Explorsearch ISSN:3048-815X(Online) Vol. 03, No. 03, July-September 2026, 16-24

Original Article | **Peer Reviewed** | **Open Access**

Exploresearch

Impact Factor (Cosmos: 6.262& I2OR: 3.585)

www.mgmpublications.com

Crossref

An Extensive Survey of Data Mining Methods and their New Applications in Loan Default Detection: Challenges and Future Research Directions

Shama Rani*

Associate Professor of Commerce, GCG Palwal, Kurukshetra and Research Scholar, University School of Management, Kurukshetra University Kurukshetra.

***Corresponding author:** shamagovtcollege@gmail.com*

***Citation**: Rani, S. (2026). An Extensive Survey of Data Mining Methods and their New Applications in Loan Default Detection: Challenges and Future Research Directions. Exploresearch, 03(03), 16–24.*

Article History:

Received: 05 August 2026
Revised: 08 August 2026
Accepted: 09 August 2026
Published: 20 August 2026

Keywords:

Data Mining, Machine Learning, Classification, Decision Making, Financial Risk, Loan Default Prediction.

Abstract: *Data mining is an effective analytical tool that has become an emerging technique to discover meaningful patterns, relationships, and prediction from large-scale data sets. Data mining is being used in various fields including healthcare, education, business, cybersecurity, agriculture, and finance to enhance decision-making processes in the face of the fast-growing digital data. Different machine learning and statistical methods such as classification, clustering, association rule mining as well as predictive modelling have been extensively utilized in knowledge discovery and automated decision support. Financial risk assessment is one of the several applications that have been in the spotlight because of the availability of customer transaction and credit related data. Prediction of default is a very critical issue for financial institutions as improper credit appraisal can lead to financial risk and economic loss. While multiple data mining techniques have been attempted for credit risk analysis, there are still some issues to be studied, namely: imbalance of the data sets, interpretability of the model, new borrower behavior, and limitations for using the models in real scenarios. This paper provides full overview of the current applications of data mining and discusses how they have helped in intelligent decision-making systems. Moreover, the study suggests that loan default detection is a potential research field where advanced data mining techniques can be developed for an accurate, reliable and explainable loan default prediction model. The review also identifies potential avenues for further research and development in the field of loan default prediction, which include the incorporation of hybrid algorithms, deep learning techniques, explainable artificial intelligence, and real-time financial data analysis.*

Introduction

Digital technologies have progressed at an amazing pace, generating an unprecedented amount of data from many different sources such as financial transactions, healthcare systems, educational platforms, social networks, industrial processes and government services. As digital information is growing continuously, data is now a useful strategic resource for organisations that are looking to create effective decision-making mechanisms. But, the massive amount of data, data velocity, and complexity have made it difficult to extract meaningful knowledge from data through traditional data-processing techniques. Thus, powerful analytical methods that are able to extract information from large data sets to reveal hidden patterns, relationships, and insights for prediction have become key in today's world of computation (Han et al., 2022).

Data mining is one of the most important research fields that allows one to discover useful knowledge from massive amounts of data by combining database systems, statistics, machine learning and artificial intelligence. Han et al.,2022 defined data mining as discovering interesting patterns and knowledge from large scale data, systematically with computational techniques. Data mining is different from conventional statistical methods, primarily because it is not for testing hypotheses, but is for automated pattern extraction, prediction and knowledge generation from large and multidimensional data (Tan et al., 2019). Data mining techniques have become very important tools in intelligent decision-support systems, because they can generate knowledge from raw data.

Data mining techniques have been well implemented in many fields in recent years to enhance the efficiency of the operations and the accuracy of decision making. Healthcare provides examples of the use of data mining methods for disease diagnosis, medical image analysis, prediction of patient risk and personalized treatment planning (Kaur & Rani, 2016). In the education domain, predictive data mining techniques have been used to mine student performance data to discover learning patterns and forecast students' results (Romero and Ventura,2020). In the same way, applications of data mining have been used in the agriculture industry for crop prediction, soil analysis and precision farming, where environmental and agriculture data is combined (Benos et al.,2021).

Data mining is applied in the field of business for the customer relationship management, market segmentation, recommendation systems and sales forecasting. The algorithms of association rule mining, clustering, and classification enable the organizations to understand the consumer's behaviour and gain strategies based on the data (Witten et al.,2017). In addition, cyber security applications use data mining techniques for intrusion detection, anomalous behavior detection, and prevention of malicious behaviors by discovering abnormal patterns in large network datasets (Ahmed et al.,2016).

The financial industry is one of the most significant industries for the use of data mining techniques, as it provides a lot of structured and transactional data. There are vast amounts of data that financial institutions produce that are associated with their customers' profile, their credit history, the loan transactions they undertake and their repayment behaviour. Data mining techniques are effective solutions of credit scoring, fraud detection, risk assessment, customer profiling and investment analysis (Lessmann et al.,2015). These techniques' ability to detect intricate patterns within monetary characteristics helps businesses make better decisions and reduce economic risks.

It is an important research issue in financial data analytics to predict loan default. Banking instability, bank losses and impaired credit structures are all consequences of loan defaults. The traditional credit evaluation techniques are mostly based on the financial indicators that are predefined and history data of the credit information; but these cannot be fully able to capture the interactional effect among various borrower features. Hence, data mining and machine learning methods have been considered as alternative methods to build predictive credit-risk models that can detect potential risks to default prior to the occurrence of financial losses (Baesens et al.,2003).

A wide range of classification and prediction algorithms such as decision tree, logistic regression, support vector machine, artificial neural network, random forest and ensemble learning have been explored on credit risk assessment and prediction of loan default. Lessmann et al. showed that the use of advanced classification methods to credit scoring can enhance the performance of the credit scoring process by considering non-linear relationships in the financial data. In the same manner, the utilization of the data-driven model to predict credit risk using historical financial information was found to be effective by Yeh and Lien (2009).

Although considerable efforts have been made, there are still a number of issues that are still unsolved in the field of loan default prediction and hence it remains a challenging research area. In many datasets, as is the case in finance, defaulting cases are of lower proportions than cases of non-default. Moreover, financial institutions have to interpret and explain the models, since the consequences of lending decisions directly affect individuals and organisations. With the rise of explainable artificial intelligence, the need to build predictive models that can explain the decisions made by the models has become more critical than just having high accuracy in predictions (Arrieta et al.,2020). Additionally, adaptive models are needed to handle changing economic conditions, evolving customer behaviour, and dynamic financial environments to ensure accurate predictions over time.

While the use of data mining techniques has been widely proven in several areas such as healthcare, education, business, agriculture and cybersecurity, making robust, explainable and adaptive models for loan default detection is still a huge research space. Most of the current works have centered on optimizing the predictive performance of traditional data sets, with few investigations of integrated methods that incorporate cutting-edge data mining algorithms, explainability, real-time financial behaviour analysis and handling of imbalanced data.

Hence, this is a review paper, which intends to discuss the usage of data mining techniques in various real-world areas and how they have helped in bringing intelligence to decision making systems. In addition, loan default detection is recognized as a new area of research and the current problems, limitations and future prospects in the development of efficient financial risk prediction models are discussed. The results of this review give a glimpse of the potential of advanced data mining techniques in reliable, transparent and adaptable decision-making systems in the modern data-driven world.

Methodology Systematic Literature Review

The PRISMA 2020 guidelines were used to generate a complete literature search based on scientific databases (Page et al., 2021) such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, and SpringerLink**.**

The search strategy consisted of various combinations that involved the keywords “data mining,” “machine learning,” “artificial intelligence,” “data-driven decision making,” “credit risk” and “loan default prediction.” Peer-reviewed publications (journal articles and conference papers) were included in the study. Duplicate papers, non-English publications and papers not related to data mining applications were filtered out.

The chosen studies were evaluated according to research domain, data mining techniques, Datasets used, Evaluation methods and Key findings. Literature was divided into the five major areas of use: healthcare, education, agriculture, business, cybersecurity and finance. Financial applications, especially credit risk assessment and loan default prediction, were focussed to find current approaches, limitations and future research, to be carried out.

A selection process was conducted using PRISMA, which included 4 stages: 1) identification, 2) screening, 3) eligibility assessment and 4) inclusion of relevant studies. Results were analysed qualitatively to gain insights into the development of data mining techniques, their role in intelligent decision-making and the new research needs in the field of loan default detection.

Data Mining Applications Review

As digital data proliferates, so have the capabilities of data mining techniques spread in several fields for knowledge mining, prediction and smart decision making. Data mining includes a mix of statistical analysis, machine learning, and computational methods to discover patterns that are not apparent in vast amounts of data. There are many previous studies that have shown how the data mining applications are beneficial in enhancing operational efficiency and decision-support systems in healthcare, education, agriculture, business, cyber security, and finance.

- **Healthcare Applications**

One of the most well researched areas for data mining is healthcare, with electronic health records, medical images, and clinical datasets being readily available. Data Mining techniques have been used to predict, diagnose, classify patients and optimize treatment. The classification methodology like Decision Tree, SVM, ANNs and Ensemble models have effectively shown the pattern identification, which is helpful to make clinical decisions**.**

Koh and Tan (2005) summarized healthcare data mining applications and the ability for healthcare data mining to enhance medical decision-making. Kaur and Rani (2016) also explained about the application of classification and clustering in disease prediction and healthcare analytics. More recently, Rieke et al., (2020), the application of artificial intelligence and data-driven approaches to develop personalized healthcare systems has been emphasized.

- **Educational Applications**

To analyse and understand the student learning behaviour and optimise the educational management, educational data mining (EDM) has become an important field of research. Predicting academic performance, detecting at-risk students, analysing learning patterns and formulating adaptive learning systems are accomplished using data mining approaches.

Romero and Ventura (2020) have reviewed comprehensive educational data mining and emphasized on the classification, clustering and association rule mining technique for extracting informative content from educational data. Likewise, Baker and Inventado (2014) showed that data mining techniques can be used for early prediction of student outcomes and personalization of learning strategies.

- **Applications of Business and Marketing**

Data mining techniques are very useful for business organizations for analysing customers' behaviour, optimising their marketing strategies and for supporting managerial decisions. Major applications include customer segmentation, recommendation systems, market basket analysis and sales forecasting.

Linoff and Berry (2011) discussed the role of data mining for Customer Relationship Management (CRM) and Business Intelligence (BI). Han et al. (2022) emphasized that clustering, classification and association rule mining are very popular in finding consumer behavior and enhancing strategy making.

- **Agricultural Applications**

Data mining has been used in agriculture by combining environmental, climatic and crop related data. For crop yield prediction, crop disease identification, soil analysis and precision agriculture, data mining and machine learning models have been used.

Studies in machine learning in agriculture have been carried out in Liakos et al. (2018), of which they concluded that machine learning applications in the agriculture sector enhance the productivity and management of the resources. Additionally, Benos et al. (2021) highlighted how AI and data analysis can contribute to sustainable agriculture.

- **Cybersecurity Applications**

With the growing complexity of the cyber threat has come a growing acceptance of data mining techniques for intrusion detection, anomaly identification and security monitoring. By analyzing network data on a large scale, abnormal patterns and potential cyber attacks can be discovered.

Ahmed et al. (2016) surveyed anomaly detection techniques and concluded that data mining is a good method for detecting abnormal behaviours in the network. Buczak and Guven (2016) pointed out the usage of machine learning and data mining techniques in cyber security threat detection.

- **Financial Applications**

The financial industry is one of the largest fields of application of data mining due to the ample amount of data related to transactions, customers and credit. Data Mining techniques have been well used in credit scoring, fraud detection, risk assessment, financial forecasting and customer analysis.

Hand and Henley (1997) formulated statistical methods of credit scoring and which later became the financial prediction based on machine learning. In their work, Baesens et al. (2003) presented a comparison of classification algorithms on credit scoring and showed the success of advanced computational methods. Lessmann et al. (2015) presented an in-depth analysis of the state-of-the-art classifiers for credit risk prediction and illustrated the growing relevance of machine learning techniques in financial decision-making.

Predicting loan defaults has become a relevant research line in financial data mining due to the fact that under or over estimating borrower risk may entail enormous economic losses. Many algorithms were studied such as logistic regression, decision trees, support vector machines, random forests, neural networks and the ensemble model to predict default behaviour (Yeh & Lien, 2009; Abdou & Pointon, 2011). But issues concerning the problem of imbalanced data, the transparency of models and dynamic economic environments and actual deployment into the real world still underpin further research.

Classification of Data Mining Techniques

In the literature, data mining techniques were classified based on various learning strategies: data analysis goals and knowledge discovery capabilities. Some of the initial works segment the data mining techniques into descriptive and predictive techniques. Descriptive approaches are used to discover hidden patterns and predictive approaches are used to create models for future estimation and data-founded decision making (Fayyad et al., 1996). Later studies added new ones which included machine learning, statistical learning, artificial intelligence-based approaches, etc. methods for analyzing complex data (Kotsiantis, 2007; Wu et al., 2008). Due to these well-recognized types, the important data mining methods are classification, regression, clustering, association rule mining, anomaly detection, ensemble learning and deep learning methods.

Category	Learning Type	Major Algorithms	Primary Objective	Representative References
Classification	Supervised learning	Decision Tree, SVM, Naïve Bayes, ANN	Predict predefined classes	Han et al. (2022); Witten et al. (2017)
Regression	Supervised learning	Linear Regression, Regression Trees	Predict continuous values	Tan et al. (2019)
Clustering	Unsupervised learning	K-means, Hierarchical clustering, DBSCAN	Discover hidden patterns	Han et al. (2022)
Association Mining	Unsupervised learning	Apriori, FP-Growth	Identify relationships among variables	Agrawal et al. (1993)
Anomaly Detection	Semi-supervised/Unsupervised	Isolation Forest, One-Class SVM	Detect abnormal behaviour	Ahmed et al. (2016
Ensemble Learning	Supervised learning	Random Forest, Boosting, XGBoost	Improve prediction performance	Lessmann et al. (2015)
Deep Learning	Representation learning	ANN, CNN, RNN	Extract complex patterns	LeCun et al. (2015)

The classification presented in Table 2 summarizes the major data mining approaches reported in the literature based on their learning mechanisms and analytical objectives. Traditional techniques such as classification, regression, clustering, and association rule mining remain widely used for pattern discovery and predictive analysis, while advanced approaches including ensemble learning and deep learning have gained importance due to their improved capability in handling complex and large-scale datasets. Classification and ensemble-based methods are popular approaches largely used in the research of financial risk assessment and loan default prediction because they are effective in modelling the complex relationships present in financial structured data.

Comparative Analysis of Data Mining Techniques

After presenting the classification of data mining techniques, a comparative analysis is necessary to understand the strengths, limitations, and suitability of different approaches for real-world decision-making applications. Previous studies have demonstrated that the performance of data mining algorithms depends on factors such as dataset characteristics, feature complexity, class distribution, and application requirements (Wu et al., 2008; Lessmann et al., 2015).

Different techniques provide distinct advantages. Traditional statistical and machine learning methods such as logistic regression, decision trees, and support vector machines offer interpretability and are widely adopted in structured data analysis. However, complex datasets with nonlinear relationships often require advanced approaches such as ensemble learning and deep learning models to improve predictive capability (Breiman, 2001; LeCun et al., 2015).

Comparative Analysis of Data Mining Techniques

Technique	Major Algorithms	Strengths	Limitations	Common Applications
Classification	Decision Tree, SVM, Naïve Bayes, ANN	Effective for prediction and decision-making with labelled data	Requires labelled datasets; sensitive to feature quality	Disease diagnosis, credit scoring, fraud detection
Regression	Linear Regression, Regression Trees, SVR	Simple interpretation and suitable for numerical prediction	Limited capability for complex nonlinear relationships	Forecasting, financial prediction
Clustering	K-means, Hierarchical Clustering, DBSCAN	Identifies hidden patterns without predefined classes	Requires appropriate parameter selection and validation	Customer segmentation, behavioural analysis
Association Mining	Apriori, FP-Growth	Discovers relationships among variables and frequent patterns	May generate redundant or large numbers of rules	Market analysis, recommendation systems
Anomaly Detection	Isolation Forest, One-Class SVM, Autoencoders	Effective for detecting rare and abnormal events	Difficult evaluation due to limited abnormal samples	Fraud detection, cybersecurity
Ensemble Learning	Random Forest, AdaBoost, XGBoost	Provides high predictive accuracy and improved robustness	Reduced interpretability and increased computational complexity	Credit risk prediction, classification problems
Deep Learning	ANN, CNN, RNN	Learns complex patterns from high-dimensional datasets	Requires large datasets and computational resources	Image analysis, advanced prediction systems

From the comparative analysis, it can be seen that traditional data mining techniques are still relevant because of their interpretability and lower computational demands; however, advanced data mining approaches like ensemble learning and deep learning show better prediction performance in the complex data sets. In structured data applications, ensemble methods such as Random Forest and gradient boosting have been successful, such as in financial risk analysis. But issues of interpretability, computational complexity and data needs are important concerns for practical implementations.

- **Application-Oriented Comparison**

Choosing the right data mining technique is really a function of research objectives and characteristics of data. For the applications that demand preliminary decisions to be made (e.g. diagnosis, fraud in finance, or credit risk) classification approaches are still dominating. When the analysis is interested in determining some grouping (behaviors) without the presence of a set of categories, clustering techniques are used. Association rule mining is primarily used to discover relations while anomaly detection concentrating in detecting unusual observation.

In many prediction tasks, ensemble learning methods have successfully been employed to create a set of models and weaken the effect of the limitations of each of them, so that they can provide competitive prediction results based on recent studies (Breiman, 2001; Chen &Guestrin, 2016). In a similar fashion, deep learning techniques have proven themselves to be powerful tools for identifying complex patterns in high dimensional data, but are usually too expensive to be widely adopted and too difficult to be interpreted (LeCun et al., 2015).

- **Relevance to Financial Decision-Making**

Classification and ensemble learning (both are data mining techniques) were well studied area in the field of financial applications because most of the credit-related decisions may be classified as a classification problem. In the credit scoring, we have found that various research projects have proven machine learning models such as random forest, gradient boosting, support vector machines and neural network can improve predictive capability over traditional statistical models (Lessmann et al., 2015).

But even if the predictive power increased, replacing traditional models there would be issues of transparency, an imbalance of data, and how well models would cope with shifting economic conditions. The restrictions suggest that there is a need to further understand strong explainable data mining models which have the potential to identify these loan defaults.

Data Mining Techniques in Loan Default Detection: A Literature Review

One of the significant applications of data mining in financial decision making is loan default prediction, which helps the financial institution to avoid risky loans in order to minimize the credit losses. The earlier credit assessment models primarily used statistical methods and with the advent of profusion in financial and customer behaviour data, the scope has expanded and the use of machine learning and data mining for automated credit risk prediction is increasingly becoming accepted.

The very first studies in credit scoring showed that statistical classification techniques can be used very well to predict loan default. Hand and Henley (1997) discussed the statistical methods that can be used for credit scoring and introduced the relevance of predictive modelling to financial decision systems. Some comparison of the different classification algorithms for credit scoring was done by Baesenset al., (2003) who demonstrated that machine learning techniques could match the performance of traditional statistical methods. The machine learning methods for loan default prediction.

- **The machine learning methods for loan default prediction**.

Classification algorithms were started to perform loan default behaviour prediction several studies have been applied. Since the problem of loan default prediction can be usually formulated as a binary classification problem, decision trees, support vector machines, artificial neural networks and ensemble models have been widely investigated.

Yeh and Lien (2009) showed the performance of various data mining methods in credit default prediction and they confirmed the effectiveness of machine learning methods in financial classification problems. Abdou and Pointon (2011) analyzed the Credit Scoring methods and concluded that the Credit Scoringbased on artificial intelligence is a potential alternative with respect to the classical statistical models.

- **An Ensemble Based Credit Risk Prediction**

Ensemble learning techniques have become more popular in recent studies for their capability of enhancing the prediction performance and dealing with the intricate association in financial settings. The Random Forest, Gradient Boosting and XGBoost model performs well in credit risk prediction as they are able to model the nonlinear interactions between borrower attributes.

For the credit scoring task, Lessmann et al. (2015) performed a large-scale benchmarking study with various classification algorithms, and found advanced machine learning models to be competitive to traditional ones for credit scoring. Ensemble method was also shown to be effective for credit risk prediction data that consist of imbalanced class labels, Brown and Mues (2012).

- **Advanced Learning and Data Mining Techniques**

Credit risk analysis has been an interest in deep learning as data on credit risk can be expressed in terms of large volumes of financial data. Models based on neural networks are capable of detecting a tricky pattern based on high-dimensional financial data. But in the case of financial applications, predictive quality is not the only parameter that is important; they need to be transparent and explainable. Thus, newer studies have dedicated efforts to adding explainable artificial intelligence methods to machine learning models to enhance trust and regulatory acceptance.

- **Challenges identified in the previous research on loan default prediction**

As identified by the literature reviewed, there are multiple challenges that are yet to be addressed:

Challenge	Literature-Based Observation
Class imbalance	Default cases are usually limited compared with non-default cases, affecting model learning (Brown & Mues, 2012)
Interpretability	Complex models often provide limited explanations for financial decisions (Arrieta et al., 2020)
Dataset dependency	Model performance varies according to dataset characteristics and financial environments (Lessmann et al., 2015)
Changing borrower behaviour	Historical patterns may not always represent future repayment behaviour
Regulatory requirements	Financial institutions require transparent and explainable prediction models

Conclusion and Future Perspectives

As discussed in this review, data mining methods have played a huge role in the transformation of data-driven decision-making in various fields such as healthcare, education, agriculture, business, cybersecurity and finance. The literature shows that old data mining methods like classification, regression, cluster model, and association rule mining are still the foundation to discovery of knowledge and the sophisticated methods include use of ensemble learning and deep learning which have enhanced predictive ability against complicated data.

The review of literature reveals that there has been a growing interest in finance applications in general, especially loan default prediction because of the need to have an accurate risk of credit analysis. Ensemble-based and machine learning have shown prospective performance in spotting possible default risks, but various complications have not been neglected, such as unbalanced financial data, insufficient model interpretability, model reliant on a particular data set, and model adjustability to evolving economic circumstances.

Future directions in developing data mining frameworks should be aimed at robust, explainable, and adaptive data mining in detecting loan default. Predictive models can be more reliable and usable, with the potential for the incorporation of explainable artificial intelligence (XAI), hybrid learning approaches, real-time financial data analysis, and cross-institutional datasets. Moreover, future research needs to focus on finding a balance between predictive accuracy and transparency, fairness, and regulation to promote credible financial decision-making.

Data mining has proven to hold potential in many application fields; default detection on the loans is still a developing research stream demanding novel techniques to offer accurate, meaningful and scalable solutions to present-day financial systems.

References

1. Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management, 18*(2–3), 59–88.
2. Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19–31.
https://doi.org/10.1016/j.jnca.2015.11.016
3. Baker, R. S., &Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics: From research to practice* (pp. 61-75). New York, NY: Springer New York.

4. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., &Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, *54*(6), 627-635.

5. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.

   https://doi.org/10.1016/j.inffus.2019.12.012

6. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors, 21*(11), Article 3758. https://doi.org/10.3390/s21113758

7. Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39*(3), 3446–3453.

8. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cybersecurity intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176.

9. Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.

10. Kaur, H., & Rani, P. (2016). A review: Data mining techniques in healthcare. *International Journal of Computer Applications, 144*(7), 34–38.

11. Koh, H. C., & Tan, G. (2005). Data mining applications in healthcare. *Journal of Healthcare Information Management, 19*(2), 64–72.

12. Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

13. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), Article 2674. https://doi.org/10.3390/s18082674

14. Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.

15. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

16. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M. J., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine, 3*, Article 119. https://doi.org/10.1038/s41746-020-00323-1

17. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), Article e1355. https://doi.org/10.1002/widm.1355

18. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.

19. Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications, 36*(2), 2473–2480. https://doi.org/10.1016/j.eswa.2007.12.020.

❑❑❑