grant credit and the size of the credit line. Techniques such as LR, probit regression, nonparametric smoothing, Markov chain model, and recursive segmentation have all replaced the widely used discriminant analysis and linear regression methods [8].

Analytics of large data is typically linked to cloud computing. Real-time analysis of large datasets requires deploying a framework to distribute work to many computers. Understanding ML algorithms in the context of big data is crucial to advance the development of the nation, businesses, and society. The processing of today's massive amounts of data requires ML. ML overcomes limitations imposed by the human factor, efficiently processes data using deep learning (DL), NN, and decision trees, and optimizes data operations [9]. Based on current values, regression makes predictions. Regression uses well-known statistical techniques like linear regression in its most basic form. But in reality, there are a lot of problems that can't solve with simple linear projections of prior values. For example, complex interactions between different predictor parameters can make it challenging to anticipate product sales volumes, stock prices, and failure rates. Using the same model types for classification and regression is frequently possible [10].

To effectively manage credit risk, financial institutions rely on transparent information mechanisms. Big data can impact how organizations and consumers use the market-based credit system by fusing the advantages of cloud computing and information technology. The main objective of the research is to use ML techniques to predict and analyze the credit risk of financial management. Although this is a small idea, the forecast is more accurate than the current methodology, used with a considerable amount of data, one can predict risk and prevent significant losses for the company. The use of this strategy is appropriate for both small and large businesses. The following are this paper's major contributions.

- ❖ This main research objective is to provide an important aspects of project method for classify the Financial risk
- ❖ The method for classifying financial risk from input parameters using a large scale datasets for devolving the speed of classification process.
- ❖ XG Boost, KNN are the machine learning are proposed for financial risk management with IOT deployement.
- ❖ Curve, overall accuracy, precision, true positive rate, and true negative rate are used to determine the performance score.

2. Literature review

Systemic risk study is very recent and closely follows the developments of the financial Crisis. An exhaustive analysis is given below. Specific systemic risk measurements are covered in further detail.

In 2021,Carl. et al. [11], proposed a Reinforcement Learning in Economics and Finance. By lots of practice, incentive learning algorithms explain how an agent can learn the best course of action to take in a progressive decision-making process. We propose applications in economics, game theory, operation research, and finance using the most recent restoration learning approach. The long-term aim of deep learning is to identify an ideal policy, or a mapping from the world's states to a set of actions, in order to enhance cumulative reward.

In 2019,GKou. [12], proposed a machine learning method for systemetic risk analysis in financial sectors. An important topic in role of financial systems is financial systemic risk. Researchers have increasingly used machine learning techniques in their quest to identify and address systemic risk using the huge volumes of data collected in financial markets and systems. This paper's major goals are to explain recent work on financial systemic risk using machine learning techniques and to suggest future research possibilities.

In 2021, Yuegang. [13], introduced a Financialization Risk Assessment for Risk Control based on Machine Learning. The goal is to fully

utilise big data and machine learning to enhance the capacity of trade finance efforts to improve the danger of excessive securitization. Data from particular examples are used to examine the precision. The data is gathered and processed using genetic algorithm (GA), neural network, and principal component analysis (PCA) techniques, and a hazard analysis model of excessive securitization of financial companies is then developed.

In 2021,M.clintworth. et al. [14], proposed a Financial risk evaluation in shipping: a comprehensive machine learning approach. All stakeholders, including regulators and banks, who depend on reasonable assessments of default risk for both credit institutions and bank loan portfolios, place a high value on corporate financial distress (FD) prediction models. The cluttered and unclear nature of financial statement data from transport companies has not received enough consideration.

In 2021, Boning Huang. et al. [16], introduced an Evaluation of Risk Control Using Machine Learning. An essential safeguard for an enterprise's healthy development is scientific significant risk. The discipline of parameter estimation and vulnerability analysis has benefited greatly from machine learning technology's continued progress and maturity. The enterprise's risk management scales, which completely represent the numerous threats presented by the company through a variety of elements, are first created in the particular implementation.

In 2021,Kristian. [17], created an Uncertainty absorption and amplification in machine learning-driven finance. Financial products are affected by confusion about market changes and their effects. Machine learning is being used more frequently as a method to take this unknown and turn it into controllable threat. It makes sense logically and economically to want to use ML methods to decrease uncertainty.

In 2021, S jomthanachi. [18], developed a Risk Management Strategy Using Data Envelopment Analysis and Machine Learning. This research proposes a risk management approach that combines DEA with machine learning. An ML method is used in the risk control and treatment procedures to forecast the level of current risk based on simulated data that matches the hazard analysis assumption.

In 2020, LEI Shimin. [19], created a platform based on Xgboost for detecting financial fraud. This study presented our solution for e-commerce operators to detect fraud. This approach mixes manual and automatic classifications, which is different from many other efforts. Researchers and engineers may be motivated to develop and implement online transaction systems as a result of this work.

In 2020,Y zhang. et al. [20], proposed an Applying the Xgboost Model to Identify Customer Transaction Fraud. The identification of client data fraud is a critical topic for both the common person and lenders, and it is a hotly debated subject in both science and commerce. In this paper, we proposed a content and graphical transaction identity verification model based on Xgboost.

In 2020, Elorlova. [15], designed a Methods of Credit Resource Management Decision-Making By using Machine Learning and Efficiency. The banks' core business is credit operations, which also contribute significantly to their revenue. Financial distress are rising as a result of a rise in financial workloads. The purpose of the study is to support and create innovative technology and organizational models for bank lending that lower credit problems and improve lending performance.

In 2019, G Kou. et al. [12], designed a systemic risk analysis methods using machine learning in financial sector. The role of financial systemic risk to businesses and financial systems cannot be overstated. Several researchers have been using machine learning techniques more regularly in an effort to identify and resolve hazard using the rising amounts of data collected in financial markets and systems.

From the above literature review, the fundamental issue in finance is the pace of innovation by information technology. It covers cryptocurrency development, online transactions, crowdfunding, trading on various platforms, and mobile banking services. Thousands of gigabytes of data are generated by transactions made on these sites. The most crucial factor to consider in this situation is how such massive amounts of data are handled because they contain sensitive personal and financial data that, if improperly handled, could fall into the wrong hands and have disastrous effects on numerous businesses and industries. The inadequacy of standard models to produce accurate credit risk projections because they only contain identity and demographic information (such as ID, name, age, marriage status, and education level). The situation, as mentioned earlier, significantly limits financial organizations' ability to seek new clients. Consolidating heterogeneous, disconnected data from numerous sources is challenging, given the models and statistical approaches currently used in financial risk management.

3. Proposed model

Demand for financing fluctuates from loan provider to finance provider in a cycle. Financing must involve both the supply and demand for the asset to secure the progress of the asset. One type of change-over exercise is financing, which includes cash loans on quick exchanges. From the perspective of asset demanders, venture financing is a financial movement that completes the development of operational property exercises and employs various budgeting strategies, such as the trading of defenses and the total allocation of resources. Based on their development needs, distinct enhancement strategies will be applied to various tiers of ventures. Promoting the practical and visually stimulating activities they hold is the first step in implementing the funding model. Banks are reluctant to lend more money, and early efforts of a firm cannot afford considerable human costs. The buildup of specialized capital is the primary driver of venture finance. The architecture of the proposed model is sketched in Fig. 2.

Before training and analyzing three models, KNN, LR, and XG boost, for their capacity to predict loan defaults and their likelihood, this research explored and pre-processed the large-scale data. The models' capacity to forecast class labels is evaluated using precision, recall, F1, and ROCAUC. The reliability plot and the brier score are assessed using the proposed model. (see Fig. 3).

3.1. Large-scale data

Massive data and complicated data kinds represent big data. A popular framework for processing large batches of data is Hadoop. Its ecosystem includes functional components like Hadoop Distributed File System (HDFS), MapReduce, and HBase. Large-scale data processing tasks are broken down and sent to other computing nodes to be finished. Among them, MapReduce implements task decomposition and scheduling and coordinates computing tasks across multiple machines in parallel operations [1]. With increasing relevance and utility in the IoT era, big data is a new phenomenon that results from collecting massive amounts of complex and diverse data generated at any time and from any location. Despite being of utmost relevance, big data needs a clear and accepted definition. The measures of variability, variety, volume, velocity, veracity, visualization, and value are often used to describe it. The data analysis is made using recent technological advancements that enable the high-velocity capture of variable and complicated (semi--structured, unstructured, and structured data), making it easier to administer, distribute, and store that data [15].

3.2. LR

Based on current values, regression makes predictions. Regression uses well-known statistical techniques like linear regression in its most basic form. But in reality, there are a lot of problems that can't be solved with simple linear projections of prior values. For example, complex interactions between different predictor parameters can make it challenging to anticipate product sales volumes, stock prices, and failure rates. More complicated algorithms might be needed to predict future information. Regression and classification commonly use the same model types.

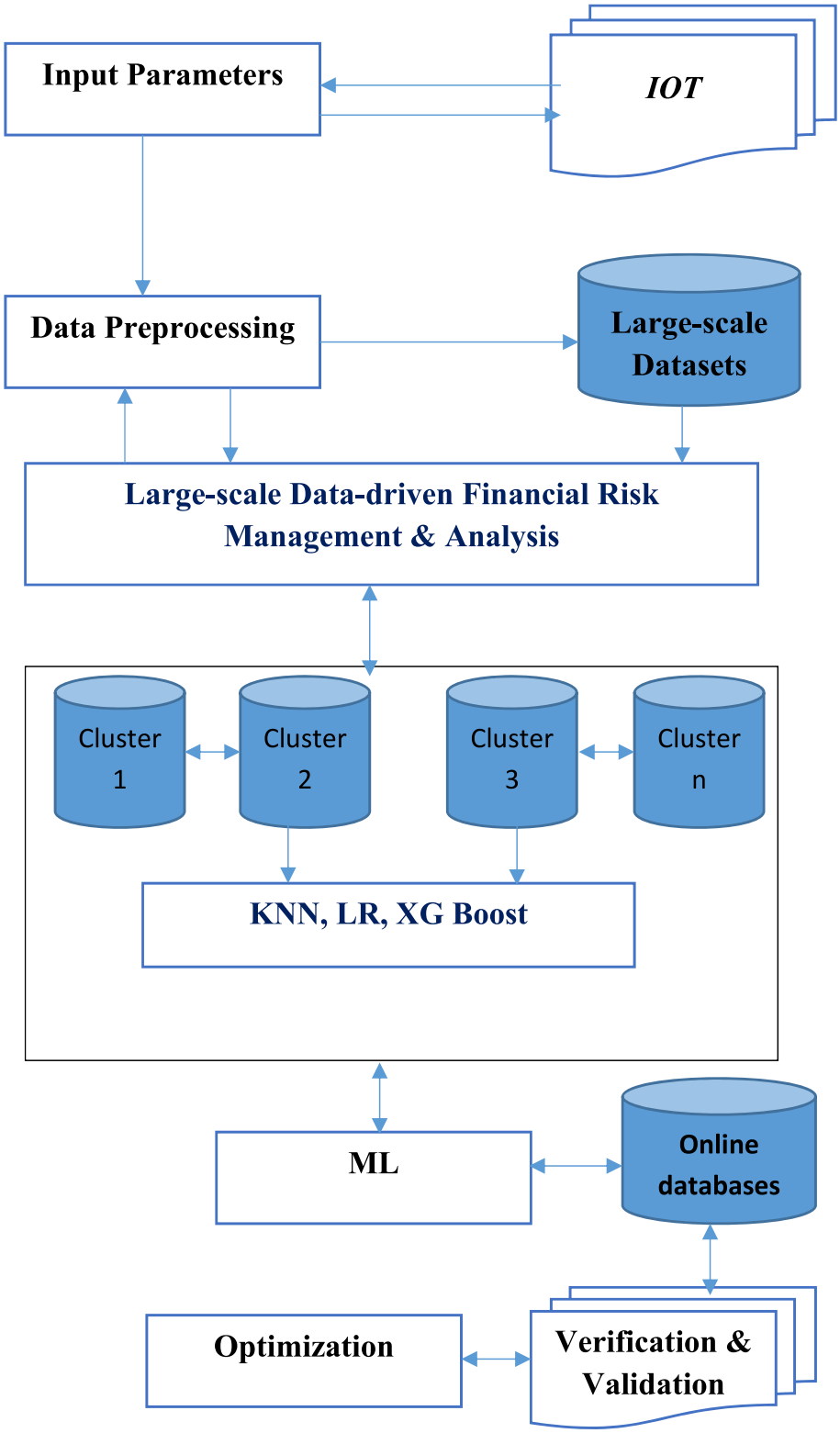

Fig. 2. Proposed architecture.

One of the most popular statistical techniques for categorizing and linking circumstances is LR. Unlike other statistical methods, the LR model estimates utilizing a different distribution function, making it excellent for credit rating issues. Many methods for creating a binary LR model have been provided to increase the model's precision and adaptability. It seems logical to use LR when the dependent variable contains discrete values. As previously stated, the creditworthiness rating, discrete zero, and one independent variable [7]. The popularity of LR is influenced by the structure and kind of behavior of the logistics function as defined:

$$f(u) = \frac{1}{1 + e^{-u}} \tag{1}$$

The range of this function is a closed range between 0 and 1, and its

domain is real numbers, R. Calculate the logistic function's value $f(u)$ to fit an LR model. A statistical analysis method known as LR predicts a data value based on prior observations from a data set. In ML, the significance of LR has increased. The technique enables an ML application's algorithm to classify incoming input using primary data. An LR model predicts a dependent variable by investigating the relationship between one or more preexisting independent variables. The results of LR can be interpreted probabilistically and can be regularised to avoid overfitting.

3.3.

A fundamental classification supports the categorization of multi-valued data. It analyses the closest neighbors in each category by using samples from the training set to determine the distance of new samples to each data point. Because the fresh samples must be compared to the complete training data set, the procedure of this technique may be computationally expensive if the training set is extensive in size and scope. Furthermore, inaccurate classifications will result if the training data contains inaccuracies. KNN is extensively used because it is easy to use, train, and get precise results. This method is employed in various search apps, such as those that suggest related products. KNN is extensively used since it is simple to understand and takes little time to calculate. The parameter k in this process needs to be carefully selected. Two parameters that must be available for different k values are the training and validation error rates. The KNN method's application to text classification has an extensive range. For instance, a system for categorizing public complaints was developed to achieve good governance and democracy and involve citizens in municipal growth [16].

Fig. 2 shows the KNN decision rule for two classes of samples with $k = 1$ and $k = 4$. Fig. 1(a) illustrates the classification of an unknown sample using a single known sample, whereas Fig. 1(b) illustrates the classification of an unknown sample using several known samples. The parameter k is set to 4 in the final scenario, meaning that the four closest to the unknown sample are utilized to classify it. Only one is from the other class, while three are from the same one. The KNN algorithm is shown defined in algorithm 1 and graphically depicted in Fig. 2.

Algorithm1. KNN

1: Classify the identified (i) and unidentified samples.
2: For all the known samples j do:
3: Find the separation between i and j.
4: Find the k-shortest distance.
5: Select the samples and apply the classifiers.

Numerous changes have been made to the KNN approach. It was shown that text categorization performance is enhanced by a flexible KNN approach that combines a weighting technique and a K-variable algorithm. Combining some other ML types and KNN classification, which increases performance and classification accuracy, is another refinement of the KNN method. By merging model and evidence theory, a novel KNN classification algorithm helps to get beyond some of the drawbacks of traditional KNN approaches, namely time-consuming learning.

3.4. XG boost

XG boost is a sharp gradient boosting implementation that may be applied to regression predictive modeling. Test an XG boost regression model using the best practice method of repeated k-fold cross-validation. Whether the task at hand involves classification or regression, the ML algorithm is most frequently used. The fact that it outperforms all other ML algorithms is well known. The XG boost package includes an implementation of the gradient-boosting decision tree technique. This method goes by the titles gradient boosting, stochastic gradient boosting, multiple additive regression trees, and gradient boosting machines. It is an open-source implementation of gradient-boosted trees that is well-liked and effective. Gradient boosting is a supervised learning technique that combines the predictions of several smaller, weaker models to predict a target variable with some degree of accuracy.

Excellent tree-boosting methods include XG boost, also known as eXtreme Gradient Boosting. According to Tianqi Chen, regularisation to accommodate sparse data and a weighted quantile sketch for tree learning are two features of XG boost. They also offer insights that help create a quick and scalable tree-boosting method. Some of these revelations are data compression, sharding, and cache access patterns. Because of these methods and insights, XG boost outperforms most other ML algorithms in terms of speed and accuracy. For engineers or data scientists, using XG boost systems in a distributed system or GPU machine is particularly practical. The loss function is normalized using gradient descent algorithm to attain better prediction score. Training data helps these models learn over time, and the cost function within gradient descent specifically acts as a barometer, gauging its accuracy with each iteration of parameter updates.

4. Results & analysis

In financial risk prediction, accuracy and error rates are critical indicators of categorization systems. Measures of the receiver operating characteristic curve area under the curve, overall accuracy, precision, true positive rate, and true negative rate are used to determine the performance score.

Accuracy is the proportion of cases that are correctly categorized. It is one of the most well-liked performance metrics for categorization.

$$Accuracy = (TN + TP) / (TP + FP + FN + TN) \quad (2)$$

These terms are used to describe true positivity (TP), negative positivity (TN), false positivity (FN), and truthful positivity (FN). FP is the percentage of situations incorrectly labeled as fault-prone but is nonetheless included in the category. The number of incidents mistakenly classified as non-fault-prone is known as FN. The percentage of positive or abnormal events that are actually positive is known as precision.

$$Precision = TP / (TP + FP) \quad (3)$$

Operating characteristic (ROC), an acronym for the trade-off between transmission and reception (TP/FP) rates, is used here. The

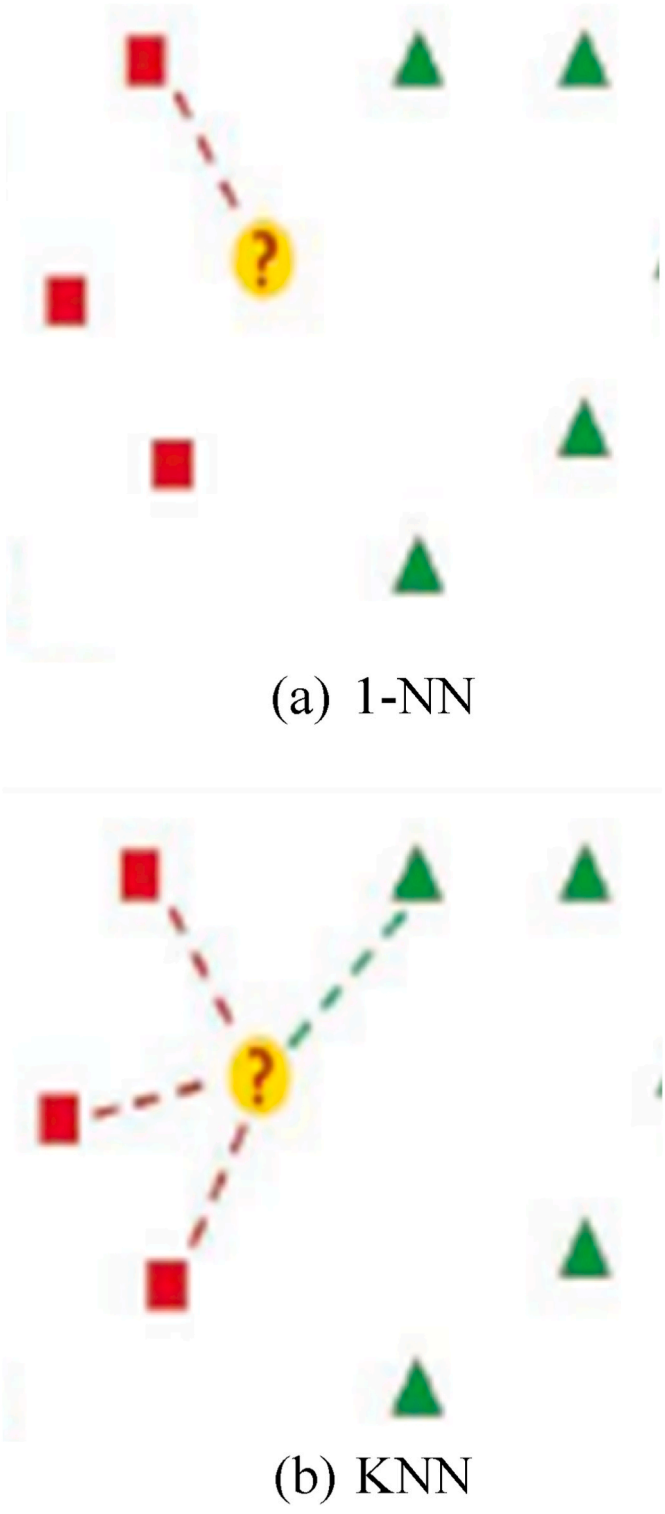

Fig. 3. Decision rules.

accuracy of a classifier can be measured using the ROC curve (AUC). Having a broader area to classify improves the classifier's performance.

4.1. Weight of evidence

There are many ways to determine how the rates of good and bad loans vary between categories. However, the acceptance criteria that is frequently used is the strength of the evidence. Its weight of evidence generally indicates a dependent variable's ability to be predicted by an independent variable. The ratio of observations of the first type of dependent variable outcome that fall into the appropriate category of the independent variable to observations of the second type of dependent variable outcome that fall into the appropriate category of the independent variable is the natural logarithm of the weight of evidence formula.

$$WOE = \ln\left((good\ distribution) / (bad\ distribution)\right) \tag{4}$$

In the analysis, the index should be sorted from small to large before calculating the appropriate WOE. The greater the positive index, the higher the reverse index; the higher the WOE value, the lower the low index. If the reaction index is more positive and the disaster value is less positive, the index has a better chance of being identified. The closeness of its value to a straight line indicates a need for more capacity to interpret indicators. The index is judged uneconomical and needs to be deleted if the positive index has a negative correlation with the catastrophe [24].

4.2. Results analysis

The proposed architecture is simulated using large-scale online datasets for the past few years. The results are analyzed for the optimized data-driven risks. The simulation results conclude that the proposed model works better in parametric search over the existing techniques. The average stock return for some of the recent months is drawn in Fig. 4. This datasets gather from FNCE5313. This split into training and testing. For training 30% and testing 70% (see Table 1).

The proposed model outcomes are compared with the other methods and depicted in Fig. 5. The simulation results prove that more than 74% of the purchase of total assets in a single iteration yields good performance. The investment comparison with some other approaches is plotted in Fig. 5. The wealth versus years comparison and optimal consumption are plotted in Figs. 6 and 7, respectively. The inferences conclude that when there is a gradual increase in wealth, its trajectory has ups and downs since some tangible purchase is sometimes required. The investor's wealth proportion measure of the proposed model ranges from 0.02 to 0.09. Applying the value-at-risk strategy, the optimal consumption stability not exceeded 5% of the total investment wealth. The performance metrics comparison of the proposed model with other techniques such as FABC [35] IFABC [36] and QFABC [37]is shown in Table 2(seeFig. 8).

The following are the essential inferences obtained from the simulation.

- The distribution of investment is better than other models.
- The proposed strategy is applied to sparse problems.
- The proposed method controls the global measures.
- The complexity and risks are reduced when applying large-scale investments.

5. Conclusions & future work

Competition in the banking industry means that service quality during credit risk assessment is essential. The bank should review the credit request as soon as feasible to gain a competitive advantage. Because of the importance of financial risk analysis, financial institutions produce the majority of approaches and models to determine whether or not to give credit. KNN is a popular classifier since it is simple to use and effective. The most challenging aspect of the KNN method is determining the appropriate value of k, which denotes the number of neighbors. KNN, LR, and XG boost fared the best on all measures. Thus, this research analyzed which attributes were most important to the main predictions using information gain. The control parameter determines the accuracy and costs trade-off to predict the observation strength. The simulation results prove that the proposed representative works better

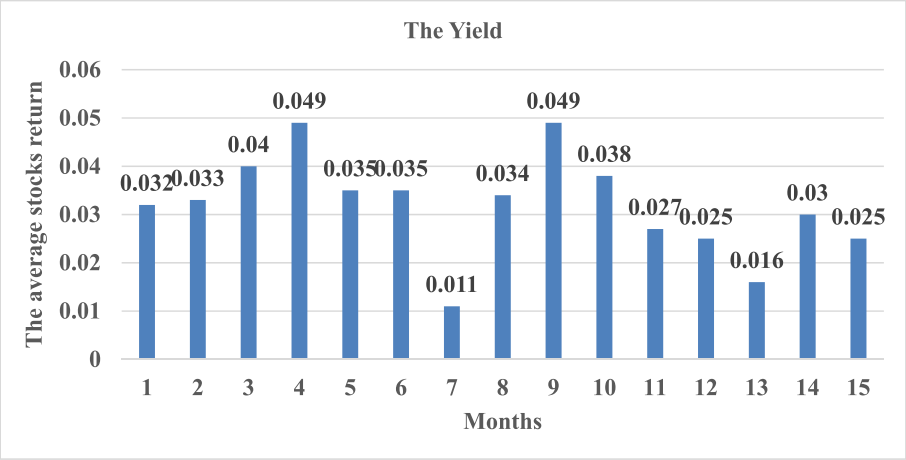

Fig. 4. The average stocks return.

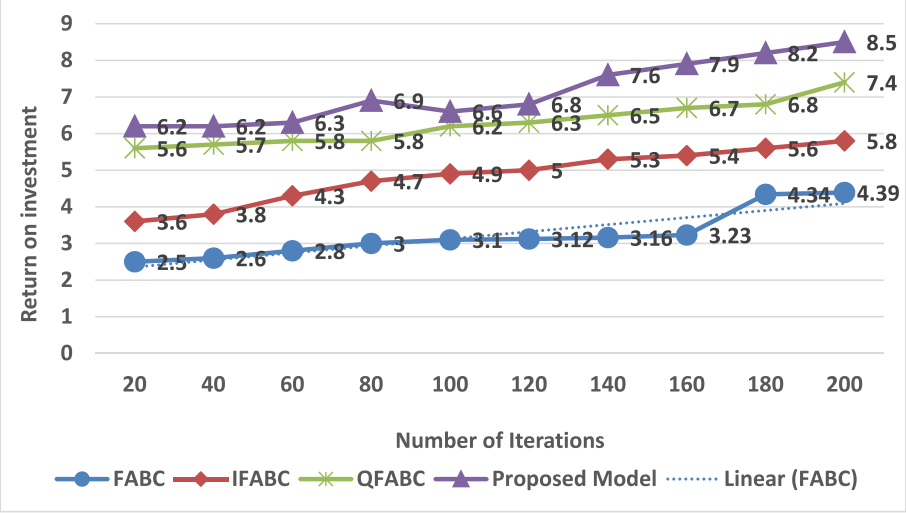

Fig. 5. Performance with other iterative algorithms.

Fig. 6. Financial securities investment comparison.

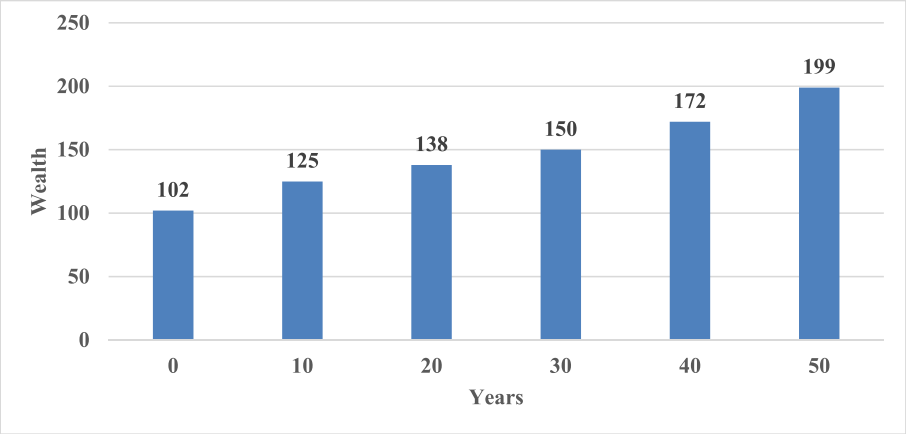

Fig. 7. Years vs. wealth.

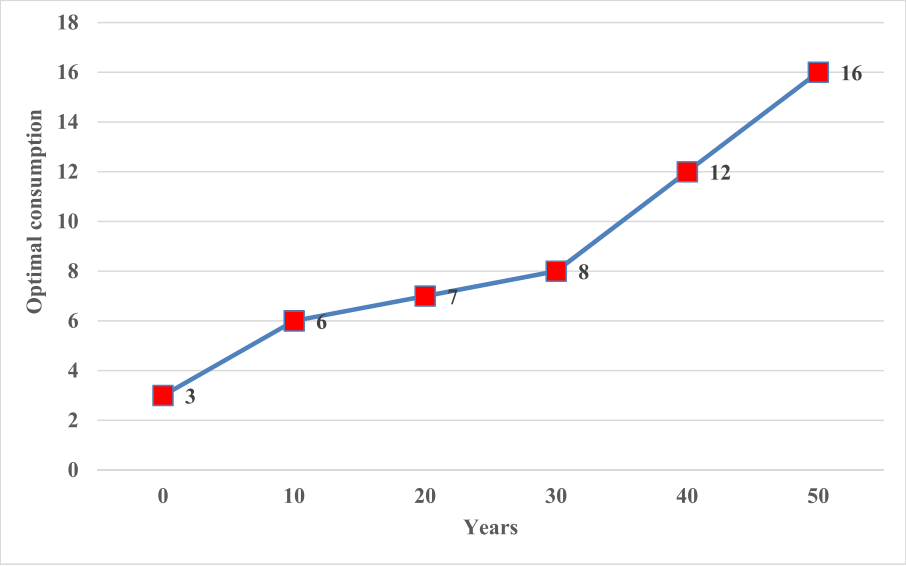

Fig. 8. Years vs. optimal consumption.

than the state-of-the-art methods. The investor's wealth proportion measure of the proposed model ranges from 0.02 to 0.09. Applying the value-at-risk strategy, the optimal consumption stability not exceeded 5% of the total investment wealth. The simulation results of the proposed model obtained better results of large-scale data-driven financial risks over the state-of-the-art methods.

The following are the future enhancements.

- New data analytics constraints will be added [34,35].
- New factors are to be analyzed to reduce further the financial risks [31–33].
- New soft computing strategies will be designed for higher dimensional analysis to reduce the risks [21–23,25–30].

Table 1
Features used in the proposed model.

Predictor Variables	Description
funded_amnt	The loan amount.
emp_length	Employment duration in years. The possible values range from 0 to 10, with 0 indicating less than one year and 10 indicating ten or more years.
annual_inc	Employment duration in years The possible values range from 0 to 10, with 0 indicating less than one year and 10 indicating ten or more years.
last_pymnt_amnt	The most recent total payment received.
mort_acc	Mortgage account count
int_rate	The loan's interest rate.
mo_sin_old_rev_tl_open	Months since the first revolving account was opened.
avg_cur_bal	All accounts' average current balance.
acc_open_past_24 mths	The number of trades that have been opened in the last 24 months.
num_sats	The number of satisfied customers.

Funding

Authors did not receive any funding.

Table 2
Performance metrics comparison with other techniques.

Metrics	XG boost model	LR models	knn model
System vulnerabilities	**2% to 5%**	2.6%–5%	2.8%–5.2%
Average detection time	**< 7 ms**	6–9 ms	8 ms
Average response time	**< 8 ms**	5–12 ms	14millisecond
Higher access users	**1000 to 2500**	1200 to 1250	1300–2600
Mean absolute error	**2% to 7%**	5%–9%	10%–13%
Accuracy	**91% to 96%**	87%–95%	97%–98%

Availability of data and material

Available on request.

Authors' contributions

M.Senthil Murugan, Dr.T. Sree Kala conceptualized the model, collected the Routing protocol and Route maintenance index details and reviewed the manuscript.

CRediT authorship contribution statement

M. Senthil Murugan: Conceptualization. **Sree Kala T:** Funding acquisition, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The author with a deep sense of gratitude would thank the supervisor for his guidance and constant support rendered during this research.

References

[1] S. Mohammed, Research on Financial Risk Prevention and Control Methods Based on Big Data, 2019.
[2] H. Zhou, G. Sun, S. Fu, J. Liu, X. Zhou, J. Zhou, A big data mining approach of PSO-based BP neural network for financial risk management with IoT, IEEE Access 7 (2019) 154035–154043.
[3] G. Du, Z. Liu, H. Lu, Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment, J. Comput. Appl. Math. 386 (2021), 113260.
[4] N. Zhang, Big data simulation for financial risk assessment of real estate bubble based on embedded system and artificial intelligence algorithm, Microprocess. Microsyst. 82 (2021), 103941.
[5] P. Cerchiello, P. Giudici, Big data analysis for financial risk management, J. Big Data 3 (1) (2016) 1–12.
[6] Y. Bayar, H.F. Sezgin, Ö.F. Öztürk, M.Ü. Şaşmaz, Financial literacy and financial risk tolerance of individual investors: multinomial logistic regression approach, Sage Open 10 (3) (2020), 2158244020945717.
[7] M.J. Ershadi, D. Omidzadeh, Customer validation using hybrid logistic regression and credit scoring model: a case study, Calitatea 19 (167) (2018) 59–62.
[8] S. Fan, Y. Shen, S. Peng, Improved ML-based technique for credit card scoring in internet financial risk control, Complexity 2020 (2020).
[9] R. Hou, Y. Kong, B. Cai, H. Liu, Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning, Neural Comput. Appl. 32 (10) (2020) 5399–5407.
[10] S.B. Imandoust, M. Bolandraftar, Application of k-nearest neighbor (knn) approach for predicting economic events: theoretical background, Int. J. Eng. Res. Afr. 3 (5) (2013) 605–610.
[11] A. Charpentier, R. Elie, C. Remlinger, Reinforcement learning in economics and finance, Comput. Econ. (2021) 1–38.
[12] G. Kou, X. Chao, Y. Peng, F.E. Alsaadi, E. Herrera-Viedma, Machine learning methods for systemic risk analysis in financial sectors, Technol. Econ. Dev. Econ. 25 (5) (2019) 716–742.
[13] Y. Song, R. Wu, The impact of financial enterprises' excessive financialization risk assessment for risk control based on data mining and machine learning, Comput. Econ. 60 (4) (2022) 1245–1267.
[14] M. Clintworth, D. Lyridis, E. Boulougouris, Financial Risk Assessment in Shipping: a Holistic Machine Learning Based Methodology, Maritime Economics & Logistics, 2021, pp. 1–32.
[15] E.V. Orlova, Decision-making techniques for credit resource management using machine learning and optimization, Information 11 (3) (2020) 144.
[16] B. Huang, J. Wei, Y. Tang, C. Liu, Enterprise Risk Assessment Based on Machine Learning vol. 2021, Computational Intelligence and Neuroscience, 2021.
[17] K.B. Hansen, C. Borch, The absorption and multiplication of uncertainty in machine-learning-driven finance, Br. J. Sociol. 72 (4) (2021) 1015–1029.
[18] S. Jomthanachai, W.P. Wong, C.P. Lim, An application of data envelopment analysis and machine learning approach to risk management, IEEE Access 9 (2021) 85978–85994.
[19] L.E.I. Shimin, X.U. Ke, Y. Huang, S.H.A. Xinye, An Xgboost based system for financial fraud detection, in: E3S Web of Conferences, vol. 214, EDP Sciences, 2020, p. 2042.
[20] Y. Zhang, J. Tong, Z. Wang, F. Gao, Customer transaction fraud detection using xgboost model, in: 2020 International Conference *on Computer Engineering And Application (ICCEA)*, IEEE, 2020, March, pp. 554–558.
[21] R. Marappan, G. Sethumadhavan, Solution to Graph Coloring Problem using Evolutionary Optimization through Symmetry-Breaking Approach, International Journal of Applied Engineering Research 10 (10) (2015) 26573–26580.
[22] Raja Marappan, Sethumadhavan. Gopalakrishnan, Solving Fixed Channel Allocation using Hybrid Evolutionary Method MATEC Web of, Conferences 57 (2016) 02015, https://doi.org/10.1051/matecconf/20165702015.
[23] R. Marappan, G. Sethumadhavan, Solution to graph coloring problem using divide and conquer based genetic method, in: 2016 International Conference on Information Communication and Embedded Systems (ICICES), 2016, pp. 1–5, https://doi.org/10.1109/ICICES.2016.7518911.
[24] R. Marappan, G. Sethumadhavan, Solution to graph coloring using genetic and tabu search procedures, Arabian J. Sci. Eng. 43 (2018) 525–542, https://doi.org/10.1007/s13369-017-2686-9.
[25] R. Marappan, G. Sethumadhavan, Complexity analysis and stochastic convergence of some well-known evolutionary operators for solving graph coloring problem, Mathematics 8 (2020) 303, https://doi.org/10.3390/math8030303.
[26] R. Marappan, G. Sethumadhavan, Solving graph coloring problem using divide and conquer-based turbulent particle swarm optimization, Arabian J. Sci. Eng. (2021), https://doi.org/10.1007/s13369-021-06323-x.
[27] S. Bhaskaran, R. Marappan, Design and analysis of an efficient machine learning based hybrid recommendation system with enhanced density-based spatial clustering for digital e-learning applications, Complex Intell. Syst. (2021), https://doi.org/10.1007/s40747-021-00509-4.
[28] G. Sethumadhavan, R. Marappan, "A genetic algorithm for graph coloring using single parent conflict gene crossover and mutation with conflict gene removal procedure", in: 2013 IEEE International Conference on Computational Intelligence and Computing Research, 2013, pp. 1–6, https://doi.org/10.1109/ICCIC.2013.6724190.
[29] R. Marappan, S. Bhaskaran, New evolutionary operators in coloring DIMACS challenge benchmark graphs, Int. j. inf. tecnol. (2022), https://doi.org/10.1007/s41870-022-01057-x.
[30] S.V. Ilkevich, E.Y. Listopad, N.V. Malinovskaya, P.P. Rostovtseva, N. N. Drobysheva, A.V. Borisov, Financial risk and profitability management in Russian insurance companies in the context of digitalization, Risks 10 (2022) 214, https://doi.org/10.3390/risks10110214.
[31] B. Gładysz, D. Kuchta, Sustainable metrics in project financial risk management, Sustainability 14 (2022), 14247, https://doi.org/10.3390/su142114247.
[32] A.A. Prabhawa, I. Harymawan, Readability of financial footnotes, audit fees, and risk management committee, Risks 10 (2022) 170, https://doi.org/10.3390/risks10090170.
[33] I. Kalina, V. Khurdei, V. Shevchuk, T. Vlasiuk, I. Leonidov, Introduction of a corporate security risk management system: the experience of Poland, J. Risk Financ. Manag. 15 (2022) 335, https://doi.org/10.3390/jrfm15080335.
[34] E. Pecina, D. Miloš Sprčić, I. Dvorski Lacković, Qualitative analysis of enterprise risk management systems in the largest European electric power companies, Energies 15 (2022) 5328, https://doi.org/10.3390/en15155328.
[35] R. Marappan, G. Sethumadhavan, Divide and conquer based genetic method for solving channel allocation, in: 2016 International Conference on Information Communication and Embedded Systems (ICICES), 2016, pp. 1–5, https://doi.org/10.1109/ICICES.2016.7518914.