

Symbolic decision stumps in individual credit scoring

Marcin Pełka*

Submitted: 7 January 2019. Accepted: 15 October 2019.

Abstract

Polish bank law defines credit ability as the ability to repay a credit and interest according to terms that have been set in the credit agreement. Credit scoring is a crucial element for any bank with a fundamental impact on its future financial condition. Credit scoring can be calculated with the application of statistical methods. The main aim of this paper is to present the possibility of an ensemble of symbolic decision stumps in credit scoring where two real data sets are used. Results show that symbolic decision stumps can be applied in individual credit scoring.

Keywords: credit ability, symbolic data, decision stumps

JEL: C39, C53, C58, C63, C88

* Wrocław University of Economics and Business, Department of Econometrics and Computer Science;
e-mail: marcin.pelka@ue.wroc.pl.

1 Introduction

In banking law, creditworthiness is understood as the likelihood that a borrower will repay the loan on time, with interest and according to the payment dates set out in a credit agreement. Additionally, Polish banking law states that a borrower should provide the bank with all the information that is necessary for assessing this likelihood (Article 70 of the Banking Act of 29 August 1997, Journal of Laws, No. 140, item 939, as amended).

A credit assessment is not a one-off procedure; rather, it should take place continually to verify that the borrower can keep repaying a loan with interest and on time. This assessment and the period in which the principal or the interest is paid is related to the classification of credit exposure and, in turn, to the related level of reserve funds (Journal of Laws 2008, No. 235, item 1589, as amended).

According to reports by the the Polish Financial Supervision Authority (KNF), at the end of 2017 the banks in Poland had 21.8 million consumption loans, with a total value of 155.6 billion zloty and an average value of 7,200 zloty (see Kotowicz 2018a, p. 21). On the other hand, the share of non-performing loans in various areas of consumer consumption lending was as high, on average, as 9.68% of the total credit extended (see Kotowicz 2018a, p. 23). At the same time, in spite of the heightened pace of growth in lending activity, the quality of the consumer credit portfolio generally remained stable (see Kotowicz 2018b, pp. 7–9).

From a bank's point of view, then, a key issue is the analysis and assessment of creditworthiness of an agreement with a customer at each stage of its life. The most common approaches in assessment methods involve quality or credit scoring analyses.

Quality analysis is carried out on the basis of the borrower's individual characteristics. Key issues here include the borrower's age, marital status, number of years of employment and residential status, as well as the number of persons in the borrower's household. Credit scoring, on the other hand, involves assigning a given number of points to specific variants of measureable and non-measureable characteristics. For example, a score of "10" may be assigned to a borrower who provides three potential sources of collateral; "5" if two sources are presented; "2" for one source; "0" for none.

In banking practice, behavioural and utility-based scoring are used. Behavioural scoring assesses the bank's long-term customers. Here, the assessment is not based on a credit application, but on the customer's history of dealing with the bank. This allows the bank to propose new products to the customer, or to change the existing ones. Utility-based scoring, which applies to new customers, involves the assessment of a submitted credit application.

Reviews of quantitative methods (in the broader sense of the term) of credit scoring, including those involving machine learning, have been presented by authors such as Wójciak (2007), Kuryłek (2000), Feruś (2006), Witkowska, Chrzanowska (2006), Migut (2003), Pisula (2013), Hoffmann (2009), Munkhdalai et al. (2019), Leo, Sharma, Maddulety (2019), Louzada, Ara, Fernandes (2016), and Lessmann et al. (2015). Generally speaking, various methods and approaches are used in assessing creditworthiness.

Among those that are most commonly used and presented, the following stand out: logistic regression, support vector machines, MARS-type regression, decision trees, random forests, artificial networks (including multilayer perceptrons and GNG-type networks), the *k*-nearest neighbour method, multi-model approaches, hybrid approaches, the fuzzy logic approach, linear discriminant analysis

(cf., for example, Wójciak 2007; Kuryłek 2000; Feruś 2006; Witkowska, Chrzanowska 2006; Migut 2003; Pisula 2013; Hofmann 2009; Munkhdalai et al. 2019; Leo, Sharma, Maddulety 2019; Louzada, Ara, Fernandes 2016; Lessmann et al. 2015), and survival analysis (Leo, Sharma, Maddulety 2019, p. 16).

The results of comparative studies on the application and usefulness of different methods and approaches to credit scoring – reported in the works of, among others, Munkhdalai et al. (2019), Leo, Sharma, Maddulety (2019); Louzada, Ara, Fernandes (2016), and Lessmann et al. (2015) – warrant the view that, generally, methods related to neuron networks, logistic regression, support vector machines and fuzzy logic provide somewhat better results (in the sense of such measures as relevance and sensitivity) than other proposed solutions; cf. Munkhdalai et al. (2019, pp. 15–16), Lessman et al. (2015, p. 32), and Louzada, Ara, Fernandes (2016, pp. 19, 25).

Common multi-model messages, such as boosting and XGBoosting, also lead to good results; see, for example, Munkhdalai et al. (2019, pp. 17–18), Lessmann et al. (2015, pp. 36–38).

On the other hand, when symbolic data is involved in the assessment of creditworthiness, special problems arise that only the works of Dudek (2013) and Pełka (2018) have addressed. Both use kernel discriminant symbolic data analysis, symbolic decision trees, the k -nearest neighbour method and the multi-model approach. In both articles, the error of the model is used to assess the model. The lowest error value was obtained for random forests, and then for a decision tree. In these papers, the least suitable method turned out to be the k -nearest-neighbour method for symbolic data.

In analyzing the possible benefits arising from the application of symbolic data to the assessment of creditworthiness, it is certainly worth noting the possibility of a full description of items with the use of symbolic data of a different type. It is possible, for example, to apply a number interval (symbolic interval-valued variable) to the period of the loan, or the value of the loan itself, or to apply multivariate variables to proposed types of insurance. This allows a bank representative to assess the creditworthiness not only for a single loan amount or period, but also for a number of variants. The use of this method is also advantageous in that a variety of methods are available that may serve to assess creditworthiness. On the other hand, taking into account possible problems, it is worth noting that in the case of symbolic data we have only a set of symbolic data that describes a thousand borrowers from German banks. Thus, we do not have very many possibilities to compare different methods for various data sets, a procedure reported in the works of Munkhdalai et al. (2019), Leo, Sharma, Maddulety (2019), Louzada, Ara, Fernandes (2016), and Lessman et al. (2015). Another problem may be the need to prepare data properly for symbolic data analysis in the form of specially prepared files. For this type of data, such preparation requires special knowledge and skills. A final problem may be the need to use the R program for calculations, which may in turn require certain knowledge and experience.

Decision trees for classic data are a useful tool for assessing creditworthiness (see, for example, Wójciak 2007; Witkowska, Chrzanowska 2006; Kuryłek 2006; Migut 2003; Pisula 2013; Baesens et al. 2003; Xia et al. 2017; Zekic-Susac, Sarlija, Bencic 2004; Zhang et al. 2010; Hand, Henley 1997; Yobas, Crook, Ross 2000; Munkhdalai et al. 2019; Leo, Sharma, Maddulety 2019; Louzada, Ara, Fernandes 2016; Lessman et al. 2015).

Regardless of whether we are discussing classical or symbolic data, this work will focus mainly on typical decision trees, decision stumps for classic data (cf., for example, Ben-David, Frank 2009; Paleologo, Elisseeff, Antonini 2010) and the multi-model approach using such trees. On the other hand, there is a gap in the subject literature in the area of the use of symbolic data in the presentation and use of single-level decision trees (decision stumps).

The main purpose of this article is to present the use of a multi-model approach to the analysis of symbolic data using single-level decision trees in assessing the creditworthiness of physical persons. The work has adopted the hypothesis that decision stumps of symbolic data may be a useful tool for assessing creditworthiness.

The results obtained in the empirical part will be compared to those obtained with the help of kernel discriminant analysis¹ of symbolic data and classic trees for symbolic data, based on optimal intervals. The `symbolicDA` package for R software (Dudek, Pełka, Walesiak 2018) as well as the author's script in the R program have been used for calculations.

2 The multi-model approach to the analysis of symbolic data and one-level decision trees for symbolic data²

In the analysis of symbolic data, objects may be described with the help of the following variables (see, for example, Bock, Diday 2000, pp. 2–3; Billard, Diday 2006, pp. 7–30; Diday, Noirhomme-Fraiture 2008, pp. 10–19):

- nominal,
- ordinal,
- interval,
- ratio,
- symbolic interval,
- symbolic multivariate (or lists of categories) – also known as symbolic multinomial,
- symbolic multivariate with weights (or lists of categories with weights) – also known as symbolic multinomial with weights,
- symbolic histogram (or number intervals with weights).

A broader description of symbolic items, and the similarities and differences of this type of data in comparison with classical data has been provided by, among others, Bock, Diday (2000, pp. 2–8), Billard, Diday (2006, pp. 7–66), Diday, Noirhomme-Fraiture (2008, pp. 3–30), Dudek (2013, pp. 42–43).

For every researcher, the proper choice of tools for solving a problem constitutes an essential question. For example, in the case of assessing creditworthiness many different models may be used, such as decision trees, neuron networks, logistic regression and discriminant analysis. The choice of a single suitable tool (a model) is not always obvious and simple. This is why instead of applying one model it is better to use several diverse models and combine their results into one. This practice is called “ensemble approach”.

The ensemble approach is nothing more than a combination or aggregation of results obtained from many (M) base models into one aggregated model (see Figure 1). The purpose of combining models is to improve model prediction, because the aggregate number is more accurate (it has fewer errors) than any one of the constituting models (cf. Gatnar 2008, p. 62).

¹ Kernel discriminant analysis for symbolic data has been more widely described by, among others, Dudek (2013, pp. 143–168); and in Gatnar i Walesiak (2011, pp. 280–291).

² This theoretical section, dealing with decision trees, has been prepared to a large extent with reference to the publication Gatnar and Walesiak (2011).

In the case of the multi-model approach the essential issue is to connect the models (represented by Σ in Figure 1). Many approaches in this area have been proposed in the subject literature; among others, bagging, boosting, and stacking (see, for example, Gatnar 2008, pp. 138–168; Polikar 2006; 2007).

In this article, the bootstrap aggregating method (bagging) has been used, as proposed by Breimana in 1996. This method uses the idea of the architecture of parallel aggregated models, which assumes the independent operation of each of the base models – see Figure 2 (see, for example, Gatnar 2008, p. 140; Breiman 1996, p. 123).

An algorithm for the bagging method may be summarized in the following steps (see Polikar 2006; 2007; Gatnar 2008):

1) establishing the number of base models M ,

2) division of the original data set into M subsets and sub-samples, called bootstrap samples. The items are randomly assigned to subsets, with returns.

In the case of the bagging method, about 37% of observations from the first data set usually are not placed in any of the subsets. These items form a set of observations called out-of-bag (OOB), which are often used as a supplementary test set.

1. The construction of a base model for each of the bootstrap samples (for example, decision trees, support vector machines, and so on). In effect we receive results for M base models.

2. Combining the results obtained with the help of many base models into one aggregate model, D^* . One of the most often used methods for combining partial results in the case of discriminant (classification) methods is the majority voting method, where the item is assigned to the class in which the majority of the M base models classifies it (Gatnar 2008, p. 114).

In the case of issues related to regression, one of the most frequently used methods is the averaging of results (cf. Gatnar 2008, p. 140; Kuncheva 2004, p. 204).

Single-level decision trees for symbolic data, also called decision stumps, are specific types of decision trees with a depth equal exactly one; i.e. with only one node (see Iba, Langley 1992).

In the case of decision trees for symbolic data in a data set, it is necessary to have at least one nominative variable (a dependent variable) whose realizations depend on a set of symbolic variables (independent variables), which may include symbolic interval-valued, multinomial or classic variables (see Gatnar, Walesiak 2011, pp. 282–285).

The subsequent steps of the algorithm include (see Gatnar, Walesiak 2011, pp. 282–285):

1. Data collection and the construction of a symbolic data table.

2. For a symbolic multinomial, nominal or ordinal variable a frequency table is built. In the case of symbolic multinomial, nominal or ordinal variables, the table presents a count of the frequency with which a given category is observed for particular items. On the other hand, for interval symbolic variables, it is essential to calculate the arithmetical mean for each possible combination of upper and lower bounds of the variable ranges.

3. The *a priori* establishment of a limit value n^* for the node size (number of items in the node) and the limit value W^* for the quality criterion of the tree division. If the node size is less than the n^* value, this is the final node. If the quality criterion is greater than W^* for a particular binary question, the question may be used for division.

4. The construction of binary questions for each of the variables, and the calculation of the probability $p_k(l)$ of assigning the item k to the left node.

A. For symbolic interval and quotient variables, we use the mean of the distributions calculated in point 2 above. They will constitute the so-called cutting values c . Thereafter, depending on whether the established c value is found in the symbolic interval variable distribution – above the lower bound, or below the upper bound – the probability of assigning this item to the left node is established (see Gatnar, Walesiak 2011, pp. 283–284; Dudek 2013, p. 153):

– if the value c is found within the symbolic interval variable distribution, then

$$p_k(l) = \frac{c - \underline{v}_{kj}}{\bar{v}_{kj} - \underline{v}_{kj}} \quad (1)$$

where:

$k = 1, \dots, n$ – number of the symbolic object,

\underline{v}_{kj} – lower bound of a symbolic variable,

\bar{v}_{kj} – upper bound of a symbolic variable;

– if the value c is found below the lower bound for the symbolic interval variable, then

$$p_k(l) = 0 \quad (2)$$

– if the value c is found above the upper bound for the symbolic interval variable, then

$$p_k(l) = 1 \quad (3)$$

B. For ordinal, nominal and multivariant variables, the cutting value c constitutes a particular variable data category (excluding the last category). For each item, the frequencies of variable values that are lower than c , and those that are higher, should be added. Likewise, the c -value constitutes a distinct variable category for nominal variables. For a given item, the c -value is equal to the frequency of the category.

5. The probability of assigning an item to the right node is $p_k(r) = 1 - p_k(l)$.

6. The calculation of the quality criterion for assigning the W node for each c point (see Gatnar, Walesiak 2011, p. 284; Dudek 2013, p. 154):

$$W_j(t, c) = \log \prod_{k=1}^n [p_k(l)P_l(s) + p_k(r)P_r(s)] \quad (4)$$

where:

$j = 1, \dots, m$ – variable number,

t – node number,

c – cutting value,

$p_k(l)$ – probability of assigning the k item to the left node,

$p_k(r) = 1 - p_k(l)$ – probability of assigning the k item to the right node,

$P_l(s)(P_r(s))$ – the conditional probability of observing the class to which the k -th item belongs on the left node (this is the quotient of the sum of probabilities of assigning all items to this node).

7. The choice of the highest W values for each variable.

8. The choice of the W variable that is simultaneously greater than W^* and splitting the node according to the method appropriate for the given variable, on condition that the size of node n

is greater than that of n^* . Where one of these conditions has not been fulfilled, the node may not be divided further, and is thus a final node.

9. Steps 6–8 should be repeated for each node until the final nodes are obtained. At a later stage of tree construction, the questions dealt with in earlier steps are not addressed.

10. Visualizing and interpreting the results.

In the empirical part, the `decisionTree.SDA` function from the `symbolicDA` package (see Dudek, Pełka, Walesiak 2018) is used. This assures control both over the depth of the construction of the classification tree, and the random selection of variables for tree construction, similarly to what occurs in the random forest algorithm (see, e.g., Ho 1998).

3 Characteristics of the data set

For the purposes of empirical research, a data set has been used containing information on a thousand borrowers from German banks. Dudek prepared a table of symbolic data for the purposes of his monograph (see Dudek 2013, p. 162). It contains symbolic first-order items described by 17 symbolic variables of a various types (see Table 1). In a further part of this work, they will be denoted as D-B. For the purposes of the multi-model approach, this set has been randomly divided into a training set of 700 items and a test set of 300 items.

4 Results of the empirical research

For the purposes of comparing the effectiveness of a multi-model approach to the use of one-level decision trees in the assessment of creditworthiness, we also present the results obtained with the help of a single symbolic data decision tree, an aggregate model using symbolic data decision trees, kernel discriminant analysis, as well as analysis of symbolic data neural networks (multilayer perceptron).

On the other hand, logistic regression of symbolic interval data has not been used. As this model can be applied only with symbolic interval data, its use here would entail the removal of multivariate variables from the first data set and thus a loss of part of the information.

Generally speaking, kernel discriminant analysis of symbolic data makes use of the idea of the intensity estimator, which makes it possible to provide an estimate of the number of items of a given class that are close to what is to be classified. More on kernel discriminant analysis for symbolic data was written by, among others, Dudek (2013, pp. 143–168) and Gatnar and Walesiak (2011, pp. 280–291).

In the case of the multilayer perceptron for symbolic data, the first step is to change symbolic data to classic data, where symbolic interval variables are represented by a logarithm of their length, while symbolic multinomial, nominal and ordinal variables (or m variants) are represented by m dichotomous variables (in which 1 signifies the presence of the category, and 0 not). On the other hand, symbolic multivariate data with weights (or m variants) are represented by m variables, which appear as the weights of particular symbolic variable variants (categories). Thus the original data matrix is much larger than the symbolic data table. Multilayer perceptrons and symbolic data were described more broadly by Diday and Norimhomme-Fraiture (2008, pp. 373–391).

Table 2 presents the most important parameters used in the construction of single and aggregate models.

Figure 3 presents the results obtained for a single classification tree constructed on the basis of all variables.

The most important variables are savings, the loan period, as well as the borrower's employment status and the amount of the loan.

Table 3 summarizes the averaged parameters (from 20 repetitions for each approach) that enable evaluation and comparison of the values obtained from single models, and the multi-layer approach using decision stumps.

The database used in this article was also analyzed in the work by Dudek (see Dudek 2013, pp. 161–168) and Peřka (see Peřka 2018). Both these publications address model error only in the discussion of results and there is no other information allowing to compare them with the results provided in Table 5. Both Dudek's (Dudek 2013) and Peřka's (Peřka 2018) works use decision trees, random forests, kernel discriminant analysis of symbolic data, and the k -nearest neighbour method. In the case of a single decision tree, the error was 6.94% (Dudek 2013, p. 166; Peřka 2018, p. 206); for random forest, the error was 5.556% (Dudek 2013, p. 166); and where the k -nearest neighbour method was used, the error was 16% for the individual model and 12% for the aggregate model (Peřka 2018, p. 206).

Comparing these results with those obtained for the multi-method approach using a decision stump, it may be said that the proposed solution is similar in terms of error to random forests, and to a single decision tree divided into a test set of 928 items, or a test set of 72 items (Dudek 2013, pp. 165–166).

5 Summary

The ensemble approach and the single model using symbolic data of various types may be used successfully to assess the creditworthiness of physical persons.

The ensemble approach used in assessing creditworthiness makes it possible to obtain more precise data (in terms of smaller model errors) than the single model. On the other hand, in the case of the single model, a significant advantage appears when the results are interpreted on the basis of the obtained decision trees. Both typical decision trees and decision stumps provided similar results in a multi-model approach.

In the case of the G-B set, the most important variables are the borrower's savings, the loan period, the borrower's employment status and the amount of the loan.

Presenting the possible advantages, limitations and problems arising from the use of the multi-model approach for symbolic data in assessing the creditworthiness of physical persons, it is essential to indicate the possibility of a more complete description of the items, using symbolic data of various types. Symbolic data make it possible to assess the creditworthiness for more than one loan amount or period, and for more than one type of collateral or loan purpose. Symbolic data allow creditworthiness to be assessed for a range of amounts (symbolic interval variables) various credit periods within a specified range, various forms of collateral, and various purposes for which the loan is to be extended. Another advantage of the use of a symbolic approach is the wide availability of models and methods which operate in a similar way to models and methods developed for classical data.

Among the problems and limitations, it is worth pointing out that in the case of symbolic data there is only one symbolic data set describing a thousand borrowers from German banks. This means that there is not much opportunity to compare different methods for different data sets. Further, it is necessary to have a certain knowledge in the area of data preparation, import and use in the R³ program.

Comparing the results described in the present article with those presented in the subject literature for this data set, it may be stated that the proposed solution is similar, from the point of view of error, to random forests or decision stumps divided into a test set of 928 items or a test set of 72 items (Dudek 2013, pp. 165–166).

The goal of further research should be to apply methods developed with the analysis of symbolic data in mind to well-known classic data sets (for example, from the Australian market) and to compare the obtained results with those from classic methods.

References

- 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.
- Bock H.-H., Diday E., eds. (2000), *Analysis of Symbolic Data. Explanatory Methods for Extracting Statistical Information from Complex Data*, Springer Verlag.
- Billard L., Diday E. (2006), *Symbolic Data Analysis. Conceptual Statistics and Data Mining*, John Wiley & Sons.
- Ben-David A., Frank E. (2009), Accuracy of machine learning models versus “hand crafted” expert systems – a credit scoring case study, *Expert Systems with Applications*, 36(3), 5264–5271.
- Breiman L. (1996), Bagging predictors, *Machine Learning*, 24, 123–140.
- Diday E., Noirhomme-Fraiture M. (2008), *Symbolic Data Analysis. Conceptual Statistics and Data Mining*, Wiley.
- Dudek A. (2013), *Metody analizy danych symbolicznych w badaniach ekonomicznych*, Wyd. UE we Wrocławiu.
- Dudek A., Pełka M., Walesiak M. (2018), *The symbolicDA package for R software*, <https://cran.r-project.org/web/packages/symbolicDA/index.html>.
- Feruś A. (2006), Zastosowanie metody DEA do określania poziomu ryzyka kredytowego przedsiębiorstw, *Bank i Kredyt*, 7, 44–59.
- Gatnar E. (2008), *Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN.
- Gatnar E., Walesiak M., red. (2011), *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck.
- Hand D.J., Henley W.E. (1997), Statistical classification methods in consumer credit scoring: a review, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523–541.

³ A meaningful solution to this problem is described in Gatnar and Walesiak (2011). This work contains a description of methods and models, as well as examples of their application.

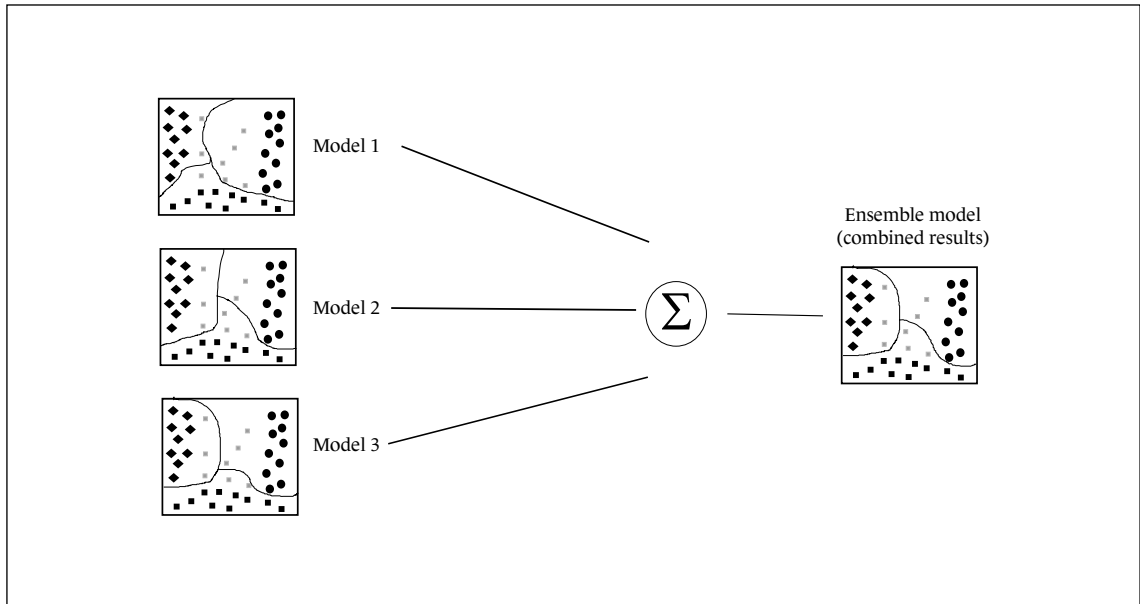
- Ho Tk (1998), The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Hoffmann W. (2009), Metoda AHP w ocenie stosowalności metod oceny zdolności kredytowej w systemie ekspertowym, *Metody Ilościowe w Badaniach Ekonomicznych*, 10(1), 99–106.
- Iba W., Langley P. (1992), Induction of one-level decision trees, in: D.H. Sleeman, P. Edwards (eds.), *Proceedings of the Ninth International Workshop on Machine Learning*, Morgan Kaufmann Publishers.
- Kotowicz A. (2018a), *Wyniki badania portfela kredytów mieszkaniowych i konsumpcyjnych gospodarstw domowych według stanu na koniec 2017 r.*, Urząd Komisji Nadzoru Finansowego, Warszawa, https://www.knf.gov.pl/publikacje_i_opracowania.
- Kotowicz A. (2018b), *Raport o sytuacji banków w 2017 r.*, Urząd Komisji Nadzoru Finansowego, Warszawa, https://www.knf.gov.pl/publikacje_i_opracowania.
- Kuncheva L.I. (2004), *Combining Pattern Classifiers. Methods and Algorithms*, Wiley.
- Kuryłek W. (2000), Credit scoring – podejście statystyczne, *Bank i Kredyt*, 6, 72–77.
- Leo M., Sharma S., Maddulety K. (2019), Machine learning in banking risk management, *Risks*, 7(1), 1–22.
- Louzada F., Ara A., Fernandes G.B. (2016), Classification methods applied to credit scoring: A systematic review and overall comparison, *Surveys in Operations Research and Management Science*, 21(2), 117–134.
- Lessmann S., Baesens B., Seow H.-V., Lyn C.T. (2015), Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research, *European Journal of Operational Research*, 247(1), 124–136.
- Migut G. (2003), *Modelowanie ryzyka kredytowego*, Statsoft Polska, [media.statsoft.nazwa.pl > _old_dnn > downloads > migut](http://media.statsoft.nazwa.pl/_old_dnn/downloads/migut).
- Munkhdalai L., Munkhdalai T., Namsrai O.-E., Lee J.Y. Ryu K.H. (2019), An empirical comparison of machine-learning methods on bank client credit assessments, *Sustainability*, 11(3), <https://doi.org/10.3390/su11030699>.
- Paleologo G., Elisseeff A., Antonini G. (2010), Subagging for credit scoring models, *European Journal of Operational Research*, 201(2), 490–499.
- Pełka M. (2018), *Podejście wielomodelowe analizy danych symbolicznych w ocenie zdolności kredytowej osób fizycznych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 507, 200–207.
- Polikar R. (2006), Ensemble based systems in decision making, *IEEE Circuits and Systems Magazine*, 6(3), 21–45.
- Polikar R. (2007), Bootstrap inspired techniques in computational intelligence: ensemble of classifiers, incremental learning, data fusion and missing features, *IEEE Signal Processing Magazine*, 24(4), 59–72.
- Pisula T. (2013), *Metodyczne aspekty zastosowania modeli skoringowych do oceny zdolności kredytowej z wykorzystaniem metod ilościowych*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 323, 265–288.
- Witkowska D., Chrzanowska M. (2006), *Drzewa decyzyjne jako metoda grupowania klientów banku*, Prace Naukowe Akademii Ekonomicznej w Katowicach, Modelowanie preferencji a ryzyko.
- Wójciak M. (2007), *Metody oceny ryzyka kredytowego*, Polskie Wydawnictwo Ekonomiczne.

- Zekic-Susac M., Sarlija N., Bencic M. (2004), *Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models*, 26th International Conference on Information Technology Interfaces, IEEE.
- Zhang D., Zhou X., Leung S.C.H., Zheng J. (2010), Vertical bagging decision trees model for credit scoring, *Expert Systems with Applications*, 37(12), 7838–7843.
- Xia Y., Liu C., Li Y.Y., Liu N. (2017), A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring, *Expert Systems with Applications*, 78, 225–241.
- Yobas M.B., Crook J.N., Ross P. (2000), Credit scoring using neural and evolutionary techniques, *IMA Journal of Management Mathematics*, 11(2), 111–125.

Appendix

Figure 1

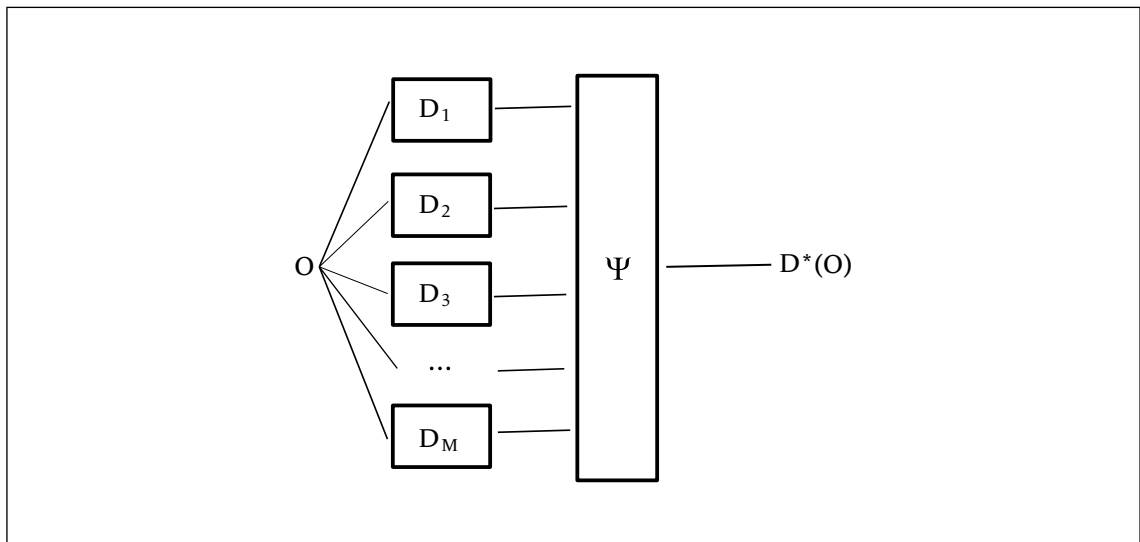
The general idea of the ensemble approach



Source: own elaboration, on the basis of Gatnar (2008, pp. 62–67), Polikar (2006; 2007).

Figure 2

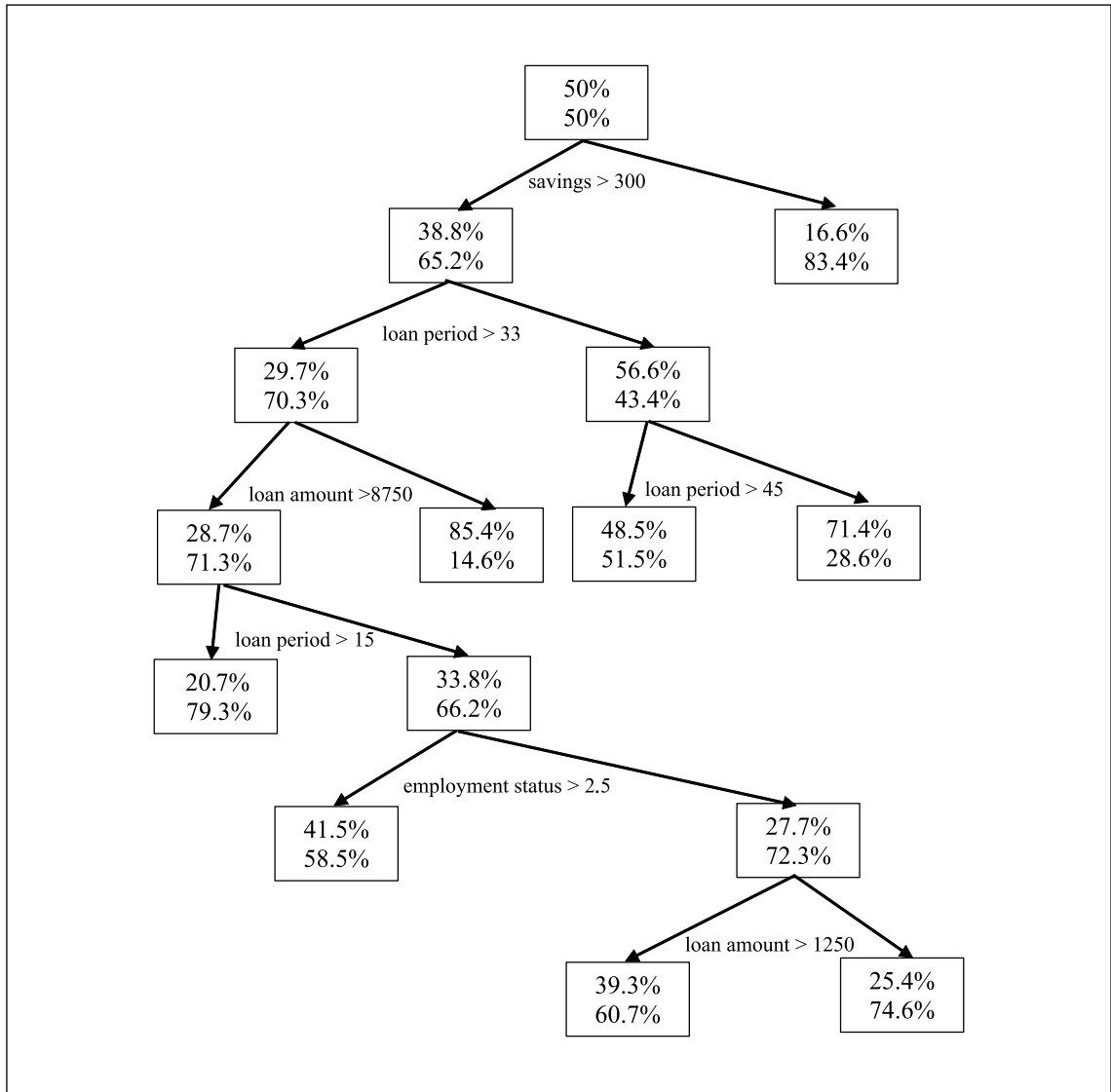
The idea of parallel architecture in the construction of aggregate models



Source: the author's own work, on the basis of Gatnar (2008, p. 140) and Breiman (1996, p. 123).

Figure 3

Classification tree for German bank customers



Source: own elaboration using the R program.

Table 1
Description of the variables for the second data set

No.	Variable name	Variable type	No.	Variable name	Variable type
1	Assignment to Class 1 (on-time repayment) or Class 2 (repayment with problems)	nominal	10	Sureties	multinomial
2	Loan period	interval	11	Most valuable assets	multinomial
3	Information about other loans	multinomial	12	Age	interval
4	Purpose of the loan	multinomial	13	Information about other loans	multinomial
5	Loan amount	interval	14	Type of ownership of premises	multinomial*
6	Savings	interval	15	Previous loans	multinomial*
7	Employment status	interval	16	Description of occupation and type of employment	multinomial
8	Installment payments as a percentage of income	interval	17	Foreigner	multinomial*
9	Sex	multivariant*			

* Multivariant variable, for which only one of its variants is possible.

Source: own elaboration.

Table 2
Parameters adopted in single models and in a multi-model approach

Criteria	Single model			Multi-model approach		
	decision tree	kernel discriminant analysis	multi-level perceptron	decision tree	one-level decision tree	multi-level perceptron
Model parameters	$n^* = 2$ $W^* = -1e10$	Bandwidth $h = 1.2$ normalized Ichino-Yaguchi distance measure (U_3)	8 hidden layers	$n^* = 2$ $W^* = -1e10$ 50 models	$n^* = 2$ $W^* = -1e10$ 50 models random selection of variables within the models	8 hidden layers 20 models

n^* – minimum number of items in the node; W^* – minimum function-criterion value

Source: own elaboration.

Table 3

Averaged results of parameters enabling the assessment of received model

Parameter	Model	Method			
		kernel discriminant analysis	decision trees	multi-layer perceptron	decision stump
Sensitivity	single	0.74	0.79	0.79	a
	aggregate	b	0.885	0.86	0.875
Specificity	single	0.94	0.945	0.95	a
	aggregate	b	0.975	0.97	0.97
Precision	single	0.87	0.89	0/90	a
	aggregate	b	0.95	0.93	0.94
Model error	single	0.13	0.11	0.10	a
	aggregate	b	0.05	0.07	0.06

a – indicates that the method was not used as a single model.

b – indicates that the method was not used in a multi-model approach.

Source: own elaboration based on results obtained from R software.

