

STATISTICS IN TRANSITION new series, June 2016
Vol. 17, No. 2, pp. 265–280

HETEROSCEDASTIC DISCRIMINANT ANALYSIS COMBINED WITH FEATURE SELECTION FOR CREDIT SCORING

Katarzyna Stapor¹, Tomasz Smolarczyk², Piotr Fabian³

ABSTRACT

Credit granting is a fundamental question and one of the most complex tasks that every credit institution is faced with. Typically, credit scoring databases are often large and characterized by redundant and irrelevant features. An effective classification model will objectively help managers instead of intuitive experience. This study proposes an approach for building a credit scoring model based on the combination of heteroscedastic extension (Loog, Duin, 2002) of classical Fisher Linear Discriminant Analysis (Fisher, 1936, Krzyśko, 1990) and a feature selection algorithm that retains sufficient information for classification purpose. We have tested five feature subset selection algorithms: two filters and three wrappers. To evaluate the accuracy of the proposed credit scoring model and to compare it with the existing approaches we have used the German credit data set from the study (Chen, Li, 2010). The results of our study suggest that the proposed hybrid approach is an effective and promising method for building credit scoring models.

Key words: heteroscedastic discriminant analysis, feature subset selection, variable importance, credit scoring model.

1. Introduction

Credit scoring models are the basis for financial institutions like retail and consumer credit banks. The purpose of these models is to evaluate the likelihood of credit defaulting applicants in order to decide whether to grant them credit. The set of decision models and their underlying methods that serve lenders in granting consumer credits are called credit scoring (CS) (Zhang et al. 2010).

¹ Institute of Computer Science, Silesian University of Technology, Gliwice, Poland.
E-mail: katarzyna.stapor@polsl.pl.

² Institute of Computer Science, Silesian University of Technology, Gliwice, Poland.
E-mail: tomasmo356@student.polsl.pl.

³ Institute of Computer Science, Silesian University of Technology, Gliwice, Poland.
E-mail: piotr.fabian@polsl.pl.

Since customer demand for personal loans has increased in the last decades, the consumer credit market evolved to become an important sector in the financial field and today represents a high-volume business. These developments in the retail credit market requires automatic, fast and consistent decisions and processes to handle the huge amount of applications. The use of credit scoring models is now a key component in retail banking. The development of the so-called scorecards therefore represents the core competence of a retail bank's risk management when assessing the creditworthiness of an individual. Since the market is changing rapidly, new statistical and mathematical methods are required to optimize the scoring problem to decide on the question of whom to offer credit to.

Discriminant analysis, linear regression, logistic regression, neural networks, k-nearest neighbours, support vector machines and classification trees cover the range of different surveys on CS models (Thomas et al., 2005). An overview of publications is given in Thomas (2000) and Crook et al. (2007).

Many credit scoring models have been widely developed by reducing redundant features through feature selection to improve the accuracy of credit scoring models during the past few years. The detailed survey of the existing methods for feature selection is given in (Dash, 1997), for example. A feature subset selection algorithm can be divided into two categories: the filter approach and the wrapper approach (Dash, 1997). The filter relies on various measures like distance, information, dependency on feature evaluation which are then used for their ranking. The wrapper model usually uses the predictive accuracy of the pre-determined learning algorithm to determine the goodness of the selected feature subsets.

The use of feature selection in the construction of credit scoring models has already been reported, for example in (Chen, Li, 2010, Somol, 2005), but there are no references on the selection of variables for their use in discriminant analysis in building credit scoring models. In (Chen, Li, 2010), classical Fisher Discriminant Analysis (**FDA**) (Fisher, 1936, Fukunaga, 1990, Krzyśko, 1990) is used, but with all the input features to generate discriminators for their use in SVM classifier. No feature selection is applied to this model.

This work proposes a new method for constructing a credit scoring model which is based on the feature selection in Heteroscedastic Discriminant Analysis (**HDA**) (Loog, Duin, 2002). HDA is the extension of FDA for dealing with the case of unequal covariance matrices in populations, the situation that occurs very often in practice and in our experiments, too.

In our experiments, for the evaluation of the accuracy of our proposed credit scoring model, we have used the German credit data set, the same that was used in the (Chen, Li, 2010) study. Using classical FDA for feature extraction, we have obtained very poor results (prediction accuracy defined as the number of correct classifications divided by the total number of classifications was about 30%), suggesting that probably the covariance matrices in the two classes are not equal. This was the main reason for the usage of heteroscedastic extension of FDA in

our proposed model. Using HDA as feature extraction combined with the input feature subset selection causes the prediction accuracy to improve up to 76%. This proves that the proposed model is better fitted to the data.

The prediction accuracy of the credit scoring model based on FDA from study (Chen, Li, 2010) is the same as in our case (i.e. 75%). However, this accuracy was achieved in (Chen, Li, 2010) by using nonlinear SVM classifier (with Gaussian kernel), making the learning process more complex – one should estimate the parameters of the SVM classifier in the separate validation procedure which requires the additional data set and is a computationally intensive process (the grid method). Moreover, their credit scoring model uses all the input variables which makes its usage less economic and less intuitive for the interpretation. Additionally, the necessity of specifying the values of the parameters of the SVM classifier in the separate validation procedure will cause worse generalizability of the model.

Thus, our proposed credit scoring model, by using feature selection, proper model for feature extraction as well as the simpler classifier, does not have the above mentioned disadvantages of the model from (Chen, Li, 2010) study.

The valuable step in our proposed credit scoring model is the variable importance analysis, a very useful process of analysing attributes in the context of their significance in the discrimination of good and bad credit consumers.

This paper is organized as follows. Section 2 and 3 shortly present the heteroscedastic extension of the classical FDA which is based on the notion of distance directed matrices (Loog, Duin, 2002) and the feature subset selection algorithms used in the construction of our credit scoring model, respectively. Section 4 describes the proposed methodology for building credit scoring models, while section 5 – experimental results together with the variable importance analysis. Section 6 presents the conclusions and suggestions for future research and practice of our new credit scoring model based on the heteroscedastic extension of FDA.

2. Two-class Heteroscedastic Discriminant Analysis

Fisher Discriminant Analysis (FDA) (Fisher 1936; Krzyśko 1990; Fukunaga 1990) is a multivariate technique to classify study instances into groups and/or describe group differences. Discriminant analysis is widely used in many areas such as biomedical studies, banking environment (for credit evaluation), financial management, bankruptcy prediction, marketing, and many others.

There are many formulations of FDA, a typical one for pattern recognition community is given below (according to (Fukunaga, 1990)).

FDA is concerned with the search for a linear transformation that reduces the dimension of a given n -dimensional statistical model to d ($d < n$) dimensions, while maximally preserving the discriminatory information for the several classes

within the model. It determines a linear mapping A , a $d \times n$ matrix A , that maximizes the so-called Fisher criterion J_F :

$$J_F(A) = \text{tr} \left((AS_W A^T)^{-1} (AS_B A^T) \right) \quad (1)$$

Here, $S_B = \sum_{i=1}^c \frac{n_i}{n} (m_i - \bar{m})(m_i - \bar{m})^T$ and $S_W = \sum_{i=1}^c \frac{n_i}{n} S_i$ are the between-class and the average within-class scatter matrices, respectively; c is the number of classes, m_i is the mean vector of class i , n_i is a number of samples in class i ,

$n = \sum_{i=1}^c n_i$, and the estimated overall mean equals $\bar{m} = \sum_{i=1}^c \frac{n_i}{n} m_i$,

$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - m_i)(X_{ij} - m_i)^T$ is the within-class covariance matrix of class i .

Optimizing (1) comes down to determining an eigenvalue decomposition of $S_W^{-1} S_B$, and taking the rows of A equal to d eigenvectors corresponding to d largest eigenvalues (Fukunaga, 1990).

For the two-class case we have:

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \text{ and } S_W = p_1 S_1 + p_2 S_2, \quad p_2 = 1 - p_1.$$

A limitation of FDA is that it merely tries to separate class means as good as possible and it does not take the discriminatory information, which is present in the difference of the covariance matrices, into account. It is incapable of dealing explicitly with heteroscedastic data, i.e., data in which classes do not have equal covariance matrices.

For building our credit scoring model we have used one of the existing heteroscedastic generalizations of the Fisher criterion (1), namely that based on the Chernoff criterion (Loog, Duin, 2002). The heteroscedastic extension in (Loog, Duin, 2002) is based on the notion of Distance Directed Matrices (DDM) which capture not only the difference in means between two classes, but also describe their difference in covariance in a certain way. (Loog, Duin, 2002) proposed DDM based on the Chernoff distance between two probability density functions d_1, d_2 :

$$\partial_c = -\log \int d_1^\alpha(x) d_2^{1-\alpha}(x) dx \quad (2)$$

where $\alpha \in (0,1)$.

Another interesting approach to heteroscedastic linear discriminant analysis can be found in (Krzyśko, Wołyński, 1996), where authors proposed the optimal classification rules based on linear functions which maximize probabilistic distances: the Chernoff or the Morisita or the Kullback-Leibler ones.

For two normally distributed densities, the DDM is a positive semi-definite matrix S_C :

$$S_C = S^{-\frac{1}{2}}(m_1 - m_2)(m_1 - m_2)^T S^{-\frac{1}{2}} + \frac{1}{p_1 p_2} (\log S - p_1 \log S_1 - p_2 \log S_2) \quad (3)$$

where $\alpha = p_1$, $S = p_1 S_1 + p_2 S_2$. The trace of S_C is the Chernoff distance ∂_C between those two densities. Determining transformation A by an eigenvalue decomposition of S_C means that we determine a transform which preserves as much of the Chernoff distance in the lower dimensional space as possible. The heteroscedastic two-class Chernoff criterion J_C is defined as:

$$J_C(A) = \text{tr} \left(\left(A S_W A^T \right)^{-1} A (m_1 - m_2)(m_1 - m_2)^T A^T - A S_W^{-\frac{1}{2}} \frac{p_1 \log \left(S_W^{-\frac{1}{2}} S_1 S_W^{-\frac{1}{2}} \right) + p_2 \log \left(S_W^{-\frac{1}{2}} S_2 S_W^{-\frac{1}{2}} \right)}{p_1 p_2} S_W^{-\frac{1}{2}} A^T \right) \quad (4)$$

This is maximized by determining an eigenvalue decomposition of:

$$S_W^{-1} \left(S_B - S_W^{-\frac{1}{2}} \frac{p_1 \log \left(S_W^{-\frac{1}{2}} S_1 S_W^{-\frac{1}{2}} \right) + p_2 \log \left(S_W^{-\frac{1}{2}} S_2 S_W^{-\frac{1}{2}} \right)}{p_1 p_2} S_W^{-\frac{1}{2}} \right) \quad (5)$$

and taking the rows of the transform A equal to d eigenvectors corresponding to the d largest eigenvalues.

3. Feature subset selection methods

In the proposed methodology for building credit scoring models, we have used five feature selection methods. Three of them are wrapper-based and use different search strategies for finding a suboptimal set of features: the Sequential Floating Forward Search (SFFS) method (Pudil, et al., 1994), the method using Memetic Algorithms (MA) (Moscato, 2002) and the method that utilizes the Greedy Randomised Adaptive Search Procedure (GRASP) (Feo, Resende, 1989). The two filter-based methods use different techniques for scoring individual features which are then used for their ranking and selecting the top best features: Correlation-based Filter Selection (CFS) (Hall, 1997) and Fisher Score (FS) (Duda, Hart, Stork, 2001).

3.1. SFFS

SFFS is an enhanced version of the Sequential Forward Search (SFS) algorithm (Pudil, et al. 1994). Besides adding the most significant feature in each step, SFFS searches for the least significant feature in the current subset and checks whether removing it will result in the increased performance of the classifier. If so – the feature is removed and the algorithm repeats the procedure of searching and removing unnecessary features. The stopping criterion is the number of added features that did not increase the performance (set to 2 in this research).

3.2. GRASP

GRASP constructs solutions – feature subsets - based on the greedy algorithm and the controlled randomization. It starts with an empty initial solution and in each iteration a list of candidate variables with the best performance is generated from which the algorithm selects at random one variable and add it to the current solution. The level of randomization is controlled by the α parameter ($0 \leq \alpha \leq 1$). Each solution is improved by a simple local search procedure in which the current solution is replaced by a better properly defined neighbouring solution. The stopping criterion is defined as the maximum number of iterations – 30 in our study, and parameter α was set to 0.8.

3.3. MA

MA (sometimes called hybrid Genetic Algorithms) are a class of stochastic global search heuristics in which evolutionary algorithm-based approaches (Goldberg, 1989) are combined with problem-specific solvers. The later might be implemented, for example, as a local search heuristics techniques. The hybridization is meant to either accelerate the discovery of good solutions, or to reach solutions that would otherwise be unreachable by evolution or a local method alone. A single solution (a chromosome) is the vector with length equal to the number of all features composed of zeros and ones (zero means that the feature is not present in the subset). During selection phase, 50% of the best chromosomes are selected for later breeding. Population size was set to 30, crossover rate to 0.05, mutation rate to 0.05, and the number of iterations – 20 in our case – was the stopping criteria.

3.4. CFS

In CFS the goodness of a given feature is measured by the degree of association between a feature and a class, and is estimated based on the information theory (Cover, Thomas, 1991) as:

$$\text{symmetrical uncertainty} = 2.0 \times \left[\frac{H(Y) + H(X) - H(X, Y)}{H(Y) + H(X)} \right] \quad (6)$$

where $H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y)$ is the entropy of Y before observing X and

$H(Y | X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y | x) \log_2 p(y | x)$ is the entropy of Y after observing X

(X and Y are discrete random variables). Necessary preprocessing was accomplished as in (Hall, Smith, 1997).

3.5. FS

The FS is a measure of how a given feature is efficient for discrimination. It is defined by between-class and within-class scatter matrices S_B and S_W :

$$\text{FisherScore} = \frac{|S_B|}{|S_W|} \quad (7)$$

where $|\cdot|$ is a determinant. The larger the FisherScore value the more likely for the feature to be discriminative.

4. The proposed methodology for building Credit Scoring model

Figure 1 presents the proposed methodology for building the CS model which is then evaluated in this research. Feature selection is conducted as the wrapper or filter-based approach. Then, based on the selected features, the extraction methods are applied: the classical Fisher Discriminant Analysis (FDA) and heteroscedastic extension of FDA – i.e. FDA with Chernoff Criterion (FDA_Cher). For the classification of the samples in the new discriminant space (i.e. the space spanned by the extracted features - eigenvectors), the Fisher classifier is used, which is the nearest centroid method (Duda, Hart, Stork, 2011, Stapor, 2011) but in the new discriminant space. The prediction accuracy is calculated on the test data as the ratio of correct predictions to the number of all test cases (Duda, Hart, Stork, 2011).

For the filter-based methods, the importance of each feature is evaluated individually for each feature by determining the value of the criterion function (which is specific to a particular filter – see formulas (6) and (7)). Features are then ranked in order of descending values of this criterion function. The number of the top best features is the one which gives the best classification performance. In the wrapper approach, variable importance is calculated as the frequency of the selection of each feature in 10 iterations (using FDA_Cher variant of feature extraction).

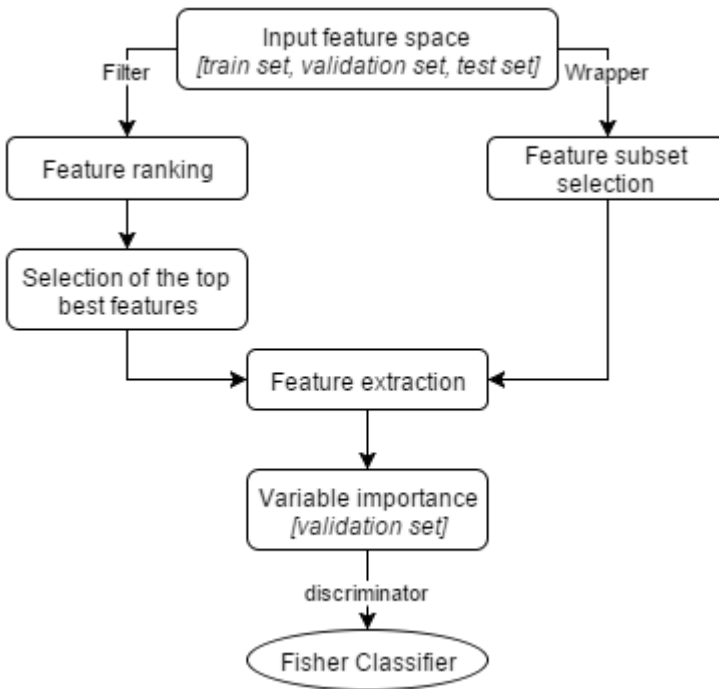


Figure 1. The methodology for building CS model (for learning stage)

5. Experimental analysis

5.1. Data set description

For the evaluation of the prediction accuracy of our proposed CS model we have used the real-world data set, the German credit data set which was also used in Chen and Li research published in *Expert Systems with Applications* (Chen, Li, 2010). The German data set consists of 700 instances of creditworthy borrowers and 300 of bad borrowers. It is composed of 20 numeric and nominal features containing information about credit duration, history, purpose, amount, savings, age, job and other personal information (a detailed structure of this data set is given in the Appendix 1). Nominal features were replaced by binary features, each one representing one of its possible states. Preprocessed data set contained 59 attributes.

5.2. Experimental results

The obtained results are summarized in Table 1. For the German data set, the classification accuracy of the two extraction methods (with all 59 features) achieved $30.00\% \pm 0.00\%$ (FDA), $59.40\% \pm 10.43\%$ (FDA_Cher) for extraction on one directions and $59.70\% \pm 11.18\%$ for extraction on 3 dimensions. For FDA

extraction, SFFS feature selection was the best approach with average $58.50\% \pm 3.06\%$ classification accuracy, and the median selected attributes were equal to 2. The FDA_Cher extraction achieved significantly better results. The best feature selection algorithm (Fisher Score) was able to achieve $75.10\% \pm 3.38\%$ accuracy rate with 18 attributes selected and 3 directions.

Table 1. Results summary with 10-fold cross validation for German data set

Algorithm \ data set	FDA		
	Accuracy rate (%)		Number of selected features
	Avg.	Std.	Median
All features	30.00%	0.00%	59
CFS	34.00%	12.65%	1
FS	55.50%	18.77%	23
SFFS	58.50%	3.06%	2
GRASP	57.90%	7.37%	2
MA	30.00%	0.00%	27

Algorithm \ data set	FDA_Cher				FDA_Cher (1 direction)		
	Accuracy rate (%)		Number of selected features	Number of directions	Accuracy rate (%)		Number of selected features
	Avg.	Std.	Median		Avg.	Std.	
All features	59.70%	11.18%	59	3	59.40%	10.43%	59
CFS	73.90%	4.95%	28	2	73.10%	5.13%	24
FS	75.10%	3.38%	18	3	74.60%	2.99%	18
SFFS	67.00%	6.04%	6	3	66.60%	5.23%	6
GRASP	67.90%	5.43%	17	3	65.70%	6.06%	12
MA	65.80%	8.68%	28	3	61.20%	22.01%	26

Using FDA_Cher model, which does not require the homoscedasticity of the data, increased the average accuracy of prediction. The feature selection algorithm helped to decrease the number of features taken into the model and in some cases significantly increased the accuracy.

In a two-class problem, FDA reduces dimensionality to one direction, since only one eigenvalue is different than 0 and there is no discriminatory information in other directions. In heteroscedastic extension of FDA, DDMs have more than one nonzero eigenvalue. Those extra directions capture, in general, the heteroscedasticity in the data. In our research, we have examined between 1 to 5 dimensions/directions to which features could be extracted by FDA_Cher. Table 1 in the “Number of directions” column presents the best results and dimensions for which this result is obtained. In all cases, the best result was achieved in more than one dimension, which demonstrates that higher dimensions also contain discriminatory information that was not captured in the first direction.

5.3. Attribute importance analysis

In credit scoring, it is very important to know which attributes (features) characterizing a consumer introduced to the model are relevant, i.e. more significant in the classification task, and which are of less importance. Generally, variable importance measures can be divided into two groups: those that use the model information and those that do not. Our proposed method for the analysis of importance (i.e. the classification effectiveness) of the attributes belongs to the second group.

Table 2. Feature rankings

	Top 10										Last 10											
CFS	11	15	29	50	42	16	18	48	28	59	...	46	31	58	12	9	45	2	25	1	8	
FS	11	8	1	15	25	2	50	12	29	45	...	54	14	44	32	19	43	52	7	4	39	
Importance	Most important										↔		Least important									

Table 2 shows the results of filter-based variable importance analysis: rankings created by CFS and FS measures on the entire data set. Both algorithms selected as the most important feature 11 indicate that the customer does not have a checking account. Attribute 8, the status of the existing checking account, was selected by FS as the second most important attribute. However, CFS moved this attribute to the last place. Delay in paying off in the past is the second most important feature (attribute 15) for CFS and fourth for FS. For FS, the third most important feature was duration in month (attribute 1), which was ranked 58th for CFS.

Table 3. Feature selection frequency – attribute number (frequency [%])

	Top 7							Last 3			
SFFS	1 (100%)	8 (80%)	11 (60%)	25 (50%)	50 (50%)	15 (40%)	9 (30%)	...	57 (0%)	58 (0%)	59 (0%)
GRASP	12 (70%)	8 (60%)	35 (60%)	24 (50%)	40 (50%)	3 (40%)	5 (40%)	...	36 (0%)	51 (0%)	57 (0%)
MA	12 (80%)	13 (80%)	52 (80%)	8 (70%)	21 (70%)	44 (70%)	1 (60%)	...	46 (20%)	58 (20%)	7 (10%)
Importance	Most important							↔	Least important		

Table 3 shows the results of wrapper-based variable importance analysis (in percentage). The more frequently the feature was selected, the better evaluation it got during the feature selection step. Attribute 1 (duration in month) was selected in each case by SFFS algorithm and in 60% of the cases in MA. The most frequently selected attribute by GRASP and MA was attribute 12 describing a customer’s credit history (no credits taken/all credits paid back duly). Attribute 8 was selected in 80% by SFFS, 70% by MA and 60% by GRASP, and it was also highly ranked by Fisher Score. Attribute 11 (the most important for filter features) was selected in 60% of cases by SFFS, 40% by MA and 10% by GRASP. In 10 iterations, SFFS algorithm was selected only from a subset of 14 attributes. One of the least frequently selected attribute was attribute 7 – the number of people being liable to provide maintenance for (0% by SFFS, 10% by GRASP, 10% by MA).

6. Conclusions

This work proposes a new method for constructing credit scoring models which is based on the feature selection in Heteroscedastic Discriminant Analysis, which is the extension of the classical linear Fisher Discriminant Analysis for dealing with the case of unequal covariance matrices in populations.

The prediction accuracy of our proposed credit scoring model is the same as in the best models currently proposed in the literature, but this accuracy is achieved using a linear (i.e. simpler) model, which implies better generalization properties.

We have proved that using heteroscedastic extension of the classical linear Fisher Discriminant Analysis results in better prediction accuracy. Moreover, this accuracy can be further improved by feature selection algorithms.

Not all information stored in the databases is relevant to predict customer behaviour and feature selection methods together with the feature extraction are crucial in reducing the dimensionality of the feature space, which is important from computational and economical point of view as well as because of the curse of dimensionality phenomenon.

Furthermore, thanks to the applied variable importance analysis, we can specify the most relevant variables for the classification task, which could be useful for the analysis of a given customer and for a better understanding of the credit scoring problem.

REFERENCES

- CHEN, F., LI, F., (2010). Combination of feature selection approaches with SVM in credit scoring, *Expert Systems with Applications*, Vol. 37, pp. 4902–4909.
- COVER, T., THOMAS, J., (1991). *Elements of information theory*. John Wiley & Sons, New York, NY.
- CROOK, J. N., EDELMAN, D. B., THOMAS, L. C., (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research* 183 (3), pp. 1447–1465.
- DASH, M., LIU, H., (1997). Feature selection for classification. *Intelligent Data Analysis*, 1, pp. 131–156.
- DUDA, R., HART, P., STORK, D., (2001). *Pattern Classification*. John Wiley & Sons, New York, 2 ed.
- FEO, T. A., RESENDE, M. G. C., (1995). Greedy randomized adaptive search procedures. *J. Global Optim.* 2, pp. 1–27
- FISHER, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7, pp. 179–188.
- FUKUNAGA, K., (1990). *Introduction to statistical pattern recognition*. New York: Academic Press.
- GOLDBERG, D., (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA: Addison-Wesley Professional.
- HALL, M., SMITH, L., (1997). Feature subset selection: a correlation based filter approach, in *International Conference on Neural Information Processing and Intelligent Information Systems*, Berlin.
- KRZYŚKO, M., (1990). *Discriminant analysis*, WNT, Warszawa (in Polish).

- KRZYŃSKO, M., WOŁYŃSKI, W., (1996). Discriminant rules based on distances, *Tatra Mountains Math. Publ.* 7(1996), pp. 289–296.
- LOOG, M., DUIN, R., (2002). Non-iterative heteroscedastic linear dimension reduction for two-class data: from Fisher to Chernoff. *Proc. 4th Int. Workshop S+SSPR*, pp. 508–517.
- MATUSZCZYK, A., (2012). *Credit scoring*. Warszawa: CeDeWu Sp. z o.o.
- MOSCATO, P., (2002). Memetic algorithms. In Pardalos, P.M., Resende, M. (eds.): *Handbook of Applied Optimization*. Oxford: Oxford University Press, pp. 157–167.
- PACHECO, J., et al., (2006). Analysis of new variable selection methods in discriminant analysis, *Computational Statistics & Data Analysis*, Vol. 51, 3, pp. 1463–1478.
- PUDIL, P., et al., (1994). Floating search methods in feature selection, *Pattern Recognition Letters*, Vol. 15, 11, pp. 1119–1125.
- SOMOL, P., et al., (2005). Filter- versus Wrapper-based Feature Selection For Credit Scoring, *International Journal of Intelligent Systems*, Vol. 20 (10), pp. 985–999.
- SPENCE, C., SAJDA, P., (1998). Role of feature selection in building pattern recognizers for computer-aided diagnosis, in *Medical Imaging 1998: Image Processing*, San Diego.
- STĄPOR, K., (2011). *Classification methods in computer vision*. PWN, Warszawa (in Polish).
- STĄPOR, K., (2015) Better alternatives for stepwise discriminant analysis. *Acta Universitatis Lodziensis, Folia Oeconomica, Multivariate Statistical Analysis in Theory and Practice*, nr 1(311), Lodzianis University Press, pp. 9–15.
- THOMAS, L. C., (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. *International Journal of Forecasting* 16 (2), pp. 149–172.
- THOMAS, L. C., OLIVER, R. W., HAND, D. J., (2005). A survey of the issues in consumer credit modelling research. *Journal of the Operational Research Society* 56 (9), pp. 1006–1015.
- ZHANG, D., X., ZHOU, S., LEUNG, C. H., ZHENG, J., (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications* 37 (12), pp. 7838–7843.

APPENDIX

The structure of the German credit data set

Attribute	Description	Values
1.	Status of existing checking account (qualitative)	A11 : ... < 0 DM A12 : 0 <= ... < 200 DM A13 : ... >= 200 DM /salary assignments for at least 1 year A14 : no checking account
2.	Duration in month (numerical)	
3.	Credit history (qualitative)	A30 : no credits granted/all credits paid back duly A31 : all credits at this bank paid back duly A32 : existing credits paid back duly until now A33 : delay in paying off in the past A34 : critical account/other credits existing (not at this bank)
4.	Purpose (qualitative)	A40 : car (new) A41 : car (used) A42 : furniture/equipment A43 : radio/television A44 : domestic appliances A45 : repairs A46 : education A47 : (vacation - does not exist?) A48 : retraining A49 : business A410 : others
5.	Credit amount (numerical)	
6.	Savings account/bonds (qualitative)	A61 : ... < 100 DM A62 : 100 <= ... < 500 DM A63 : 500 <= ... < 1000 DM A64 : .. >= 1000 DM A65 : unknown/ no savings account

7.	Present employment since (qualitative)	A71 : unemployed A72 : ... < 1 year A73 : 1 <= ... < 4 years A74 : 4 <= ... < 7 years A75 : .. >= 7 years
8.	Instalment rate in percentage of disposable income (numerical)	
9.	Personal status and sex (qualitative)	A91 : male : divorced/separated A92 : female : divorced/separated/married A93 : male : single A94 : male : married/widowed A95 : female : single
10.	Other debtors / guarantors (qualitative)	A101 : none A102 : co-applicant A103 : guarantor
11.	Present residence since (numerical)	
12.	Property (qualitative)	A121 : real estate A122 : if not A121 : building society savings agreement/life insurance A123 : if not A121/A122 : car or other, not in attribute 6 A124 : unknown / no property
13.	Age in years (numerical)	
14.	Other instalment plans (qualitative)	A141 : bank A142 : stores A143 : none
15.	Housing (qualitative)	A151 : rent A152 : own A153 : for free
16.	Number of existing credits at this bank (numerical)	

17.	Job (qualitative)	A171 : unemployed/ unskilled - non-resident A172 : unskilled - resident A173 : skilled employee / official A174 : management/ self-employed/highly qualified employee/ officer
18.	Number of people being liable to provide maintenance (numerical)	
19.	Telephone (qualitative)	A191 : none A192 : yes, registered under the customer's name
20.	Foreign worker (qualitative)	A201 : yes A202 : no