Available online at www.sciencedirect.com

ScienceDirect

The Journal of Finance and Data Science 8 (2022) 180–201

http://www.keaipublishing.com/en/journals/jfds/

Credit scoring methods: Latest trends and points to consider

Anton Markov*, Zinaida Seleznyova, Victor Lapshin [1]

National Research University Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia

Received 28 May 2022; accepted 26 July 2022
Available online 7 August 2022

Abstract

Credit risk is the most significant risk by impact for any bank and financial institution. Accurate credit risk assessment affects an organisation's balance sheet and income statement, since credit risk strategy determines pricing, and might even influence seemingly unrelated domains, e.g. marketing, and decision-making. This article aims at providing a systemic review of the most recent (2016–2021) articles, identifying trends in credit scoring using a fixed set of questions. The survey methodology and questionnaire align with previous similar research that analyses articles on credit scoring published in 1991–2015. We seek to compare our results with previous periods and highlight some of the recent best practices in the field that might be useful for future researchers.

JEL classification: **G320** Financing Policy; Financial Risk and Risk Management; Capital and Ownership Structure; Value of Firms; Goodwill; **G210** Banks; Depository Institutions; Micro Finance Institutions; Mortgages; **C440** Operations Research; Statistical Decision Theory; **C650** Miscellaneous Mathematical Tools; **C830** Survey Methods; Sampling Methods; **C450** Neural Networks and Related Topics

Keywords: Credit scoring; Survey; Statistics; Machine learning; Data mining; Performance assessment

Contents

* Corresponding author.
E-mail address: amarkov@hse.ru (A. Markov).
Peer review under responsibility of KeAi.

[1] The main body of this work has been carried out together by all of the authors listed herein in the period prior to 14.08.2022. However, Victor Lapshin has not worked on and/or contributed to this work in any way since 14.08.2022.

https://doi.org/10.1016/j.jfds.2022.07.002

1. Introduction

Credit risk assessment is a sensitive subject for any bank and financial institution for several reasons. Firstly, credit risk is subject to external evaluation, since central banks and auditors rigorously monitor how financial institutions comply with Basel and International Financial Reporting Standards (IFRS) requirements. Secondly, precise credit risk estimation is key to an organisation's profitability. If the bank fails to estimate a risk correctly, it either overprices loans and loses its market share, or sets interest rates too low to cover the expected losses, which leads to poor financial results. Finally, since credit risk is a vital part of the net present value (NPV) of financial instruments, it could be incorporated into various client-oriented recommendation systems and marketing campaigns, which the organisation might implement.

There is no general agreement in the literature on when credit scoring became a subject of scientific interest. Louzada[1], for example, date it back to the work of Durand,[2] whereas Dastile[3] suggest that modern credit scoring started in the 1950's with the judgmental approach commonly known as the 5C's approach (Character, Capital, Collateral, Capacity, and Condition). Linear Discriminant Analysis (LDA), which is now considered as one of the first instruments of credit risk analysis, was introduced by Fisher[4] for taxonomic classification in biology.

All points considered, scientific credit risk analysis started somewhere in the middle of the 20th century and slowly developed until the 1990–2000's, when Basel Accords and the world financial crisis attracted the researchers' attention to the problem. The rapid development of computational capacities has also added fuel to the fire: as a result, during the last three decades, the number of publications on the topic has been exponentially growing.[1] This coincides with the general exponential growth of science outlined by.[5]

During the last five years, credit scoring methodology has been affected by multiple factors. Firstly, the introduction of the International Financial Reporting Standard 9 Financial Instruments in 2018 changed the way credit scoring models are assessed from the accounting perspective. Second, the prudential regulation also changed over time. One might provide the examples of the Asset Quality Review conducted by the European Central Bank in 2018 and the widely discussed Basel IV updated requirements. Finally, COVID-19 had a dramatic impact on all spheres of scientific research, including credit scoring, and human life in general. Due to all these changes, we expect the demand for a review of the recent developments in credit scoring.

If we consider previous works that review credit scoring literature,[6] (published in 1997) may serve as a good starting point. The authors provide a remarkable classification of the approaches to credit scoring and discuss issues that may arise in practice. For example, the authors mention the reject inference problem and briefly cover other problems, e.g. missing values in the data. Despite being published in 1997, the work continues to be relevant today. However, with the expansion of the credit market, new more sophisticated models and approaches were proposed.[7] 25 years ago the authors concluded that *"there is normally little to choose between the results of sensitive and sophisticated use of any of the methods"*. Today's researchers are more likely to declare that more complicated approaches achieve better performance, as demonstrated by the results of our systematic survey.

Many researchers conducted credit scoring surveys, concentrating only on specific subsegments of the field. For example,[8] provided a thorough review of soft computing methods, i.e. neural networks, support vector machines, and evolutionary computation, while[9] focused on 'multiple criteria optimization-based data mining methods'[10] conducted one of the first systematic surveys in credit scoring, analysing 140 papers published between 2000 and 2013. The authors provide remarkable insights into the common approaches to training and performance assessment of credit scoring models. The paper places emphasis on the validation techniques and performance metrics employed in the literature, laying a foundation for the following systematic surveys in the field[1] significantly enriched the systematic survey methodology presented in García.[10] In addition to analysing the validation and performance evaluation

techniques,[1] review what model architectures were utilised, if the authors employ one or more datasets, if the datasets are public or private, etc. Besides, the paper draws our attention to data preparation stages and indicates how papers used variable selection and missing values imputation methods.

In,[1] the researchers conduct a simulation study to conclude what are the most and least effective credit scoring architectures. The objective of this paper is to answer the same question based on the systematic review of the most recent studies in the field. By design, this will better reflect various perspectives and the consensus existing in the literature. Besides, we introduced other changes to the methodology to match the latest trends. For example, we divided a more general 'ensemble' category into four subgroups: bagging, boosting, stacking, and other ensemble techniques, following the work of Dastile.[3] Finally, we devoted more attention to the datasets employed in the literature; we provide a full list of data sources with links in Appendix B.

All points considered, in this paper, we aim to cover the latest trends in credit scoring which one might find in the literature. We set up our research on the methodology proposed by Louzada et al,[1] who provided an extensive systematic review of papers published between 1991 and 2015. We seek to enrich the results of this and other surveys of credit risk literature with a comparable review of articles published between 2016 and 2021 (the latest available period). We provide a summary of the modern trends and best practices, highlighting those that emerged only in the last years. Our paper adopts a more methodological and procedural perspective on credit scoring process. Thus, we do not provide detailed theoretical descriptions of the techniques per se, relying on a recent noteworthy survey[3] and other works. Our work might serve as a guideline for developers of credit scoring procedures at financial institutions and researchers in this sphere.

This systematic review is based on the 110 (150 before filtration) most relevant articles obtained from Science Direct. The analysis of the articles led us to several recommendations we might offer to future researchers and analysts regarding.

This paper is structured as follows: Section 2 presents the approach we adopt to literature analysis. Our findings are presented in Section 3, and credit scoring best practices are covered in Section 4. Final comments in Section 5 draw conclusions and provide the potential for future research.

2. Survey methodology

2.1. Survey algorithm

The method of selecting articles is key to conducting a systematic article review. In this research, slightly controversial aims are pursued:

1. **Ensure comparability with previous research**. For this reason, we followed the methodology outlined by.[1] Since the authors reviewed articles published between 1992 and 2015, we adopted a similar questionnaire and will, later, present our findings in relation to the results of.[1]
2. **Illustrate the changes** that have occurred over the past five years and **provide additional insights**, which might not always be comparable with.[1]

We would like to note that the exponential growth of papers has continued over the last five years: 326 articles alone have been found in Science Direct[a], which is comparable with two previous decades[1]: provide a comprehensive review of 437 articles[b], published betwaeen 1992 and 2015 (articles obtained from five databases), mainly from Science Direct. As a result, this study covers only the TOP-150 relevant articles in the credit scoring methodology area, obtained from Science Direct.

The relevance is assessed automatically within the Science Direct search engine. This allows us to analyse publications that are most relevant to our keywords and leads to fewer costs associated with manual article classification. In addition, other papers usually analyse approximately 60–100 articles in comparable systematic literature reviews of finance and operational research fields.[3,9,11,12]

These several survey principles formed our approach to conducting the literature review as follows.

[a] Used keywords are 'machine learning', 'data mining', 'classification', 'statistic', 'deep learning', and 'credit'.
[b] Used keywords are 'machine learning', 'data mining', 'classification', 'credit scoring', and 'statistic' ([1], p. 119).

Step 1. Collect the TOP-150 articles from ScienceDirect published between 2016–June, 2021 (the latest month available), sorted by relevance to the following list of keywords: 'machine learning', 'data mining', 'classification', 'statistic', 'deep learning', and 'credit'.

Step 2. Filter out books, conference and thesis papers, papers in press, leaving only regular journal articles.

Step 3. Filter out articles that are not related to credit scoring and credit risk assessment.

Step 4. For each of the remaining articles, we identified their title, authors, journal name, publication date and filled in the questionnaire, similar to the one provided by Louzada et al.[1] The minor changes include further discretisation of ensemble methods that have become more widespread lately. In addition, we identified the best- and worst-performing models in our analysis. Finally, we dropped some of the less critical questions (see Appendix A for details).

2.2. Aspects under analysis

Due to the specifics of credit risk modelling, the information collected on the articles could be grouped into five main clusters:

- **Metadata.** Title, authors, journal, year of publication.
- **Dataset information.** Number of datasets, type of datasets, use of popular datasets.
- **Model architecture.** Data preparation and variable selection techniques.
- **Model testing.** Validation approach.
- **Model performance results.** Cost-criteria, models used in the literature, list of best- and worst-performing models.

For the complete list of questions and possible answers, see Appendix A.

Note. We have faced a problem of model classification due to a notable intersection between model categories. For example, the model proposed by Pławiak[13] is a genetic ensemble of Support Vector Models (SVM), k-Nearest Neighbours (kNN), and fuzzy systems, thus falling into at least five categories. Generally speaking, we have encountered an increased number of complicated models, which we would usually classify to the 'Other' group. However, less complicated models would be classified according to the 'add-in' model, e.g. a fuzzy logistic regression proposed by Sohn,[14] would fall into the 'Fuzzy' category.

3. Review findings

This section provides a statistical summary of the conducted review for the years January, 2016–June, 2021. We provide, where possible, a comparison with the results of,[1] where the authors answered similar questions for the following periods:

I. Years $\leq$2006
II. Years 2007–2010,
III. Years 2011–2012,
IV. Years 2012–2015.

Our results will be aggregated into the group:

V. January, 2016–June, 2021.

We have structured our results according to the five categories of data which we have collected.

3.1. Metadata

We observe an ongoing growth of publications on credit scoring. In addition to considerable business needs, research activities might be encouraged by other institutional changes, e.g. recent activities towards Basel IV, or the introduction of IFRS 9 Financial Instruments on January 1, 2018.

Table 1
Distribution of reviewed papers according to the journal title in January, 2016–June, 2021.

Journal	No. Articles	Proportion
Expert Systems with Applications	35	31.82%
Applied Soft Computing Journal	16	14.55%
European Journal of Operational Research	9	8.18%
Information Sciences	8	7.27%
Knowledge-Based Systems	6	5.45%
Decision Support Systems	6	5.45%
Engineering Applications of Artificial Intelligence	4	3.64%
Journal of Computational and Applied Mathematics	2	1.82%
Journal of Banking and Finance	2	1.82%
Electronic Commerce Research and Applications	2	1.82%
Others	20	18.18%
Total	**110**	**100%**

The scientific journals: Expert Systems with Applications, European Journal of Operational Research, and Knowledge-Based Systems remain in the top list of journals which publish articles on credit scoring. The Applied Soft Computing Journal and Information Sciences have demonstrated a notable increase in coverage of the topic of credit scoring. A more detailed distribution of credit scoring articles between journals is provided below in Table 1.

Regarding authors, we have identified 300 different co-authors, four of whom managed to publish more than three publications on the topic (see details in Table 2). Except for Cristián Bravo, who appeared in both lists, we could not directly compare this result with.[1]

3.2. Dataset information

What dataset should one use to prove their point? How many data sources are recommended to employ in the study? Even though the answers might differ from case to case, some general trends persist. Below, we analyse some of the recent trends in the literature.

Firstly, most research papers between 2016 and 2020 employed two or more datasets with an average of 2.75 (see Table 3). At the same time, the use of the public datasets, in comparison with private data, has been on the increase during the years (see Fig. 1, the sum of 'Public' and 'Both' categories), growing from 43% before 2006 to 65% between 2016 and 2021. One possible reason might be a higher number of datasets becoming publicly available through the years. This corresponds to the trend identified in Louzada et al.[1]

Three University of California Irvine (UCI) datasets — the Australian, German, and Japanese datasets — remain highly popular. Over the past five years, the Taiwanese and Polish datasets have also been added to the UCI datasets. These databases differ in the number of observations and features. In addition, the proportion of 'bad' borrowers varies between datasets (some datasets are less balanced in terms of default rate), which enabled some researchers to illustrate some approaches to modelling, e.g. under- and oversampling.[28-32] If a model proves to be efficient on both balanced and unbalanced data, it might be of greater use in practice.

Peer-to-Peer (P2P) lending databases — the Lending Club dataset being the most popular, have been included in several studies.[31-40] The use of P2P data appears to be one of the most recent trends, since all datasets were published over the last five years, and no similar trend was mentioned earlier.[1] In addition, P2P datasets are convenient for researchers due to their size (all datasets contain more than 100,000 observations), which enables researchers to develop more sophisticated credit scoring models.

As we stated in the introduction, we initially expected that credit scoring literature would be greatly affected by the COVID-19 pandemic. However, few articles published until June, 2021 assessed these effects on credit scoring modelling. As of June, 2022 (article acceptance date), we would highlight a research paper published by Saudi Central Bank[15] that covers the changes in the credit scores of borrowers in Saudi Arabia. Some other recent publications on credit scoring modelling[16,17] mention the changes caused by the pandemic, but the datasets employed by the authors might not yet include the COVID-19 period. We expect that more research papers on the topic will be published as soon as enough data is accumulated.

If we take a broader perspective on the topic, we might find numerous articles assessing how banking and credit risk management were affected by the pandemic,[18–21] to mention a few. A significant number of papers analysed the

Table 2
Distribution of reviewed papers according to the author/co-author.

Author	Affiliation, Country	
Zhang, Wenyu	Zhejiang University of Finance and Economics, China	7
Bravo, Cristián	Western University, Canada	6
Maldonado, Sebastián	University of Southampton, United Kingdom	5
He, Hongliang	Zhejiang University of Finance and Economics, China	4
Others	198 authors	301
Total		**323**

Table 3
Statistical summary of the number of used datasets.

Time period	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St.dev.
I	1.00	1.00	1.00	1.80	2.00	8.00	1.69
II	1.00	1.00	2.00	2.05	2.00	10.00	1.40
III	1.00	1.00	2.00	2.55	3.00	8.00	1.85
IV	1.00	1.00	1.00	2.31	3.00	10.00	2.32
V	1.00	1.00	2.00	2.75	3.00	13.00	2.56

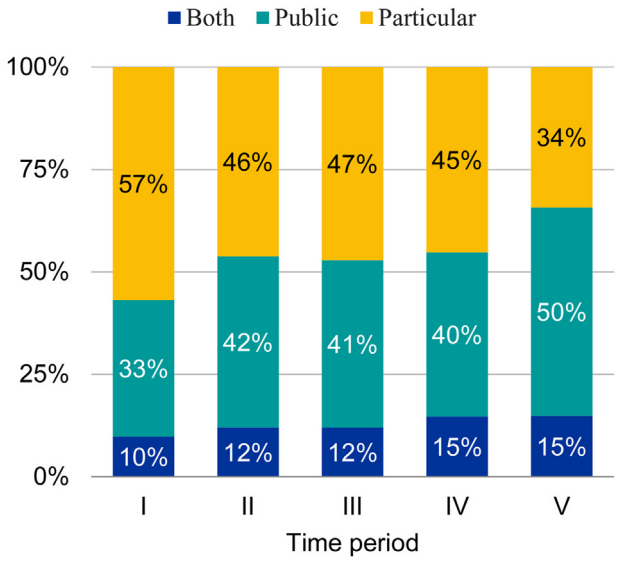

Fig. 1. Use of dataset types.

government policies aimed at mitigating the distress caused by COVID-19.[22–24] Finally, several papers discussed the impact of the pandemic on sovereign credit risks.[25–27]

See Appendix B for the complete[c] list of public datasets under our analysis.

Particular datasets have also been employed in many studies and involve different types of borrowers: shipping companies,[41] small and medium-sized enterprises,[42,43] and retail customers.[44-47] In addition, some papers illustrated their results on simulated data.[48,49]

From all information above, we might **generally recommend employing public datasets** for model testing and performance comparisons, since it follows the latest trends in the literature (see Fig. 1) and ensures the reproducibility of research. Rules are made to be broken, though: in many cases, using a particular dataset in an academic article provides additional insights otherwise unavailable. For example,[50-53] employ personal information on private borrowers to increase model performance.[54] analyse textual descriptions of borrowers using natural language processing (NLP) techniques. Such data might be potentially employed on borrowers with no credit history as a means of

[c] We mentioned the datasets that could be downloaded directly, and provided the respective links.

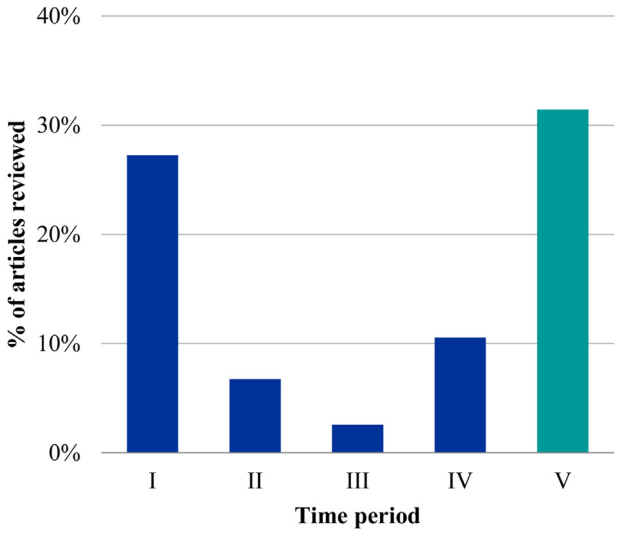

Fig. 2. Use of missing data imputation techniques.

alternative credit scoring. Similar techniques have already been adopted by various fintech companies, such as Branch International, Channel VAS, Credolab, Lendo, and Jumo.

Another good example of using private datasets might be the works of,[55,56] who analyse the data on all the Italian and Turkish companies, respectively. The data are supplied by the national regulatory institutions and provide a perspective of the whole industry rather than a loan portfolio of a single bank.

Finally,[57] generalise the credit scoring methods for the collection rate prediction. In this case, the use of a private dataset is almost inevitable due to the scarcity of non-simulated public information.

3.3. Model architecture

In general, data **preprocessing** has not changed significantly over the past four years. It usually includes the following: imputation of missing values, feature selection and transformation, and rebalancing (resampling) of datasets in use. It should be noted that all these stages are rarely used simultaneously in one study. The following describes the key changes at various stages of data preparation.

Missing values imputation. Between 2016 and 2020, many researchers have demonstrated an increasing interest in the problem of missing data (See Fig. 2). Dropping observations with missing data remains the most popular approach in the field.[39,58-66] This might lead to a number of issues, such as bias in the remaining data. It is recommended in the literature to impute the specific values for non-accessible data. However, this might be challenging, since different data and variables require different missing value imputation techniques.[67]

Possible missing value imputation techniques include:

- mean/mode imputation – for continuous/discrete variables,[30,68,69]
- incorporation of missing values into a separate category – for discrete and categorical variables,[30,63,70]
- weight of evidence (WOE) transformation,[71]
- XGBoost,[40,72] CatBoost[73] and others,
- Bayesian network iterative imputation[d] (BNII),[74]
- the division into three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) – used as a reject inference method.[75,76]

[d] The original work[73] does not provide the full name of the method, and it was obtained from the context.

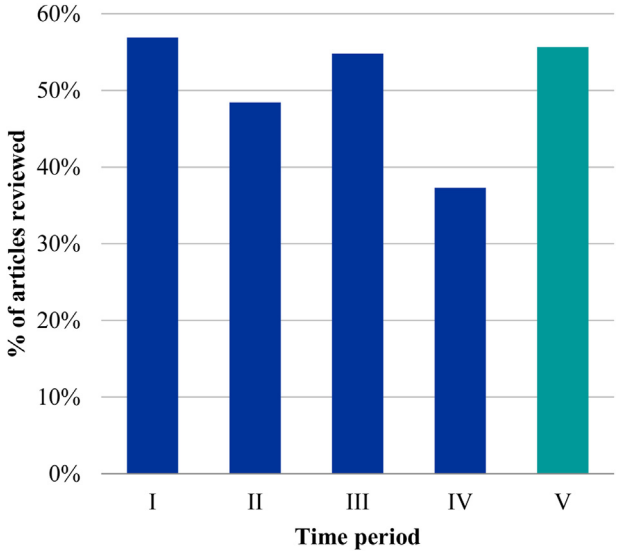

Fig. 3. Use of feature selection techniques.

While XGBoost–, CatBoost– and BNII–based approaches are complex and appear to provide better results than straightforward mean/mode imputation, these approaches deteriorate the interpretability of scoring results. On the contrary, WOE transformation and the approach employed by[75,76] allow both to enhance model performance and maintain its interpretability, which might be helpful in business applications.

Feature transformation is commonly used to prepare data for various credit scoring methods. The most popular approaches include:

- feature standardisation,[30,68,69,77]
- division of each value of the variable by a maximum value of the variable,[78]
- mapping data onto [0, 1] interval through the following transformation[e] [38,72,79]:

$$x_i^{'} = \frac{x_i - min(x_i)}{max(x_i) - min(x_i)}$$

Various authors[78,79] note that these techniques lead to more adequate results for various credit scoring methods (e.g. LR, SVM). Nevertheless, few publications cover the application of the more advanced feature transformation techniques to credit scoring.

Feature engineering and selection pursue several objectives, such as:

- the reduction of the number of explanatory variables,
- addressing the problem of multicollinearity,
- model performance improvement,
- the simplification of the final model.

Interest in the topic of feature engineering and selection has remained relatively stable over the last three decades. However, researchers appear to hold mixed opinions on this subject. Some researchers[3] provide evidence that feature engineering does not improve the predictive performance of credit scoring methods. Others[80] advocate the importance of feature engineering in credit scoring; some researchers[30,33,42,69,81-84] include feature engineering and selection

[e] This approach is designated as scaling,[37] normalisation,[78] etc. In this article we will refer to this method as 'scaling', and will assume that 'normalisation' is a synonym to 'standardisation'.

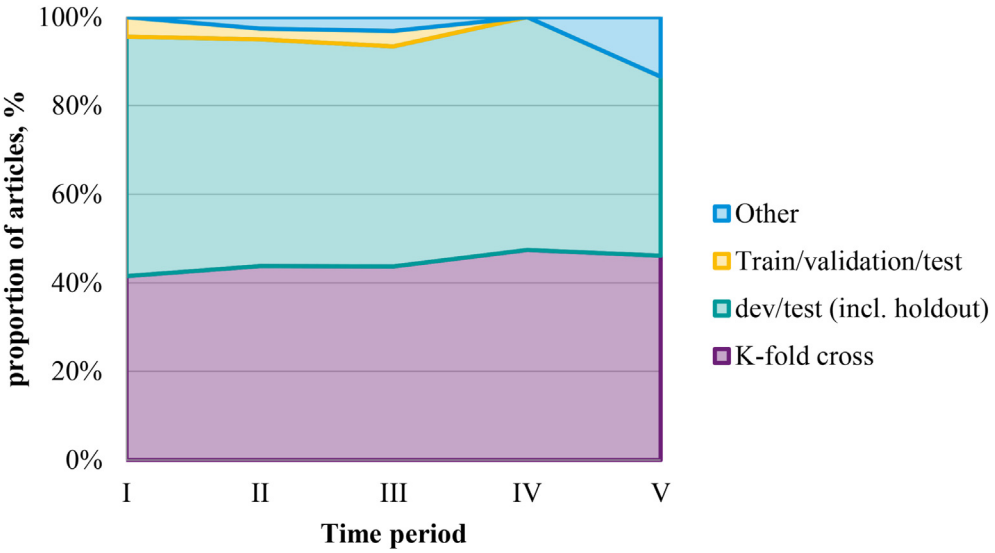

Fig. 4. Approaches to validation through time.

methods in various ensemble models. 56% of research articles in the scope of our analysis employed various feature selection techniques (See Fig. 3), and 10% of the papers were devoted solely to proposing a new method for feature selection.

We would like to note that feature engineering might be critical for the development of simplified baseline models which are otherwise unable to account for more complex data interconnections. In addition, feature engineering is almost inevitable if a non-conventional (transactional, graph-based) data source is employed.

Common practice for data preparation.[25] proposed a data preparation method that has become a common practice.[59,60,62,65] First, the authors recommend dropping variables with more than 30% of missing values, or with more than 99% of values concentrated in a single value. The second step is to conduct a univariate analysis using Kolmogorov–Smirnov (K–S) and χ^2 -tests, but this step is often skipped in the literature.

Reject inference. Even though the idea of adjusting models for the rejected loans is not new and dates back to the previous century,[86] this issue persists in capturing the researchers' attention. In the last years, authors tend to propose more complicated approaches to tackle the problem, such as the ensemble-based models proposed by.[75,76,87] In addition, the authors often resort to private datasets to account for the rejected loans.[87,88] The Lending Club public dataset also provides additional opportunities to test the reject inference methods, since it provides information on both the accepted and rejected loan applications.[89-91] All of the authors conclude that incorporating the reject inference into the model improves its performance.

Imbalanced data. Most credit scoring datasets (see Appendix B) have a more significant proportion of 'good' cases, whereas the number of defaults is low. This might potentially deteriorate the performance of scoring models, as noted by,[28,92,93] to mention a few. We would like to highlight the synthetic minority oversampling technique (SMOTE) that appears to have become one of the most renowned approaches to tackling the issue.[94] Several papers[88,95,96] have proposed improvements to the original SMOTE. However, the results of[97] suggest that randomised undersampling (RUS) is much simpler, yet might achieve similar results.

All points considered, according to the literature, all the modelling stages mentioned in this subsection result in better credit scoring model performance. In addition, some of the stages (e.g. missing values imputation and feature engineering) are vital to enhancing the quality of simple model architectures (e.g. LR, CART).

3.4. Model testing

Validation approaches have also changed over the last four years (see Fig. 4). The key trends are the following:

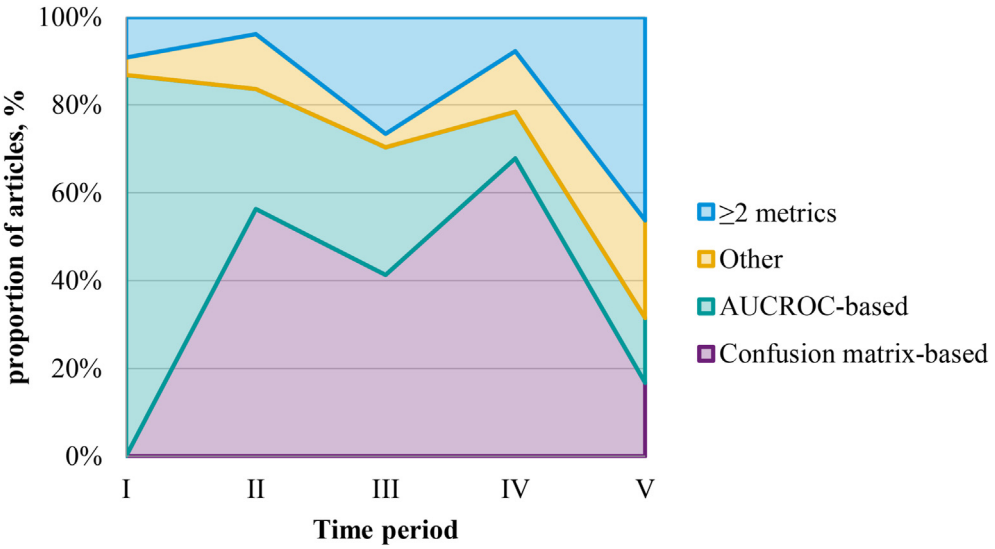

Fig. 5. Use of misclassification criteria in dynamics.

- Development (dev, or train plus validation)/test (holdout) approach and K-fold cross-validation have become the industry's best practices. However, the total proportion of these methods in use has dropped from 96.6% between 2012 and 2015 to 86.5% between 2016 and 2021 (see Fig. 4, 'dev/test' and 'K-fold' categories combined), due to the growing popularity of more 'Other' complex approaches (usually a combination of several techniques).
- Leave-one-out validation technique (technically equivalent to extreme K-fold cross-validation) has rarely been used in the literature since 2012 (periods IV and V),
- More and more articles employ more sophisticated approaches to validation.[29,63,66,69,79,98,99] These are included in the 'Other' category (see Fig. 4).

The out-of-time validation approach should also be noted. Generally speaking, it is a subcase of dev/test validation, where the holdout is not sampled out-of-sample, but out-of-time. This technique might serve as an indicator of model stability over time, since in practice, credit scoring models and their performance also depend on past data.[100] Many authors[71,95,101-103] have conducted an out-of-time validation in their work.

'Out-of-universe'[100] validation concept might also be of interest for researchers. It implies that models trained on one type of data can be adapted to and tested on different data, e.g. models trained on borrowers from one region or country could be tested on other regions and countries.

To conclude, even though all publications in the field are scrupulous at model validation, we might suggest using, where possible, the out-of-time and even the 'out-of-universe' validation techniques as industry best practices.

3.5. Model performance results

Misclassification criteria application trends have changed since 1991. During the last decade, a number of studies have pursued the objective of detecting the key advantages and disadvantages of each method.[94,104,105] In addition, several new criteria have been proposed. Below, we provide some of the key conclusions:

- Traditional **confusion matrix-based measures** are easy-to-use and easy-to-interpret measures. However, these methods depend on misclassification cost functions and thresholds (cut-offs). In addition, the confusion matrix is highly sensitive to good/bad loan imbalances in data.[106]

Due to these limitations, the proportions of articles using confusion matrix as a standalone misclassification criterion has fallen dramatically from 67.8% in period IV (2012–2015) to only 18.2% in period V (2016–June 2020) (see Fig. 5). However, these methods are often used in combination with other criteria.

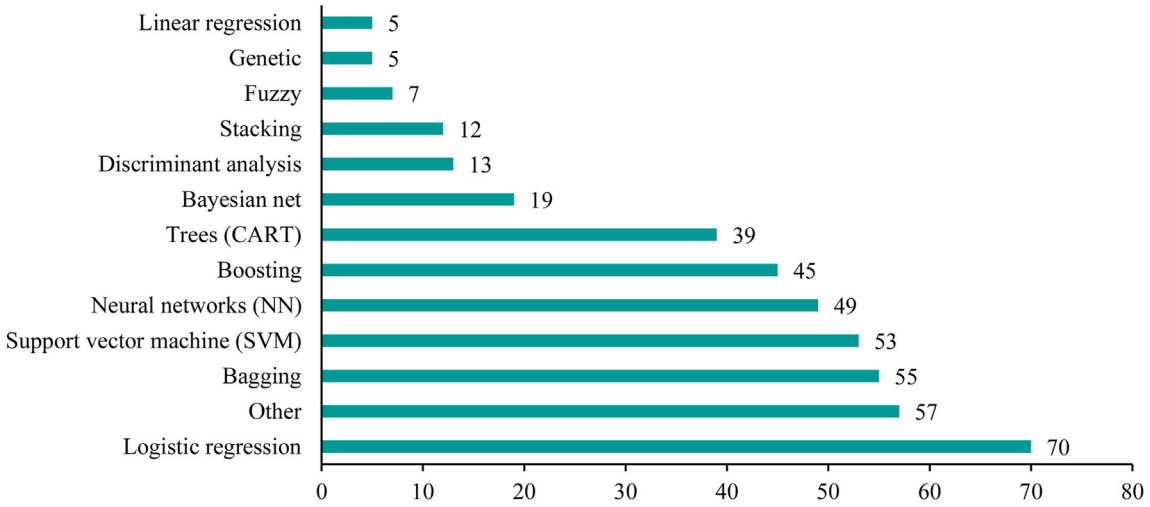

Fig. 6. Use of scoring methods, no. of papers.

- **AUCROC (area under receiver operating curve)** is another interpretable measure that does not rely on cut-off values nor data distribution between classes. Some researchers,[41,107] do not employ any other model performance criteria, but AUCROC or Gini index (which is derived from AUCROC). However,[105,108] stated that 'AUCROC uses different misclassification cost distributions for different classifiers', leading to less comparable results between models.
- **Brier score** could be utilised as a measure of calibration accuracy,[35,61,70,78] which is crucial for regression rather than classification tasks.[70]

Since all metrics reflect model performance from different viewpoints, many researchers[29,33,72,78,81] applied a combination of the criteria mentioned above to achieve a more objective assessment of the model and its robustness. As a result, the proportion of papers that use two or more metrics has dramatically increased from 7.8% in 2012–2015 (period IV) to 46.3% in 2016–2021 (period V, See Fig. 5).

Finally, several articles have been devoted to the advantages of NPV-based metrics compared to more traditional misclassification criteria. These include Maximum Profit Criterion (MPC) and Expected Maximum Profit (EMP) measures proposed in articles[109] and,[110] respectively. This approach is based on an estimate of the expected profit from the borrower, depending on the probability of default and type I and II errors. MPC and EMP have also been used in many papers.[36,60,62,98,102] Similar approaches have been used in several articles.[37,65]

Although this approach appears to be promising for business and decision-making purposes, one might find it methodologically questionable to develop credit risk models to pursue greater profits. Prudential requirements usually stipulate that risk function should be independent of the operations of the financial institution. From this point of view, it might be more appropriate, for example, to apply the H-measure proposed by.[105] This criterion was applied in the cited articles.[68,78,81,102]

To sum up, according to the latest trends in the literature, it is vital to assess model performance in several dimensions: the confusion matrix-based perspective, the discrimination of good and bad borrowers, the model calibration, and even the potential monetary effects of the model.

4. Best practices of credit scoring

The literature does not provide a single answer to what method is the most efficient means of credit scoring: various authors appear to hold different, if not opposite, opinions on the issue. Therefore, to resolve this dilemma, we would like to provide a statistical overview of model architectures that different papers claim to be the best/worst for credit scoring.

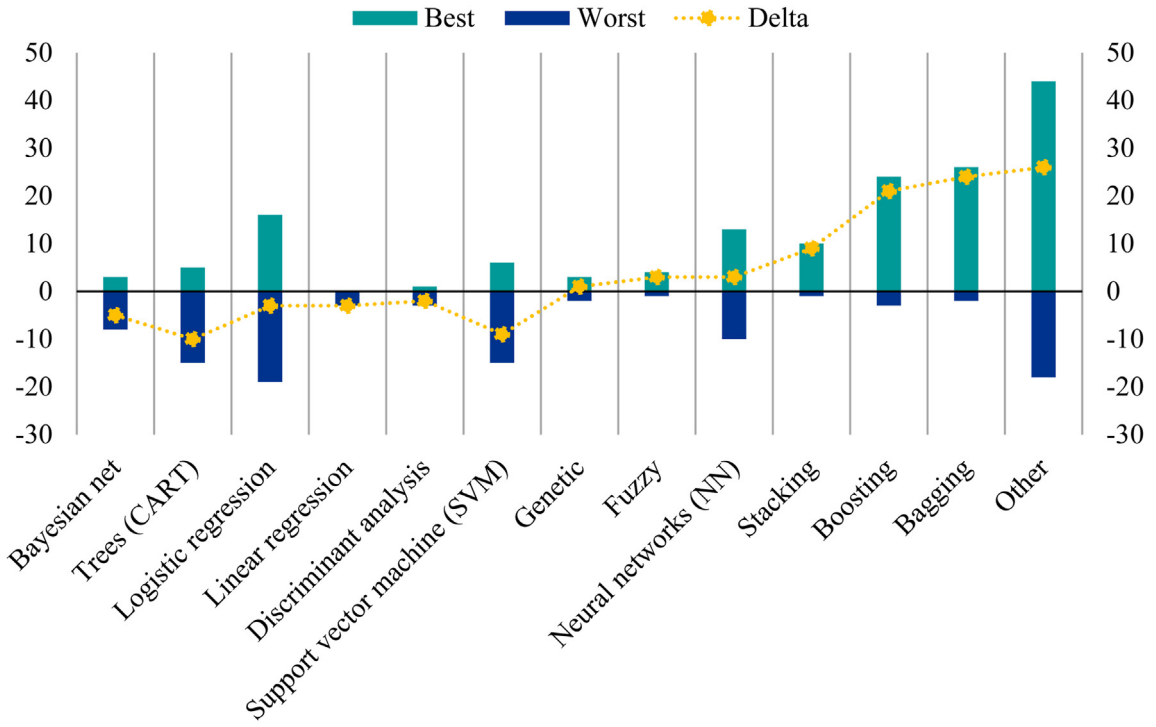

Fig. 7. Best and worst scoring methods, according to the literature No. of papers that conclude that a particular method is the best/worst in terms of credit scoring given analysed data. Left axis for delta, the right one is for the best and the worst.

Popularity. As one can see from Fig. 6, both traditional (logistic regression and decision tree) and more advanced (support vector machines, ensembles, i.e. bagging and boosting, and neural networks) techniques appear to be the most widely used credit scoring approaches.

While a decision tree (CART) and a logistic regression (LR) usually serve as baseline models, more advanced techniques (ensembles and neural networks) often become alternative 'challenging' models and usually provide better results (see Fig. 7). Interestingly enough, in various papers, SVM is used both as a baseline and an alternative model.

Efficiency. In terms of efficiency (see Fig. 7), ensemble (i.e. bagging, boosting, and stacking) and 'Other' methods are concluded to be most efficient. This result coincides with the study of Dastile et al.[3]

Even though LR, CART, and SVM appear to be worse in quality than some other methods (e.g. Genetic), they are much more common: most papers present average (not best or worst) results in terms of model performance. For example, LR was used in 70 papers, out of which it was regarded as 'best' in 16 papers and 'worst' — in 19 papers (see Fig. 7). This implies that the rest 35 papers concluded that LR demonstrates an average performance. In addition, baseline models are usually interpretable, less complex, and prone to survivorship bias: if an alternative model does not outperform its baseline, the paper is not published, and further research is conducted.

'Other' category. As mentioned above, many papers employ one or more methods from the 'other' category. Moreover, the 'other' category has the highest 'good' to 'bad' ratio.

This category generally includes three types of models:

1. Machine learning techniques:
   - k-nearest neighbours (kNN),[63,95,111-115]
   - specific types of SVM,[112] etc.
2. Highly complicated hybrid models that might be simultaneously classified into several groups (see note to Section 2.2):
   - deep genetic hierarchical network of learners (DGNHL),[13]
   - CPLE-LightGBM,[76]
   - Grabit,[116] etc.

3. Other statistical models:
   - Linear mixed model,[117]
   - Generalised additive model (GAM),[71,99]
   - Generalised extreme value (GEV) regression model,[71]
   - Tobit,[116]
   - Entropy Difference Approach (EDA),[118]
   - Multivariate adaptive regression splines (MARS),[81] etc.

The high efficiency of 'Other' techniques might be attributed to the fact that this category includes many case-specific and highly complicated models that are expected to provide better results given specific data. Examples of such models are provided in.[13,28,119,120,33,60,62,63,77-79,118]

We must note that the high efficiency might result from overfitting. For example,[13] propose a 29-layer ensemble of SVM, kNN and other kernels, trained and tested on 1000 observations. In addition, higher efficiency might be attributed to survivorship bias: less popular 'other' methods are mentioned in the datasets only when they provide better results than baseline models.

Finally, not all 'Other' techniques are highly efficient. Some of them, like kNN,[30,35,76,95,111] BOW,[77] BGEVA,[71] SVM variations,[75,76] Tobit model[116] and others appear to have the lowest values of performance measure.

To sum up, LR, CART, and SVM are usually viewed as standard baseline techniques that provide satisfactory results. Ensembles often demonstrate the best performance. The 'Other' category is quite heterogeneous: it unites seemingly less efficient traditional machine learning methods and new, highly complex approaches. Due to the efficiency of the new techniques, the overall performance of the 'Other' category is higher in comparison with all other groups. However, this might be the result of the survivorship bias.

5. Conclusions and discussion

This paper provides a systematic review of articles related to credit scoring. We examined the 150 most relevant articles, published between January, 2016–June, 2021, 110 of which met the conditions set in our study (see Section 2).

Articles, in the scope of our analysis, illustrated how various statistical and machine learning techniques could be employed at different stages of credit scoring. In addition, we reviewed papers that propose new approaches to feature selection, model testing and quality assessment.

Global IT transformation and machine learning championships have dramatically impacted the datasets employed in the literature. As a result, the volume of public data available to researchers has significantly increased. For example, Kaggle and other platforms provide the means for organising competitions in credit scoring and sharing statistical information; P2P lending platforms (Lending Club, We.com, among the best-known) show high potential for parsing large volumes of real-life data, and many researchers have taken advantage of all these opportunities in their analyses[36,39,90]). Another large dataset, provided by Fannie Mae, allows testing credit scoring techniques with application to mortgage data.

All this leads to better reproducibility and a higher overall quality of models proposed in the literature. Hence, we recommend employing several public datasets for testing credit scoring models, if possible. Appendix B provides a complete list of public datasets in the scope of our analysis, with links to data sources (where possible), for the reader's convenience (see Section 3.2).

Section 3.3 has identified an ongoing interest in data preparation, model architecture, and testing. New missing value imputation, feature engineering and selection techniques are being developed, and several standard practices are already in place.[85]

Model testing is also in the spotlight (Sections 3.4, 4). To demonstrate the quality of the proposed model, researchers continuously enrich their methodologies for model testing. In many cases, a single paper utilises several performance metrics and a combination of validation techniques. Moreover, several approaches[36,37,60,62,65,98,102] (MPC, EMP and others) have been developed to prove the financial viability of proposed models, making them more attractive to practitioners.

The analysis of the proposed models (Section 4) has shown that LR, SVM, and CART are usually regarded as baseline models and generally provide acceptable results. On the contrary, the ensembles and other more complex models (usually hard to classify and, therefore, regarded as 'Other') appear to demonstrate the best results compared to other model types. This, however, comes at a price of complexity and model risk. Besides, as we note above, complex

models might also be subject to survivorship bias: they are developed much more vigilantly than baseline models and, therefore, are bound for success.

All points considered, we might offer the following points to consider while preparing a credit scoring model:

1. It might be reasonable to employ several public datasets to test the proposed model. We have identified that an increasing number of authors resort to public datasets and utilise two or more data sources in their studies (see Section 3.2). Therefore, we suggest using at least three datasets to stay above the average and follow the trend towards the greater modelling vigilance. Appendix B provides a complete list of public datasets, mentioned in the articles in the scope of our analysis, with links to data sources (where possible) for the reader's convenience.

However, as we noted in Section 3.2, there are many cases when using a private dataset is justified; for example, when the authors seek to demonstrate how sensitive personal information and other private non-conventional data types could improve the predictions.

2. If possible, consider the potential impact of the rejected loans on the modelling results. All authors in the scope of our analysis conclude that incorporating the reject inference into the model improves its performance. Since the data on the rejected applications is not always accessible for researchers, we would like to highlight that the Lending Club public dataset provides information on both the accepted and rejected loan applications.
3. There is a persistent issue of the imbalance between the good and bad loans in credit scoring, so it might be reasonable to adjust the credit scoring model accordingly. Even though the literature provides various methods to account for data imbalances, we would like to mention the synthetic minority oversampling technique (SMOTE), as it appears to have become one of the most renowned approaches to tackling the issue.[94]
4. Data preprocessing, missing values imputation, and feature selection techniques are becoming more widespread in the literature and are recommended for use, especially to achieve higher performance of simple model architectures (See Section 3.3). Some approaches, e.g. the one outlined by,[85] are becoming standard practice.
5. It is recommended to adopt a more vigilant approach to model testing: use several misclassification criteria and thoughtful validation (see Section 3.4) to undertake more vigilant testing of the proposed model. In addition, the use of out-of-time testing samples and measuring model's financial effects and error costs[36,102] might be especially beneficial for business.
6. Logistic regressions, support vector machines (SVM), and classifier and regression trees (CART) have become the most widely used baseline models that often provide average (not best or worst) results in comparison with other developed models (especially if we consider the survivorship bias, See Section 4). Neural networks are also gaining popularity as a baseline model for other more complex credit scoring techniques.
7. Ensemble (i.e. bagging, boosting, and stacking) models are often presented as a more efficient alternative to baseline models (See Sections 3.5, 4), especially if we exclude the ‘Other’ model category due to our concerns on the overfitting and survivorship bias.

We would like to note that COVID-19 appears to have little influence on the papers published until June, 2021. One explanation might be that the crisis adjustments had an ad-hoc nature and did not affect the fundamental principles of credit scoring modelling. Another reason could be the limited time available to researchers: as of June, 2021, the researchers had approximately a year to collect the data, conduct research and publish the results. These statements hold as of June, 2022 (article acceptance date): as noted in Section 3.1, the COVID-related literature is now more focused on a broader topic of portfolio changes and government policies, rather than a direct impact of COVID-19 on credit scoring modelling.

Finally, our systematic literature review has certain limitations. Firstly, it is based on the TOP-150 relevant articles, obtained from the Science Direct database, published in English. To expand our results, researchers can resort to the Wiley Online Library, Emerald, AEA Journals and other databases, literature review articles in other languages, or analyse other time periods. Secondly, a larger sample of articles might result in the further discretisation of employed methods. Thirdly, given a larger sample of papers, the survivorship bias of alternative models and other hypotheses might be tested. However, all this requires considerable time and effort.

Future research on the topic might compare our results and the results of[1] with the articles to be published in the years to come. The potential highlights could be the impact of COVID-19 and Basel IV on credit scoring, the recent trends in

model performance assessment, model architecture, validation, and other modelling aspects. Covering new publicly available datasets would also be of use for a potential reader.

Declaration of competing interest:

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The study was financed by the grant of the Russian Science Foundation No. 22-28-01255. http://rscf.ru/en/project/22-28-01255/

Acknowledgments

The authors are grateful to *Marat Kurbangaleyev, National Research University Higher School of Economics,* for initial paper review and valuable methodological comments. The authors would like to acknowledge the valuable contributions of the Academic Writing Centre of the Higher School of Economics and, in particular, of Pete Dick and Konstantin Sheiko. Their assistance helped us to significantly improve the wording and flow of our manuscript.

Appendix A. List of questions and of possible responses to the review

The list of questions used in the article is based on a similar list in[1] for the means of comparability. However, in several cases the possible answers have been changed (see Q2, Q7, Q10; we added 'None' category, divided ensemble models into 'bagging', 'boosting' and 'stacking' categories for better differentiation), some questions have been added (see Q11, Q12) to pursue the objectives we set in our work, and some questions have been dropped if they were not directly linked to our objectives (e.g. question 'Which is the principal focus of the paper concerning the decision area? Operational research or Management science?').

1. **Which is the main objective of the paper?**
   a. Proposing a new method for rating
   b. Comparing traditional techniques
   c. Conceptual discussion
   d. Feature selection
   e. Literature review
   f. Performance measures
   g. Other issues
2. **Which is the type of the datasets used?**
   a. Public
   b. Particular
   c. Both
   d. None
3. **Does the paper perform variable selection methods?**
   a. Yes
   b. No
4. **Was missing values imputation performed?**
   a. Yes
   b. No
5. **What is the number of datasets used in the paper?**
6. **Was exhaustive simulation study performed?**
   a. Yes
   b. No
7. **What is the type of validation of the approach?**
   a. K-fold cross
   b. Handout
   c. Train/validation/test

d. Leave one out
e. Other
f. None

8. What is the type of misclassification cost criterion?
a. ROC-curve based
b. Metrics based on confusion metrics
c. Error-based
d. Others

9. Does the paper use the Australian or the German datasets?
a. Yes
b. No

10. What classification methods are tested?
a. Neural networks (NN)
b. Support vector machine (SVM)
c. Linear regression
d. Trees (CART)
e. Logistic regression
f. Fuzzy
g. Genetic
h. Discriminant analysis
i. Bayesian net
j. Boosting
k. Bagging (incl. random forest)
l. Other

11. What classification methods provide the best results, according to the authors?
a. Neural networks (NN)
b. Support vector machine (SVM)
c. Linear regression
d. Trees (CART)
e. Logistic regression
f. Fuzzy
g. Genetic
h. Discriminant analysis
i. Bayesian net
j. Boosting
k. Bagging (incl. random forest)
l. Other

12. What classification methods provide the worst results, according to the authors?
a. Neural networks (NN)
b. Support vector machine (SVM)
c. Linear regression
d. Trees (CART)
e. Logistic regression
f. Fuzzy
g. Genetic
h. Discriminant analysis
i. Bayesian net
j. Boosting
k. Bagging (incl. random forest)
l. Other

Appendix B. Description of public datasets popular in credit scoring literature

Dataset	Papers	No. of obs.	% of defaults	No. of factors
German, UCI[f]	1,3,69,72,74,78,79,81,95-98,28,112-115,118, 121-125,29,126-135,30,136,137,33,39,63,64,68	1000	30.0%	24
Australian, UCI[g]	1,3,78,79,81,82,95,97,98,112-114,30,115,118,121-124, 126,128-130,33,132-135,138,63,64,68,69,72,74	690	55.5%	14
US Lending Club, Kaggle[h]	29,31,66,68,72,75,76,83,89,90,92,125,32,126, 136,139-14133,35-40	29,909,442	3.71% (2018)	150
Japanese, UCI[i]	1,28,126,128,130,133,135,30,68,69,78,79,81,112,113	652	45.40%	15
Taiwanese, UCI[j]	29,68,129,132,136,69,72,75,96,115,120,126,128	30,000	22.12%	23
GMSC, Kaggle[k]	29,64,136,69,79,97,98,130,132,134,135	150,000	6.68%	10
PAKDD[l]	64,69,97,98,135,136	50,000	26.08%	373
UCSD, 2007 Data Mining Contest[m]	81,113	2435	24.60%	38
Th02[n]	69,97,114,132,136	1225	26.37%	14
Polish, UCI[o]	78,113,129,142	43,405	4.82%	64
AER, Kaggle[p]	69	1319	22.44%	11
Prosper, Kaggle[q]	32	28,399	30.35%	48
BLSD, Kaggle[r]	39	82,000	23.00%	19
Czech, PKDD′99 Discovery Challenge[s]	111	682	6.60%	69
hmeq[t]	98,132,136	5960	19.95%	12
Qualitative, UCI[u]	138	250	57.20%	7
ANALCAT, Brigham Young University[v]	138	50	50.00%	6
Creator, Creator Information Technology Co.	128,129	35,960	38.93%	61
FMPD, Federal National Mortgage Association[w]	143	49,307,309	depends on the definition of default	108
USmort	142	622,489	2.43%	19
CSDS-1[x]	144	315,539	12.97%	178
CSDS-2[22]		50,401	2.06%	37
CSDS-3[22]		97,226	26.07%	152
We.com, parsed by authors	33,72,76,126	Since datasets are obtained by parsing, underlying data continuously change		
Renrendai, parsed by authors	119,145			
PaiPaiDai, parsed by authors	29,32,68,79,130			

Note 1. A number of datasets – Benelux 1, 2,[123] UK,[64] Iranian,[78,113] Brazilian, Indonesian, European Credit Bureau, etc. – are mentioned in a number of papers, yet appear to be accessible by means of contacting the original authors, as noted by.[146] We did not include these papers in the list of public sources.

Note 2. Some datasets, e.g. Lending club dataset, have been updated in the past, and various authors might have utilised different versions of input data.

[f] German Credit Data. URL: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data).

[g] Australian Credit Approval. URL: http://archive.ics.uci.edu/ml/datasets/statlog+(australian+credit+approval).

[h] All Lending Club loan data (2007 – 2Q2018). URL: https://www.kaggle.com/wordsforthewise/lending-club.

[i] Japanese Credit Screening Data Set. URL: https://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening.

[j] Default of credit card clients Data Set. URL: https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.

[k] Give Me Some Credit. URL: http://www.kaggle.com/c/GiveMeSomeCredit.

[l] PAKDD 2009 Data Mining Competition. URL: https://github.com/JLZml/Credit-Scoring-Data-Sets.

[m] UCSD - Originally accessed at http://mill.ucsd.edu/, however the original dataset is no longer publicly available for download. A copy of the dataset, for example, is available from the Kennedy Kenneth at. kennedykenneth@gmail.com

[n] Thomas L, Crook J, Edelman D. Credit Scoring and Its Applications, Second Edition. URL: https://github.com/JLZml/Credit-Scoring-Data-Sets/tree/master/5. thomas.

[o] Polish companies bankruptcy data Data Set. URL: http://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data.

[p] Credit Card Data from 'Econometric Analysis'. URL: https://www.kaggle.com/dansbecker/aer-credit-card-data.

[q] Prosper Loan Dataset. URL: https://www.kaggle.com/yousuf28/prosper-loan.

[r] Bank Loan Status Dataset. URL: https://www.kaggle.com/zaurbegiev/my-dataset.

[s] 1999 Czech Financial Dataset - Real Anonymized Transactions. PKDD′99 Discovery Challenge. URL: https://data.world/lpetrocelli/czech-financial-dataset-real-anonymized-transactions.

[t] Baesens B, Roesch D, Scheule H. Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS. URL: https://github.com/JLZml/Credit-Scoring-Data-Sets/tree/master/6. Credit risk analysis/hmeq

[u] Qualitative_Bankruptcy Data Set. URL: https://archive.ics.uci.edu/ml/datasets/Qualitative_Bankruptcy.

[v] Simonoff, J S. *Analyzing Categorical Data*. Springer, 2003. URL: http://axon.cs.byu.edu/data/statlib/nominal/analcatdata_bankruptcy.arff.

[w] Fannie Mae's Single-Family Loan Performance Data. URL: http://www.fanniemae.com/portal/funding-the-market/data/loan-performance-data.html.

[x] Anonymised dataset shared by authors. URL: http://www.ppgia.pucpr.br/jean.barddal/datasets/CSDS.zip.

References

1. Louzada F, Ara A, Fernandes GB. Classification methods applied to credit scoring: systematic review and overall comparison. *Surv Oper Res Manag Sci*. 2016;21(2):117–134. https://doi.org/10.1016/j.sorms.2016.10.001.
2. Durand D. *Risk Elements in Consumer Installment Financing*. New York: National Bureau of Economic Research; 1941.
3. Dastile X, Celik T, Potsane M. Statistical and machine learning models in credit scoring: a systematic literature survey. *Appl Soft Comput J*. 2020;91, 106263. https://doi.org/10.1016/j.asoc.2020.106263.
4. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen*. 1936;7(2):179–188.
5. Price DJDS. *Little Science, Big Science*. Columbia University Press; 1963. https://doi.org/10.7312/pric91844.
6. Hand DJ, Henley WE. Statistical classification methods in consumer credit scoring: a review. *J R Stat Soc Ser A*. 1997;160(3):523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x.
7. Xu X, Zhou C, Wang Z. Credit scoring algorithm based on link analysis ranking with support vector machine. *Expert Syst Appl*. 2009;36(2):2625–2632. https://doi.org/10.1016/j.eswa.2008.01.024.
8. Lahsasna A, Ainon RN, Wah TY. Credit scoring models using soft computing methods: a survey. *Int Arab J Inf Technol*. 2010;7(2):115–123.
9. Shi Y. Multiple criteria optimization-based data mining methods and applications: a systematic survey. *Knowl Inf Syst*. 2010;24(3):369–391. https://doi.org/10.1007/s10115-009-0268-1.
10. García V, Marqués AI, Sánchez JS. An insight into the experimental design for credit risk and corporate bankruptcy prediction systems. *J Intell Inf Syst*. 2015;44(1):159–189. https://doi.org/10.1007/s10844-014-0333-4.
11. Orús R, Mugel S, Lizaso E. Quantum computing for finance: overview and prospects. *Rev Phys*. 2019;4, 100028. https://doi.org/10.1016/j.revip.2019.100028.
12. Hand DJ, Henley WE. Statistical classification methods in consumer credit scoring: a review. *J R Stat Soc Ser A (Statistics Soc.* 1997;160(3):523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x.
13. Pławiak P, Abdar M, Pławiak J, Makarenkov V, Acharya UR. DGHNL: a new deep genetic hierarchical network of learners for prediction of credit scoring. *Inf Sci*. 2020;516:401–418. https://doi.org/10.1016/j.ins.2019.12.045.
14. Sohn SY, Kim DH, Yoon JH. Technology credit scoring model with fuzzy logistic regression. *Appl Soft Comput J*. 2016;43:150–158. https://doi.org/10.1016/j.asoc.2016.02.025.
15. Bouaguel W, Alsulimani T, Alarfaj O. *The impact of the COVID-19 pandemic on the saudi credit industry: an empirical analysis using machine learning techniques to focus on the factors affecting consumer credit scoring*. 2022. Published online.
16. Yang F, Qiao Y, Huang C, Wang S, Wang X. An Automatic Credit Scoring Strategy (ACSS) using memetic evolutionary algorithm and neural architecture search. *Appl Soft Comput*. 2021;113, 107871. https://doi.org/10.1016/j.asoc.2021.107871.
17. Imteaj A, Amini MH. Leveraging asynchronous federated learning to predict customers financial distress. *Intell Syst with Appl*. 2022;14. https://doi.org/10.1016/j.iswa.2022.200064.
18. Gopalakrishnan B, Jacob J, Mohapatra S. COVID-19 pandemic and debt financing by firms: unravelling the channels. *Econ Model*. 2022;114. https://doi.org/10.1016/j.econmod.2022.105929.
19. Yin J, Han B, Wong HY. COVID-19 and credit risk: a long memory perspective. *Insur Math Econ*. 2022;104:15–34. https://doi.org/10.1016/j.insmatheco.2022.01.008.
20. Zhang D, Sogn-Grundvåg G. Credit constraints and the severity of COVID-19 impact: empirical evidence from enterprise surveys. *Econ Anal Policy*. 2022;74:337–349. https://doi.org/10.1016/j.eap.2022.03.005.
21. Ho ATY, Morin L, Paarsch HJ, Huynh KP. A flexible framework for intervention analysis applied to credit-card usage during the coronavirus pandemic. *Int J Forecast*. 2022;38(3):1129–1157. https://doi.org/10.1016/j.ijforecast.2021.12.012.
22. Norden L, Mesquita D, Wang W. COVID-19, policy interventions and credit: the Brazilian experience. *J Financ Intermediation*. 2021;48. https://doi.org/10.1016/j.jfi.2021.100933.
23. Cao Y, Chou JY. Bank resilience over the COVID-19 crisis: the role of regulatory capital. *Financ Res Lett*. 2022;48. https://doi.org/10.1016/j.frl.2022.102891.
24. Chen J, Gong RK, Cheng Z, Li J. Riding out the COVID-19 storm: how government policies affect SMEs in China. *SSRN Electron J*. 2022;75. https://doi.org/10.2139/ssrn.4000321.
25. Naifar N, Shahzad SJH. Tail event-based sovereign credit risk transmission network during COVID-19 pandemic. *Financ Res Lett*. 2022;45. https://doi.org/10.1016/j.frl.2021.102182.
26. Tran Y, Vu H, Klusak P, Kraemer M, Hoang T. Sovereign credit ratings during the COVID-19 pandemic. *Int Rev Financ Anal*. 2021;78. https://doi.org/10.1016/j.irfa.2021.101879.
27. Augustin P, Sokolovski V, Subrahmanyam MG, Tomio D. In sickness and in debt: the COVID-19 impact on sovereign credit risk. *J Financ Econ*. 2022;143(3):1251–1274. https://doi.org/10.1016/j.jfineco.2021.05.009.
28. Yu L, Zhou R, Tang L, Chen R. A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data. *Appl Soft Comput J*. 2018;69:192–202. https://doi.org/10.1016/j.asoc.2018.04.049.
29. Melo Junior L, Nardini FM, Renso C, Trani R, Macedo JA. A novel approach to define the local region of dynamic selection techniques in imbalanced credit scoring problems. *Expert Syst Appl*. 2020;152. https://doi.org/10.1016/j.eswa.2020.113351.

30. Zhang H, He H, Zhang W. Classifier selection and clustering with fuzzy assignment in ensemble model for credit scoring. *Neurocomputing*. 2018;316:210–221. https://doi.org/10.1016/j.neucom.2018.07.070.
31. Zanin L. Combining multiple probability predictions in the presence of class imbalance to discriminate between potential bad and good borrowers in the peer-to-peer lending market. *J Behav Exp Financ*. 2020;25. https://doi.org/10.1016/j.jbef.2020.100272.
32. Niu K, Zhang Z, Liu Y, Li R. Resampling ensemble model based on data distribution for imbalanced credit risk evaluation in P2P lending. *Inf Sci*. 2020;536:120–134. https://doi.org/10.1016/j.ins.2020.05.040.
33. Xia Y, Liu C, Da B, Xie F. A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Syst Appl*. 2018;93:182–199. https://doi.org/10.1016/j.eswa.2017.10.022.
34. Jiang C, Wang Z, Zhao H. A prediction-driven mixture cure model and its application in credit scoring. *Eur J Oper Res*. 2019;277(1):20–31. https://doi.org/10.1016/j.ejor.2019.01.072.
35. Teply P, Polena M. Best classification algorithms in peer-to-peer lending. *N Am J Econ Finance*. 2020;51. https://doi.org/10.1016/j.najef.2019.01.001.
36. Bastani K, Asgari E, Namavari H. Wide and deep learning for peer-to-peer lending. *Expert Syst Appl*. 2019;134:209–224. https://doi.org/10.1016/j.eswa.2019.05.042.
37. Serrano-Cinca C, Gutiérrez-Nieto B. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. *Decis Support Syst*. 2016;89:113–122. https://doi.org/10.1016/j.dss.2016.06.014.
38. Cai S, Zhang J. Exploration of credit risk of P2P platform based on data mining technology. *J Comput Appl Math*. 2020;372. https://doi.org/10.1016/j.cam.2020.112718.
39. Arora N, Kaur PD. A Bolasso based consistent feature selection enabled random forest classification algorithm: an application to credit risk assessment. *Appl Soft Comput J*. 2020;86. https://doi.org/10.1016/j.asoc.2019.105936.
40. Ma X, Sha J, Wang D, Yu Y, Yang Q, Niu X. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. *Electron Commer Res Appl*. 2018;31:24–39. https://doi.org/10.1016/j.elerap.2018.08.002.
41. Kavussanos MG, Tsouknidis DA. Default risk drivers in shipping bank loans. *Transport Res Part E Logist Transp Rev*. 2016;94:71–94. https://doi.org/10.1016/j.tre.2016.07.008.
42. Li K, Niskanen J, Kolehmainen M, Niskanen M. Financial innovation: credit default hybrid model for SME lending. *Expert Syst Appl*. 2016;61:343–355. https://doi.org/10.1016/j.eswa.2016.05.029.
43. Guégan D, Hassani B. Regulatory learning: how to supervise machine learning models? An application to credit scoring. *J Financ Data Sci*. 2018;4(3):157–171. https://doi.org/10.1016/j.jfds.2018.04.001.
44. Fonseca DP, Wanke PF, Correa HL. A two-stage fuzzy neural approach for credit risk assessment in a Brazilian credit card company. *Appl Soft Comput J*. 2020;92. https://doi.org/10.1016/j.asoc.2020.106329.
45. Kvamme H, Sellereite N, Aas K, Sjursen S. Predicting mortgage default using convolutional neural networks. *Expert Syst Appl*. 2018;102:207–217. https://doi.org/10.1016/j.eswa.2018.02.029.
46. Abdou HA, Tsafack MDD, Ntim CG, Baker RD. Predicting creditworthiness in retail banking with limited scoring data. *Knowl Base Syst*. 2016;103:89–103. https://doi.org/10.1016/j.knosys.2016.03.023.
47. Zhang T, Zhang W, Xu W, Hao H. Multiple instance learning for credit risk assessment with transaction data. *Knowl Base Syst*. 2018;161:65–77. https://doi.org/10.1016/j.knosys.2018.07.030.
48. Li Y, Wang X, Djehiche B, Hu X. Credit scoring by incorporating dynamic networked information. *Eur J Oper Res*. 2020;286(3):1103–1112. https://doi.org/10.1016/j.ejor.2020.03.078.
49. Silva DMB, Pereira GHA, Magalhães TM. A class of categorization methods for credit scoring models. *Eur J Oper Res*. 2021. https://doi.org/10.1016/j.ejor.2021.04.029. Published online.
50. Roa L, Correa-Bahnsen A, Suarez G, Cortés-Tejada F, Luque MA, Bravo C. Super-app behavioral patterns in credit risk models: financial, statistical and regulatory implications. *Expert Syst Appl*. December 2020;2021:169. https://doi.org/10.1016/j.eswa.2020.114486.
51. Zhou J, Wang C, Ren F, Chen G. Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. *Decis Support Syst*. 2021, 113611. https://doi.org/10.1016/j.dss.2021.113611. Published online.
52. Jiang J, Liao L, Lu X, Wang Z, Xiang H. Deciphering big data in consumer credit evaluation. *J Empir Finance*. 2021;62:28–45. https://doi.org/10.1016/j.jempfin.2021.01.009.
53. Djeundje VB, Crook J, Calabrese R, Hamid M. Enhancing credit scoring with alternative data. *Expert Syst Appl*. 2021;163, 113766. https://doi.org/10.1016/j.eswa.2020.113766.
54. Stevenson M, Mues C, Bravo C. The value of text for small business default prediction: a Deep Learning approach. *Eur J Oper Res*. 2021. https://doi.org/10.1016/j.ejor.2021.03.008. Published online.
55. Yıldırım M, Okay FY, Özdemir S. Big data analytics for default prediction using graph theory. *Expert Syst Appl*. 2021:176. https://doi.org/10.1016/j.eswa.2021.114840.
56. Moscatelli M, Parlapiano F, Narizzano S, Viggiano G. Corporate default forecasting with machine learning. *Expert Syst Appl*. 2020;161, 113567. https://doi.org/10.1016/j.eswa.2020.113567.
57. Nazemi A, Rezazadeh H, Fabozzi FJ, Höchstötter M. Deep Learning for Modeling the Collection Rate for Third-Party Buyers. *Int J Forecast*. 2021. https://doi.org/10.1016/j.ijforecast.2021.03.013. Published online.
58. Sousa MR, Gama J, Brandão E. A new dynamic modeling framework for credit risk assessment. *Expert Syst Appl*. 2016;45:341–351. https://doi.org/10.1016/j.eswa.2015.09.055.
59. Maldonado S, Pérez J, Bravo C. Cost-based feature selection for Support Vector Machines: an application in credit scoring. *Eur J Oper Res*. 2017;261(2):656–665. https://doi.org/10.1016/j.ejor.2017.02.037.

60. López J, Maldonado S. Profit-based credit scoring based on robust optimization and feature selection. *Inf Sci*. 2019;500:190–202. https://doi.org/10.1016/j.ins.2019.05.093.
61. de Castro Vieira JR, Barboza F, Sobreiro VA, Kimura H. Machine learning models for credit analysis improvements: predicting low-income families' default. *Appl Soft Comput J*. 2019;83. https://doi.org/10.1016/j.asoc.2019.105640.
62. Maldonado S, Bravo C, López J, Pérez J. Integrated framework for profit-based feature selection and SVM classification in credit scoring. *Decis Support Syst*. 2017;104:113–121. https://doi.org/10.1016/j.dss.2017.10.007.
63. Bao W, Lianju N, Yue K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Syst Appl*. 2019;128:301–315. https://doi.org/10.1016/j.eswa.2019.02.033.
64. Xiao J, Zhou X, Zhong Y, Xie L, Gu X, Liu D. Cost-sensitive semi-supervised selective ensemble model for customer credit scoring. *J Knowledge Based Syst*. 2020;189, 105118. https://doi.org/10.1016/j.knosys.
65. Maldonado S, Peters G, Weber R. Credit scoring using three-way decisions with probabilistic rough sets. *Inf Sci*. 2020;507:700–714. https://doi.org/10.1016/j.ins.2018.08.001.
66. Mancisidor RA, Kampffmeyer M, Aas K, Jenssen R. *Deep generative models for reject inference in credit scoring*. 196. 2020, 105758. https://doi.org/10.1016/j.knosys.
67. Molenberghs G, Fitzmaurice G, Kenward MG, Tsiatis A, Verbeke G. *Handbook of Missing Data Methodology*. CRC Press; 2014. https://doi.org/10.1201/b17622.
68. He H, Zhang W, Zhang S. A novel ensemble method for credit scoring: adaption of different imbalance ratios. *Expert Syst Appl*. 2018;98:105–117. https://doi.org/10.1016/j.eswa.2018.01.012.
69. Feng X, Xiao Z, Zhong B, Qiu J, Dong Y. Dynamic ensemble classification for credit scoring using soft probability. *Appl Soft Comput J*. 2018;65:139–151. https://doi.org/10.1016/j.asoc.2018.01.021.
70. Vanneschi L, Horn DM, Castelli M, Popovič A. An artificial intelligence system for predicting customer default in e-commerce. *Expert Syst Appl*. 2018;104:1–21. https://doi.org/10.1016/j.eswa.2018.03.025.
71. Mushava J, Murray M. An experimental comparison of classification techniques in debt recoveries scoring: evidence from South Africa's unsecured lending market. *Expert Syst Appl*. 2018;111:35–50. https://doi.org/10.1016/j.eswa.2018.02.030.
72. Xia Y, Liu C, Li YY, Liu N. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst Appl*. 2017;78:225–241. https://doi.org/10.1016/j.eswa.2017.02.017.
73. Jabeur S Ben, Gharib C, Mefteh-Wali S, Arfi W Ben. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol Forecast Soc Change*. 2021;166(October 2020), 120658. https://doi.org/10.1016/j.techfore.2021.120658.
74. Lan Q, Xu X, Ma H, Li G. Multivariable data imputation for the analysis of incomplete credit data. *Expert Syst Appl*. 2020;141. https://doi.org/10.1016/j.eswa.2019.112926.
75. Liu Y, Li X, Zhang Z. A new approach in reject inference of using ensemble learning based on global semi-supervised framework. *Future Generat Comput Syst*. 2020;109:382–391. https://doi.org/10.1016/j.future.2020.03.047.
76. Xia Y, Yang X, Zhang Y. A rejection inference technique based on contrastive pessimistic likelihood estimation for P2P lending. *Electron Commer Res Appl*. 2018;30:111–124. https://doi.org/10.1016/j.elerap.2018.05.011.
77. Zhang W, Xu W, Hao H, Zhu D. Cost-sensitive multiple-instance learning method with dynamic transactional data for personal credit scoring. *Expert Syst Appl*. 2020;157. https://doi.org/10.1016/j.eswa.2020.113489.
78. Ala'raj M, Abbod MF. Classifiers consensus system approach for credit scoring. *Knowl Base Syst*. 2016;104:89–105. https://doi.org/10.1016/j.knosys.2016.04.013.
79. Zhang W, He H, Zhang S. A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: an application in credit scoring. *Expert Syst Appl*. 2019;121:221–232. https://doi.org/10.1016/j.eswa.2018.12.020.
80. Brown I, Mues C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst Appl*. 2012;39(3):3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033.
81. Ala'raj M, Abbod MF. A new hybrid ensemble credit scoring model based on classifiers consensus system approach. *Expert Syst Appl*. 2016;64:36–55. https://doi.org/10.1016/j.eswa.2016.07.017.
82. Pławiak P, Abdar M, Rajendra Acharya U. Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring. *Appl Soft Comput J*. 2019;84. https://doi.org/10.1016/j.asoc.2019.105740.
83. Papouskova M, Hajek P. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decis Support Syst*. 2019;118:33–45. https://doi.org/10.1016/j.dss.2019.01.002.
84. Nalić J, Martinović G, Žagar D. New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers. *Adv Eng Inf*. 2020;45, 101130. https://doi.org/10.1016/j.aei.2020.101130.
85. Bravo C, Maldonado S, Weber R. Granting and managing loans for micro-entrepreneurs: new developments and practical experiences. *Eur J Oper Res*. 2013;227(2):358–366. https://doi.org/10.1016/j.ejor.2012.10.040.
86. Hand DJ, Henley WE. Can reject inference ever work? *IMA J Manag Math*. 1993;5(1):45–55. https://doi.org/10.1093/imaman/5.1.45.
87. Shen F, Zhao X, Kou G. Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. *Decis Support Syst*. 2020;137, 113366. https://doi.org/10.1016/j.dss.2020.113366.
88. Kang Y, Jia N, Cui R, Deng J. A graph-based semi-supervised reject inference framework considering imbalanced data distribution for consumer credit scoring. *Appl Soft Comput*. 2021;105, 107259. https://doi.org/10.1016/j.asoc.2021.107259.
89. Li Z, Tian Y, Li K, Zhou F, Yang W. Reject inference in credit scoring using semi-supervised support vector machines. *Expert Syst Appl*. 2017;74:105–114. https://doi.org/10.1016/j.eswa.2017.01.011.
90. Anderson B. Using Bayesian networks to perform reject inference. *Expert Syst Appl*. 2019;137:349–356. https://doi.org/10.1016/j.eswa.2019.07.011.

91. Mancisidor RA, Kampffmeyer M, Aas K, Jenssen R. *Deep generative models for reject inference in credit scoring*. 196. 2020, 105758. https://doi.org/10.1016/j.knosys.
92. Tian Y, Bian B, Tang X, Zhou J. A new non-kernel quadratic surface approach for imbalanced data classification in online credit scoring. *Inf Sci*. 2021;563:150–165. https://doi.org/10.1016/j.ins.2021.02.026.
93. Engelmann J, Lessmann S. Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning. *Expert Syst Appl*. 2021;174, 114582. https://doi.org/10.1016/j.eswa.2021.114582.
94. López V, Fernández A, García S, Palade V, Herrera F. An insight into classification with imbalanced data: empirical results and current trends on using data intrinsic characteristics. *Inf Sci*. 2013;250:113–141. https://doi.org/10.1016/j.ins.2013.07.007.
95. Shen F, Zhao X, Li Z, Li K, Meng Z. A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation. *Phys A Stat Mech its Appl*. 2019;526. https://doi.org/10.1016/j.physa.2019.121073.
96. Shen F, Zhao X, Kou G, Alsaadi FE. A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Appl Soft Comput*. 2021;98, 106852. https://doi.org/10.1016/j.asoc.2020.106852.
97. Xiao J, Wang Y, Chen J, Xie L, Huang J. Impact of resampling methods and classification models on the imbalanced credit scoring problems. *Inf Sci*. 2021;569:508–526. https://doi.org/10.1016/j.ins.2021.05.029.
98. Kozodoi N, Lessmann S, Papakonstantinou K, Gatsoulis Y, Baesens B. A multi-objective approach for profit-driven feature selection in credit scoring. *Decis Support Syst*. 2019;120:106–117. https://doi.org/10.1016/j.dss.2019.03.011.
99. Fitzpatrick T, Mues C. An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *Eur J Oper Res*. 2016;249:427–439. https://doi.org/10.1016/j.ejor.2015.09.014.
100. Baesens Bart, Rösch D, Scheule H. *Credit Risk Analytics. Measurement Techniques, Applications, and Examples in SAS*. John Wiley & Sons; 2016.
101. Butaru F, Chen Q, Clark B, Das S, Lo AW, Siddique A. Risk and risk management in the credit card industry. *J Bank Finance*. 2016;72:218–239. https://doi.org/10.1016/j.jbankfin.2016.07.015.
102. Garrido F, Verbeke W, Bravo C. A Robust profit measure for binary classification model evaluation. *Expert Syst Appl*. 2018;92:154–160. https://doi.org/10.1016/j.eswa.2017.09.045.
103. Barboza F, Kimura H, Altman E. Machine learning models and bankruptcy prediction. *Expert Syst Appl*. 2017;83:405–417. https://doi.org/10.1016/j.eswa.2017.04.006.
104. Aman S, Simmhan Y, Prasanna VK. Holistic measures for evaluating prediction models in smart grids. *IEEE Trans Knowl Data Eng*. 2015;27(2):475–486. https://doi.org/10.1109/TKDE.2014.2327022.
105. Hand DJ. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach Learn*. 2009;77(1):103–123. https://doi.org/10.1007/s10994-009-5119-5.
106. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. *Learning from Class-Imbalanced Data: Review of Methods and Applications*. 73. Elsevier; 2017:220–239. https://doi.org/10.1016/j.eswa.2016.12.035.
107. Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B. The value of big data for credit scoring: enhancing financial inclusion using mobile phone data and social network analytics. *Appl Soft Comput J*. 2019;74:26–39. https://doi.org/10.1016/j.asoc.2018.10.004.
108. Hand DJ, Anagnostopoulos C. A better Beta for the H measure of classification performance. *Pattern Recogn Lett*. 2014;40(1):41–46. https://doi.org/10.1016/j.patrec.2013.12.011.
109. Verbeke W, Dejaeger K, Martens D, Hur J, Baesens B. New insights into churn prediction in the telecommunication sector: a profit driven data mining approach. *Eur J Oper Res*. 2012;218(1):211–229. https://doi.org/10.1016/j.ejor.2011.09.031.
110. Verbraken T, Verbeke W, Baesens B. A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Trans Knowl Data Eng*. 2013;25(5):961–973. https://doi.org/10.1109/TKDE.2012.50.
111. Neto R, Jorge Adeodato P, Carolina Salgado A. A framework for data transformation in credit behavioral scoring applications based on model driven development. *Expert Syst Appl*. 2017;72:293–305. https://doi.org/10.1016/j.eswa.2016.10.059.
112. Gorzałczany MB, Rudziński F. A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability. *Appl Soft Comput J*. 2016;40:206–220. https://doi.org/10.1016/j.asoc.2015.11.037.
113. Abellán J, Castellano JG. A comparative study on base classifiers in ensemble methods for credit scoring. *Expert Syst Appl*. 2017;73:1–10. https://doi.org/10.1016/j.eswa.2016.12.020.
114. Bequé A, Lessmann S. Extreme learning machines for credit scoring: an empirical evaluation. *Expert Syst Appl*. 2017;86:42–53. https://doi.org/10.1016/j.eswa.2017.05.050.
115. Jadhav S, He H, Jenkins K. Information gain directed genetic algorithm wrapper feature selection for credit rating. *Appl Soft Comput J*. 2018;69:541–553. https://doi.org/10.1016/j.asoc.2018.04.033.
116. Sigrist F, Hirnschall C. Grabit: gradient tree-boosted Tobit models for default prediction. *J Bank Finance*. 2019;102:177–192. https://doi.org/10.1016/j.jbankfin.2019.03.004.
117. Pérez-Martín A, Pérez-Torregrosa A, Vaca M. Big Data techniques to measure credit banking risk in home equity loans. *J Bus Res*. 2018;89:448–454. https://doi.org/10.1016/j.jbusres.2018.02.008.
118. Carta S, Ferreira A, Reforgiato Recupero D, Saia M, Saia R. A combined entropy-based approach for a proactive credit scoring. *Eng Appl Artif Intell*. 2020;87. https://doi.org/10.1016/j.engappai.2019.103292.
119. Luo C, Wu D, Wu D. A deep learning approach for credit scoring using credit default swaps. *Eng Appl Artif Intell*. 2017;65:465–470. https://doi.org/10.1016/j.engappai.2016.12.002.
120. Fang F, Chen Y. A new approach for credit scoring by directly maximizing the Kolmogorov–Smirnov statistic. *Comput Stat Data Anal*. 2019;133:180–194. https://doi.org/10.1016/j.csda.2018.10.004.
121. Wang D, Zhang Z, Bai R, Mao Y. A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring. *J Comput Appl Math*. 2018;329:307–321. https://doi.org/10.1016/j.cam.2017.04.036.

122. Xiao H, Xiao Z, Wang Y. Ensemble classification based on supervised clustering for credit scoring. *Appl Soft Comput J*. 2016;43:73–86. https://doi.org/10.1016/j.asoc.2016.02.022.
123. Hayashi Y. Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective. *Oper Res Perspect*. 2016;3:32–42. https://doi.org/10.1016/j.orp.2016.08.001.
124. Luo J, Yan X, Tian Y. Unsupervised quadratic surface support vector machine with application to credit risk assessment. *Eur J Oper Res*. 2020;280(3):1008–1017. https://doi.org/10.1016/j.ejor.2019.08.010.
125. Ashofteh A, Bravo JM. A conservative approach for online credit scoring. *Expert Syst Appl*. 2021;176(July 2020), 114835. https://doi.org/10.1016/j.eswa.2021.114835.
126. Liu W, Fan H, Xia M. Step-wise multi-grained augmented gradient boosting decision trees for credit scoring. *Eng Appl Artif Intell*. 2021;97(May 2020), 104036. https://doi.org/10.1016/j.engappai.2020.104036.
127. Lappas PZ, Yannacopoulos AN. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Appl Soft Comput*. 2021;107, 107391. https://doi.org/10.1016/j.asoc.2021.107391.
128. Zhang W, Yang D, Zhang S. A new hybrid ensemble model with voting-based outlier detection and balanced sampling for credit scoring. *Expert Syst Appl*. 2021;174(December 2020), 114744. https://doi.org/10.1016/j.eswa.2021.114744.
129. Zhang W, Yang D, Zhang S, Ablanedo-Rosas JH, Wu X, Lou Y. A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Syst Appl*. 2021;165, 113872. https://doi.org/10.1016/j.eswa.2020.113872.
130. Xia Y, Zhao J, He L, Li Y, Niu M. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Syst Appl*. 2020;159, 113615. https://doi.org/10.1016/j.eswa.2020.113615.
131. Trivedi SK. A study on credit scoring modeling with different feature selection and machine learning approaches. *Technol Soc*. 2020;63, 101413. https://doi.org/10.1016/j.techsoc.2020.101413.
132. Gunnarsson BR, vanden Broucke S, Baesens B, Óskarsdóttir M, Lemahieu W. Deep learning for credit scoring: do or don't? *Eur J Oper Res*. 2021;295(1):292–305. https://doi.org/10.1016/j.ejor.2021.03.006.
133. Tripathi D, Edla DR, Kuppili V, Bablani A. Evolutionary Extreme Learning Machine with novel activation function for credit scoring. *Eng Appl Artif Intell*. 2020;96(September), 103980. https://doi.org/10.1016/j.engappai.2020.103980.
134. Jeyasothy A, Ramasamy S, Sundaram S. Meta-neuron learning based spiking neural classifier with time-varying weight model for credit scoring problem. *Expert Syst Appl*. 2021;178(October 2020), 114985. https://doi.org/10.1016/j.eswa.2021.114985.
135. Tsai C-F, Sue K-L, Hu Y-H, Chiu A. Combining feature selection, instance selection, and ensemble classification techniques for improved financial distress prediction. *J Bus Res*. 2021;130(300):200–209. https://doi.org/10.1016/j.jbusres.2021.03.018.
136. Engelmann J, Lessmann S. Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning. *Expert Syst Appl*. 2021;174, 114582. https://doi.org/10.1016/j.eswa.2021.114582.
137. Zhou J, Zhang Q, Li X. Fuzzy factorization machine. *Inf Sci*. 2021;546:1135–1147. https://doi.org/10.1016/j.ins.2020.09.067.
138. Uthayakumar J, Vengattaraman T, Dhavachelvan P. Swarm intelligence based classification rule induction (CRI) framework for qualitative and quantitative approach: an application of bankruptcy prediction and credit risk analysis. *J King Saud Univ - Comput Inf Sci*. 2020;32(6):647–657. https://doi.org/10.1016/j.jksuci.2017.10.007.
139. Moscato V, Picariello A, Sperlí G. A benchmark of machine learning approaches for credit score prediction. *Expert Syst Appl*. 2021;165, 113986. https://doi.org/10.1016/j.eswa.2020.113986.
140. Li Z, Zhang J, Yao X, Kou G. How to identify early defaults in online lending: a cost-sensitive multi-layer learning framework. *Knowl Base Syst*. 2021;221, 106963. https://doi.org/10.1016/j.knosys.2021.106963.
141. Lee JW, Lee WK, Sohn SY. Graph convolutional network-based credit default prediction utilizing three types of virtual distances among borrowers. *Expert Syst Appl*. 2021;168, 114411. https://doi.org/10.1016/j.eswa.2020.114411.
142. Maldonado S, López J, Vairetti C. Time-weighted fuzzy support vector machines for classification in changing environments. *Inf Sci*. 2021;559:97–110. https://doi.org/10.1016/j.ins.2021.01.070.
143. Chen S, Guo Z, Zhao X. Predicting mortgage early delinquency with machine learning methods. *Eur J Oper Res*. 2021;290(1):358–372. https://doi.org/10.1016/j.ejor.2020.07.058.
144. Barddal JP, Loezer L, Enembreck F, Lanzuolo R. Lessons learned from data stream classification applied to credit scoring. *Expert Syst Appl*. 2020;162(March), 113899. https://doi.org/10.1016/j.eswa.2020.113899.
145. Pang PS, Hou X, Xia L. Borrowers' credit quality scoring model and applications, with default discriminant analysis based on the extreme learning machine. *Technol Forecast Soc Change*. 2021;165, 120462. https://doi.org/10.1016/j.techfore.2020.120462.
146. Kennedy K. *Credit Scoring Using Machine Learning [PhD Thesis]*. Technological University Dublin; 2013.