Knowledge-Based Systems 51 (2013) 72–84

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Feature selection in corporate credit rating prediction

Petr Hajek [a,*], Krzysztof Michalak [b]

[a] Institute of System Engineering and Informatics, Faculty of Economics and Administration, University of Pardubice, Studentská 84, Pardubice, Czech Republic
[b] Department of Information Technologies, Wrocław University of Economics, Komandorska 118/120, Wrocław, Poland

ARTICLE INFO

Article history:
Received 20 June 2012
Received in revised form 3 July 2013
Accepted 13 July 2013
Available online 19 July 2013

Keywords:
Feature selection
Credit rating
Classification
Wrapper
Mixed feature selection method

ABSTRACT

Credit rating assessment is a complicated process in which many parameters describing a company are taken into consideration and a grade is assigned, which represents the reliability of a potential client. Such assessment is expensive, because domain experts have to be employed to perform the rating. One way of lowering the costs of performing the rating is to use an automated rating procedure. In this paper, we assess several automatic classification methods for credit rating assessment. The methods presented in this paper follow a well-known paradigm of supervised machine learning, where they are first trained on a dataset representing companies with a known credibility, and then applied to companies with unknown credibility. We employed a procedure of feature selection that improved the accuracy of the ratings obtained as a result of classification. In addition, feature selection reduced the number of parameters describing a company that have to be known before the automatic rating can be performed. Wrappers performed better than filters for both US and European datasets. However, better classification performance was achieved at a cost of additional computational time. Our results also suggest that US rating methodology prefers the size of companies and market value ratios, whereas the European methodology relies more on profitability and leverage ratios.

1. Introduction

There are many instances when the credibility of a company needs to be measured. Bond investors, debt issuers, governmental officers, and companies that provide credit use credit ratings to assess the investment risk. These ratings are a basis for important decisions, and therefore there is a need for them to be as accurate as possible. Unfortunately, this usually means that many details of a company's profile have to be taken into consideration. Such a detailed analysis can be performed by experts, but this is often costly and time-consuming.

Because manual analysis of a company's profile is slow and costly, currently significant emphasis is placed on computational methods of credit rating assessment. The methods applied to this task can be broadly divided into two groups: traditional statistical methods (e.g. [36]) and artificial intelligence (AI) methods (e.g. [6,32,26]). Using traditional statistical methods is difficult because of the complexity of dependencies between various factors that influence the final rating. Nevertheless, methods such as multiple discriminant analysis (MDA) and linear regression (LR) have been applied to credit rating prediction in the literature.

AI methods are often employed when the relationships between input parameters and the outcome are too complex to describe analytically. In the case of credit rating prediction, the most promising group of methods are various classifiers that can be trained on examples – a dataset describing companies previously labeled by experts. A previously trained classifier is a model that represents dependencies between input parameters and object classification (in the case of corporate credit rating prediction – a grade representing the credibility of a company). This type of classifier has a generalization ability – it can produce ratings for yet unseen companies.

Apart from generating credit ratings for companies, corporate credit rating prediction models can be used as early warning indicators. In fact, such models may detect the start of financial crises. In addition, automatic corporate credit rating prediction can be effectively used by supervisory institutions for determining regulatory capital requirements [21].

Companies for which a credit rating is performed are described by a number of variables (features) that are used as input data for modeling. These parameters mainly reflect various aspects of economic and financial performance of a company.

An important problem in modeling credit ratings is the selection of the most appropriate set of variables. The influence of various parameters on the classification result must be quantified. Thus, only important variables can be used for the credit rating process. There is a wide variety of feature selection methods that

* Corresponding author. Tel.: +420 466 036 074; fax: +420 466 036 010.
E-mail addresses: petr.hajek@upce.cz (P. Hajek), krzysztof.michalak@ue.wroc.pl (K. Michalak).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.07.008

can be used to select the appropriate set of variables. Sensitivity analysis (stepwise procedure) may be used to select features that have the most influence on the classification result.

In this paper a different method was used, whereby features are selected using the wrapper approach (an iterative selection procedure that selects features based on the evaluation of performance of the classifier using selected features). In addition, our research is aimed at comparing feature selection methods used for credit rating prediction. To the best of our knowledge, no previous study has attempted to compare filters and wrappers as feature selection methods within the classification process in this or related business domains. We hypothesized that wrapper approaches would contribute to a higher classification accuracy of the employed classifiers compared to filter approaches. In our research we compared filter and wrapper methods in order to verify this hypothesis.

Since two datasets were used, namely data from the US and Europe, we also address the impact of country-related determinants. Thus, our results may lead to a richer understanding of the role of individual input variables and the categories of input variables, respectively, in corporate credit rating prediction.

This paper is structured as follows. In Section 2, details of the credit rating process are presented and methods currently used in the literature for corporate credit rating prediction are described. Section 3 contains data description and a discussion of input variables used for classification. In Section 4 the feature selection process is described and four feature selection methods are introduced, namely two wrappers and two filters. The results obtained in the experiments using various classifiers are presented in Section 5. Section 6 concludes the paper.

2. Credit rating prediction – literature review

Corporate credit rating is a process in which a grade $\omega \in \Omega$ from a predefined rating scale Ω is assigned to a company. Rating agencies, such as Standard & Poor's (S&P's), Moody's, and Fitch have their own rating scales. For example, the rating scale of the S&P's is $\Omega = \{AAA, AA, A, BBB, BB, B, CCC, CC, C, D\}$ – a total of 10 grades (rating classes) that are ordered from AAA, the most promising for investors, to D, the most risky one.

Prior studies on credit rating prediction vary with respect to assessed objects, input variables used, and the set of rating classes Ω. Traditional statistical methods and AI methods have been previously employed in the literature for corporate credit rating prediction.

Studies comparing traditional statistical methods showed that the most successful methods of that type are the ordered logistic regression (OLR) and ordered probit model (OPM) [38]. The two methods have outperformed other statistical methods such as linear regression (LR) and multiple discriminant analysis (MDA) [36,37]. This may be due to the fact that OLR and OPM take the ordering of rating classes into consideration.

To use statistical methods, one has to first choose a model with a predefined structure to represent observations. Then, the parameters of the model are estimated to fit the model to the observational data. The advantage of such an approach is that the models are relatively easy to explain. However, statistical models require various assumptions to be theoretically valid.

Another approach is to use AI methods. The AI methods differ from traditional statistical methods in that they allow learning the model from data [32]. The advantage of such an approach is that AI methods usually do not require specific assumptions on the distribution of data.

Using concepts from the machine learning paradigm, the problem of credit rating prediction can be formulated as a classification problem in which rating classes used by a particular rating agency are known in advance. A typical classification procedure is performed as supervised learning. This learning type requires a sample of companies that were initially assigned proper ratings. A classifier of a chosen type is first trained using this sample. Then, the trained classifier can be used to predict the ratings of previously unseen companies.

Neural networks (NNs) are commonly used in the literature for credit rating prediction. NNs were found to be significantly more accurate than traditional statistical methods in previous studies (e.g. [6]). Hajek [26] compared the performance of a variety of NNs. Radial basis function neural networks (RBF) and probabilistic neural networks (PNNs) significantly outperformed methods such as multilayer perceptron (MLP), group method of data handling (GMDH), MDA, and LR. Because of high generalization ability, support vector machines (SVMs) also produced good results in terms of classification accuracy (e.g. [32,47]). For a small proportion of labeled companies, kernel-based approaches with semi-supervised learning [29] have provided better results than supervised learning methods.

Fuzzy logic based classifiers, such as adaptive fuzzy rule based systems (AFRBs), fuzzy decision trees (FDTs) and the Wang–Mendel algorithm have been employed by Hajek [28]. The main advantage of these classifiers is the fact that the model obtained can be interpreted in terms of membership functions and fuzzy if-then rules. However, the model can be very complex in the case of credit ratings. As a result, hundreds of fuzzy if-then rules have to be generated in order to obtain an accurate prediction model.

Other AI methods used for credit rating prediction include artificial immune systems (AISs) (e.g. [13]), case-based reasoning (CBR) (e.g. [43,65,47]), evolutionary algorithms [5], and ant colony optimization [54].

Table 1 summarizes the literature on corporate credit rating prediction. In all presented studies, data used in the tests were obtained from US companies. Rating classes were provided by two rating agencies: S&P's or Moody's. Nevertheless, it would be inappropriate to compare these studies among themeselves as they are based on different datasets (the companies included might not be the same and data were obtained from different time periods, and thus they describe companies operating in different macroeconomic conditions).

In addition to studies focused on US data, some studies have explored data from other countries and rating agencies. The assessments of Korean or Japanese rating agencies have been used in the following studies. Methods such as NNs [47], SVMs [1,2], CBR [65], Bayesian networks [72,10], and hybrid methods combining AI methods [43,4,23] have been used in these studies. Furthermore, credit ratings of sub-sovereign and municipal entities have been studied recently (e.g. [20,27]).

Considering the process of feature selection, statistical tests (one-way analysis of variance [ANOVA], Kruskal–Wallis test), factor analysis, and stepwise procedure have been used previously in various combinations [32,65,43,47,42,58]. Huang et al. [32] used one-way ANOVA to find statistically significant input variables for two datasets: Korean and US. In the case of the US dataset, 14 out of 19 features were selected with a $P < 0.1$. In particular, liquidity ratios did not have a significant impact with regard to the credit rating decision. Shin and Han [65] applied a two-stage feature selection process. At the first stage, 27 variables were selected using factor analysis, one-way ANOVA, and Kruskal–Wallis test (for qualitative variables). In the second stage, they used a stepwise procedure of MDA to reduce the dimensionality to 12 final input variables. A wide range of variables' categories was included in the resulting set of variables. In a similar manner, Kim and Han [43] performed one-way ANOVA and Kruskal–Wallis in the first stage and factor analysis with stepwise procedure of CBR in the second and third stage, respectively. Thus, the original set of 129

Table 1
List of prior studies on corporate credit rating prediction.

Study	Feature selection	q	m	n	Method: CA_{test} [%]
Shin and Han [65]	ANOVA + stepwise	5	3886	168	MDA: 60.0, CBR: 62.0, GA-CBR: 75.5
Kim and Han [43]	ANOVA + stepwise	5	2971	329	LVQ + CBR: 69.1, SOM + CBR: 67.1, CBR: 61.1, MDA: 55.0
Brennan and Brabazon [6]	–	2	600	8	MLP: 84.0
Brennan and Brabazon [6]	–	5	791	8	MLP: 52.7
Delahunty and OCallaghan [13]	–	2	791	8	AIS: 72.5
Huang et al. [32]	ANOVA	5	265	5	MLP: 80.0, SVM: 78.9
Huang et al. [32]	ANOVA	5	265	12	MLP: 79.3, SVM: 80.0
Kim [44]	–	4	1080	26	ALN: 83.8
Barbazon and O'Neill [5]	–	2	791	8	GE: 84.9, MLP: 83.3, MDA: 85.2
Cao et al. [8]	–	6	237	17	SVM: 84.6, FFNN: 80.3, LR: 77.9
Lee [47]	ANOVA + stepwise	5	3017	297	SVM: 67.2, CBR: 63.4, FFNN: 59.9, MDA: 58.8
Hwang et al. [36]	–	3	736	24	OLR: 72.8
Hajek [25]	CBF	2	852	6	PNN: 88.5, RBF: 85.6, SVM: 87.4
Hajek [26]	CBF	9	852	11	PNN: 58.5, RBF: 58.3, SVM: 55.6
Hwang et al. [37]	–	3	779	4	OPM: 76.0, OSPM: 81.1
Kim and Ahn [42]	Stepwise	5	1295	14	SVM + OPP: 68.0, SVM: 67.3, FFNN: 65.7
Hajek and Olej [29]	CBF	2	1021	6	GCM: 83.7, HGM: 85.8
Hajek [28]	CBF	9	852	11	AFRBS: 59.6
Yeh et al. [71]	RF	3	2470	18	RS: 93.4, DT: 84.0, SVM: 74.4

q is the number of rating classes, n is the number of input variables, m is the number of objects in data set, CA_{test} [%] is classification accuracy on testing data set, ALN is adaptive learning network, LVQ is learning vector quantization NN, RBES is rule based expert system, PNN is probabilistic NN, AFRB is adaptive fuzzy rule based system, OPM is ordered probit model, OSPM is ordered semiparametric probit model, GCM is global consistency model (refferred results hold for 30% of labeled data), HGM is harmonic Gaussian model (refferred results hold for 60% of labeled data), RS are rough sets, OPP is ordinal pairwise partitioning, CBF is correlation-based filter, GR is gain ratio, IG is information gain, RF is random forest, DT is decision tree.

input variables was significantly reduced to only 13 variables. Lee [47] used a combination of one-way ANOVA and stepwise procedure of MDA to reduce the initial 297 financial ratios to the final set of 10 concerning leverage ratios in particular.

Huang [35] predicted the credit ratings of Taiwanese companies using an integration of a nonlinear graph-based dimensionality reduction scheme with SVMs. This feature extraction method provided a significant improvement of classification accuracy achieved by SVMs compared to the traditional feature extraction methods (principal component analysis and independent component analysis) and recursive feature elimination method, respectively.

Recently, it has been demonstrated that the classification performance of AI methods can be significantly improved using a correlation-based filter [30] for corporate credit ratings [26,28]. Here, the optimum set of input variables was constructed considering the correlations between the input variables and rating classes. In addition, the impact of inter-correlated input variables was limited.

Yeh et al. [71] used random forests to select the 18 most important features which caused the highest mean decrease in classification accuracy when removed.

In related works dedicated to bankruptcy prediction, several methods for feature selection have also been applied. For example, Tsai [68] performed a comparison of the following feature selection methods: correlation matrix, t-test, factor analysis, principal component analysis, and stepwise procedure. MLP was used as a classifier. The t-test method provided the best results with regard to classification accuracy, while stepwise procedure provided the highest feature reduction rate.

3. Datasets

Rating agencies do not reveal the parameters used to perform credit rating assessment. In order to have the best possible description of the status of companies that are to be rated, we use as many measures of economic and financial performance as possible. In particular, we include the parameters that have been used by other authors in previous studies.

Our datasets contain information on 852 US and 244 European non-financial companies, respectively. In the European dataset, only companies from the EU states were represented, and specifically the UK with 52 companies, Germany with 35, France with 30, and others. We handled the two dataset separately because financial ratios of companies outside the US are not directly comparable with statistics published for US industries. There are several reasons for this approach. The companies outside the US differ in the treatment of goodwill, asset valuation practices, contingent liabilities reporting, accounting techniques, and other factors. We did not consider country risk, which is thought to be important primarily for emerging markets.

For each variable, we used the mean values calculated over the years 2006–2008. The reason of using a three year average follows a process known as "rating through the cycle", which is adopted by rating agencies to achieve rating stability and to minimize the business cycle effect [3,64,19]. This longer-term perspective is usually implemented by considering the three-year averages of relevant financial ratios. Rating stability has been confirmed empirically by Mizen and Tsoukas [57]. In the study presented by these authors, credit ratings showed a high autocorrelation and previous years' ratings were a key input variable when predicting current credit ratings.

We used ratings assigned by the S&P's rating agency in the year 2009 as target classes. Rating agencies take into account many diverse national considerations, but express their ratings on a single scale. This allows the debt-holders to compare issues of equivalent credit quality.

Parameters used in our study for describing companies can be divided into two main categories: business position and financial indicators.

The business position of a company can be described using factors such as industry risk, size, character, management skills, and others. The size of a company can be expressed by measuring market capitalization, assets, equity, cash flow, and others. The ability of a company to pay off its loans is determined, among other factors, by company size. Company size is also correlated with diversification and market power. The character (reputation) of a company is difficult to measure. To some extent this factor can be inferred from information about insiders' and institutional holdings. Industry risk represents the sensitivity of companies in a

Table 2
Overview of the input variables used for corporate credit rating modeling in prior studies.

Study	Input variables
Delahunty and OCallaghan [13]	Current ratio, retained earnings/total assets, interest coverage, total debts/total assets, net margin, market to book value, total assets, return on total assets
Brennan and Brabazon [6]	Current ratio, retained earnings/total assets, interest coverage, total debts/total assets, net margin, market to book value, total assets, return on total assets
Huang et al. [32]	Total assets, total liabilities, long-term debts/total capital, total debts/total assets, operating margin, return on equity
Kim [44]	Total assets, current ratio, return on total assets, total debts/total assets, sales/fixed assets, operating margin, interest coverage, long-term debts/total capital, cash flow/current liabilities
Barbazon and O'Neill [5]	Current ratio, retained earnings/total assets, interest coverage, total debts/total assets, net margin, market to book value, total assets, return on total assets
Hwang et al. [36]	KMV-Merton default probability, market equity value, earnings, total assets, total debt/(earnings before interest and taxes increased by depreciation and amortization), total assets/equity, long-term debts/total capital, short-term debts/total capital, interest coverage, (earnings before interest and taxes increased by depreciation and amortization)/interest, cash flow, interest, net income, return on capital, return on equity, return on total assets, operating margin, retained earnings/total assets, current ratio, quick ratio, cash ratio
Hajek and Olej [29]	Size class, SIC code, market debt/total capital, high/low stock price, correlation of stock returns with market index, dividend yield
Hajek [28]	Size class, market capitalization, shares held by mutual funds, effective tax rate, fixed assets/total assets, intangible assets/total assets, market debt/total capital, beta regression coefficient, high/low stock price, correlation of stock returns with market index, dividend yield

particular industry or market to external business factors, such as macroeconomic changes. Finally, reputation and industry risk have only been involved to a very small extent in previous studies.

Financial indicators are the second important category of factors taken into account in the corporate credit rating process. Financial indicators can further be divided into several subcategories: profitability ratios, activity ratios, liquidity ratios, leverage ratios, and market value ratios.

Profitability ratios measure the influence of asset management, financing, and liquidity on the profit of a company. Profitability ratios used by other authors (e.g. [36]) include the absolute size of profit, the effect of return on total assets, return on equity, return on sales, operating margin, and net margin.

Activity ratios measure the effectiveness of asset management. However, the influence of asset management on the credit rating of a company is rather indirect, because asset management belongs to common financial decision-making areas. Consequently, only sales to net worth and fixed assets were used in previous studies (e.g. [44]).

Most authors used current ratio as a representative of liquidity ratios. There are also other parameters, such as quick ratio and cash ratio but these have been rarely used in the literature.

Leverage ratios (indebtedness) have been represented by total debts to total assets and long-term debts to total assets, respectively, in the literature (e.g. [6]). The assessment of the capability of a company to pay off debt from the generated profit was also an important factor in credit rating prediction. This aspect can be measured using different approaches, such as with the interest coverage parameter.

Market value ratios reflect how the past activity of a company and its future outlook are perceived by the market. The impact of market value ratios on corporate credit rating was demonstrated by Hajek [28] and Hajek and Olej [29]. In these studies, the correlation of stock returns with market index, high/low stock price, and dividend yield were regarded as important factors of corporate rating process.

Table 2 shows the list of input variables used in previous studies. Since our dataset only covers US and European companies, we did not report input variables used in predicting Korean and Japanese credit ratings in this table. In general, most authors used the size of a company as well as its profitability, liquidity, and leverage ratios as input variables to credit rating prediction.

The choice of input variables for our experiments is presented in Table 3. The US dataset contained 81 input variables obtained from the Value Line Database and S&P's database, while the European dataset covers only those input variables marked in italics (i.e., 43 input variables extracted from the Bloomberg and Capital IQ databases). The input variables are divided into 9 categories: size of a company, corporate reputation, profitability ratios, activity ratios, asset structure, business situation, liquidity ratios, leverage ratios, and market value ratios. In this paper we have performed a feature selection step to select a subset of parameters from the sets presented in Table 3. Companies are classified into 9 output rating classes $\omega \in \Omega$, $\Omega = \{\text{AAA}, \text{AA}, \ldots, \text{D}\}$ (Fig. 1).

We performed a principal component analysis to closely assess the nature of the data. Only 14 principal components with eigenvalues greater than 1 were extracted both from the original set of 81 variables (US dataset) and 43 variables (European dataset), respectively. The first principal component explained 45.12% and 15.38% of the total variance, respectively (see the Pareto charts in Figs. 2 and 3). For the US dataset, it represents the input variables from several categories that correlated with the size of companies. The second component shows the capital market position of a company. For the European companies, the first component can be labeled as the capital market position, the second as the size of companies, and so forth. We can conclude that there are several significantly intercorrelated variables in the dataset, which could have a strong impact on the results of the feature selection process [12].

4. Feature selection

The set of features should be reduced before they are used for classification. Performing this step not only makes the calculations faster and the parameters easier to collect, but also may improve classification accuracy [73]. The benefits of feature selection are as follows: the dimensionality of the feature space is reduced, learning algorithms operate faster and classification accuracy can be improved. Recently, several feature selection have been proposed which are based on computational intelligence methods such as genetic algorithms [15,48], rough sets [9], memetic algorithms [41], SVMs [52,53], or hybrid algorithms [39,22].

The variety of feature selection methods is usually divided into filters, wrappers, and embedded feature selection methods [40]. Filters perform feature selection based on the characteristics of data itself (e.g., by employing statistical measures). Filters operate independently of any learning algorithm by estimating the usefulness of features using an evaluation function [63]. Features that are not expected to provide valuable information for classification are filtered out of the dataset before classification starts. This process can be performed, among the others, using correlation ranking [17], a two-sample t-test [66], penalized pseudo-likelihood [7,18], and mutual information [16].

Embedded methods measure the importance of features while building a model, and therefore the feature selection step is an inherent part of the training process (e.g. [69]).

Wrappers use some type of enumeration algorithm to explore the space of feature subsets. The enumeration algorithm may be

Table 3
Input variables used for credit rating prediction.

Size of company		Business situation	
TA	Total assets	*ETR*	*Effective tax rate*
TC	*Total capital*	S gr	Growth in sales last year
S	*Sales (last year)*	S exp	Expect. growth in S (next 5 years)
TS	*12-Mth trailing sales*	SGAE	SG&A expenditures
CF	*Cash flow*	Liquidity ratios	
E	Equity	CR	Current ratio
EV	Enterprise value	CaR	Cash ratio
FV	*Firm value*	*Cash/FV*	*Cash to firm value*
CE	*Capital expenditures*	*Cash*	*Cash*
SC	Size class	NCWC	Non-cash working capital
MC	*Market capitalization*	Leverage ratios	
TV	Trading volume	*BV/E*	*Book value to equity*
NS	No. of shares outstanding	*BD/TC*	*Book debt to total capital*
Corporate reputation		*EV/TC*	*Enterprise value to total capital*
IH	*Shares held by mutual funds*	EV/BV	Enterprise value to book value
InH	Shares held by insiders	MC/TD	Market capitalization to total debt
Profitability ratios		*TD*	*Total debt*
EBIT	Earnings before interest and taxes	*CF/TD*	*Cash flow to total debt*
EAT	*Earnings after taxes*	*MD/E*	*Market debt to equity*
NI	*Net income*	*MD/TC*	*Market debt to total capital*
TNI	12-Mth trailing NI	NG	Net gearing
NM	*Net margin*	*MD/EBITDA*[a]	*Market debt to EBITDA*
OM	*Operating margin*	Market value ratios	
ROA	Return on total assets	*P var*	*3-Year stock price variation*
ROE	*Return on equity*	*Beta*	*Beta regression coefficient (3 year)*
ROC	*Return on capital*	VLB	Value line beta
EBITDA	*EBIT increased by depreciation and amortization*	*Cor*	*The correlation of stock returns with market index*
EV/EBITDA			*Enterprise value to EBITDA*
HiLo	*High/low stock price*		
EV/EBIT	*Enterprise value to EBIT*	Div	Dividends
RE/TA	Retained earnings to total assets	*Div/P*	*Dividends to stock price*
Activity ratios		EPS	Earnings per share
EV/S	*Enterprise value to sales*	EPS gr	Growth in earnings per share (last 5 years)
NCWC gr	*Growth in NCWC*	*EPS exp*	*Expected growth in earnings per share (next 5 years)*
EV/TS	Enterprise value to trailing sales	P/CF	Stock price to cash flow
S/NW	Sales to net worth	*P/E*	*Stock price to earnings*
S/TA	Sales to total assets	TP/E	12-Mth trailing stock price to earnings
OR/TA	Operating revenue to total asses	FP/E	Forward stock price to earnings
WC/S	Working capital to sales	*PEG*	*Stock price to earnings to EPS growth*
Cash/S	Cash to sales	*PBV*	*Price to book value ratio*
NCWC/S	Non-cash working capital to sales	RE	Retained earnings
Asset structure		*RR*	*Reinvestment rate*
FA/TA	Fixed assets to total assets	*PR*	*Payout ratio*
IA/TA	Intangible assets to total assets	*PS*	*Stock price to sales*
WC/TA	Working capital to total assets	*P*	*Stock price*
Dep	Depreciation		

[a] This input variable was not used in the US dataset.

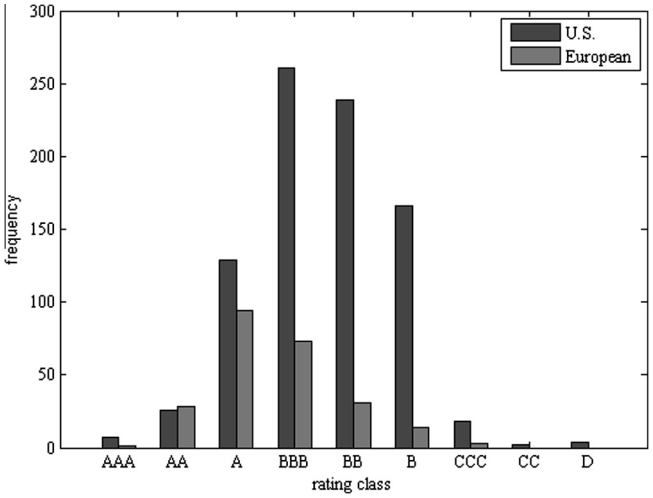

Fig. 1. Frequencies of companies in rating classes.

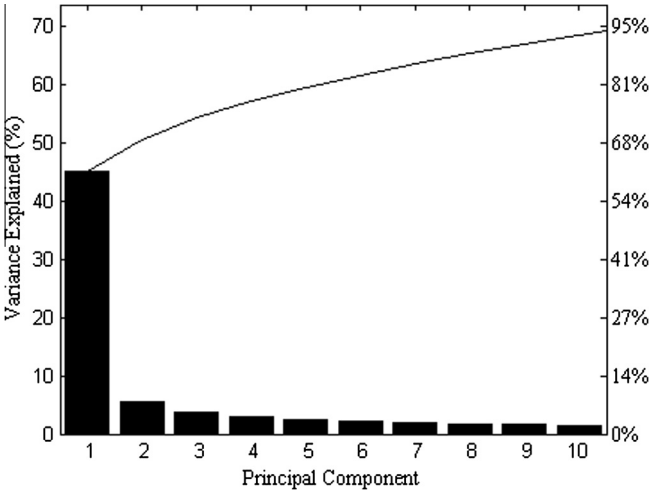

Fig. 2. Variance explained using principal component analysis – US dataset.

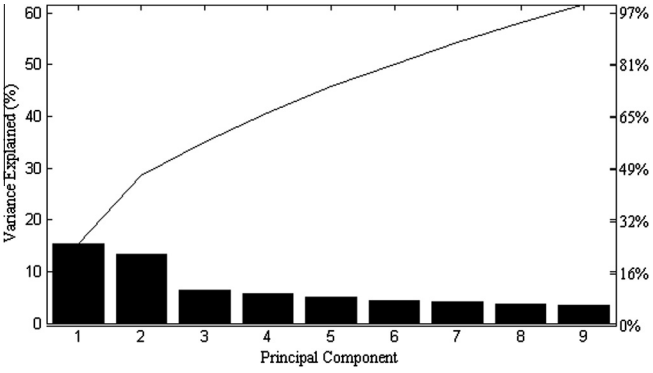

Fig. 3. Variance explained using principal component analysis – European dataset.

a simple sequential search [14], but more sophisticated methods such as genetic algorithms [34,59,70] or Particle Swarm Optimization are also used [51]. Performance of the classification algorithm is tested on each enumerated feature set and the performance measure is used as evaluation of the subset quality. This approach may be slow since the classifier must be retrained on all candidate subsets of the feature set and its performance must also be measured. On the other hand, wrappers may produce better results than filters, because specific requirements of the classifier can be taken into account. In comparative studies, wrappers have

performed better than filters when the sample size was sufficiently large [33].

Wrapper feature selection requires enumerating feature subsets, because in general, the number of all possible feature subsets is very large. Therefore, it is necessary to employ a search procedure that only iterates over a portion of all of the possible subsets.

Since the number of possible subsets of the feature set is too large to evaluate all of them in most cases, iterated search is most often employed. For example, a method proposed by Kittler [45] is based on the following step-wise procedure. The procedure starts with an empty set of features. In the next steps the feature set is expanded using the variables that provide the largest improvement in classification accuracy. The procedure is stopped when the addition of more variables provides no improvement in classification accuracy.

If features are selected one-by-one as it is done in the method described above, then it is possible that important dependencies between features are not taken into account. Therefore, multivariate approaches are sometimes employed.

Evaluating features in pairs has been proposed in the literature [61,31] in order to take into account at least some of the dependencies between features. The pairwise approach results in quadratic complexity with respect to the number of features, which is clearly far better than evaluating all possible feature sets.

Quadratic complexity is still quite high when the number of features is large. Michalak and Kwasnicka [55,56] proposed a mixed feature selection (MFS) method that evaluates features individually or in pairs. The decision on how to evaluate features is made based on a quantitative criterion that measures the level of dependence between features. As only a fraction of features is evaluated in pairs the complexity of the MFS method is lower than the complexity of the pairwise method.

The overview of the Mixed Feature Selection Method is as follows:

$F_0 = \varnothing$
$i = 1$
while $|F_{i-1}| < n_{\max}$ do
  if $|F_{i-1}| < n_{\max} - 1$ then
    F_i = SelectMixed(F_{i-1})
  else
    add one, the best feature to F_{i-1}
  end if
  $i = i + 1$
end while

In the above algorithm, i is the iteration number, $n_{\max}$ is the number of features required, and F_i are iteratively constructed feature sets. The SelectMixed procedure performs each feature selection iteration using evaluation function δ to decide whether to evaluate each feature individually or in pairs with all other features. The SelectMixed procedure is implemented as follows:

$Q_{\max} = 0$
for each $f_k \notin F_{i-1}$
  if $\exists f_p \cdot \delta(f_k, f_p) > \theta$ then
    Q_{test} = the quality of the best feature set $F_{i-1} \cup \{f_k, f_q\}$,
   where $f_k \notin F_{i-1}$
    $F_{\text{test}} = \{f_k, f_q\}$
  else
    Q_{test} = the quality of the feature set $F_{i-1} \cup \{f_k\}$
    $F_{\text{test}} = \{f_k\}$
  end if
  if $Q_{\text{test}} > Q_{\max}$
    $F_i = F_{i-1} \cup F_{\text{test}}$
  end if
end for

In the above procedure δ is an evaluation function that numerically represents the level of dependence between features. If the estimated level of dependence between a given feature f_k and at least one other feature f_p exceeds a predefined threshold θ, then the feature f_k is evaluated in pairs with all (not yet selected) features and from that evaluation step the best pair $\{f_k, f_q\}$ is selected.

In this paper, feature selection was performed using individual feature selection (IFS) proposed by Kittler [45] and the MFS method proposed by Michalak and Kwasnicka [55,56]. An absolute value of correlation coefficient was used for the evaluation function ($\delta = |\rho_{k,p}|$). The threshold was set to $\theta = 0.5$. Furthermore, we compared the performance of the presented wrapper approaches with three filter methods: correlation-based filter (CORF), which was developed by Hall [30], consistency-based filter (CONF), which was proposed by Liu and Setiono [49], and genetic algorithm filter (GAF) [50].

In the CORF method the quality of a set of variables is measured based on the individual predictive ability of each feature together with the degree of redundancy between them. The objective function $f(\lambda)$, based on Pearson's correlation coefficient, can be expressed as:

$$f(\lambda) = \frac{\lambda \times \zeta_{cr}}{\sqrt{\lambda + \lambda \times (\lambda - 1) \times \zeta_{rr}}}, \tag{1}$$

where λ is the subset of features, ζ_{cr} is the average feature to class correlation, and ζ_{rr} is the average feature to feature correlation. The numerator in Eq. (1) expresses the predictive power of a set of features λ. The denominator represents the degree of redundancy among the features in λ. Irrelevant features have low correlation with the class, and therefore they receive a low evaluation. Redundant features are discriminated against because they are usually highly correlated with one or more of the other features, and therefore produce high values of the denominator. The correlations between features ζ_{cr} and ζ_{rr} are computed from the discretized values of the features. Symmetrical uncertainty (SU) is used as a correlation measure to estimate the degree of association between discrete features (X and Y):

$$\text{SU} = 2.0 \times \left[\frac{H(X) + H(Y) - H(X,Y)}{H(X) + H(Y)}\right]. \tag{2}$$

After computing a correlation matrix, CORF applies a heuristic search strategy to find a good subset of features according to Eq. (1).

The CONF uses an inconsistency criterion that specifies to what extent the dimensionality of reduced data can be accepted. The inconsistency rate γ of selected features is checked against a prespecified rate $\gamma_{\max}$. The inconsistency rate is calculated based on the count of inconsistent instances (matching instances except for their class labels). The CONF algorithm can be written as follows:

for $i = 1$ to MAX-TRIES
  F_i = random feature set
  n_i = number of features in F_i
  $n_{\max}$ = number of features
  n_{best} = the best number of features
  if $n_i < n_{\max}$
    if $\gamma < \gamma_{\max}$
      $F_{\text{best}} = F_i$
      $n_{\text{best}} = n_i$

(continued on next page)

else if ($n_i = n_{best}$) and ($\gamma < \gamma_{max}$)
end if
end for

The GAF was originally developed with the aim of choosing the neurons in the hidden layer of a hybrid NN [50]. The objective function $\theta = V - P$ of this method is based on both the mutual information between feature and class $I(X,Y)$ and the mutual information between features $I(X_i,X_j)$:

$$I(X_i, Y) = H(X_i) + H(Y) - H(X_i, Y), \quad (3)$$

$$I(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j). \quad (4)$$

$$V = \frac{1}{n}\sum_{i=1}^{n} I(X_i, Y), \quad (5)$$

$$P = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(X_i, X_j). \quad (6)$$

The GAF algorithm can be expressed as follows:

Input:

N_{pop} = population size (the number of random feature sets F_i)
N_{gen} = number of generations
n_i = number of features in F_i
n_{max} = number of features
α = selective pressure
$I(X,Y)$, $I(X_i,X_j)$ = mutual information

Output:

{i} – selected feature set

Algorithm:

generate initial population
for $j = 1$ to N_{gen}
for $i = 1$ to N_{pop}
$\theta_i = V_i - P_i$
end for
rank the individuals according to their fitness θ_i
store the genes of the best individual in $\{i\}$
$k = 0$
for $j = 1$ to N_{pop}
$k = k + 1$
ϑ_m = random number from [0,1] with uniform distribution, $m = (1,2)$
$parent_m = round(N_{pop}(e^{\alpha\vartheta m} - 1)/(e^{\alpha} - 1))$
store the indexes which are absentees in both parents in $\{i_{abs}\}$
for $i = 1$ to n_i
randomly select a parent ($parent_1$ or $parent_2$) to give the ith gene for the kth individual of the new generation
if there is a duplication of indexes then pick up a new index from $\{i_{abs}\}$
end if
end for
end for

5. Experimental results

We used supervised learning to train various classifiers and perform credit rating prediction. The experiments were conducted using the following classifiers: Multilayer Perceptron (MLP) neural network, Radial Basis Function (RBF) neural network, Support Vector Machines (SVM), Naive Bayes (NB), Random Forest (RF), Linear Discriminant Classifier (LDC), and Nearest Mean Classifier (NMC).

The U.S. and European datasets consisted of 852 and 244 vectors, respectively, representing companies described by 81 parameters (43 parameters) introduced in Section 3. The missing values of each feature were replaced by the median value calculated using existing values of the same feature. The data were standardized to mean zero and unit standard deviation. To avoid overfitting, 60% of the data were used for training the classifier and 40% were treated as unseen data and used for testing. Rating classes $\omega \in \Omega$, $\Omega = \{AAA, AA, \ldots, D\}$ assigned by the S&P's rating agency were known for all companies, so it was possible to train the classifier and calculate the classification accuracy on the test set.

The MLP neural network classifier used in the experiments had the following parameters. The number of input neurons was equal to the number n of selected features. The models were tested for different numbers h of hidden layer neurons $h = \{5, 10, 15\}$. The neurons used logistic activation function. The learning process was performed using scaled conjugate gradient algorithm with the learning rate of 0.1, momentum of 0.2, and the number of epochs of 2000.

The RBF neural network classifier used in the experiments had the following parameters. The number of neurons in the hidden layer was modified in the range $h = \{q \times 2, q \times 3, q \times 4\}$, i.e. 2, 3 and 4 neurons for each of the $q = 9$ classes. The radius of the RBF was set to 0.2.

The training of SVM was performed using the sequential minimal optimization algorithm proposed by Platt [62]. Kernel functions of the SVM were represented by RBFs with parameter gamma set to 0.3. The complexity parameter C was chosen from the range $C = \{1.0, 2.0, 3.0\}$.

The remaining classifiers were tested with the following parameters. The Naive Bayes classifier used the normal distribution estimator for numeric features. In the case of Random Forest, 10 classification trees were generated in each experiment.

The NMC does not require any parameters. It uses one separate mean vector for each of the classes, and one variance estimation for all of the classes and prior class probabilities. All of these parameters are estimated from the training sample during the classifier learning process.

The LDC also does not require any parameters. Apart from mean vectors and prior probabilities of the classes it uses a single covariance matrix for all of the classes. The covariance matrix is estimated as a mean of covariance matrices for all classes.

For each classifier and each feature selection method 10 runs of the entire training–testing cycle were completed. Classification accuracy CA_{test} [%], Type I error rate [%] [defined as (1 – specificity) × 100], and misclassification cost (MCC) on the test set were recorded in each of the 10 iterations. Type I error rate (a false acceptance of the rating class) was calculated as a weighted average of all classes. Similarly, it is also possible to calculate Type II error rates (false rejections of the correct rating class), which can be expressed as $100 - CA_{test}$ [%] in multiclass problems. Both Type I and Type II error rates have been considered important measures of credit rating prediction models [11]. In addition, we calculated MCC to take into account the fact that the rating classes are ordered from the best to the worst, so it is a more serious mistake to classify a D-class company as AAA than to give the same AAA

Table 4
Misclassification cost matrix.

Rating class	Predicted					
	AAA	AA	A	...	D	
Actual	AAA	0	1	2	...	8
	AA	1	0	1	...	7
	A	2	1	0	...	6
	...	...	...	...	...	...
	D	8	7	6	...	0

Table 5
Experimental results – classification accuracies, type I errors, misclassification costs, and numbers of selected features for the US dataset.

Feature selection	MLP	RBF	SVM	NB	RF	LDC	NMC
None							
CA_{test} [%]	56.16 ± 4.91	47.59 ± 2.57	54.73 ± 1.76	40.71 ± 5.57	**58.64 ± 5.44**	**56.13 ± 2.05**	54.31 ± 1.67
Type I err. [%]	15.04 ± 1.54	17.32 ± 1.11	15.88 ± 0.87	16.00 ± 1.43	**13.90 ± 1.55**	**12.98 ± 1.15**	13.76 ± 1.01
MCC	0.52 ± 0.05	0.64 ± 0.03	0.54 ± 0.02	0.87 ± 0.16	**0.49 ± 0.02**	**0.58 ± 0.03**	0.60 ± 0.01
n	81	81	81	81	81	81	81
IFS							
CA_{test} [%]	**58.24 ± 3.13**	**56.43 ± 2.43**	**56.10 ± 4.83**	**43.79 ± 1.73**	**58.22 ± 3.77**	**54.78 ± 2.11**	50.67 ± 3.56
Type I err. [%]	**12.73 ± 0.84**	**13.99 ± 1.86**	**14.75 ± 1.85**	20.25 ± 1.86	14.69 ± 1.40	**13.54 ± 0.96**	16.21 ± 1.92
MCC	0.58 ± 0.05	0.57 ± 0.04	0.54 ± 0.05	**0.67 ± 0.04**	**0.49 ± 0.07**	0.60 ± 0.02	0.68 ± 0.05
n	14.20 ± 4.84	23.00 ± 3.08	7.75 ± 5.27	3.40 ± 1.37	13.00 ± 7.81	69.00 ± 4.85	13.20 ± 7.67
MFS							
CA_{test} [%]	**57.45 ± 3.07**	54.21 ± 3.75	**57.98 ± 4.65**	**44.08 ± 1.54**	**59.39 ± 4.81**	**55.66 ± 2.64**	46.86 ± 3.03
Type I err. [%]	14.82 ± 0.71	15.29 ± 1.69	**14.77 ± 1.25**	20.47 ± 1.71	**13.68 ± 1.56**	**13.18 ± 0.94**	18.42 ± 1.61
MCC	**0.50 ± 0.04**	0.60 ± 0.04	**0.50 ± 0.04**	**0.66 ± 0.03**	**0.47 ± 0.04**	0.60 ± 0.03	0.70 ± 0.02
n	9.33 ± 4.11	9.25 ± 2.95	9.00 ± 7.84	2.00 ± 0.00	12.25 ± 6.42	65.60 ± 5.87	9.20 ± 6.05
CORF							
CA_{test} [%]	**56.66 ± 2.24**	54.43 ± 3.18	53.61 ± 3.42	39.94 ± 3.61	**58.42 ± 1.27**	51.61 ± 4.90	54.13 ± 2.11
Type I err. [%]	14.67 ± 1.53	**13.78 ± 1.10**	16.56 ± 1.74	**14.51 ± 1.82**	**13.39 ± 1.09**	14.30 ± 1.28	13.82 ± 1.22
MCC	**0.51 ± 0.02**	0.58 ± 0.03	0.54 ± 0.04	1.02 ± 0.13	**0.47 ± 0.02**	0.62 ± 0.09	**0.58 ± 0.03**
n	23.00 ± 1.79	23.00 ± 1.79	23.00 ± 1.79	23.00 ± 1.79	23.00 ± 1.79	23.00 ± 1.79	23.00 ± 1.79
CONF							
CA_{test} [%]	51.26 ± 2.71	48.94 ± 2.78	38.36 ± 9.21	38.30 ± 3.15	55.60 ± 3.14	45.28 ± 3.44	46.74 ± 4.10
Type I err. [%]	17.64 ± 0.37	18.25 ± 1.94	22.20 ± 3.50	19.03 ± 1.48	14.46 ± 1.10	18.60 ± 1.85	18.13 ± 1.95
MCC	0.58 ± 0.04	0.63 ± 0.04	0.70 ± 0.12	1.03 ± 0.11	0.52 ± 0.03	0.70 ± 0.06	0.68 ± 0.05
n	14.80 ± 1.72	14.80 ± 1.72	14.80 ± 1.72	14.80 ± 1.72	14.80 ± 1.72	14.80 ± 1.72	14.80 ± 1.72
GAF							
CA_{test} [%]	**57.51 ± 2.68**	**55.98 ± 2.63**	**56.30 ± 1.40**	42.79 ± 4.04	57.51 ± 2.15	**55.72 ± 2.28**	**56.42 ± 1.66**
Type I err. [%]	14.28 ± 1.31	**14.26 ± 1.27**	15.49 ± 0.82	**14.74 ± 2.43**	**13.89 ± 0.70**	**13.05 ± 1.47**	**13.04 ± 0.94**
MCC	**0.49 ± 0.03**	**0.52 ± 0.03**	**0.50 ± 0.01**	0.87 ± 0.11	0.50 ± 0.03	**0.56 ± 0.04**	**0.55 ± 0.02**
n	18.60 ± 2.20	18.60 ± 2.20	18.60 ± 2.20	18.60 ± 2.20	18.60 ± 2.20	18.60 ± 2.20	18.60 ± 2.20

Table 6
Experimental results – classification accuracies, type I errors, misclassification costs, and numbers of selected features for the European dataset.

Feature selection	MLP	RBF	SVM	NB	RF	LDC	NMC
None							
CA_{test} [%]	54.18 ± 4.02	50.51 ± 2.55	53.37 ± 2.81	32.27 ± 16.65	50.51 ± 2.34	32.65 ± 21.33	47.14 ± 3.73
Type I err. [%]	18.43 ± 2.54	19.42 ± 2.30	21.17 ± 1.16	20.08 ± 2.63	20.65 ± 1.14	**13.91 ± 4.33**	**17.35 ± 2.60**
MCC	0.59 ± 0.06	0.63 ± 0.03	0.56 ± 0.04	0.67 ± 0.15	0.63 ± 0.04	1.26 ± 0.71	0.81 ± 0.12
n	43	43	43	43	43	43	43
IFS							
CA_{test} [%]	**63.06 ± 2.73**	61.33 ± 3.27	**60.20 ± 1.71**	**67.14 ± 2.95**	**56.43 ± 3.85**	**51.33 ± 2.70**	**50.82 ± 4.35**
Type I err. [%]	**15.66 ± 1.86**	**15.37 ± 1.56**	**18.40 ± 2.06**	**13.34 ± 1.28**	**18.43 ± 1.85**	16.85 ± 1.62	19.70 ± 2.34
MCC	**0.47 ± 0.03**	**0.48 ± 0.06**	**0.50 ± 0.03**	**0.40 ± 0.05**	**0.55 ± 0.06**	**0.67 ± 0.05**	**0.67 ± 0.08**
n	19.70 ± 5.87	16.50 ± 8.82	24.20 ± 7.65	13.70 ± 5.83	11.80 ± 5.10	37.40 ± 4.40	24.00 ± 5.35
MFS							
CA_{test} [%]	**62.35 ± 3.08**	**64.39 ± 3.61**	**60.82 ± 1.84**	**68.78 ± 2.42**	53.57 ± 2.05	**50.71 ± 3.42**	**49.59 ± 3.71**
Type I err. [%]	**16.07 ± 1.73**	**14.61 ± 1.95**	**18.45 ± 2.09**	**12.38 ± 1.43**	**18.87 ± 1.20**	17.40 ± 2.09	20.71 ± 2.88
MCC	**0.49 ± 0.06**	**0.45 ± 0.05**	**0.48 ± 0.03**	**0.39 ± 0.04**	**0.59 ± 0.05**	**0.67 ± 0.04**	**0.67 ± 0.07**
n	18.30 ± 5.59	13.00 ± 3.55	20.70 ± 8.56	14.30 ± 5.18	14.20 ± 6.54	33.70 ± 5.62	23.10 ± 6.17
CORF							
CA_{test} [%]	48.67 ± 5.68	48.06 ± 3.64	48.47 ± 6.66	36.63 ± 4.53	50.41 ± 2.38	45.20 ± 15.64	**51.02 ± 5.22**
Type I err. [%]	24.52 ± 3.96	20.60 ± 1.89	26.45 ± 4.24	17.40 ± 1.97	19.83 ± 1.28	22.06 ± 5.89	23.02 ± 3.92
MCC	0.68 ± 0.10	0.67 ± 0.07	0.70 ± 0.13	0.85 ± 0.08	0.63 ± 0.03	0.85 ± 0.52	**0.64 ± 0.07**
n	7.90 ± 1.45	7.90 ± 1.45	7.90 ± 1.45	7.90 ± 1.45	7.90 ± 1.45	7.90 ± 1.45	7.90 ± 1.45
CONF							
CA_{test} [%]	49.18 ± 6.20	48.67 ± 3.39	51.23 ± 6.94	35.41 ± 5.88	49.39 ± 4.59	34.69 ± 22.34	48.47 ± 7.25
Type I err. [%]	23.49 ± 4.89	19.04 ± 2.24	24.15 ± 4.83	18.59 ± 2.06	20.49 ± 1.20	18.63 ± 7.62	23.43 ± 4.49
MCC	0.67 ± 0.11	0.72 ± 0.09	0.63 ± 0.12	0.84 ± 0.10	0.63 ± 0.07	1.17 ± 0.75	**0.65 ± 0.09**
N	12.10 ± 1.64	12.10 ± 1.64	12.10 ± 1.64	12.10 ± 1.64	12.10 ± 1.64	12.10 ± 1.64	12.10 ± 1.64
GAF							
CA_{test} [%]	51.94 ± 3.83	49.08 ± 3.43	53.57 ± 2.34	43.16 ± 8.45	50.51 ± 3.85	49.39 ± 3.57	**49.80 ± 2.91**
Type I err. [%]	19.74 ± 3.19	18.93 ± 2.78	21.94 ± 1.82	17.21 ± 3.64	20.38 ± 2.04	18.99 ± 2.35	19.03 ± 1.76
MCC	0.64 ± 0.05	0.69 ± 0.08	0.59 ± 0.05	0.74 ± 0.13	0.63 ± 0.07	**0.69 ± 0.08**	**0.66 ± 0.07**
N	15.40 ± 4.74	15.40 ± 4.74	15.40 ± 4.74	15.40 ± 4.74	15.40 ± 4.74	15.40 ± 4.74	15.40 ± 4.74

Table 7
Features appearing in at least 40% of the 10 sets using wrappers – US dataset.

Method	IFS	MFS
MLP	Size of company (NS, TC, TA) Activ. ratios (S/NW, NCWC gr) Asset structure (IA/TA) Liquid. ratios (Cash, NCWC) Lever. ratios (BD/TC, CF/TD) Market value ratios (P, EPS, PEG, HiLo, PR)	Size of company (SC) Asset structure (FA/TA, WC/TA) Lever. ratios (MD/TC) Market value ratios (P)
RBF	Size of company (TV, CE) Profit. ratios (EV/EBITDA, NM, OM) Activ. ratios (EV/S, EV/TS, NCWC/S, Cash/S, S/NW, S/TA, WC/S) Lever. ratios (BD/TC) Market value ratios (PS, Beta, VLB, HiLo)	Size of company (MC, TC) Corporate reputation (InH) Profit. ratios (EV/EBITDA) Liquid. ratios (Cash, NCWC) Lever. ratios (BD/TC) Market value ratios (P, RR)
SVM	Size of company (SC) Profit. ratios (ROA)	Size of company (SC, MC, EV) Asset structure (IA/TA) Market value ratios (Cor, PS, Div, P)
NB	Size of company (SC) Asset structure (IA/TA)	Lever. ratios (BD/TC) Market value ratios (Div, Div/P)
RF	Size of company (CF) Corporate reputation (InH) Profit. ratios (EBIT, RE/TA) Activ. ratios (NCWC gr, NCWC/S) Asset structure (IA/TA) Business situation (S exp) Lever. ratios (BV/E) Market value ratios (PR, RR, Div, Div/P)	Size of company (MC, E) Corporate reputation (InH) Profit. ratios (EV/EBITDA) Liquid. ratios (Cash, NCWC) Market value ratios (Div, P, RR)
LDC	Size of company (TA, TC, S, TS, CF, E, FV, CE, SC, MC, TV, NS) Corporate reputation (IH, InH) Profit. ratios (EBIT, EAT, NI, TNI, NM, OM, ROA, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT, RE/TA) Activ. ratios (EV/S, NCWC gr, EV/TS, S/NW, S/TA, OR/TA, Cash/S, NCWC/S) Asset structure (FA/TA, IA/TA, WC/TA, Dep) Business situation (ETR, S gr, S exp, SGAE) Liquid. ratios (CR, CaR, Cash/FV, Cash, NCWC) Lever. ratios (BV/E, BD/TC, EV/TC, EV/BV, MC/TD, TD, CF/TD, MD/E, MD/TC, NG) Market value ratios (P var, Beta, VLB, Cor, HiLo, Div, Div/P, EPS, EPS gr, EPS exp, P/CF, P/E, TP/E, FP/E, PEG, PBV, RE, RR PR, PS, P)	Size of company (TA, TS, CF, E, EV, FV, CE, SC, MC, TV, NS) Corporate reputation (InH) Profit. ratios (EBIT, EAT, NI, TNI, NM, OM, ROA, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT, RE/TA) Activ. ratios (EV/S, NCWC gr, EV/TS, S/NW, S/TA, OR/TA, Cash/S, NCWC/S) Asset structure (FA/TA, IA/TA, WC/TA) Business situation (ETR, S gr, S exp, SGAE) Liquid. ratios(CR, CaR, Cash/FV, Cash, NCWC) Lever. ratios (BV/E, BD/TC, EV/TC, MC/TD, TD, CF/TD, MD/E, MD/TC) Market value ratios (P var, Beta, VLB, Cor, HiLo, Div, Div/P, EPS, EPS gr, EPS exp, P/CF, P/E, TP/E, FP/E, PEG, PBV, RE, RR PR, PS, P)
NMC	Size of company (TC, TS, CF, E, EV, FV, CE, SC, MC, TV, NS) Corporate reputation (InH) Profit. ratios (EBIT, EAT, NI, TNI, NM, OM, ROA, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT, RE/TA) Activ. ratios (EV/S, NCWC gr, EV/TS, S/NW, S/TA, WC/S, Cash/S, NCWC/S) Asset structure (FA/TA, IA/TA, WC/TA, Dep) Business situation (ETR, S gr, S exp, SGAE) Liquid. ratios (CR, CaR, Cash/FV, Cash, NCWC) Lever. ratios (BV/E, EV/TC, EV/BV, MC/TD, TD, CF/TD, MD/E, MD/TC, NG) Market value ratios (P var, VLB, Cor, HiLo, Div, Div/P, EPS, P/CF, P/E, TP/E, FP/E, PEG, PBV, RE, RR PR, PS, P)	Size of company (CF, SC, MC, TV, NS) Corporate reputation (InH) Profit. ratios (NI, TNI, NM, OM, ROA, ROE, ROC, EV/EBITDA, EV/EBIT, RE/TA) Activ. ratios (EV/S, NCWC gr, EV/TS, S/NW, WC/S, Cash/S, NCWC/S) Asset structure (FA/TA, IA/TA, WC/TA, Dep) Business situation (S gr) Liquid. ratios (CR, CaR, Cash/FV, Cash, NCWC) Lever. ratios (BV/E, EV/TC, EV/BV, MC/TD, TD, CF/TD, MD/E, NG) Market value ratios (P var, Div, EPS, P/CF, P/E, TP/E, FP/E, PEG, PBV, RE, RR PR, PS, P)

rating to an AA-class company. The corresponding cost matrix is highlighted in Table 4. The greater the difference between actual and predicted class is, the higher the MCC.

Tables 5 and 6 summarize the results of the experiments. In these tables, average values from the 10 runs are presented along with standard deviations. For each classification method, the best and statistically similar results (at $P < 0.05$ using a paired t-test) are separately marked in bold. The average numbers of selected features n (with standard deviations) for the 10 runs are also provided.

For the CORF and CONF, we applied the best first procedure that searches the space of feature subsets by greedy hillclimbing augmented with a backtracking facility. The level of backtracking was controlled by setting the number of consecutive non-improving nodes allowed to 5. We used forward search in both filters. For the GAF, we used population size $N_{pop} = 200$, maximum number of generations $N_{gen} = 8$, and selective pressure $\alpha = 6$. Besides that, the GAF requires the setting of the target number of selected features. We examined this number n over the range $n = \{1, 2, \ldots, n_{max}\}$. The final number of features n was selected based on the shape of the objective function θ. In the area behind n there was no significant decrease in the objective function θ.

In the case of the US dataset, Random Forest classifier employing the MFS method outperformed the rest of the methods. For this learning scheme, 12.25 ± 6.42 features were selected (out of 81 original features; 15.1% on average). For the European dataset, the NB classifier with 14.30 ± 5.18 features (33.3% of all features) selected by the MFS worked best. The wrapper feature selection strategies improved the classification accuracies for the remaining classification methods in both datasets. For the MLP, the IFS worked best for both datasets with 17.5% and 45.8% of all features, respectively. The wrappers significantly improved the classification performance of the RBF and SVM as well.

In the case of the US dataset, the reason for a worse performance of the NB classifier lies in the nature of the data. In addition to mutual correlations between features, we also tested the correlations between features and predicted class. Only weak correlations were identified, and all were less than 0.1. As the NB classifier is a linear separator unable to represent disjunction or conjunction, it performs worse in cases when the correlated

Table 8
Features appearing in at least 40% of the 10 sets using wrappers – European dataset.

Method	IFS	MFS
MLP	Size of company (TC, S, TS, FV, MC) Profit. ratios (NI, OM, ROC, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S) Business situation (ETR) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, BD/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, Div/P, P/E, PEG, PBV, PS, P)	Size of company (TC, S, TS, FV, MC) Corporate reputation (IH) Profit. ratios (NI, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S) Liquid. ratios (Cash) Lever. ratios (BD/TC, EV/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, P/E, PS, P)
RBF	Size of company (TC, CF, FV, MC) Corporate reputation (IH) Profit. ratios (EAT, NM, OM, ROE) Activ. ratios (EV/S) Liquid. ratios (Cash) Lever. ratios (BD/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, P/E, RR, PS)	Size of company (TC, FV, MC) Profit. ratios (ROE) Activ. ratios (EV/S) Liquid. ratios (Cash) Lever. ratios (TD, MD/E, MD/EBITDA) Market value ratios (P var, Beta, Div/P, P/E)
SVM	Size of company (S, TS, FV, MC) Profit. ratios (NI, OM, ROC, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S, NCWC gr) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, BD/TC, EV/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, EPS exp, P/E, PEG, PBV, RR, PS, P)	Size of company (TS, FV, MC) Profit. ratios (EAT, NI, NM, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S) Business situation (ETR) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, BD/TC, EV/TC, TD, CF/TD, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, Div/P, EPS exp, P/E, PEG, RR, PS, P)
NB	Corporate reputation (IH) Profit. ratios (NM, OM, EV/EBITDA, ROC) Activ. ratios (EV/S) Lever. ratios (BD/TC, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, PS, Div/P)	Corporate reputation (IH) Profit. ratios (EV/EBITDA, ROC) Activ. ratios (EV/S) Business situation (ETR) Lever. ratios (BD/TC, MD/TC, TD, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, P/E)
RF	Size of company (S, MC, FV) Profit. ratios (EAT) Activ. ratios (NCWC gr) Liquid. ratios (Cash/FV) Lever. ratios (CF/TD, MD/TC, MD/E, MD/EBITDA) Market value ratios (P var, PEG, PR)	Size of company (S, TS, MC, FV) Profit. ratios (EAT, NM, EBITDA, EV/EBITDA, EV/EBIT) Liquid. ratios (Cash) Lever. ratios (MD/TC, MD/EBITDA) Market value ratios (P var, HiLo, P/E, PEG, PS)
LDC	Size of company (TC, S, TS, CF, FV, CE, MC) Corporate reputation (IH) Profit. ratios (EAT, NI, NM, OM, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S, NCWC gr) Business situation (ETR) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, BD/TC, EV/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, Div/P, EPS exp, P/E, PEG, PBV, RR, PR, PS, P)	Size of company (TC, S, TS, CF, FV, CE, MC) Corporate reputation (IH) Profit. ratios (EAT, NI, NM, OM, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S, NCWC gr) Business situation (ETR) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, BD/TC, EV/TC, TD, CF/TD, MD/E, MD/TC, MD/EBITDA) Market value ratios (P var, Beta, Cor, HiLo, Div/P, EPS exp, P/E, PEG, PBV, RR, PR, PS, P)
NMC	Size of company (CF, FV, CE, MC) Profit. ratios (NI, NM, ROE, ROC, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S, NCWC gr) Liquid. ratios (Cash/FV, Cash) Lever. ratios (BV/E, TD, CF/TD, MD/E, MD/EBITDA) Market value ratios (P var, Beta, HiLo, Div/P, EPS exp, P/E, PEG, PBV, RR, PR, PS, P)	Size of company (TC, CF, FV, CE, MC) Corporate reputation (IH) Profit. ratios (EAT, NI, NM, OM, ROE, ROC, EBITDA, EV/EBITDA, EV/EBIT) Activ. ratios (EV/S) Liquid. ratios (Cash/FV, Cash) Lever. ratios (EV/TC, TD, CF/TD, MD/E, MD/EBITDA) Market value ratios (P var, Beta, Cor, Div/P, EPS exp, P/E, PBV, RR, PR, PS, P)

feature is not included. This classifier achieved a much better performance with the European dataset, where the correlations between features (especially profitability and leverage ratios) and predicted class were statistically significant.

High correlations between the features themselves and the results of the Kruskal–Wallis ANOVA test also imply that many features are redundant and irrelevant when applied alone, respectively. It is the suitable combination (subset) of features that enables the classifiers to perform well on our datasets. However, as pointed out by Guyon and Elisseeff [24], adding features that are presumably redundant may help to reduce noise and consequently to obtain a better class separation. In addition, a feature that is completely useless when used individually may provide a significant performance improvement when used together with other features. In this respect, filters have a disadvantage, since they are unable to include irrelevant features that may actually help performance [46]. Concerning the filters, the GAF (with 23.0% of all features on average) generally outperformed both the CORF (28.4% of all features) and the CONF (18.3% of all features) for both datasets. In the case of the European dataset, the filters provided worse classification results in comparison with the wrappers. Nevertheless, only 18.4% (CORF) to 35.8% (GAF) of all features was selected using the filters, in contrast to wrappers: 45.6% (MFS) and 48.9% (IFS) of all features on average.

In addition to the measures of prediction accuracy, we recorded computation time as well. In our experiments, the feature selection using filters was completed in less than 0.5 s for both datasets. Moreover, the LDC and NMC were trained in less than 0.2 s. In contrast, feature selection using wrappers was slower. It took more than 1.4 s for the IFS to perform feature selection using both the LDC and the NMC in the case of European data. For the MFS, it was 13.8 s (LDC) and 23.1 s (NMC), respectively. In the analysis of US data, it took even more time to select the features. For instance, feature selection using the MFS method required 591 s and 261 s when used with the LDC and the NMC, respectively.

For most of the classifiers, the wrappers selected less consistent sets of features (with higher standard deviations). In Tables 7–9, we provide an overview of the most frequently used features for

Table 9
Features appearing in at least 40% of the 10 sets using filters.

Method	Features
US data	
CORF	Size of company (SC, MC, FV, EV, CF) Profit. ratios (ROE, ROA, NI, TNI) Business situation (S gr, ETR, SGAE) Liquidity ratios (Cash) Leverage ratios (BV/E, MD/E, MD/TC, BD/TC, MC/TD) Market value ratios (P, EPS, PS, TP/E, HiLo, P var, Cor, PR, Div, P/CF)
CONF	Size of company (MC, FV, EV, TV) Profit. ratios (NI) Business situation (S gr) Liquidity ratios (Cash, NCWC) Leverage ratios (MD/E) Market value ratios (P, EPS, EPS gr, Beta, VLB, TP/E, P var, PR, Div)
GAF	Size of company (SC, MC) Profit. ratios (NI, TNI) Activ. ratios (Cash/S) Business situation (ETR) Liquidity ratios (Cash, NCWC) Leverage ratios (MD/E, MD/TC, CF/TD) Market value ratios (P, TP/E, EPS, HiLo, P dev, Cor, Div, PEG)
European data	
CORF	Size of company (MC, FV, TC) Profit. ratios (EAT) Liquidity ratios (Cash) Leverage ratios (BV/E, MD/TC) Market value ratios (PS)
CONF	Size of company (MC, FV, TC, S, TS) Profit. ratios (EAT, NI) Activ. ratios (EV/S) Liquidity ratios (Cash) Leverage ratios (BV/E, MD/TC, TD, MD/EBITDA) Market value ratios (PS)
GAF	Size of company (FV, TC) Corporate reputation (IH) Profit. ratios (EV/EBIT, ROC, NM) Activ. ratios (EV/S) Liquidity ratios (Cash) Leverage ratios (MD/EBITDA, FCFF/MD, MD/E, EV/TC, BV/E) Market value ratios (P, Cor, P var, P/E, PS, PEG, PR, RR)

all of the feature selection methods. Only those features that occurred in at least 4 runs of 10 were reported in these tables. For the wrapper methods, the results show that different learning algorithms perform better with different core features, even if using the same training set, which is in agreement with previous findings [46]. There are also only few coincidences among the features selected by the filter and wrapper approaches. The filter approaches select features that are closely related with the class label while the wrappers take the predictive capacity of features into account. Even if considered in the objective function, several highly correlated features were selected using the CORF method (e.g. in the size of company category). This finding indicates another disadvantage of filter methods, as they are unable to remove correlated features that may hurt performance [46].

6. Conclusions

In this study we carried out an automatic corporate credit rating assessment using several classification methods. Because rating agencies do not publicly reveal what parameters are used in the credit rating process, we decided to collect as many parameters describing the companies as possible. The US and European datasets consisted of 81 features and 43 features, respectively, which included various financial and non-financial factors. In the training of the classifiers as well as during testing on previously unseen data, we used credit ratings previously assigned to the companies by a credit rating agency.

To reduce the complexity of the classification process and improve the accuracy of classification process, we performed feature selection prior to classification. The feature selection was done using two different wrapper algorithms. Our approach maintained or even improved the accuracy after removing redundant or irrelevant features that may degrade the classification accuracy. In most of the cases, the classification accuracy was between 45% to 65%, which is quite good for a 9-class problem. Feature selection using IFS and MFS methods improved the classification accuracy in most of the cases. This improvement can lead to a substantial reduction of the cost related to credit risk management. As pointed out by Tong et al. [67], with sizable loan portfolios, even a slight improvement in the accuracy of credit evaluations can reduce the creditors' risk and can also be translated into considerable savings in the future.

Feature selection using wrapper methods seems to be a very useful preprocessing step for an automatic corporate credit rating assessment performed using classification methods. The number of features required for classification has been significantly reduced, and in most of the cases, was reduced to less than 25 features. Thus, we conclude, that a relatively small enumeration of features decides the rating class assignment. In agreement with Huang et al. [32], this occurrence appears consistently in various financial markets. Since, in general, it is unknown what parameters are used by credit rating companies, the usual approach is to include as many parameters as possible. Feature selection makes it possible to choose the parameters that should actually be used. Additionally, lowering the dataset dimensionality may improve classification accuracy, because the influence of the so called "curse of dimensionality" is reduced. It is worth noting that the best classification accuracies (59.39% for the Random Forest classifier – the US companies, and 68.78% for the NB classifier – the European companies) were achieved for a feature set reduced using the MFS method and not for a set of all of the available features.

The filter methods were outperformed by the wrapper methods in this specific problem. The reason for this might be the ability of the wrappers to make use of both correlated and uncorrelated features for a particular classifier. There are several problems with wrappers that have been described in the literature, especially overfitting and the large amount of CPU time required. We did not observe problems with overfitting, since it mainly concerns small training sets. In our experiments the wrappers tended to produce smaller feature subsets with higher classification accuracies. Although the wrapper approaches exhibited more accurate behavior than the filter methods, this improvement comes at a price of a high computational load.

Regarding the selected features with the wrapper approaches, it was confirmed that different algorithms have different biases and the optimal subsets of features differed greatly.

Our results suggest that both the size of a company and market value ratios were the most important parameters in the US rating methodology. In contrast, the rating process of the European companies to a large extent relied on the profitability and leverage ratios. To the best of our knowledge, in the only other market comparative study, Huang et al. [32] demonstrated that the size of banks is the most important determinant of U.S. bond raters, whereas Taiwan raters focus more on profitability.

The use of feature selection algorithms seems to be a promising approach in corporate credit rating prediction. The algorithms allow for a reduction in problems arising from the fact that credit rating agencies do not publicly reveal the parameters they use to assess the credibility of companies. Further work in this area may include using feature selection methods specialized for particular classifiers (such as salience-based methods for NNs). We also

envision the use of wrappers and filters while automatically fixing the number of features. This study should also encourage the expansion of the wrapper approaches to related business domains, such as bankruptcy prediction, stock price trend prediction, and credit card fraud detection. In our work, we selected a 1-year prediction horizon. In view of the fact that this horizon may vary from short to long, appropriate accuracy measures, such as Harrell's *C* [60], could be developed and incorporated into the AI classifiers.

Acknowledgments

We gratefully acknowledge the help provided by constructive comments of the anonymous referees. This work was supported by a grant provided by the Ministry of Interior of the Czech Republic No. VF20112015018 and by the scientific research project of the Czech Sciences Foundation Grant No: 13-10331S.

References

[1] H. Ahn, K.J. Kim, Combining pairwise SVM classifiers for bond rating, in: KMIS International Conference, 2005, pp. 586–590.
[2] H. Ahn, K.J. Kim, Corporate credit rating using multiclass classification models with order information, World Academy of Science, Engineering and Technology 60 (2011) 95–100.
[3] E.I. Altman, H.A. Rijken, The Effects of Rating Through the Cycle on Rating Stability, Rating Timeliness and Default Prediction Performance, Working paper, 2005.
[4] J.K. Bae, J. Kim, Combining models from neural networks and inductive learning algorithms, Expert Systems with Applications 38 (5) (2011) 4839–4850.
[5] A. Brabazon, M. O'Neill, Credit classification using grammatical evolution, Informatica (Ljubljana) 30 (3) (2006) 325–335.
[6] D. Brennan, A. Brabazon, Corporate bond rating using neural networks, International Conference on Artificial Intelligence IC-AI'04 1 (2004) 161–167.
[7] E. Candes, T. Tao, The Dantzig selector: statistical estimation when p is much larger than n (with discussion), Annals of Statistics 35 (6) (2007) 2313–2404.
[8] L. Cao, L.K. Guan, Z. Jingquing, Bond rating using support vector machine, Intelligent Data Analysis 10 (3) (2006) 285–296.
[9] Y. Chen, D. Miao, R. Wang, A rough set approach to feature selection based on ant colony optimization, Pattern Recognition Letters 31 (3) (2010) 226–233.
[10] D.Y. Choi, K.C. Lee, N.Y. Jo, Approach to corporate credit grading prediction using Bayesian Network model, Information 14 (9) (2011) 3143–3153.
[11] J. Cornaggia, K.J. Cornaggia, Does the Bond Market Want Informative Credit Ratings? Working paper, 2011.
[12] S. Das, Filters, wrappers and a boosting-based hybrid for feature selection, in: The 18th International Conference on Machine Learning, 2001, pp. 74–81.
[13] A. Delahunty, D. OCallaghan, Artificial Immune Systems for the Prediction of Corporate Failure and Classification of Corporate Bond Ratings, University College Dublin, Dublin, 2004.
[14] K. Dunne, P. Cunningham, F. Azuaje, Solutions to Instability Problems with Sequential Wrapper-Based Approaches to Feature Selection, Technical Report TCD-CD-2002-28, Department of Computer Science Courses, Trinity College, Dublin, 2002.
[15] M.E. ElAlami, A filter model for feature subset selection based on genetic algorithm, Knowledge-Based Systems 22 (5) (2009) 356–362.
[16] P.A. Estévez, M. Tesmer, C.A. Perez, J.M. Zurada, Normalized mutual information feature selection, IEEE Transactions on Neural Networks 20 (2) (2009) 189–201.
[17] J. Fan, J. Lv, Sure independence screening for ultra-high dimensional feature space (with discussion), Journal of Royal Statistical Society, Series B 70 (2008) 849–911.
[18] J. Fan, R. Samworth, Y. Wu, Ultrahigh dimensional feature selection: Beyond the linear model, Journal of Machine Learning Research 10 (2009) 2013–2038.
[19] F. Fei, A.M. Fuertes, E. Kalotychou, Credit rating migration risk and business cycles, Journal of Business Finance and Accounting 39 (1–2) (2012) 229–263.
[20] N. Gaillard, The determinants of Moody's sub-sovereign ratings, International Research Journal of Finance and Economics 31 (1) (2009) 194–209.
[21] A.P.M. Gama, H.S.A. Geraldes, Credit risk assessment and the impact of the New Basel Capital Accord on small and medium-sized enterprises: an empirical analysis, Management Research Review 35 (8) (2012) 727–749.
[22] I.A. Gheyas, L.S. Smith, Feature subset selection in large dimensionality domains, Pattern Recognition 43 (1) (2010) 5–13.
[23] X. Guo, Z. Zhu, J. Shi, A corporate credit rating model using support vector domain combined with fuzzy clustering algorithm, Mathematical Problems in Engineering (2012) 1–20.
[24] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.
[25] P. Hajek, Credit Rating Modelling by Neural Networks, Nova Science, New York, 2010.
[26] P. Hajek, Probabilistic neural networks for credit rating modelling, in: Proceedings of the International Conference on Fuzzy Computation and International Conference on Neural Computing, vol. 1, Valencia, 2010, pp. 289–294.
[27] P. Hajek, Municipal credit rating modelling by neural networks, Decision Support Systems 51 (1) (2011) 108–118.
[28] P. Hajek, Credit rating analysis using adaptive fuzzy rule-based systems: an industry-specific approach, Central European Journal of Operations Research 20 (3) (2012) 421–434.
[29] P. Hajek, V. Olej, Credit rating modelling by kernel-based approaches with supervised and semi-supervised learning, Neural Computing & Applications 20 (6) (2011) 761–773.
[30] M.A. Hall, Correlation-Based Feature Subset Selection for Machine Learning, Universtiy of Waikato, Hamilton, 1998.
[31] A. Harol, C. Lai, E. Pekalska, R.P.W. Duin, Pairwise feature evaluation for constructing reduced representations, Pattern Analysis and Applications 10 (1) (2007) 55–68.
[32] Z. Huang, H. Chen, C.J. Hsu, W.H. Chen, S. Wu, Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Systems 37 (4) (2004) 543–558.
[33] J. Hua, W.D. Tembe, E.R. Dougherty, Performance of feature-selection methods in the classification of high-dimension data, Pattern Recognition 42 (3) (2009) 409–424.
[34] Ch.-L. Huang, Ch.-J. Wang, A GA-based feature selection and parameters optimization for support vector machines, Expert Systems with Applications 31 (2) (2006) 231–240.
[35] S.Ch. Huang, Integrating nonlinear graph based dimensionality reduction schemes with SVM for credit rating forecasting, Expert Systems with Applications 36 (4) (2009) 7515–7518.
[36] R.C. Hwang, K.F. Cheng, C.F. Lee, On multiple-class prediction of issuer credit ratings, Applied Stochastic Models in Business and Industry 25 (5) (2009) 535–550.
[37] R.C. Hwang, H. Chung, C.K. Chu, Predicting issuer credit ratings using a semiparametric method, Journal of Empirical Finance 17 (1) (2010) 120–137.
[38] R.C. Hwang, Forecasting credit ratings with the varying-coefficient model, Quantitative Finance (ahead-of-print) (2013) 1–19.
[39] R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Transactions on Fuzzy Systems 17 (4) (2009) 824–838.
[40] G.H. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in: W.W. Cohen, H. Hirsh (Eds.), Machine Learning, Morgan Kauffman, 1994. vol. 129, pp. 121–129..
[41] S.S. Kamar, N. Ramaraj, A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm, Knowledge-Based Systems 23 (6) (2010) 580–585.
[42] K. Kim, H. Ahn, A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach, Computers & Operations Research 39 (8) (2012) 1800–1811.
[43] K.S. Kim, I. Han, The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases, Expert Systems with Applications 21 (3) (2001) 147–156.
[44] K.S. Kim, Predicting bond ratings using publicly available information, Expert Systems with Applications 29 (1) (2005) 75–81.
[45] J. Kittler, Feature set search algorithms, in: C.H. Chen (Ed.), Pattern Recognition and Signal Processing, Sijthoff and Noordhoff, Alphen aan den Rijn, 1978, pp. 41–60.
[46] R. Kohavi, G. John, Wrappers for feature subset selection, Artificial Intelligence 97 (12) (1997) 273–324.
[47] Y.C. Lee, Application of support vector machines to corporate credit rating prediction, Expert Systems with Applications 33 (1) (2007) 67–74.
[48] J. Leng, C. Valli, L. Armstrong, A wrapper-based feature selection for analysis of large data sets, in: Proceedings of the 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010), 2010, pp. 167–170.
[49] H. Liu, R. Setiono, A probabilistic approach to feature selection – a filter solution, in: Proceedings of the 13th International Conference on Machine Learning ICML'1996, Bari, 1996, pp. 319–327.
[50] O. Ludwig, U. Nunes, Novel maximum-margin training algorithms for supervised neural networks, IEEE Transactions on Neural Networks 21 (6) (2010) 972–984.
[51] M. Macas, L. Lhotska, E. Bakstein, D. Novak, J. Wild, T. Sieger, P. Vostatek, R. Jech, Wrapper feature selection for small sample size data driven by complete error estimates, Computer Methods and Programs in Biomedicine 108 (1) (2012) 138–150.
[52] S. Maldonado, R. Weber, A wrapper method for feature selection using support vector machines, Information Science 179 (13) (2009) 2208–2217.
[53] S. Maldonado, R. Weber, J. Basak, Simultaneous feature selection and classification using kernel-penalized support vector machines, Information Sciences 181 (1) (2011) 115–128.
[54] D. Martens, T. Van Gestel, M. De Backer, R. Haesen, J. Vanthienen, B. Baesens, Credit rating prediction using ant colony optimization, Journal of the Operational Research Society 61 (4) (2009) 561–573.
[55] K. Michalak, H. Kwasnicka, Correlation-based feature selection strategy in classification problems, International Journal of Applied Mathematics and Computer Science 16 (4) (2006) 503–511.
[56] K. Michalak, H. Kwasnicka, Correlation based feature selection method, International Journal of Bio-Inspired Computation 2 (5) (2010) 319–332.

[57] P. Mizen, S. Tsoukas, Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model, International Journal of Forecasting 28 (1) (2012) 273–287.
[58] H. Ogut, H.M. Doganay, N.B. Ceylan, R. Aktas, Prediction of bank financial strength ratings: the case of Turkey, Economic Modelling 29 (3) (2012) 632–640.
[59] A.L.I. Oliveira, P.L. Braga, R.M.F. Lima, M.L. Cornelio, GA-based method for feature selection and parameters optimization for machine learning regression applied to software effort estimation, Information and Software Technology 52 (11) (2010) 1155–1166.
[60] W. Orth, The predictive accuracy of credit ratings: measurement and statistical inference, International Journal of Forecasting 28 (1) (2012) 288–329.
[61] E. Pekalska, A. Harol, C. Lai, R.P.W. Duin, Pairwise selection of features and prototypes, in: Proceedings of 4th International Conference on Computer Recognition Systems CORES'05, Advances in Soft Computing, Springer, 2005, pp. 271–278.
[62] J. Platt, Fast training of support vector machines using sequential minimal optimization, Advances in Kernel Methods – Support Vector Learning, MIT Press, 1999. pp. 185–208..
[63] N. Sánchez-Marono, A. Alonso-Betanzos, M. Tombilla-Sanromán, Filter methods for feature selection – a comparative study, in: H. Yin, P. Tino, E. Corchado, W.M. Byrne, X. Yao (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2007, Lecture Notes in Computer Science, vol. 4881, 2007, pp. 178–187.
[64] Ch.H. Shen, Y.L. Huang, I. Hasan, Assymetric Benchmarking in Bank Credit Rating, Bank of Finland Research Discussion Papers 13, 2012.
[65] K.S. Shin, I. Han, A case-based approach using inductive indexing for corporate bond rating, Decisions Support Systems 32 (1) (2001) 41–52.
[66] R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Class prediction by nearest shrunken centroids, with applications to DNA microarrays, Statistical Science 18 (1) (2003) 104–117.
[67] L.I. Tong, Ch.H. Yang, H.P. Yu, Credit rating analysis using general regression neural network, in: Proceedings of the 2nd International Conference on Innovations of Management and Interdisciplinary, 2005, pp. 1–11.
[68] Ch.F. Tsai, Feature selection in bankruptcy prediction, Knowledge-Based Systems 22 (2009) 120–127.
[69] E. Tuv, A. Borisov, G. Runger, K. Torkkola, Feature selection with ensembles, artificial variables, and redundancy elimination, Journal of Machine Learning Research 10 (2009) 1341–1366.
[70] L.D. Vignolo, D.H. Milone, J. Scharcanski, C. Behaine, An evolutionary wrapper for feature selection in face recognition applications, in: Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2012), 2012, pp. 1286–1290.
[71] Ch.Ch. Yeh, F. Lin, Ch.Y. Hsu, A hybrid KMV model, random forests and rough set theory approach for credit rating, Knowledge-Based Systems 33 (2012) 166–172.
[72] P. Wijayatunga, S. Mase, M. Nakamura, Appraisal of companies with Bayesian networks, International Journal of Business Intelligence and Data Mining 1 (3) (2006) 329–346.
[73] B.K. Wong, V.S. Lai, J. Lam, A bibliography of neural network business applications research: 1994–1998, Computers and Operations Research 27 (11–12) (2000) 1045–1076.