Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Economics, Management and Humanities

COMPANY BANKRUPTCY MODELS

Doctoral Thesis

Ing. Július Bemš

Prague, March 2015

Doctoral Study Programme: Electrical Engineering and Information Technology
Branch of study: Business Management and Economics

Supervisor: Prof. Ing. Oldřich Starý, CSc.

Declaration

I hereby declare that I have completed this master thesis independently and that I have listed all the literature and publications used.

Prague, March 17th, 2015 ....

Ing. Július Bemš

Acknowledgement

I would like to acknowledge my supervisor prof. Ing. Oldřich Starý, CSs. for his guidance my doctoral thesis. I would like to thank my colleagues from the Department of Cybernetics who assisted me with machine learning oriented calculation. Last but not least I want to thank my colleagues from the Department of Economics, Management and Humanities for valuable comments and suggestions during the completing of my doctoral thesis. Special thanks belongs to SAS Institute Czech republic for providing me academic license of SAS software.

Table of Contents

Abstract

This thesis introduces the new scoring method for company failure (default) prediction and performance evaluation. The method is based on usage of the modified *magic square* (spider diagram with four perpendicular axes) which is standardly used to evaluate economic performance of countries. Evaluation is quantified by the area of polygon whose vertexes are points lying on the axes. This new method deals with magic square shortcomings e.g. axis zero point not in the axes origins. The polygon area represents the company score when applied on company financial indicators. The main goal is to find out the parameters such as axes order, parameters weights and angles between axes in the way to achieve best scoring performance. Input data are financial ratios from financial statements (debt ratio, return on costs etc.) and binary (true/false) information whether company defaulted or bankrupted. The aim is to achieve good ordering performance (defaulted companies should have lower score than non-defaulted companies) which is represented by Gini and Kolmogorov-Smirnov indexes. In addition, threshold value for score dividing default and non-default companies is calculated because it is necessary for the default prediction, true positive rate and true negative rate calculation. Training and validation processes were performed on two independent and disjunct datasets. Results show that performance of this new method is comparable to other methods such as logistic regression, neural networks, Bayesian models etc. Thesis includes comprehensive comparison of the proposed method with competitive methods. Polygon area can also be a tool for relative company comparison. The result of relative comparison of electricity distribution companies by proposed method was compared to the results from the performance index based on data envelopment analysis. The real advantage of the proposed method over other methods is in graphical representation in diagram. User of this model does have graphical tool, which gives him the ability of examining company performance by the shape of the polygon and the ability to determine an impact of individual factors on final performance. Beside of this graphical tool, the standard statistics are available as it is common in other methods.

1 Introduction

The term *magic square* (depicted on figure 1) was firstly used in macroeconomics sphere by German economist and former minister of finance Karl Schiller [1]. It is the diagram with four perpendicular axes on which are depicted a country's main macroeconomics indicators: gross domestic product (GDP) growth rate, consumer prices inflation rate, unemployment rate and balance of trade expressed as a portion of GDP [2]. The area of quadrangle is used for relative comparison between countries. The higher quadrangle area means the better economic performance of the examined country. The idea of economic performance evaluation based on the stated indicators was introduced by Nicolas Kaldor (born Miklós Káldor), Hungarian economist [3].

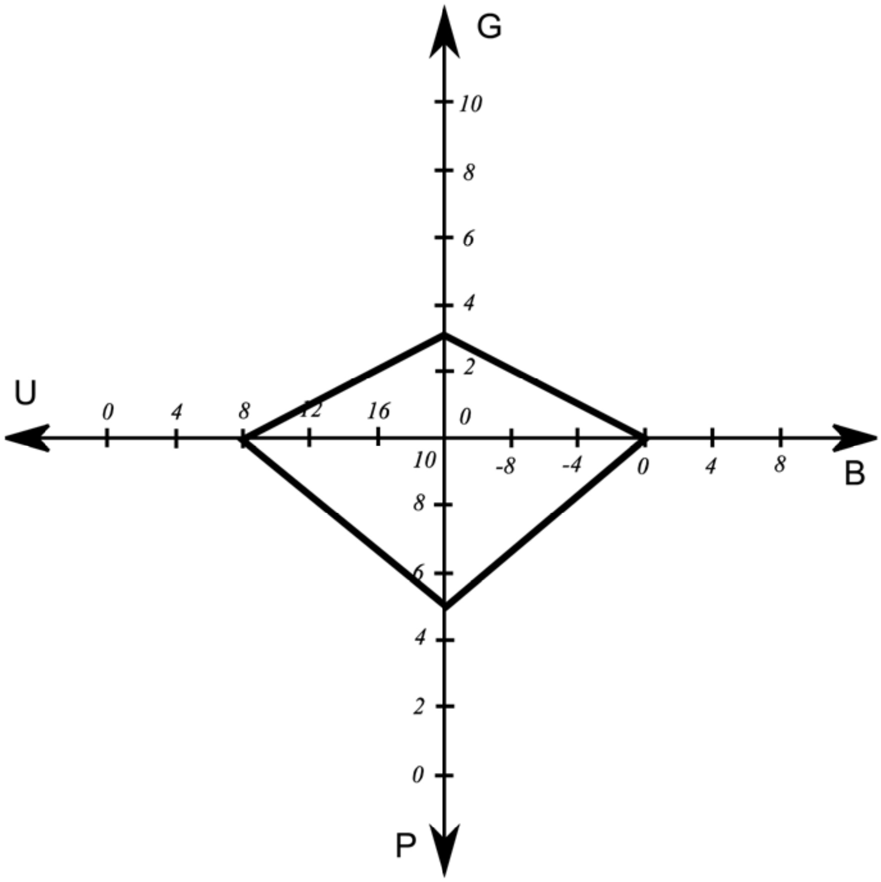


Figure 1: The example of Magic Square (values in percent)
(G - GDP growth rate, U - unemployment rate, P - inflation rate, B - trade balance)

This chart (figure 1) is the special case of a spider diagram (radar chart) where the axes count can be higher than four. The disadvantage is that axes do not have the zero point in intersection and the quadrangle area is changing based on zero location on each axis. The example of another disadvantage is that consumer prices inflation rate below zero (deflation) is not desirable [4], but the quadrangle area is growing in this case.

The similar principle of evaluation and comparison can be used also in case of companies. There will be several axes with performance indicators (figure 2) and a company with higher polygon area has better performance (is rated higher) than company with lower polygon area. To avoid above stated disadvantages all axes have the same origin (zero point) and values are be transformed to the numbers in the 0 – 1 interval according to "the higher is the better" principle. The final polygon area is going to be used as a company score in the same way as Altman did in his original work [5].

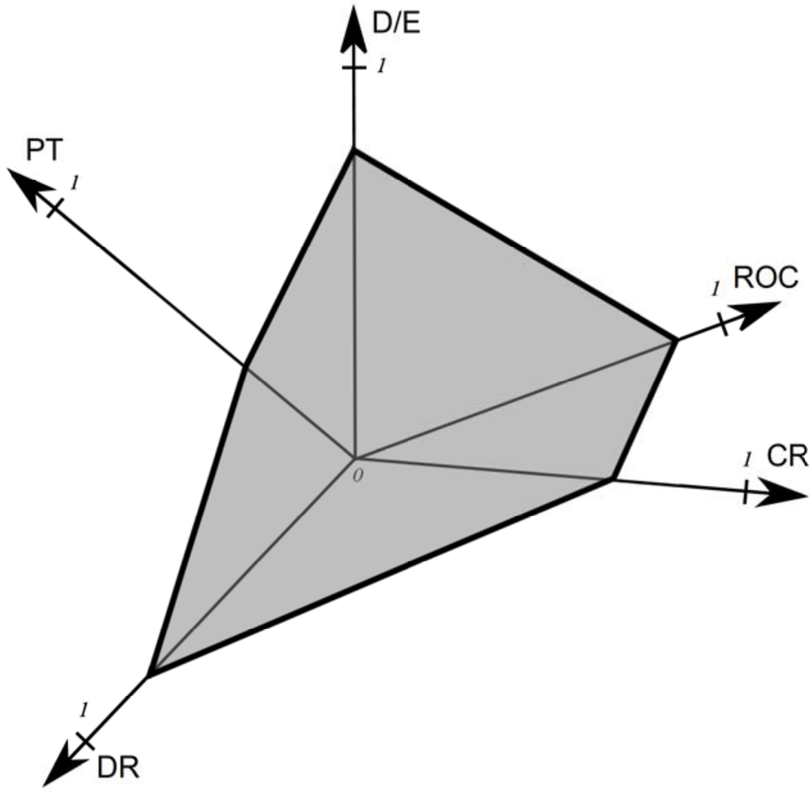


Figure 2: The example of the diagram for the company. The greyed area represents company score. (PT – payables turnover, D/E – debt to equity, ROC – return on costs, CR – current ratio, DR – debt ratio)

The ideal scoring method evaluates each defaulted[1] company[2] by score value lower than any non-defaulted company. This leads to the ordering performance evaluation of the scoring method that is quantified by the Gini and Kolmogorov-Smirnov indexes (deeper explained in *Performance Tests* section on page 33). The other option is to find out the score threshold value, which means that companies with the score below this value are treated as default and companies with score above this value are treated as non-default companies. Subsequently the true positive rate (sensitivity, p. 40) and true negative rate (specifity, p. 40) can be calculated. Threshold value determination is inescapable for default prediction. The difference between this scoring approach and classification models such as logistic regression or neural network is that Gini and Kolmogorov-Smirnov indexes will be used as objective function as well as the measurement of model ordering performance.

There exist three groups of parameters that affect the polygon area (score).

1. Order of axes. The area of polygon created by more than three indicators depends on the order of axes.

[1] The definition of defaulted company varies. It can be a company that is not able to fulfil its obligations in specified time (e.g. 90 days) or it is a bankrupted company. Anyway this is not important for this paper because the aim to present new scoring method.
[2] A company is not defaulted in time of evaluation, but gets defaulted during the observation period.

2. Angles between axes. It is not necessary that axes are perpendicular like in figure 1. The only requirement is that sum of all angles must equal to 360°. Polygon area is changing by changing angles between axes.
3. Weight of indicators. Each indicator's value is between 0 – 1. The weight can be any non-negative number by which the indicator value will be multiplied.

The main goal of this work will be design a new scoring model. The values of model parameters will be changing in order to achieve highest Gini and Kolmogorov-Smirnov index values. This work will try to find the answer on the following **research questions.**

1. Does axes order have significant impact on Gini and Kolmogorov-Smirnov index values when the weights and angles are constant and equal?
2. Is the performance of proposed scoring method competitive with other methods such as logistic regression or neural networks?
3. Is it necessary to change all groups of parameters or some of them (weights, angles) can remain constant?
4. Is the relative order of electricity distribution companies obtained by proposed method different from assessing by the *Y performance index* used in thesis [6]?

These configurations (scenarios) will be evaluated.

A. Order of the indicators will be determined, other parameters will be constant.
B. The weights will have the identical constant value. Order of the indicators and angles between their axes will be determined.
C. The angles will have the identical constant value. Order of the indicators and their weights will be determined.
D. All of the parameters will be determined (order, weights, angles).

Indicators that enter the calculation are calculated from publicly available data[3]. Proposed method is trained and validated on disjunct datasets. Training in this context means to find out the values of parameters (order, weights and angles) that meet the constraints and achieve maximal Gini (or Kolmogorov-Smirnov) index value. Validation [7] is the process where the values of parameters obtained in training phase are applied on data that are not the part of training dataset. Gini and Kolmogorov-Smirnov indexes values from training and validation phase will be compared. In case when the drop of values between training and validation data sets is not significant[4] one can say that the used method is not fitted just to training data and can be used in general analysis. The results are compared with other widely used methods. Proposed method can be used for relative performance (benchmarking) comparison between companies due the fact that the method output is score, which can be

[3] There is plenty of data sources and information about the companies. Some countries (such as Czech Republic, Slovakia and a few others European countries) have publicly available financial statements of all registered companies free of charge. There is also public insolvency register, debtors register and other free or paid sources. Data gathering differs from country to country and it is dependent on provided data exchange interfaces. This process is time consuming and requires advanced computer knowledge. Due this fact, it is possible to order aggregated data from specialized companies or institutions.

[4] Significance is relative term in this context and the difference in values will be compared with differences in other methods.

used for construction of relative order. The relative order of performance of electricity distribution companies obtained by proposed method is compared to the order from [6]. Author [6] analysed data using different approach and therefore the comparison has sense.

1.1 Thesis Sections

The core of this dissertation is covered in the *Proposed Method* section where the new approach is described and the *Results and Discussion* section where the results are presented.

The *Methods Background* section covers all theory needed for good understanding of the problem and describes the current state of research. The largest part of this section is devoted to description of statistical background needed for calculations and financial theory background required for understanding the financial statements data.

The data description and theory necessary for understating company financial ratios is present in section *Data Description*. Data transformation into the desired format and data analysis are already the part of Proposed Method section.

The *Introduction* also contains the research question and necessary information about whole research. Answers on research questions are the part of *Conclusions* section. Others sections contains some minor formalities not necessary for the core understanding of a problem.

Appendix section contains additional and deeper information related to the calculations and conclusion from the main part of the thesis.

2 Motivation and Goals

Motivation of this research was preparation of effective method mainly used for company failure prediction. Beneficiary of this information can be creditors who need to assess their business partners. Idea was to use statistical tools and combine them with the *magic square* to obtain the method graphical tool that can be used for company failure prediction and performance prediction.

One can find plenty of various scoring methods, classifiers and other tools that can assess the company based on historical data or based on analysis with the similar companies. The result is mainly only one number, which is comfortable because many input indicators can be "transformed" to only one simple number. The shortcoming of this approach is big generalization and loss of information that can be essential. It can be understand the same as arithmetic average of a group of number. The result is only one number, but part of information about this group is lost. Mentioned graphical tool can help in decision-making process and it can be used for easier visualisation.

The *magic square* is nice and simple tool that can assess the performance of countries' economy. Shortcoming of this tool is described in Introduction and State of the Art section. These shortcomings are relievable and *magic square* idea can be transferred to the field of companies.

The goal was introducing a new method and graphical tool for company assessment. The advantage of the graphical tool is in representativeness of results. Standard statistical methods' output is a number. This tool output is not the only number, but also a diagram that can represent more than just one number. User can easily see the pros and cons of individual company factors based on the visual display of individual factors. The final diagram shape can be also useful mostly when assessing the companies from one industry sector. One can assume that each industry sector have a typical diagram and comparison of one company diagram to typical sector diagram can be made.

When a new model and tool is designed and built its output can be also used for relative companies' performance assessment. Research in a field of benchmarking was performed in doctoral thesis [6] at the Czech Technical University in Prague. The logical goal was to compare the results of proposed method to benchmarking research results.

3 State of the Art

State of the art is partially covered in the Introduction and Methods sections. Therefore, this section is only brief description of current research state and references to other sections where detailed state of art is performed with the references to scientific background. A design of a new method for assessment of companies' performance is the main result of this thesis. Several different concepts are joint into a new one. A *magic square*, which is simple tool for a comparison of countries' performance is modified and used for a comparison and assessment of companies' performance. GDP growth rate, unemployment rate, inflation rate, trade balance values are depicted on the diagram (figure 1) and the quadrangle area represented by the values is used for relative comparison of countries performance. Higher area means higher performance and vice versa.

Magic square is well described in the Introduction section with adequate references [1-4]. Two main issues have to be solved for correct use of *magic square*. The first one is to have one common zero point for each input variable and another one rests in having higher-better logic for input parameters. The methods described in Variables Transformation and Selection section are suitable and widely used for transformation of input parameters' values into the required values logic.

Methods used for bankruptcy (failure, default) prediction are based on statistical analysis of subjects' quantitative indicators and information about subjects' failure. The oldest widely used tool for bankruptcy prediction is Altman score [8] described on page 14. Altman score is the result of discriminant analysis applied on set of companies. The three groups of companies identified based on the score: companies likely to bankrupt, companies unlikely to bankrupt and grey zone where the prediction is not accurate enough. Similar approach is used also for the new proposed method. Score threshold values are defined to separate companies probable to default from other companies. The output of the bankruptcy (failure) prediction is a real number representing score or probability. Score can be in many cases transformed into probability. Logistic regression [9-11] (page 15) model is very often used for company failure prediction. Logistic regression output represents probability of company default. Other widely used model are neural networks (page 17). There are many various neural network models [10, 12-17] (backpropagation neural network, radial basis neural network, support vector machine etc.). Output of these models can be again probability of default. Neural networks and logistic regression are widely used classifiers for company default prediction. In addition, there are other classifiers such as Bayesian models [18-21], decision trees [20, 22, 23], memory based reasoning [16, 20, 24, 25] and evolutionary algorithms [26-28] that can be also used for specified purpose. The detailed view on these methods is present in the Methods section. Actually, here described methods are used for companies' failure prediction and they are presented later in this work.

Output of all methods is score (or probability), which can be used for performance ordering. This can be also done by benchmarking methods described on page 36 [6, 29-31]. These methods can set up companies order according to specified criterions. New proposed method result is confronted with the benchmarking models results.

The accuracy of models (actually used or new ones) is measured by Gini and Kolmogorov-Smirnov indexes as well as other performance indicators (sensitivity, specificity, etc.) based on hypothesis testing techniques and logic [32-39]. This all is described in Statistical Errors section on page 37.

Briefly, actual usage and description of *magic square* was presented. Various classifying methods were also presented and current state of the problematic was elaborated in this section and mainly Introduction and Methods section. The new method is interconnecting actual knowledge from classification algorithms (machine learning) with the *magic square* tool and presents new approach that can be used not only as a tool for company failure prediction but also as a benchmarking tool.

4 Research Methodology

This dissertation is based on utilization of quantitative techniques. Used research procedures are empirical and theoretical. Empirical procedures are processing input data that describe observed objects (companies) and these procedures result in factual conclusions about observed objects. Exact methods on specific data are used to obtain factual knowledge. In addition, theoretical procedures such as deduction are used for drawing conclusions. Historical data that describe companies are used for the models creation. The specific features can be assigned to the research object based on the created model. The common knowledge is applied on individual object and conclusion about this object is made.

Primarily these quantitative methods were used in this dissertation thesis:

- correlation analysis
- descriptive statistics
- statistical modelling
- statistical transformations
- probability sampling

The data used in this work are exact facts about the companies and no subjective inputs (like experts' opinions) were used. These input data were statistically analysed, evaluated and interpreted. Statistical analysis was used for assessment of companies and synthesis is used when final model was developed and when research question were proved or rejected. Finally, comparative methods were used in order to compare results of proposed method with other methods.

5 Methods Background

The largest part of this section is devoted to various regression methods. The fundamental logic behind regression is to fit the training data in the most proper way. There is a vector of input variables ***x*** and one binary output variable y. Fitting means the process of finding parameters of model in the way that model prediction power will be maximized. The aim is not to fit the training data perfectly, but to maximize predictive power that is computed on test or validation data. The almost perfect fit can be achieved by the high number of model parameters. On the other hand, the high number of model parameters leads to perfect fit on training data, but probably will led to very bad predictive power. This problem is called overfitting and compromise between number of parameters and model error should be found. This section also covers the methods of performance evaluation such as Gini and Kolmogorov-Smirnov indexes. In addition, the hypothesis testing and error measures of binary models are covered.

5.1 Scoring Methods

The results from the new proposed method will compared to the results from Altman score calculation, logistic regression, neural networks, memory based reasoning, Bayesian models and evolutionary algorithms. Logistic regression, neural network with three hidden units and memory based reasoning calculations will be performed using SAS[5] software implementation. The calculation and deeper analysis will be performed for these three models. Other models such as Bayesian models, neural network with two hidden units and genetic algorithms will be covered briefly. Calculation for these models will be performed mostly by the Matlab[6] software. All mentioned algorithms are explained in this section.

Method	Implementation
Altman Score	Own
Logistic Regression	SAS
Neural Networks 3H	SAS
Neural Networks (other)	Matlab
Bayesian Models	Matlab
Decision Trees	Matlab
Memory Based Reasoning	SAS
Evolutionary Approach	Own

Table 1: Basic summary of methods used for the comparison with the proposed method

5.1.1 Altman Score

The first widely used companies scoring model was introduced by Edward Altman [5] who applied statistical method of discriminant analysis on a set of manufacturing companies. The result (Z-Score) is linear combination of five financial statements' indicators.

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 0.999X_5 \quad (1)$$

[5] Statistical Analysis System - software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics (http://www.sas.com/)
[6] MATtrix LABoratory – multi-paradigm numerical computing environment (http://www.mathworks.com/products/matlab/)

$$X_1 = \frac{Working\ Capital}{Total\ Assets} = \frac{Current\ Assets - Current\ Liabilities}{Total\ Assets}$$
$$X_2 = \frac{Retained\ Earnings}{Total\ Assets}$$
$$X_3 = \frac{EBIT}{Total\ Assest}$$
$$X_4 = \frac{Market\ Value\ of\ Equity}{Book\ Value\ of\ Total\ Debt}$$
$$X_5 = \frac{Sales}{Total\ Assets}$$

When the score falls below 1.8 it means that the company default is probable, score between 1.8 and 3.0 is a grey zone and companies with score above 3.0 are not likely to default. The problem for binary classification is in large group of companies, which fall into grey zone. These companies will be treated as non-defaulted companies for purpose of calculations.

Coefficients used in Altman score calculation may differ between economy sectors and countries. Here presented values are the results of Altman calculation on group of manufacturing firms in USA [5].

5.2 Classifiers

This section covers classifiers used for binary classification tasks. The most frequently used classifiers in the credit scoring are logistic regression and neural networks. Logistic regression is more preferred than neural network because it is easily explainable and understandable. The resulting model clearly describes outputs changes based on input changes. The neural network models are complex and very often unclear. Neural networks may lead to slightly better results and they are often used like benchmarks for logistic regression models. Other models are used rarely in credit scoring or are used in some very specific tasks and they will be discussed briefly. Learning process is shown in the example presented in the Bayesian models section.

5.2.1 Logistic Regression

The logistic regression [40] describes relation between one or more independent variable and one binary dependent variable. Binary variable is represented by 0 and 1 values where 0 means not defaulted company and 1 means defaulted company in the context of company scoring. There is plenty of other applications for example in medicine or machine engineering etc. [41].

The simplest example is a model with one independent and one dependent variable. The probability of having outcome equal to 1 is *p*, which is the ratio of ones count to whole sample size. The probability of having outcome equal to 0 is *1-p*, which is proportion of zeroes in the given sample.

The independent variable can be discrete or continuous. In any case, it can be easily discretised or binned (see figure 23). The example of discrete variable in financial scoring can be the number of household members and the example of continuous variable can be the debt ratio. It is obvious that company default risk rises with the growing debt ratio[7]. If this continuous variable is discretized into ten categories[8], the probability of default *p* rises from the lowest to the highest categories. The distribution of these default probabilities is in ideal case monotonous with bounds 0 and 1 [9]. Usage of linear regression for problem of this type is improper because the linear regression outputs the real numbers from $-\infty$ to $+\infty$. The desired output is in 0 to 1 interval. Moreover, the presumption of constant residuals variance is violated in case of using linear regression.

The *S* shape curve for one independent variable is defined by the equation (2). F(x) represents the probability of output equal to 1.

$$p = F(x) = \frac{e^{\alpha_0+\alpha_1 x}}{1+e^{\alpha_0+\alpha_1 x}} \tag{2}$$

$$\alpha_0 + \alpha_1 x = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{F(x)}{1-F(x)}\right) \tag{3}$$

Equation (3) is called the logit transformation. This function has many desirable properties of liner regression such as it may range from $-\infty$ to $+\infty$ and it is linear in its parameters [11]. The term *p/(1-p)* is odds of occurring default. The difference between odds and probability is that odds can have any non-negative value and probability is represented by values in 0 – 1 interval. For example, odds of having greater outcome than 2 when throwing a dice is 4:2, but probability of the same outcome is 4:6.

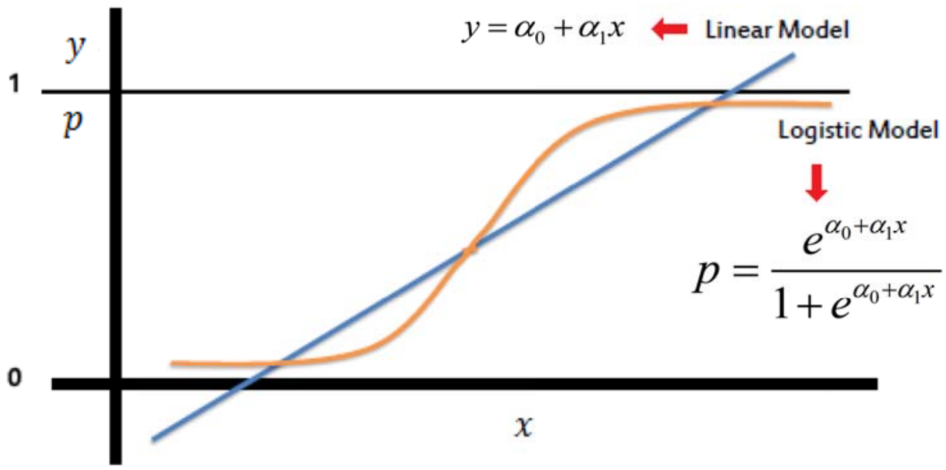


Figure 3: Graphical interpretation of linear and logistic model

The values of parameters α_0 and α_1 must be estimated to fit a logistic regression model described above. There is no exact mathematical solution for least square estimates of the parameter for logistic regression [9]. The least square approach is primarily used for linear regression models. Therefore maximum likelihood [42, 43] approach is used. Due to the

---

[7] Higher debt ratio increases risk that company will not be able to repay its debts. The default probability increases with growing debt ratio because the absolute amount of money needed for the interest payments rises with growing debt.

[8] For example it is binned into 10 categories by the debt ratio value in the meaning that debt ratio between 0 % and 10 % is the first category, between 10 % and 20 % is the second category and so on.

nonlinearity, the numerical methods are used to obtain the solutions. Likelihood function is given by the expression (4).

$$L(\boldsymbol{\alpha}) = \prod_{i=1}^{n} F(x_i)^{y_i} \left[1 - F(x_i)\right]^{1-y_i} \tag{4}$$

x_i, x_i is one observation
F(x) is conditional probability for output (y) equal to 1
[1 – F(x)] is conditional probability for y equal to 0
Vector α represents unknown parameters

L(α) must be maximized to obtain the values of vector α. It is more convenient to work with log-likelihood function, which is the natural logarithm of likelihood function. Logarithm is monotonic function and the logarithm achieves its maximums at the same points as the primary function (before logarithm takes place). To find a function maximum it is often necessary to derivate the function. Very often it is mathematically easier to derivate the logarithm of the function due to the desirable character of logarithmic function[9]. Equation (5) represents log likelihood function.

$$LL(\boldsymbol{\alpha}) = \ln(L(\boldsymbol{\alpha})) = \sum_{i=1}^{n} \left\{ y_i \ln\left[F(x_i)\right] + (1 - y_i)\ln\left[1 - F(x_i)\right] \right\} \tag{5}$$

Equation (2) is logit for one variable. Its extension for n-variables will be similar. Variables used in equations (5)(6)(7) were already described before and have the same meaning.

$$g(\mathbf{x} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n \tag{6}$$

$$p = F(\mathbf{x} = \frac{e^{g(\mathbf{x}}}{1 + e^{g(\mathbf{x}}} \tag{7}$$

The rest of the procedure is analogic to single variable procedure.

5.2.2 Neural Networks

Biological Background

The models of artificial neural networks are based on biological principle of human brain functioning. Human brain roughly consists of 100 billion neurons and each neuron is connected to 10 thousands other neurons[10] by synapses [44, 45].

[9] Logarithm of product is the sum of individual logarithm. Derivative of these sums is easier to evaluate than derivative of product.

[10] The exact facts about biological principle of neural networks are not crucial for this thesis. They are collected from various online sources that can contain simplified information.

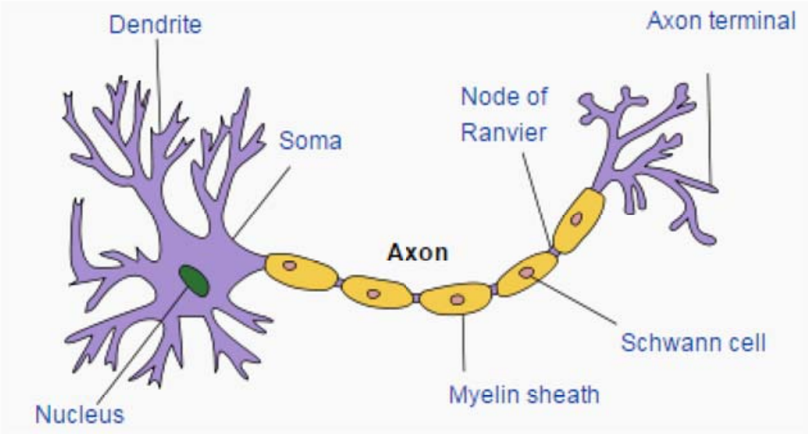


Figure 4: Neuron scheme [46]

Neuron receives information (signals) from other neurons by dendrites. It processes information and may send the signal to other neurons by axon, which can be understood as the output point. Output signal can have digital or analogue character [47]. Axon terminals are connected to other neurons' dendrites or muscles etc.

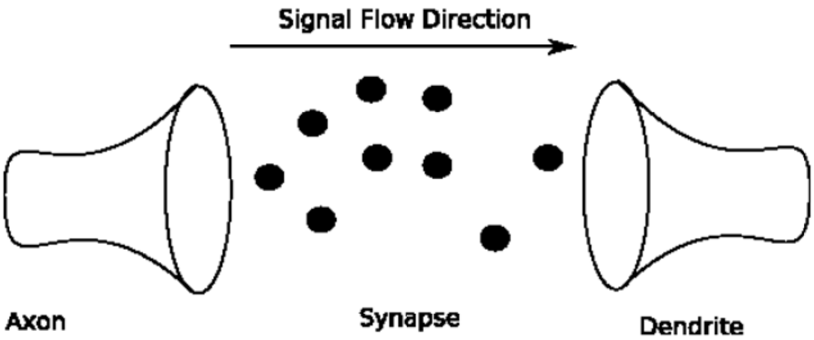


Figure 5: Synapse scheme [48]

The connection between two neurons (axon terminal and dendrite of other neuron) is called synapse, which transmits the signal between neurons. The important property of synapses is their ability of changing strength. It means that they can adopt their behaviour in time.

Here described biological neurons behaviour can be simulated by computers. Neuron soma is a function, which has input parameters represented by dendrites and one output parameter represented by axon. Synapses strength is the weight of parameters entering the function.

Artificial Neuron

An artificial neuron is a mathematical model of biological neuron. It has one or more inputs and one output. Usually the weighted sum or other type of mathematical calculation (transfer function) is used to process input values. The result transfer function is then passed through activation function [49, 50]. Activation function transforms its input into desired output. The goal of neural network training algorithms is to determine the most suitable weights values.

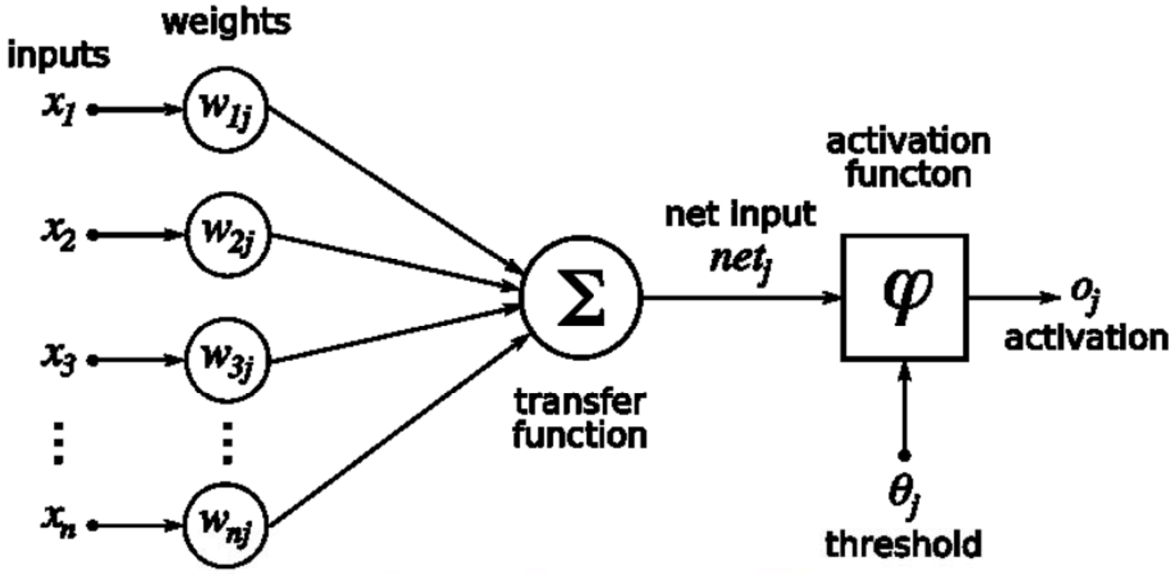


Figure 6: Artificial neuron scheme [50]
© Chrislb / Wikimedia Commons / CC-BY-SA-3.0 / GFDL

The neuron showed in the figure 6 has threshold activation function (step function), which is the simplest one. If the value from transfer function reaches threshold level, output is 1 otherwise output is 0. Neurons usually have one more input called bias. A bias value can shift the activation function and this behaviour may be critical for successful learning.

Following figures show several types of activation functions. The sigmoid activation function is the most suitable for probability prediction because it has several helpful features already described in the Logistic Regression section.

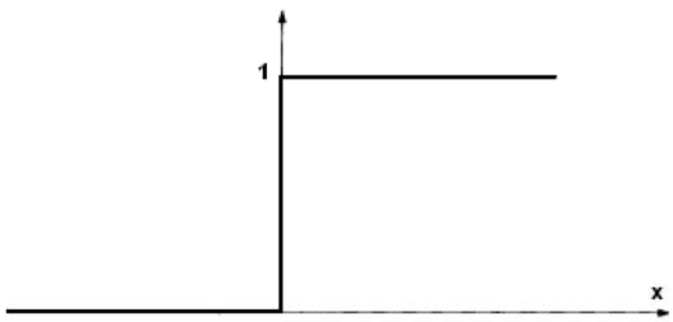


Figure 7: Threshold (step) activation function

Step activation function $f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ (8)

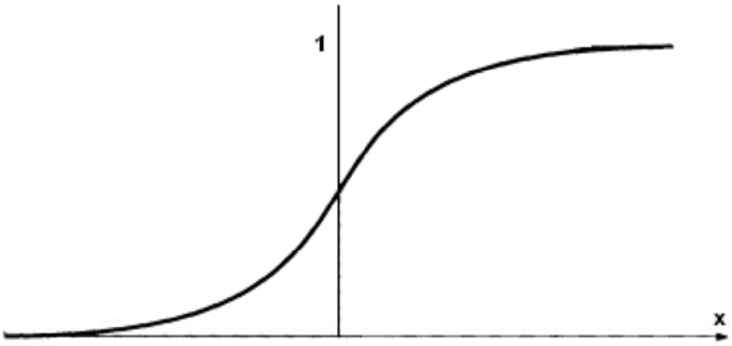


Figure 8: Sigmoid activation function, equation (7)

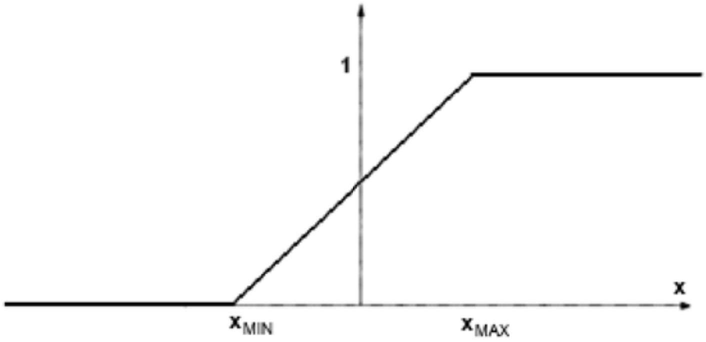


Figure 9: Piecewise linear activation function

$$\text{Piecewise linear activation function } f(x) = \begin{cases} 0, & x \leq x_{min} \\ ax + b, & x \in (x_{min}, x_{max}) \\ 1, & x \geq x_{max} \end{cases} \quad (9)$$

There are many other activation functions such as simple identity, exponential, reciprocal, square, sine, cosine, Elliot etc. [51].

Training process

There are three basic types of neural networks learning

- Supervised learning
- Unsupervised learning
- Reinforced learning

Supervised training was already indirectly described. Training dataset consists of inputs and desired outputs. Usually training data are some historical data and therefore output is known. A good example can be stock exchange data where outputs are prices of stock in time. Another example can be data regarding company failure. The inputs can be accounting data, company behavioural data, macroeconomic data influencing companies etc. and output data is information whether company failed to pay its liabilities or not. The idea of supervised learning algorithms is to find out optimal model settings. It means that model outputs will be close to given outputs in training dataset. A model learns itself based on historical data. A cost function evaluation is based on the difference between actually calculated data and given output data. Mean-squared error is often used as a cost function [15] that minimises average squared error between actual network output and desired output given by training dataset.

The backpropagation algorithm is mostly used for optimising of neural networks used for solving regression tasks. This algorithm is achieved by using gradient descent method.

Unsupervised learning principle is finding solution without telling how to do it. There is no training dataset output given for learning purposes [52]. The goal is to find some patterns, clusters or intrinsic structures. The most common unsupervised learning method is cluster analysis that is searching for patterns and groupings in data. Unsupervised learning can be understood also as a form of reinforcement learning.

Reinforcement learning [53] technique is searching for action to perform in current situation with yielding maximum reward in long term. It is similar to learning by experience. There are no given input data but these data are usually generated by agent's interaction with the environment. Agent performs some actions and generates observations and its costs. The aim is to minimize cost or maximize reward. To obtain high reward, agent must prefer actions that it has tried in the past and found to be effective in producing reward. To discover such actions, it has to try actions that it has not selected before.

There are two fundamental parameters required for proper training

- Learning rate
- Momentum

The learning rate parameter is common in many learning algorithms and represents how quickly neural network can achieve minimum. When the value is low, the convergence process can take a very long time to find final solution. On the other hand, when the learning rate value is too high, the system can oscillate around the solution or completely diverge.

The momentum parameter is required for neural network not fall into a local minimum. When the momentum value is low, system can stop at the local minimum. When the parameter value is too high, the system can jump over the minimum and this can lead to instability.

The algorithm can be stopped after some number of learning cycles. One cycle is called epoch. This parameter is set before the training process starts. The disadvantage of this approach is that one does not know optimal number of iterations and the number of cycles can be over or underestimated. If it is underestimated, the best possible solution is not found and it is necessary to resume the calculation process. If this value is overestimated, the calculation time is uselessly prolonged. Therefore exists minimal error criterion. Minimal mean squared error is calculated in each cycle and if the error falls below defined level, the calculation is stopped.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_N - O_T)^2 \tag{10}$$

MSE – mean squared error, n – number of observations,
O_N – the output from neural network, O_T – target output

These two parameters are often used jointly. When the mean square error does not achieve the required value, the algorithm is stopped after given number of cycles. Mean square error measures dispersion around the real value of parameter that is actually estimated. It is similar to variance, which measures the dispersion of estimator around the mean.

Linear perceptron

Linear perceptron is the simplest type of artificial neural network [15]. It consists only from one neuron and the activation function is linear. The architecture is shown in the figure 6.

Neural Network (Back Propagation)

An artificial neural network scheme is show in the figure 10. The green nodes represent input layer, blue nodes represent hidden layer and yellow node is output. There is no bias node in the scheme but each type of neural network can have some. Each node is represents one artificial neuron.

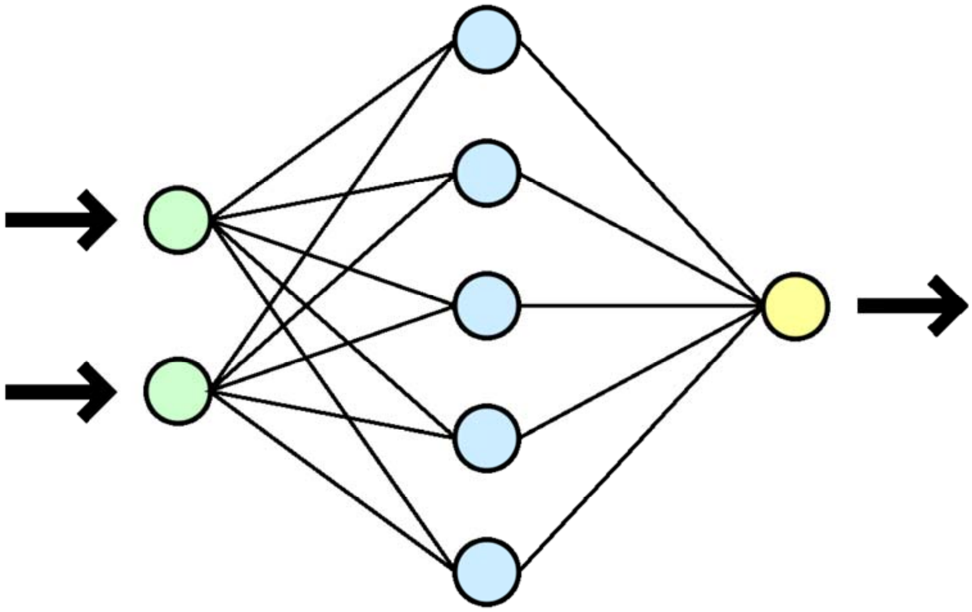


Figure 10: Scheme of artificial neural network [54]
by Dake, Mysid / CC-BY-1.0 (http://creativecommons.org/licenses/by/1.0)

The number of layers and neurons in each layer is not strictly given. Neural network can have more output neurons, not only one as shown in the figure 10.

Backpropagation neural network operation can be divided into two stages:

- Feedforward
- Backpropagation

Feedforward step was almost fully described in Artificial Neuron section. It includes the calculation of weighted sum of neurons inputs eventually with bias node. It is followed by activation function. This process is performed for each neuron in the network.

Backpropagation step [55, 56] begins from the output and the aim is to change the weights of all neurons to achieve better result. It means that error is calculated on the outputs of output neurons and is back propagated. The difference between neural network output and desired output from a training set is calculated using equation (11)

$$\Delta_k = t_k - O_k \tag{11}$$

Δ_k – the difference between neural network output and desired output
k – output node, O_k – output value of node k, t_k – desired value of node k

The error term is calculated by equation (12) where $O_k(1 - O_k)$ is the derivative of sigmoid function.

$$\delta_k = \Delta_k O_k (1 - O_k) \tag{12}$$

δ_k – error term

Following formulas are used to modify weight between two successive nodes j and k (output node). Change in the weight interconnecting nodes j and k is proportional to the error at node k, where the j node is input for the node k [56].

$$\Delta w_{j,k} = l_r \delta_k x_k \tag{13}$$

$$w_{j,k} = w_{j,k} + \Delta w_{j,k} \tag{14}$$

$w_{j,k}$ – weight between nodes j and k, $\Delta w_{j,k}$ – change in the weight between nodes j and k,
l_r – learning rate, x_k – input value to the node k

Learning rate in backpropagation is analogous to the step-size parameter that is used in gradient descent algorithm. Equation for weights change calculation will extend including momentum term μ.

$$\Delta w_{j,k}^{n} = l_r \delta_k x_k + \Delta w_{j,k}^{n-1} \mu \tag{15}$$

n – cycle (epoch) number

The error term for hidden layer can be calculated as:

$$\delta_k = \Delta_k O_k \sum (w_{j,k} \delta_k) \tag{16}$$

Another type of backpropagation algorithm is Levenberg-Marquardt algorithm [57]. It is suitable for small neural networks training where it is very quick and efficient [17, 58]. Neural network for company failure prediction on several parameters and one layer of hidden neurons is considered as small. There are many applications such as image or voice recognition where much bigger neural networks are required.

Neural Network (Radial Basis)

Radial basis network consists of three layers (input, hidden, output). It cannot have more layers as a backpropagation network can have. The major difference between radial basis function (RBF) networks and back propagation networks is that single hidden layer does not use sigmoid (or other S-shaped function) but it uses Gaussian or other basis kernel function [12, 14]. The output is linear combination of radial basis functions in hidden layer. Deeper comparison between radial basis function network and standard backpropagation network is described in [13].

Support Vector Machine

A Support Vector Machine (SVM) is classification method performed by constructing a hyperplane, which separates data into two categories. The simplest well describing example is two-dimensional space, where the classification will be performed. There are two predictor variables as shown in following figure.

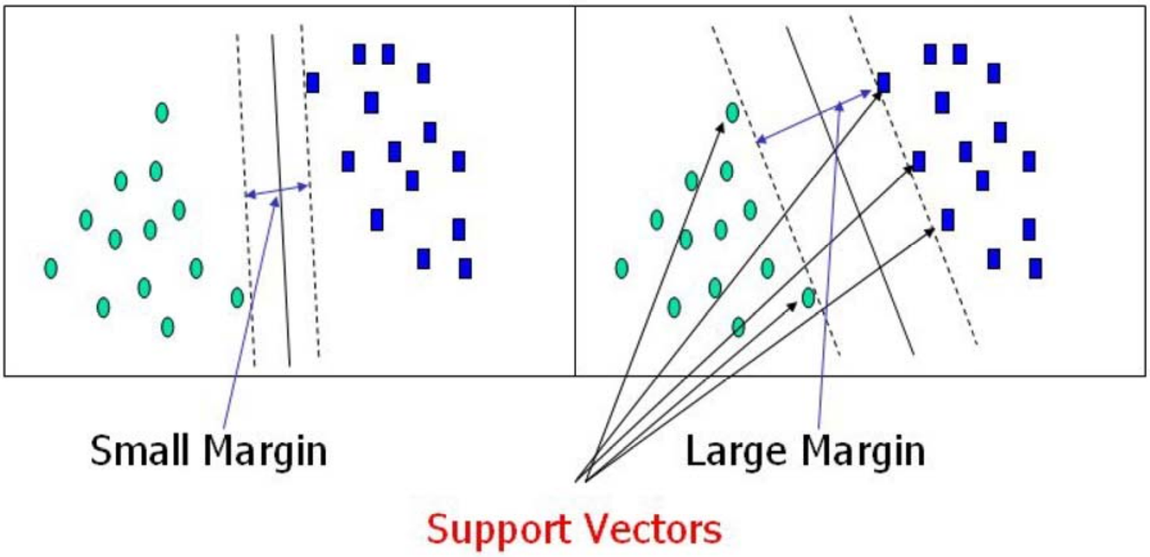


Figure 11: Support vectors lines [59]
© Phil Sherrod[11] / DTREG

This example is very simple and data are well organised, so it is trivial to find one-dimensional hyperplane. It is the line, which separates ovals and rectangles. However, there is infinite number of division lines. Two of them are shown in the figure 11. Separation line is solid and boundary lines are dashed. Boundary lines are parallel and the distance from the division line is maximal. It means that distance between the division line and nearest category point is the same as distance between division and boundary lines. Distance between division line and boundary line is called „margin“. Points that are constraining the width of margin are called „support vectors“. The best solution is the hyperplane with maximised margins.

There are much-complicated data sets than this one shown above. There can be many predictor variables unable to divide by simple line or hyperplane. In addition, there can be more than two categories. For example, figure 12 demonstrates points, which are separated by nonlinear object. There is need of nonlinear separator line or transforming data to different space where hyperplane can be used for separation.

[11] Author gave permission to reuse this graphics in dissertation, 20-th January 2015

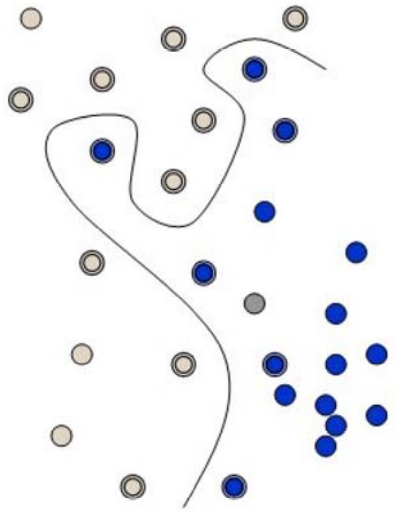


Figure 12: SVM non-linear separator [59]
© Phil Sherrod[11] / DTREG

The kernel function is used for mapping data into different space [15]. It usually transforms data into higher dimensions where separation can be made by hyperplane.

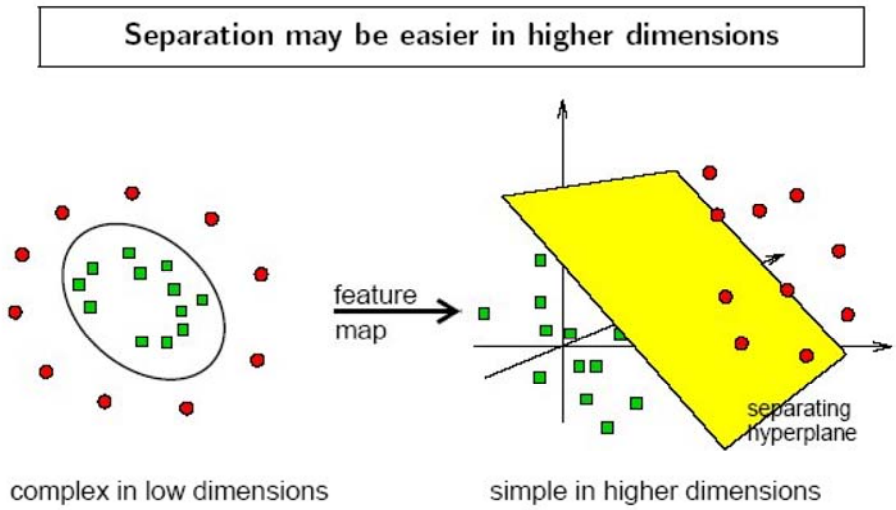


Figure 13: Separation in higher dimensions [59]
© Phil Sherrod[11] / DTREG

Predictor variable is called "attribute" and transformed variable is called "feature". The disadvantage of support vector machines is that the classification result is purely dichotomous, and no probability of class membership is given [10].

Neural Networks Settings Used in Calculation

There are several different types of artificial neural networks commonly used in pattern recognition area [15]. The simplest neural network is linear perceptron for which a perceptron training rule with training rate 0.1 was used in calculations.

Another very simple feed forward neural network with only one unit in the hidden layer was used. The training was performed by Levenberg-Marquardt procedure [15]. This popular algorithm is used because of its relatively high speed and because it is highly recommended as a first-choice supervised algorithm by Matlab Neural Network toolbox. Both the hidden and output neurons use the sigmoid function. The mean squared error was minimized and the training was stopped after one training epoch (presenting all the training samples).

Slightly more complex feed forward network was the back propagation network with two hidden units [15]. For this network, the training was stopped after the number of training epochs exceeded 50 or if the error gradient reached 10^{-6}. The back propagation algorithm with momentum and adaptive learning rate was used. The momentum parameter was 0.95. The learning rate was initially 0.01 and was multiplied by factor 1.05 or 0.7, if the error increased or decreased more than by 4%, respectively. Another neural network used was the radial basis function network with 92 Gaussian basis function units. The calculation was also performed for back propagation neural network and radial basis neural network with three hidden units. The learning rate was set to 0.1 and the momentum rate equal to zero.

Finally, support vector machine, which can be also understood as a neural network, was applied. A version with linear kernel was chosen, whose advantage is in reasonable computational time. Its training procedure uses quadratic programming to maximize so called margin around the decision boundary [60].

The source of Neural Networks Setting section is [20].

5.2.3 Bayesian Models

Bayesian classifiers use Bayes theorem to compute a posteriori probability of each class from conditional probabilities (likelihoods) and class prior probabilities [21]. The feature vector is further assigned into the class that maximizes a posteriori probability. The uniform prior probabilities were considered here. The true conditional probability distributions are not known and must be estimated from the training data. Several different types of this estimation corresponding to different types of Bayes classifier were used [20]:

- Naive Bayes classifier uses conditional independency of particular input features, which makes it possible to compute the total conditional density by simple multiplication of separate conditional densities. Those separate conditional densities were supposed to follow multivariate multinomial distribution, which is appropriate for categorical features.
- Linear Bayes classifier assumes normal conditional densities with equal covariance matrices, which leads to linear decision boundaries. The parameters of the normal densities are estimated using the sample means and covariance matrix. For this classifier, Adaboost combination of 40 linear Bayes classifiers was also used [18]. A weighted voting procedure was used to aggregate the weak classifiers.
- Quadratic Bayes assumes normal conditional densities with unequal covariance matrices, which leads to quadratic decision boundaries. The parameters of the normal densities are estimated using the sample means and covariance matrices.
- Mixture of Gaussians based Bayes classifier models the conditional density as mixture of Gaussian. In underlying experiments, mixture of only two Gaussians is considered.

Naïve Bayes classifier is the simplest one from the group of Bayesian classifiers and it is going to be explained in more details in this section. Naïve Bayes classifier is based on the Bayes theorem and conditional probabilities.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{17}$$

Equation (17) explains conditional probability. It is the probability of an event (A) given that another event (B) has occurred.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{18}$$

Bayes theorem is shown in equation (18). The interpretations are various, but basically one can say that it shows the relation between conditional probabilities P(A|B) and P(B|A). It can be understood using Bayesian terminology as (19).

$$Posterior\ probability = \frac{Likelihood \times Prior\ probability}{Evidence} \tag{19}$$

In case of more than one attribute (feature), the equation (18) will change to (20) where ***x*** is the features vector and A_i is possible outcome.

$$P(A_i|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|A_i)P(A_i)}{P(\boldsymbol{x})} \tag{20}$$

The naïve conditional independence assumption leads to product of conditional probabilities.

$$P(\boldsymbol{x}|A_i) = \prod_{j=1}^{n} p(x_j|A_i) \tag{21}$$

The final class will be chosen the one with the highest probability.

Here is one example for understanding above terminology and learning process[12]: 1,000 financial institution clients applied for a credit card. Seventy-five of them were not able to pay their debts and were classified as defaulted. Three attributes were considered during credit card approval process: age, income, education. The goal is to classify a client (whether he/she will be able to pay its debt) when attributes values are given.

The learning process in Bayesian models is done by the calculation of likelihood and evidence as show in table 2. Process is simple and does not require any numerical methods or intensive calculations (such as maximum likelihood in logistic regression or learning processes in neural networks).

[12] It is enough to show learning process on one classification method because the principle is similar in other approaches. Naïve Bayes model is simple to understand and calculation can be performed demonstratively.

Attribute	Group	Default	P Likelihood	Not Default	P Likelihood	TOTAL	P Evidence
Age	18 - 27	50	0.667	475	0.514	525	0.525
	28 - 38	20	0.267	300	0.324	320	0.320
	over 38	5	0.067	150	0.162	155	0.155
Income	0 - 1000	40	0.533	375	0.405	415	0.415
	1000 - 3000	25	0.333	300	0.324	325	0.325
	over 3000	10	0.133	250	0.270	260	0.260
Education	Primary	55	0.733	625	0.676	680	0.680
	Secondary	18	0.240	200	0.216	218	0.218
	University	2	0.027	100	0.108	102	0.102
	TOTAL	75	0.075	925	0.925	1000	

Table 2: Demonstration learning sample

Attribute or feature is the set of variables according which the default will be predicted. Attributes are categorized into groups and for each group is given the count of defaulted and non-defaulted clients. Likelihood is the ratio of actual group client count to all groups client count for each classifier. Evidence is the ratio of all clients in specific group and the total number of clients.

The question is: when a new client comes what will be the probability that the client will default? For example, client is 20 years old with low incomes (under 1,000) and with the primary education only.

$$P(Default|age =' 18\text{-}27', income =' 0-1000', education = 'primary') = \frac{0.667 \times 0.533 \times 0.733 \times 0.075}{0.525 \times 0.415 \times 0.680} = 0.132 \tag{22}$$

$$P(Not\ Default|age =' 18\text{-}27', income =' 0-1000', education = 'primary') = \frac{0.514 \times 0.405 \times 0.676 \times 0.925}{0.525 \times 0.415 \times 0.680} = 0.878 \tag{23}$$

One can see that the sum of both results is not equal to one and it should be equal to one. The problem is in the fact that data is not fully independent and the independency is presumption in Naïve Bayes classification. To get proper results it is necessary to normalize outputs. Denominators are the same in both cases and it is not necessary to use them. Numerator values are P(default) = 0.0195 and P(not default) = 0.1302. The final probability of default will then be $0.0195/(0.0195 + 0.1302) = 13\%$. The result is similar to the equation (22) but it is only by chance and the higher data dependency results in the higher difference between these two numbers.

5.2.4 Decision Trees

C4.5 decision tree classifier [20] was used with pessimistic (top-down) pruning defined by Quinlan [22]. Splitting (creation of branches) was based on the change of Gini index, which is a natural choice for the underlying task. If the change in the Gini index was less than a threshold, the split was not performed, which leads to smaller trees and can prevent overfitting.

More complex decision tree based method is Breiman's decision forest, which averages the response from 50 decision trees, each trained on a bootstrapped version of the training dataset. In each node of the tree training, during the splitting procedure, only one randomly selected feature is considered [23].

The source of Decision Trees section is [20].

5.2.5 Memory Based Reasoning

One of the memory based classification methods is the k-nearest neighbour. This method does not build a model but uses data for classification [16, 25].

Two particular memory based classification methods [20] were applied in classification. First, the k-nearest neighbour classifier simply finds three training data instances that are the most similar to the testing instance and assigns the instance into most common class amongst the three nearest neighbours. The Euclidean distance is used for similarity quantification, because it was observed to lead to good classification accuracies while keeping reasonable computational requirements [24].

Another memory-based method is the nearest mean classifier, which assigns an observation according to its nearest class mean.

The source of Memory Based Reasoning section is [20].

5.2.6 Evolutionary Algorithms

Genetic algorithms (GAs) [26] are stochastic iterative optimization methods based on principles of natural evolution. They maintain a population of candidate solutions, which are recombined and mutated to create a new generation of potential solutions. These then compete with their parents and among themselves: better solutions survive to the next generation.

Genetic programming (GP) [27] is an extension to GAs. The purpose of GP is to evolve computer programs, mathematical expressions, and other similar structures. Grammatical Evolution (GE) [28] is a particular type of GP algorithm allowing the user to prescribe the structure of all potential solutions (programs, expressions, etc.) with a context-free grammar (CFG).

The goal of GE method is to evolve a function of five variables (the economic indicators of a company from variable selection process fully described in 7.1 section p. 48) so that it can be used as a company score (and later to predict company default). Except the five above-mentioned variables, the used CFG contains two constant terms (+1 and -1) and some operators and functions: addition, subtraction, multiplication, division, unary minus, natural exponential (e^x) and natural logarithm ($\ln x$).

The maximum depth of the derivation tree in the initial population was limited to 12. Population size of 1 000 was used and algorithm ran for 50 generations. Crossover, mutation, pruning and duplication probabilities were 0.8, 0.05 (per codon), 0.1 and 0.1. Wrapping the GE genotype was not allowed during decoding process.

Experiment was executed two times, once with Gini as a fitness and one with Kolmogorov-Smirnov (KS) index. In order to lower the possibility of overfitting the training dataset was split into two disjoint subsets, A and B, of equal size. The split was random but preserving the class ratio (i.e. halves of the first class and the second class separately). The evolution was driven by the fitness (Gini or KS) on subset A, i.e. the selection and elitism work with respect to this measure. On the other hand, the so-called best-so-far solution was maintained with respect to the fitness on subset B.

Since GE is inherently stochastic, the algorithm was executed 96 times per experiment, each time with different seed of the pseudo-random number generator, and we present the results in the form of statistics over these 96 runs, on both training data (the full training set) and the testing data.

The source of Evolutionary Algorithms section is [20].

5.3 Overfitting

Overfitting is a situation when model perfectly describes training data, but it has poor predictive performance. The simple example can be show on polynomial regression. The figure 14 shows regression line, first-degree polynomial. Data are very roughly described by this line, residual variance is differs and residuals are not normally distributed. It the reason why this regression is unsatisfactory and this state is underfitting – more regression parameters is required.

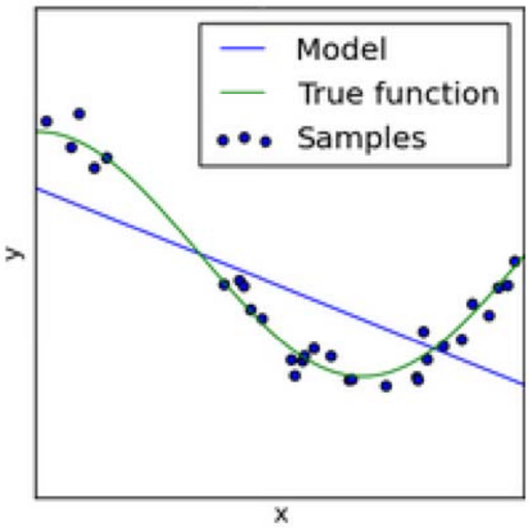


Figure 14: Underfitting, polynomial with degree 1 [61]

The figure 15 shows very good fitting to data. Residuals are distributed randomly and the residual variance looks stable. Fourth-degree polynomial was used as a regression function.

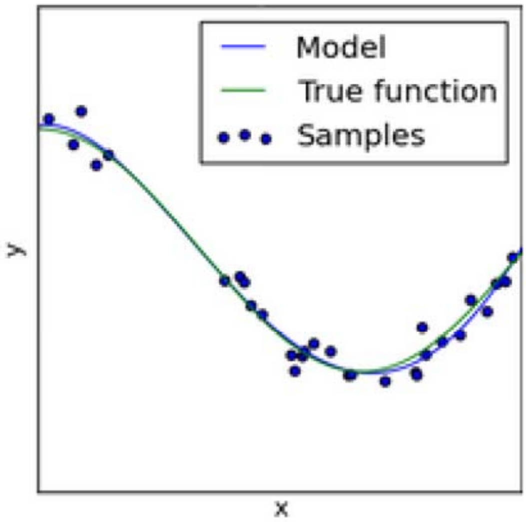


Figure 15: Very good fitting, polynomial with degree 4 [61]

Overfitting is evident on the figure 16. Regression line (15^{th} degree polynomial) perfectly describes data, but one can notice that predictive power will be very low due the model complexity. It is very probable that model will not properly fit on a new sample point around the green line (true function). This means that trade-off between fitting on training data and predictive power have to be done. The simplest possible function should be used as regression function until the regression assumptions are fulfilled (constant residual variance, normal distribution of residuals around regression function).

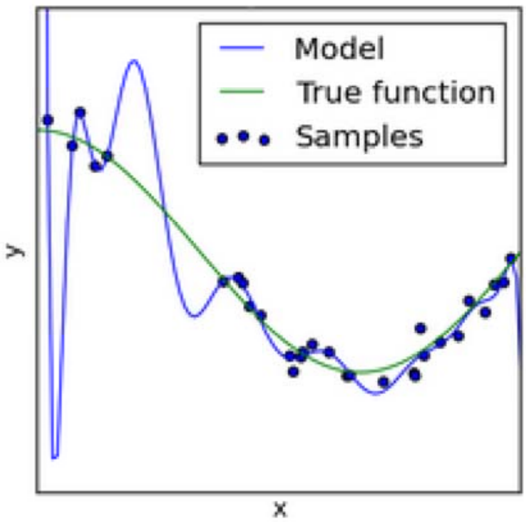


Figure 16: Overfitting, polynomial with degree 15 [61]

The figure 17 shows the dependence of model error on model complexity. The blue line represents training error. Train error (blue line in the figure 17) is decreasing with growing complexity (parameter count). On the other hand the error of testing (or validation or new) data (red line) is rising with rising model complexity. There is optimal model complexity represented by the vertical line with exclamation mark. The testing error is the lowest at this point.

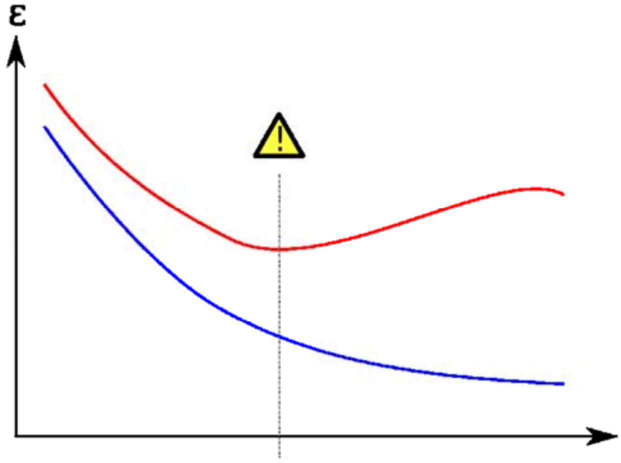


Figure 17: Training error (blue) and validation error (red) based on the model complexity.

From the methods used in this dissertation thesis, the neural networks are the most prone to overfitting. The reason is that besides of input parameters weights also other parameters (mostly in hidden nodes) are present. Regularization method [52] can be used to achieve optimal number of model parameters and avoid overfitting.

5.4 Company Financial Analysis

Financial ratios for model development were selected according to several criterions. Gini index for ratio was calculated to separate variables with weak and strong prediction power. Correlation between variables was also considered because it is not desirable to have strongly correlated input data. Financial ratios can be divided into the four groups: profitability ratios, liquidity ratios, activity ratios, debt ratios. The requirement was to have at least one ratio representing each group to achieve complex view on companies. The more detailed information about financial ratios and their selection and usage in company failure prediction is described here [8, 63].

Financial ratios used in final calculation process are: debt ratio, debt on equity, return on costs, current ratio and payables turnover ratio. These ratios were selected from a wider list, which can be found in *Data Description* section.

Debt ratio is a member of debt ratios group besides debt to equity ratio, debt service coverage ratio, capitalization ratio, interest coverage ratio etc. It is the ratio between company total liabilities and total assets. Debt to equity ratio is the ratio between company total liabilities and equity. These two indicators are highly important in default prediction because high debt value has significant impact on ability of company to service the debt.

Return on costs belongs to profitability ratios family and it is the ratio between company profit and costs. This ratio expresses how much profit brought each invested money unit. Other profitability ratios are return on assets, return or equity return on capital and others.

Current ratio belongs to the group of liquidity ratios. Liquidity ratios measure the availability of cash (or high liquid assets) to cover the debt. Current ratio (also called work capital ratio) is the ratio between current assets and current liabilities.

Payables turnover is a member of activity ratios. It is the ratio between total revenues and payables. This indicator show how many times per period the company pays its payable amount. Activity ratios measure the effectiveness of the firm's use of resources.

5.4.1 Performance Tests

The scoring model performance will be evaluated by Gini and Kolmogorov-Smirnov indexes. These two statistics are the measurement of models' performance and accuracy not only in the field of company evaluation, but generally in binary classification.

Gini Index

Gini index was firstly applied in economics as a measurement of inequality in income distribution. It was developed by Italian statistician Corrado Gini [39] and it is usually defined mathematically from Lorenz curve. The measurement of income distribution inequality problem was further well elaborated by Hugh Dalton [37].

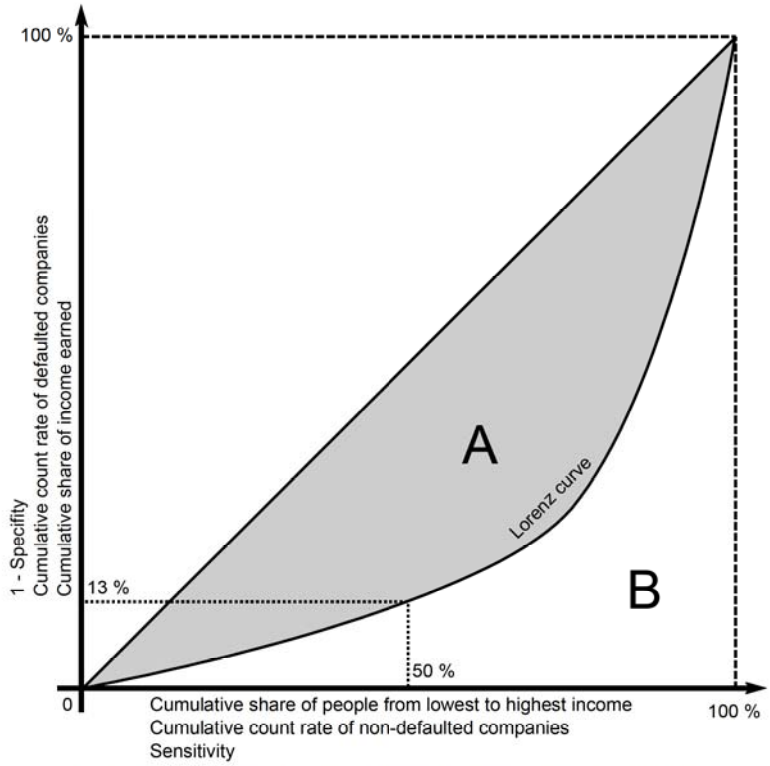


Figure 18: Lorenz curve – curve between areas A and B

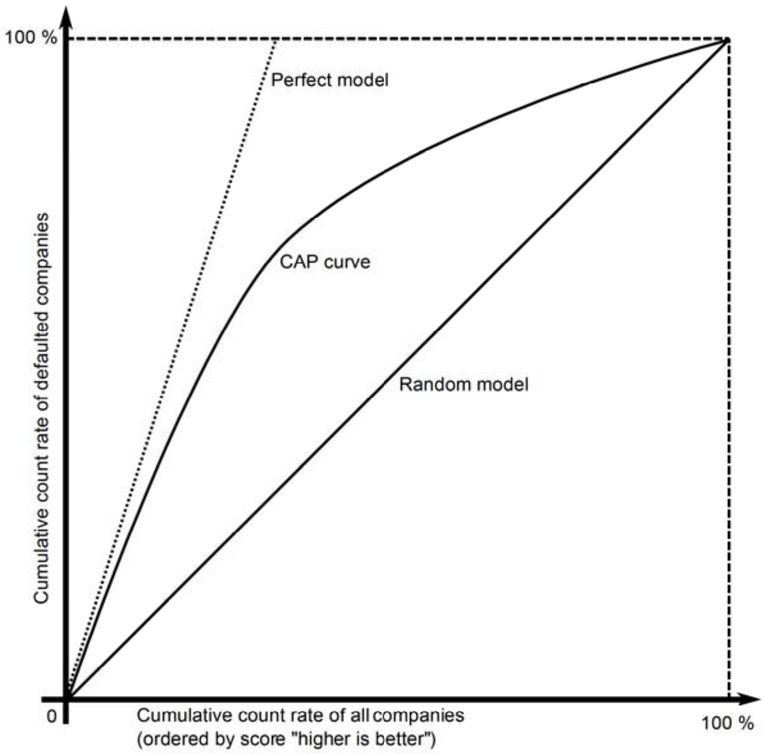


Figure 19: CAP curve

Lorenz curve on figure 18 says that poorest 50 % of population has 13 % of all incomes. In case of company scoring model when the companies are ordered by score in "lower is better" order (e.g. default probability) it says that there was 13 % of defaulted companies among 50 % of non-defaulted companies. Generally, one can use sensitivity[13] and specifity[14] of any binary classification. When the axes are swapped graph turns into receiver operating characteristics (ROC) graph. The figure 19 shows cumulative accuracy profile (CAP) [33, 38].

Gini index value for calculation of income equality was defined as proportion of the area A (area between diagonal and Lorenz curve) and whole area under diagonal (A + B) from figure 18. The extreme case when one person has all incomes leads to 100 % Gini value. For scoring model performance assessment the Gini index can be calculated [19] as the portion of area between random model line and CAP curve and area between random model line and perfect model line from the figure 19. It is equivalent to 2AUROC-1 (AUROC – Area under ROC) from receiver operating characteristics and this value is the same as area 2A in figure 18. Not only the Gini value, but also shape of ROC and CAP curve is important.

The following equation demonstrates Gini index calculation for scoring model. Companies are divided into two sets (D – defaulted, N – non-defaulted). Cartesian product of these sets is set of score pairs. For each pair when element from D set has lower score than element from N set the contribution is positive[15] and vice versa. Result is then divided by the number of pairs. The meaning is that if all defaulted companies have lower score than all non-defaulted

[13] True positive rate. Probability of a positive test if condition is present.
[14] True negative rate. Probability of a negative test if condition is not present.
[15] Depends on ordering logic – higher score is better or lower score is better

companies, model is perfect and the Gini index is 100 %. In case of random model, Gini will be equal or close to zero.

$$Gini = \frac{\sum_{\{i,j\}:D_i<N_j} 1 - \sum_{\{i,j\}:D_i>N_j} 1}{\#D\#N} \tag{24}$$

D – scores set of defaulted companies, N – scores set of non-defaulted companies,
 # – number of elements in set

Standard error of Gini index can be calculated by simplified formula (25) [32, 34, 35].

$$SE_{Gini} = \sqrt{\frac{1 - Gini^2 + (D-1)\left[\frac{4(Gini+1)}{3-Gini} - (Gini+1)^2\right] + (N-1)\left[\frac{4(Gini+1)^2}{3+Gini} - (Gini+1)^2\right]}{\#D\#N}} \tag{25}$$

Kolmogorov-Smirnov Index

Kolmogorov-Smirnov index is the maximum difference between two cumulative distributions. The figure 20 shows the cumulative count rate of defaulted and non-defaulted companies, which are ordered by score. The maximum difference (vertical gap) between these two graphs represents Kolmogorov-Smirnov index value.

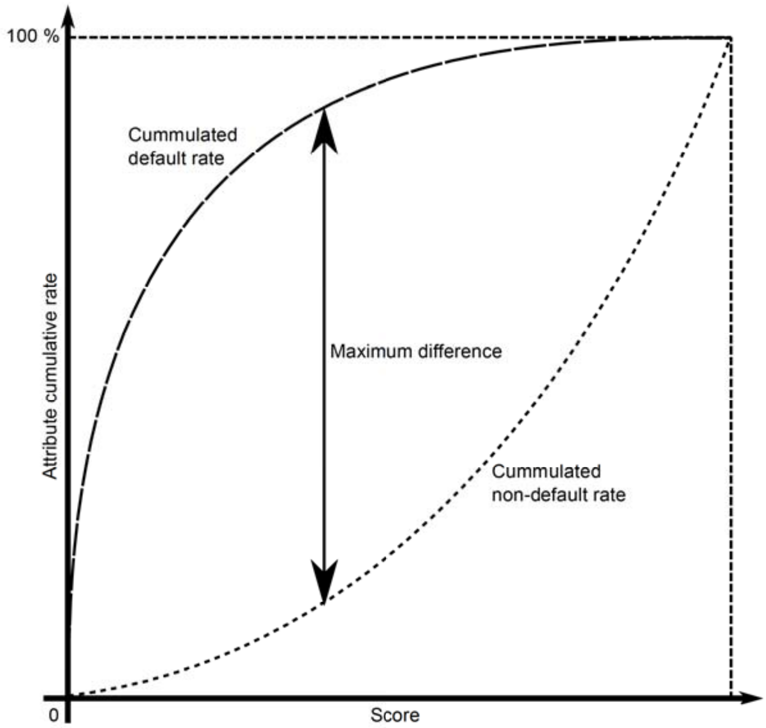


Figure 20: Kolmogorov-Smirnov Index

$$KS = \max\left|C_{D,S} - C_{N,S}\right| \tag{26}$$

$C_{D,S}$ – cumulated count rate of defaulted companies up to specific score value
$C_{N,S}$ – cumulated count rate of non-defaulted companies up to specific score value

The scoring model with high accuracy and predicting power has higher Kolmogorov-Smirnov index value because defaulted companies have lower score than non-defaulted companies.

The most of the defaulted companies are cumulated in the area of low score and non-defaulted companies are cumulated in the area of higher score. Here [36] is described information about usage of Kolmogorov-Smirnov index in ROC curve in more details.

5.5 Benchmarking Methods

This section describes benchmarking models used for relative comparison of companies. The part of this thesis is devoted to comparison of results from proposed method with the results from other dissertation thesis [6], which aim was to create a new benchmarking method. Benchmarking methods as well as the proposed methods (and other classifiers) can be used to form relative performance order of companies. This comparison is output of the research cooperation at the Czech Technical University in Prague.

5.5.1 Referential Point Method

The method [6] is based on the weighted average of relative performance indicators [31]. Each indicator can have values between 0 and 100, which represent relative performance. Score calculation is shown in the equation (27). Usage of weights vector is not compulsory. The weights can be used when there is need to have some indicators more important than the others.

$$Score = \frac{\sum_{i=1}^{n} q_i w_i}{\sum_{i=1}^{n} w_i} \tag{27}$$

q_i – indicator value, w_i – indicator weight

For each indicator a company with highest input parameter value is found. This parameter is for a found company set to 100 value. The other companies' parameter values are relatively assessed by equation (28).

$$q = \frac{x - x_{min}}{x_{max} - x_{min}} \times 100 \tag{28}$$

q – indicator value, x – actual parameter value,
x_{max} – maximal parameter value, x_{min} – minimal parameter value

This method is intuitive and simple to use. The main disadvantage of this method is influence of extreme values on the results. Therefore input values check have to be performed and possible corrections (data transformation) should be undergone upon score calculation.

5.5.2 CCR and BCC Model

The CCR (Charnes, Cooper, Rhodes) model [29] is based on data envelopment analysis methods [30]. The data envelopment analysis is linear programming method based on relative performance of decision-making units. Decision-making units can be in thesis understood individual companies. Inputs can be capital expenditures, operating costs, employers count etc. and outputs can be represented by revenues, profit, economic value added etc. This is only example regarding evaluation of companies.

$$Score = \frac{\sum_{j=1}^{J} out_j wo_j}{\sum_{i=1}^{I} in_i wi_i} \quad (29)$$

I – inputs count, wi_i – weight of the i-th input, in_i – value of the i-th input
J – outputs count, wo_j – weight of the j-th output, out_j – value of the j-th output

The equation (29) shows the calculation by the data envelopment analysis method. It is the ratio of weighted sum of inputs and weighted sum of outputs.

Own CCR model performance is evaluated by the equation (29). The aim is to find out optimal weights values for which a decision-making unit achieves the maximal performance. The maximum performance value is 1 (or 100 %) and for a given set of weights, the performance have to be lower or equal to this value. Maximizing of the equation (29) leads to the linear-fractional programming task where the numerator is maximized and the denominator is minimized. It is necessary to transform this task to the linear programming task, which is done using Charnes-Cooper transformation [6]. The fundament of CCR method is in assumption of constant returns to scale (CRS).

The BCC (Banker, Charnes, Cooper) model [30] is the modification of CCR model that makes possible assuming variable return to scale (VRS). Implementation of this model is subjected to constraint of convexity of data envelope. The CCR and BCC models are mostly used data envelopment analysis models. There is variety of extensions of these two models.

5.5.3 "Y" Performance Index

Construction of "Y" performance index was the target of dissertation thesis [6]. This index is the result of regression methods and includes six independent input variables.

$$Y = 1.087 \times \frac{1}{x_1} + 327.134 \times \frac{1}{x_2} + 77.828 \times \frac{1}{x_3} + 5.854 \times \frac{1}{x_4} + 2.437 \times x_5 + 0.322 \times x_6 \quad (30)$$

The x_1 parameter is the Customer Average Interruption Duration Index (CAIDI), which is represented by average time (in minutes) per interruption of electricity supply. The x_2 parameter is the ratio of revenues and the amount of electricity distributed (GWh). Parameter x_3 is the ratio of energy losses to distributed energy, x_4 represents operating costs without depreciations divided by long-term assets. Parameter x_5 is the ratio of the earnings before tax (EBT) and long-term assets. Finally, the x_6 is the ratio of personal costs to the number of employers.

5.6 Statistical Errors

5.6.1 Hypothesis Testing

A statistician important task is to draw inferences about population based on the sample taken from the population. These inferences are usually about some population parameters such as mean value, variance or probability of success. Statistically significant results in statistics means that it has been predicted as unlikely to occur by chance alone according to pre-determined significance level (threshold probability) [64].

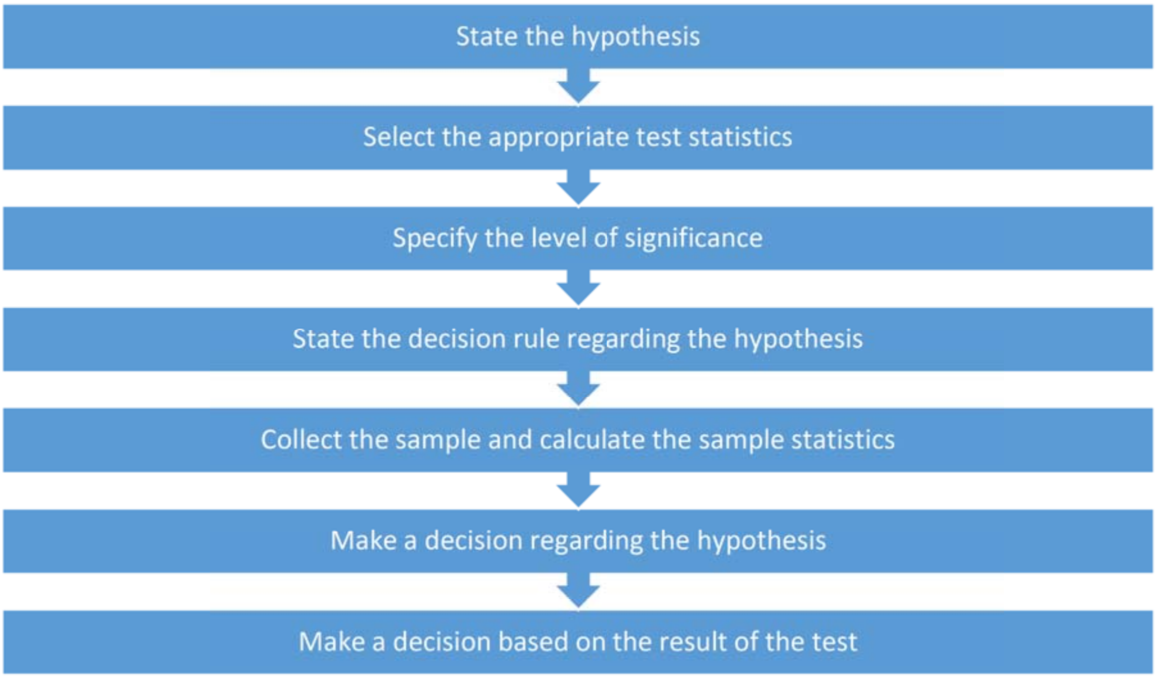


Figure 21: Hypothesis testing procedure [65]

Null and Alternate Hypothesis

An essential part of statistical testing is setting up and testing hypotheses.

Null hypothesis H_0: This is a statement being tested in a test of statistical significance. The significance test assesses the strength of evidence against the null hypotheses. This means in fact that null hypothesis should be rejected or not to be rejected but there is no sense to talk about hypothesis confirmation. When the hypothesis is not rejected, it does not automatically mean that it is accepted, but there is only not sufficient evidence against it. Null hypothesis usually represents a statement of "no effect" or "no difference" [66]. An example of null hypothesis is "The difference between heating value of new fuel and currently used fuel is not significant"

Alternate hypothesis H_A or H_1: When null hypothesis is rejected, the alternate hypothesis is adopted. It means that alternate hypothesis must differ from null hypothesis, but it does not have to be exhaustively logical negation of null hypothesis. In any case, complete logical space must be covered. An example of alternate hypothesis is "The new fuel has higher heating value than currently used fuel". In case that new fuel cannot be worse than currently used one, here stated alternate hypothesis is satisfactory. Otherwise, a logical negation should be used to cover case that new fuel heating value is lower than the currently used fuel heating value. This hypothesis would look like "The difference between heating value of new fuel and currently used fuel is significant".

The result of testing can be rejection of H_0 in favour of H_1 or not rejection of H_0. There is no testing result of rejecting H_1, accepting H_1 or H_0.

Error types

The problem occurs when true hypothesis is rejected or when false hypothesis is accepted. Both of these is problem because it may lead to incorrect decisions.

A type I error (error of the first kind) occurs in case that null hypothesis is rejected when it is true. A type II error (error of the second kind) occurs when null hypothesis is not rejected when it is false.

		Decision	
		Reject H_0	Do not reject H_0
Truth	H_0	Type I error (False positive)	Right decision (True negative)
	H_1	Right decision (True positive)	Type II error (False negative)

Table 3: Hypothesis test results

A type I error is considered as more serious and therefore it is more important to avoid type I error than type II error. The probability of type I error is called **significance level** and it is usually denoted as α (alpha). The probability of type II error is generally unknown and it is symbolized by β (beta). The probability (1 – β) of not committing type II error is called **power of the test**. These two types of errors are inversely related which means that by lowering of type I error probability the type II error probability grows and vice versa. Also exists type III error, which occurs when the conclusions drawn are not supported by the data presented. In other words type III error occurs when study provides the right answer but for the wrong research hypothesis.

Types of Statistical Tests

Statistical test is left-tailed if H_1 states that the parameter value is lower than the value claimed in H_0. In right-tailed test, H_1 states that the parameter value is lower than the value claimed in H_0. The two-tailed test is based on the H_1 statement that parameter value is not equal to the value claimed in H_0.

Decision Rules

P-Value (probability of chance)

It is the probability of incorrectly rejecting the null hypothesis if it is true. If the P-value is less or equal to significance level the null hypothesis must be rejected. The P-value is the probability of observing test statistics at its value [67]. For a statistician the P-value is much more important than just the fact whether hypothesis is rejected or not. By knowing the P-value, one can see how far the P-value is from a significance level and asses the strength of the test.

Region of Acceptance

It is the range of values where the chance of making type I error is equal to significance level. If the statistics falls within this region, the null hypothesis is not rejected. Values outside of the region of acceptance are lying in the region of rejection (critical region) and if test statistics falls within this region, the null hypothesis is rejected.

Critical Value

The critical value is the value of test statistics for which the calculated P-value corresponds to significance level. Comparison of the critical value and actual test statistics value leads to the same decision as using of P-value and significance level.

Hypothesis tests

There is plenty of various hypothesis tests. Here is the list of the most frequently used tests: proportion test, difference between proportions, mean, difference between means, variance, ratio of variances, difference between pairs, goodness of fit, homogeneity, independence, regression coefficients test etc. The aim of this thesis is not the deep research on various hypothesis tests but the most relevant is testing of errors in statistical models.

5.6.2 Errors in Binary Classification Models

Error testing is one of the most important tasks, which has to be done for newly created statistical model. Two of the most useful performance indicators for binary classification models are sensitivity and specificity.

Performance Indicators

Sensitivity refers to true positive rate (TPR). It relates to model ability for correct condition identification. Sensitivity is the ratio of true positives to all positives in population.

$$Sensitivity = TPR = \frac{true\ positives\ count}{true\ positives\ count + false\ negatives\ count} \tag{31}$$

The denominator is the sum of true positives and false negatives, where false negatives are essentially positives so the sum of these two numbers is equal to all positives count in population.

Specificity refers to the true negative rate (TNR). Higher specificity means that binary model correctly does not identify attribute when it is not present. It is the ratio of true negatives to all negatives in population.

$$Specificity = TNR = \frac{true\ negatives\ count}{true\ negatives\ count + false\ positives\ count} \tag{32}$$

The denominator is the sum of true negatives and false positives, where false positives are essentially negatives so the sum of these two numbers is equal to all negatives count in population.

False positive rate (FPR) refers to type I error and is equal to α.

$$FPR = \frac{false\ positives\ count}{false\ positives\ count + true\ negatives\ count} = 1 - Specificity = \alpha \tag{33}$$

False negative rate (FNR) refers to type II error and it is equal to β.

$$FNR = \frac{false\ negatives\ count}{false\ negatives\ count + true\ positives\ count} = 1 - Sensitivity = \beta \tag{34}$$

The power of the test is equal to (1 – β) and therefore it also equals to sensitivity value. The equation (35) represents model accuracy (ACC). It means that when no errors occur the accuracy is 100 % because the sum of true positives and true negatives must be equal to total population count. Specificity and sensitivity is directly related to type I and type II errors so these two indicators are also inversely related. It means that sensitivity is increasing when specificity is decreasing and vice versa.

$$ACC = \frac{true\ positives\ count + true\ negatives\ count}{total\ \ population\ count} \tag{35}$$

More measures of performance can be found in diagram located in appendix section at figure 1. More detailed errors description and model predictive power measurement can be found in [68-70].

6 Data Description

Data entering training and validation processes are financial ratios from financial reports:

- DR – debt ratio,
- DOE – debt on equity
- ROC – return on costs
- CR – current ratio
- PT – payables turnover
- Information whether company defaulted in one year.

The detailed description of used ratios is later in this section and in *Company Financial Analysis* section on page 32.

The training dataset size is 459 observations with 33 % of defaulted companies and validation dataset has 2 661 observations with 2.5 % of defaults. The real default rate in observed population 2.5 % looks like very small portion but it is still 30 % of all defaults available. Learning from imbalanced data is complex [71] and the 33 % default rate in training dataset was achieved by using random subsample of non-default data. The training and validation datasets are also divided by time boundary. All training data come from earlier date than the validation data. Financial ratios are input variables and company failure information is dichotomous target variable. A company will be considered as failed (defaulted) when it fails in one year from its financial statement date. This is important because various approaches can be used here.

The advantage of using observation period starting from the financial statement date is that the data from statements are actual. The disadvantage can be in fact that this date will not be the same for all companies. Most of observed companies have fiscal year starting 1st of January, but there are companies that have start of the fiscal year moved to different date. There are also companies that do not have available actual financial statements and the old statements are used. All of these issues result in time differences in observation period start date. Financial data from wide time span can make model less sensitive on quick changes. For example, it is sufficient to look at the companies before and after the last economic recession. If the model would cover financial statements from the period before recession and after recession, it would be more stable, but on the other hand, model would be less precise. Precision would be reduced due the fact that companies' behaviour before recession is different from now and the financial statements from this period would distort the model. On the other hand, if similar conditions as in the period before recession occur, the model can recognize this situation and give good prediction. When the training dataset contains data from short time span, the model will have better prediction performance than the model described before. On the other hand, it will be very inaccurate when economy conditions suddenly change. It is important to decide for how long should be model in use and choose training data according this presumption. The limiting factor is also the data availability. When there is a lack of relevant data the timespan can be broaden to have enough data.

More information about company failures and company financial ratios are in the following subsections. Information about variables selection, normalization and transformation will be covered in the *Proposed Method* section.

6.1 Company Failure

A company failure definition can be the complex task and this thesis aim is to propose new method regardless of detailed failure examination. The process of the model development is not dependent on strict failure definition. Anyway, the company failure can be understood in many ways:

- Company bankrupt - a legal finding that imposes court supervision over the financial affairs of those who are insolvent or in default.
- Insolvency - a legal term meaning that a debtor is unable to pay his or her debts.
- Appearance of company in other commercial registers.
- Default in meaning that debtor has not paid a debt which he or she is required to have paid. [72, 73]

For early failure detection, the failure to repay the debt is chronologically the first signal that warns about company financial problems. Appearance in insolvency register or possible bankrupt may occur later. It depends on usage purposes which data is most suitable for specific task. The default prediction is mainly needed for creditors that lend money to companies and other entities. These subjects are usually banks, leasing companies and other financial non-banking subjects, which are allowed to lend money. On the other hand, the companies that fail to repay debt in specified time are defaulted according to BASEL II [73] standard, but it does not mean that these companies will not pay its debt[16]. The failure prediction can be also useful for business partners. They usually do not prepare[17] own models, but they can use proprietary solutions[18]. For these purposes, other kind of default can be satisfactory. In any case, the BASEL II definition is the strictest and it deserves more attention. A default must be considered to have occurred with regard to a particular obligor when either or both of the two following events has taken place [72, 73]:

- The firm considers that the obligor is unlikely to pay its credit obligations to the firm, the parent undertaking or any of its subsidiary undertakings in full, without recourse by the firm to actions such as realising security (if held).
- The obligor is past due more than 90 days on any material credit obligation to the firm, the parent undertaking or any of its subsidiary undertakings.

6.2 Financial Ratios

This section describes fundamental financial ratios and covers all ratios that were examined during variables selection procedure. The abbreviations of ratios' names in brackets are

[16] Moreover, companies that do not pay on time but pays later are the best clients for banks because besides regular payments they pay fines and penalties. Therefore, the defaulted client according to BASEL II does not automatically mean lossmaking client.

[17] It is time consuming and the deeper knowledge of the problem is required.

[18] One of these solutions is Albertina system (http://www.albertina.cz/). It holds aggregated data about the companies from various registers. One of its functionality is the default prediction. It is actually available in the Czech Republic and Slovakia.

further used in analysis. For some ratios such as profitability ratios is favourable high value of ratio. Some ratios are tricky because it depends on the point of view whether its value is ideal. For example, low debt ratio is good for creditors to lend the money, however it is not ideal for company stakeholders. The ideal ratio value differs between industry types and therefore it is often necessary to compare individual ratio value to average characteristic for company's industrial sector. It is also important to note that there is no strict calculation procedure for these ratios. For example, the usage of profit can vary in different sources. Sometimes gross profit is used; sometimes the net profit is used. This is also typical for revenues, sales and net sales. Therefore, it is crucial to know also the calculation process of ratios when comparing to the final values.

Debt and Structure Ratios

Debt ratio shows how much are liabilities covered by assets. This indicator is important for creditors because the higher this ratio is, the higher the probability of default is.

$$Debt\ Ratio\ (DR) = \frac{Total\ Liabilities\ (Total\ Debt)}{Total\ Assets} \tag{36}$$

Interest coverage ratio is used to determine how easily a company can pay interest on debt. The lower this ratio is, the higher risky the company is and the higher probability of default is.

$$Interest\ Coverage\ Ratio\ (IC) = \frac{EBIT(Operating\ Profit)}{Interest\ Expense} \tag{37}$$

Debt on equity ratio is a measure of a company's financial leverage. This ratio presents the basic capital structure that is used for financing of assets.

$$Debt\ On\ Equity\ (DOE) = \frac{Total\ Liabilities}{Equity} \tag{38}$$

Equity to fixed assets ratio (overcapitalization) shows exposure of company owners and indirectly debt holders to the fixed assets. The lower ratio is the higher risky company is.

$$Equity\ to\ Fixed\ Assets\ Ratio\ (CAP) = \frac{Equity}{Fixed\ Assets} \tag{39}$$

Liquidity Ratios

Current ratio ascertains whether a company's short-term (current) assets are available to pay off its short-term (current) liabilities. The higher ratio value, the better.

$$Current\ Ratio\ (CR) = \frac{Current\ Assets}{Current\ Liabilities} \tag{40}$$

Quick ratio is similar to previous one, but excludes inventories. It shows ability of company to cover short-term liabilities without need of selling inventories.

$$Quick\ Ratio\ (QR) = \frac{Current\ Assets - Inventories}{Current\ Liabilities} \tag{41}$$

Cash ratio is conservative indicator that describes a company's ability to cover its short-term liabilities. The disadvantage of previous ratios is in including inventories and receivables.

Inventories and receivables cannot be used quickly for covering current liabilities. Moreover, the inventories usually cannot be sold at price they were bought so covering short-term debt by inventories leads to losses and can violate company operation.

$$Cash\ Ratio\ (CASHR) = \frac{Cash + Cash\ Equivalents}{Current\ Liabilities} \tag{42}$$

Profitability Ratios

Gross profit margin is presenting company's ability to control its production costs, which are denoted as Cost of Goods Sold (COGS). The higher the Cost of goods sold is, the lower profit margin is and the lower profit is. Maximum value of this ratio is one. This ratio does not tell directly information about company credit-risks, but the higher profit margin means higher probability of covering liabilities.

$$Gross\ Profit\ Margin = \frac{Gross\ Profit}{Revenues} = \frac{Revenues - Cost\ of\ Goods\ Sold}{Revenues} \tag{43}$$

Operating profit margin ratio indicates how much profit a company makes after paying for variable costs of production such. This ratio expresses efficiency of company core business and does not include exceptional situations.

$$Operating\ Profit\ Margin = \frac{EBIT}{Revenues} \tag{44}$$

Return on capital employed (ROCE) ratio shows what returns were made on the resources available.

$$Return\ on\ Capital\ Employed\ (ROCE) = \frac{EBIT}{Total\ Assets} \tag{45}$$

Return on assets ratio is one of the common ratios used and it expresses the utilization of assets in generating earnings. The higher ROA value means that efficiency of generating earnings by assets is good.

$$Return\ on\ Assets\ (ROA) = \frac{Net\ Income\ (EAT)}{Total\ Assets} \tag{46}$$

This ratio is similar to previous one and explains corporation's profitability by revealing how much earnings is generated by the money invested by shareholders.

$$Return\ on\ Equity\ (ROE) = \frac{Net\ Income\ (EAT)}{Total\ Equity} \tag{47}$$

Return on sales ratio is often used as a measure of company's operational efficiency. It explains how much profit is being produced per money unit of sales. The higher value is better. Sales is the sum of revenues from the goods sold and revenues from own products and services.

$$Return\ on\ Sales\ (ROS) = \frac{Net\ Income\ (EAT)}{Sales} \tag{48}$$

Return on costs shows company's profitability based on incurred costs. It is supplemental ratio to return on sales ratio.

$$Return\ on\ Costs\ (ROC) = \frac{Net\ Income\ (EAT)}{Costs} \tag{49}$$

Efficiency (Activity) Ratios

Collection period ratio shows amount of time that is needed for business to receive payments from customers and clients. The lower value means the better situation for company.

Net sales in this group of financial indicators is the sum of revenues from the goods sold, revenues from own products and services, revenues from disposals of long-term assets and material, interests and revenues from the sales of securities reduced by sales returns, allowances and discounts.

$$Collection\ Period\ Ratio = \frac{Accounts\ Receivable \times Days(365)}{Net\ Sales} \tag{50}$$

Sales to inventory ratio is also known as inventory turnover. The higher than needed state of inventories level leads to ineffective utilisation of money because the profitability of unused inventories is zero.

$$Sales\ to\ Inventory\ Ratio = \frac{Net\ Sales}{Inventories} \tag{51}$$

Net working capital is the difference between the current assets and current liabilities. Sales to net working capital also called working capital turnover is the measurement how effectively a company is using its working capital to generate sales.

$$Working\ Capital\ Turnover = \frac{Net\ Sales}{Net\ Working\ Capital} \tag{52}$$

Asset turnover ratio measures company's ability to use its assets for generating revenues. In other words, how much revenues is generated per each money unit of company total assets.

$$Assets\ Turnover\ (AT) = \frac{Revenues}{Total\ Assets} \tag{53}$$

Fixed assets turnover ratio is similar to asset turnover ratio, but only fixed assets are considered. It is due the investigation of efficiency of money generation from fixed assets. Fixed assets are mostly property, plants and equipment (PP&E) reduced by accumulated depreciations of these assets.

$$Fixed\ Assets\ Turnover\ (FAT) = \frac{Revenues}{Fixed\ Assets} \tag{54}$$

Receivables turnover measures company's ability effectively collect its receivables. Also alternative with net credit sales in numerator is often used due to the fact that cash sales do not generate receivables [74].

$$Receivables\ Turnover\ (RT) = \frac{Revenues}{Receivables} \tag{55}$$

Payables turnover ratio shows the rate at which a company pays off its suppliers. This ratio is very important for creditors because helps analyse how quickly can pay off its bills.

$$Payables\ Turnover\ (PT) = \frac{Revenues}{Payables} \tag{56}$$

Cash-Flow Ratios

The cash-flow ratios in this section were chosen for investigation how is the cash generated able to cover company's long-term debt, all liabilities, interest from debt and total debt service.

$$CFO\ To\ Long\ Term\ Debt\ (CFLTD) = \frac{CF\ From\ Operations}{Long\ Term\ Debt} \tag{57}$$

$$CFO\ To\ Liabilities\ (CFL) = \frac{CF\ From\ Operations}{Liabilities} \tag{58}$$

$$CFO\ To\ Interest\ Paid\ (CFINT) = \frac{CF\ From\ Operations}{Interest\ Paid} \tag{59}$$

$$Debt\ Service\ Coverage\ Ratio\ (DSCR) = \frac{CF\ From\ Operations}{Debt\ Payable + Interest\ Paid} \tag{60}$$

One of the most often used financial ratio is Debt Service Coverage Ratio. It say whether company is capable to fulfil its long-term obligations. Its value should be higher than one because value equal one means that all cash-flow from operations is used to cover debt. Moreover, in this case company would not have enough cash for new investments. The problem of this ratio is to find out the value of debt payable. It is not present in financial statements and difference in long-term debt between consecutive years do not guarantee its value because new long-term loans can show up. Therefore, DSCR will not be used in further analyses.

7 Proposed Method

This section covers the core of the new method preparation process. It begins with the variables analysis and selection, various transformations and finally results is the new model. Some presumptions and relations are already described in previous parts of this thesis.

7.1 Variables Transformation and Selection

The table (4) contains financial ratios that were considered for the final ratios (variables) selection. The idea was to get five final variables and have at least one variable from each group.

Profitability Ratios	
Return On Assets	ROA
Return On Equity	ROE
Return On Capital Employed	ROCE
Return On Sales	ROS
Return On Costs	ROC
Liquidity Ratios	
Current Ratio	CR
Quick Ratio	QR
Cash Ratio	CASHR
Efficiency Ratios	
Asset Turnover	AT
Fixed Assets Turnover	FAT
Receivables Turnover	RT
Payables Turnover	PT
Debt Ratios	
Debt Ratio	DR
Debt on Equity	DOE
Interest Coverage Reciprocal	IC
Over/under capitalization	CAP
CF Ratios	
CFO To Long Term Debt	CFLTD
CFO To Liabilities	CFL
CFO to Interest paid	CFINT

Table 4: financial ratios considered for use in model

Error Values Handling

All of the variables are ratios of items from financial statements. The first issue that had to be solved is zero in denominator. When this situation occurred, the final ratio value was set to maximum or minimum value according to numerator sign. For example if the costs were equal to zero in return on costs ratio, the maximum ratio value in training data was 150, minimum value was -100 and the numerator sign was negative, the -100 value (minimal negative value) was used as ratio value for this specific case. If numerator sign value was positive, maximum positive value from training data (150) would be used. There are various procedures how this situation can be solved. For example, zero value or median/average value can be used. The numerator is the number much higher than zero, denominator is zero so the limit is going to infinity. Due the practice purposes, the maximum/minimum value was chosen. In cash flow indicators, the values 10 and -10 were used in case of zero denominator. The reason was that denominator expresses debt in these indicators and when company has no debt, there is no sense to put here extreme values.

Extreme Values Handling

When looking on distribution of ratio values (Appendix Table 1) one can notice that there are extreme values. For example, ROE minimum value is -4029.5 and maximum value 2810. The values in 1st and 99th percentiles are -4.45 and 5.87 respectively. There is big difference and therefore maximum and minimum ratio value was set to 1st, 2nd, 5th or other percentile value according to values and logic for each ratio. These cut-off values for each ratio are presented in Appendix Figure 2.

Selection Process

The next step is to select variables with the highest performance. As well as for model performance measurement, Gini coefficient can be used also for each variable performance measurement. Another widely used approach of variable performance measurement for binary classification tasks is information value (IV) [75, 76].

$$IV = \sum_{i=1}^{n} (N_i - D_i) \times ln\left(\frac{N_i}{D_i}\right) \tag{61}$$

n – number of groups, N_i – portion of non-defaulted client in i-th group and all non-defaulted clients, D_i – portion of defaulted client in i-th group and all defaulted clients

The equation (61) shows information value calculation. Continuous variable has to be divided into groups to calculate information value. Gini and IV values for selected financial ratios are in the table 5.

Variable	Label	Gini Statistic	Information Value
DOE	Debt on Equity: Liabilities / Equity	33,709	0,387
T4_ETL	Book Value of Equity / Total Liabilities	33,148	0,398
DR	Debt Ratio: Liabilities / Total Assets	32,768	0,378
CAP	Over/under capitalization: Equity / Fixed Assets	30,763	0,325
CR	Current Ratio: Curr Assets / Curr Liabilities	27,954	0,271
QR	Quick Ratio: (Curr Assets - Inv) / Short Term Payables	25,544	0,224
T1_WCTA	(Current Assets – Current Liabilities) / Total Assets	25,343	0,225
ROC	Return On Costs	24,857	0,221
ROE	Return On Equity	24,275	0,203
PT	Payables Turnover: Revenues / Payables	23,351	0,242
ROA	Return On Assets	22,981	0,195
T2_RETA	Retained Earnings / Total Assets	21,804	0,158
ROS	Return On Sales	20,068	0,145
ROCE	Return On Capital Employed	19,324	0,119
RT	Receivables Turnover: Revenues / Receivables	17,671	0,116
IC	Interest Coverage Reciprocal: Interest Paid / Operating Profit	17,642	0,109
CASHR	Cash Ratio	16,143	0,084
T3_EBITTA	Earnings Before Interest and Taxes / Total Assets	15,397	0,077
CFINT	CFO to Interest paid	14,942	0,072
FAT	Fixed Assets Turnover: Revenues / Fixed Assets	13,564	0,071
CFL	CFO To Liabilities	8,206	0,024
DSCR	Debt Service Coverage Ratio from CF	7,286	0,017
CFLTD	CFO To Long Term Debt	6,945	0,016
T5_STA	Sales / Total Assets	5,734	0,011
AT	Asset Turnover: Revenues* / Total Assets	4,49	0,007

Table 5: Gini coefficient and information value for variables

Ratios in cash-flow group have low performance and they will not be used in further analysis. Five ratios were chosen with requirement of having at least one ratio from each group except of cash-flow group and variables beginning with "T" which are used in Altman score calculation. Ratios DR, DOE, ROC, CR, PT were finally chosen for model development. The correlation between variables was also considered. Appendix Table 2 shows the correlation matrix between all variables. Table 6 shows correlation between final variables. Correlations should be close to zero in ideal case.

	ROC	CR	DR	DOE	PT
ROC	1	0,035701	-0,19172	0,020875	0,119714
CR		1	-0,23356	-0,05178	0,276151
DR			1	0,061868	-0,39638
DOE				1	-0,05841
PT					1

Table 6: Correlation matrix for five final variables

The highest correlation is between debt ratio (DR) and payables turnover (PT). This can be assessed as moderate correlation and still acceptable for model development. Performance of the second best variable from efficiency ratios group is significantly lower than payables turnover. Therefore, payables turnover variable was preserved. There can probably be prepared (constructed from financial statements' items) also better performing ratios than used in this thesis but it is not the aim of this work.

Variables Grouping

Variable grouping is the process where the continuous variable is discretized. There can be many reasons to perform grouping and the original idea was not to do it in this thesis due to the data characteristic described in the next paragraph. The figure 22 shows the dependency of default rate on debt ratio value. High debt ratio value means that a company has a high debt and it is more prone to default than a company with low debt level. The function of default rate should be increasing. The data for the figure 22 are divided into 20 groups (by 5 percentiles) and the column height represents default rate (percent of defaulted companies in given percentile).

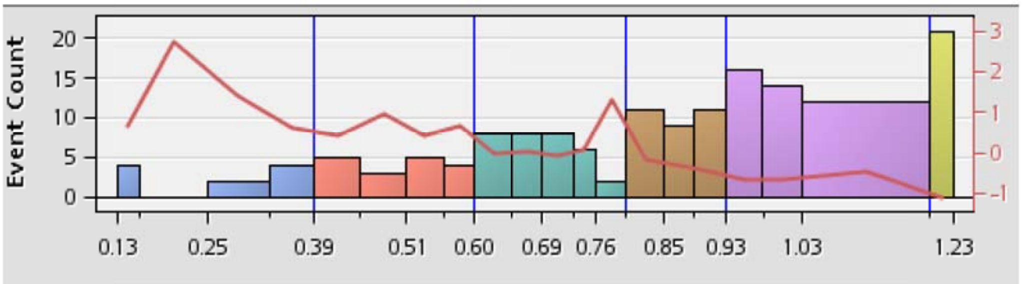


Figure 22 : Debt ratio default rate by quantiles.
Diagrams for other ratios are in appendix section (Appendix Figure 3 - Appendix Figure 6)

One can see that default rate function is not increasing. Model prepared on these type of data can behave unstably. Model would assign good score for this variable when variable value would be 0.8 and worse score in case variable value 0.4. This is not desirable and this data shortcoming is caused by not very good data quality or by relatively small training dataset. Therefore variables should be grouped into smaller number of groups and achieve the monotonic default rate function. The problem is that information from variable is reducing by the process of discretizing. Therefore, the model will be stable, but the performance of individual variables will be slightly lower than before discretizing. Optimal number of groups was selected to five and six based on the data and several trials with the number of groups for variables.

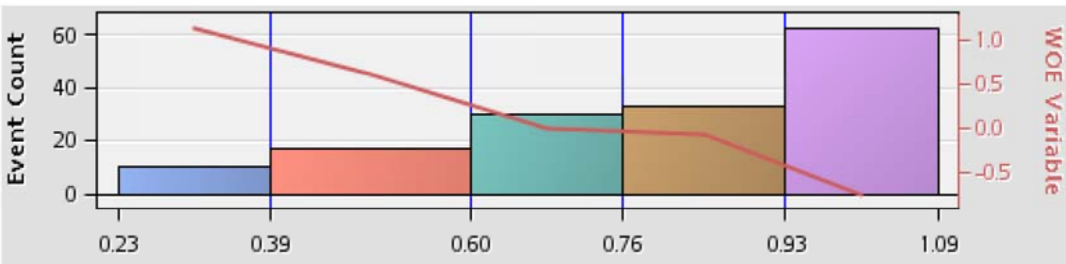


Figure 23: Debt ratio binning

Debt ratio binning looks well, the default rate is growing with growing variable value.

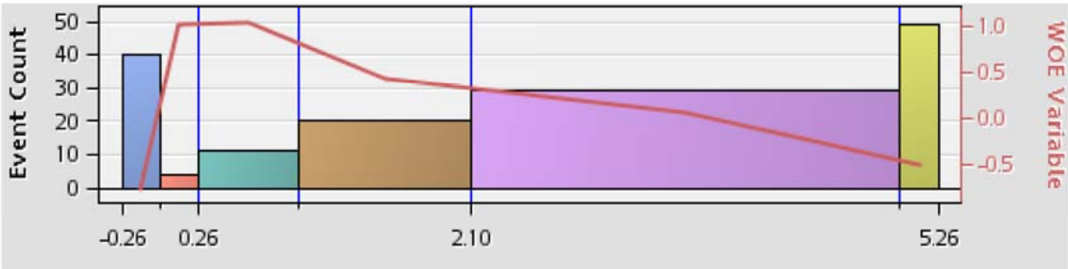


Figure 24: Debt on equity binning

Debt on equity ratio has high default rate for negative values. This is caused by the negative numerator, which represents equity value. One can assume that companies with negative equity can be more prone to default than companies with positive equity.

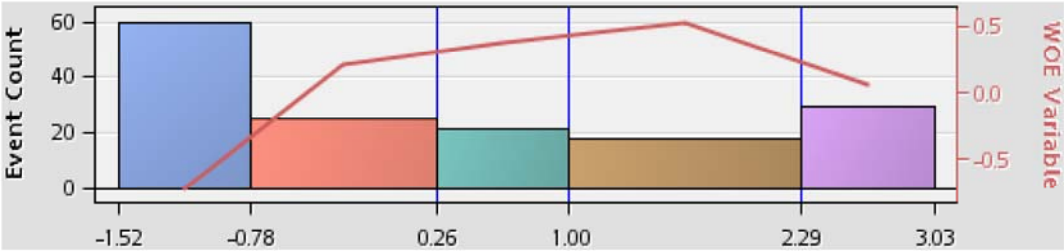


Figure 25: Return on costs binning

Return on costs has high default rate for negative values. This is caused by negative nominator, which represent profit (and loss). Companies with negative profit are more probable to default than companies with positive profit. One can notice that the last group for the highest ROC values has higher default rate than groups with lower ROC value. This can be caused for example by start-ups, which can have high profit to costs value, but these companies are less stable than companies existing for a years and therefore can be more sensitive and risky.

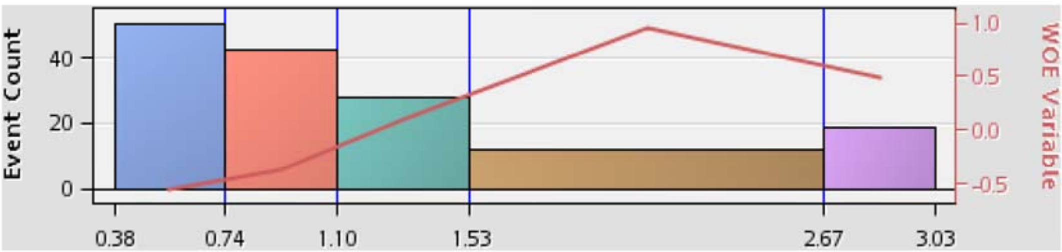


Figure 26: Current ratio binning

Current ratio also corresponds to its logic. The last group have higher default rate than the group before the last one. This can be caused by companies with low absolute value of current assets. The ratio of current assets and current liabilities can be high, but this can be achieved also when absolute values of these items are low. These companies can be more sensitive to changes of market condition and can be more prone to default. Separate models can be developed to avoid problem of companies' size. On the other hand, creating more models would shrink amount of training data for each model.

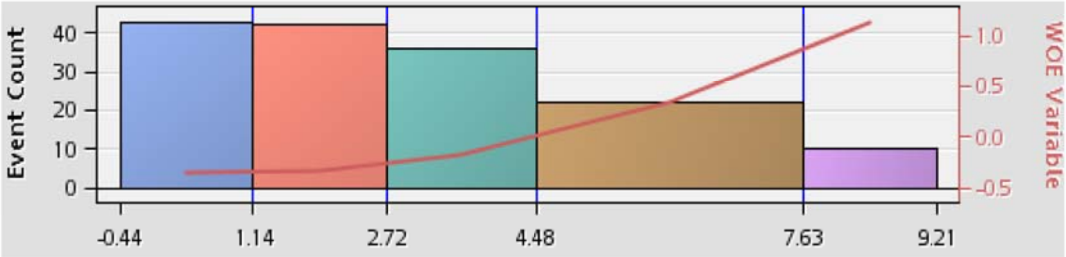


Figure 27: Payables turnover binning

The last ratio is payables turnover and its default rate is decreasing. With higher ratio value, the probability of pay obligations grows and default probability decreases.

Variables Transformation

The principle of the scoring by this method is in the computation of polygon area as it is already written in the Introduction section. All of the input parameters are financial ratios. These ratios can have value from interval (-∞; +∞). Ratio values can be very high when denominator in ratio formula is close to zero. Due this fact, the cut-off values were used to set the maximum ratio value.

The aim is to get variables transformed into 0 – 1 interval. The best values of ratios should be close to one and the worst values should be close to zero. This can be achieved using sigmoid function (62). This function transforms number from interval (-∞; +∞) into the number from interval (0; 1).

$$a = \frac{1}{1+e^{-x}} \tag{62}$$

a –transformed ratio value, x –ratio value

The transformation could be done also in linear "naïve" form by proportional shrinking of the interval borders, but this process leads to a big cluster around data mean and that would lead to low sensitivity around the mean value. The advantage of transformation (62) is that it is most sensitive for values around the zero and sensitivity is decreasing for higher and lower values. For example given equation results in 0.999954602 for argument 10 and 0.999949828 for argument 9.9. The difference is small -4.7743×10^{-6}. Results for arguments 0.5 and 0.6 are 0.622459331 and 0.645656306 respectively, which means difference equal to 0.023196975. The sensitivity is much higher for smaller values. Values entering transformation are in the tables 7 - 11 so this transformation is very useful in this case.

Another desired feature would be having variables normalized before entering the transformation process. This is desirable due relative comparison between variables. For example when two variables come from normal distribution with different parameters (mean, variance) the transformed values will also have different characteristics. Normalization can change (standardize) data to have the same distribution parameters (62). Output data have mean value equal to zero and standard deviation equal to one.

$$z = \frac{x-\mu}{\sigma} \tag{63}$$

z – standardized variable, x – ratio value, μ – arithmetic average, **σ** – standard deviation

This approach is good for monotonic variables (default rate is increasing or decreasing). Unfortunately, there are some variables such as debt on equity ratio that do not have monotonic default rate course due their characteristic. Negative values means negative equity, which is problematic for a company and companies with negative value of debt on equity ratio have very high default rate. This is visible in figure 24. This issue can be solved by Weight of Evidence [77] approach, which can assign a score (or weight of evidence) for each variable group (category) regardless of its position among other groups. For example, a group with the highest default rate will have assigned a lowest score even if it is in the middle of other groups. The monotonic assumption of default rate course does not have to be followed. On the other hand, monotonic assumption should be given attention due the logic of variable (ratio). For example, debt ratio cannot be negative because it is the ratio of debt and total capital where both number are positive. So if the monotonic presumption was violated it would lead to bad model (and decisions) even if the weight of evidence was used. However, there are variables such as debt on equity whose default rate course cannot be monotonic from the logical point of view. Weight of evidence can be very useful when nominal variables are used. Nominal variables (such as colours) cannot be ordered natively (red colour is not "higher" than blue etc.), but weight of evidence can assign to them some value (score).

$$WOE = ln\left(\frac{N_i}{D_i}\right) \tag{64}$$

N_i – portion of non-defaulted client in i-th group and all non-defaulted clients
D_i – portion of defaulted client in i-th group and all defaulted clients

One can notice that weight of evidence values will be rising with lowering default rate. Higher non-default rate means higher weight of evidence, which can be also understood as an individual score for given group. When the group contains the same rate of defaulted companies and non-defaulted ones the weight of evidence value will equal zero (logarithm of one). In practice, it means that evidence of both values in specific group is equal and this group does not have nor positive nor negative impact on score. The ratio of non-defaulted client higher than the ratio of defaulted ones means higher evidence of non-defaulted clients and therefore contribution of this category to final score is positive (in case of logic that higher score means lower default probability). In opposite case the value is negative.

The following tables 7 - 11 show data characteristics for each variable. Event count is the number of defaults in the given group; non-event count is the number of non-defaulted companies. Event rate was already mentioned as default rate and the WoE is calculated weights of evidence for each group.

Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	5		0.0	0.0	0.0	0.0	0.0
DR < 0.39	1	0.39	10.0	834.0	844.0	0.012	1.14343
0.39 <= DR ...	2	0.6	17.0	829.0	846.0	0.02	0.60678
0.6 <= DR < ...	3	0.76	30.0	801.0	831.0	0.036	0.00444
0.76 <= DR ...	4	0.93	33.0	807.0	840.0	0.039	-0.08341
DR >= 0.93	5		63.0	796.0	859.0	0.073	-0.74376

Table 7: Debt ratio grouping

Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	3		0.0	0.0	0.0	0.0	0.0
DOE < 0	1	0.0	40.0	500.0	540.0	0.074	-0.75449
0 <= DOE < ...	2	0.26	4.0	295.0	299.0	0.013	1.02046
0.26 <= DOE...	3	0.94	11.0	839.0	850.0	0.013	1.05409
0.94 <= DOE...	4	2.1	20.0	822.0	842.0	0.024	0.43579
2.1 <= DOE < 5	5	5.0	29.0	816.0	845.0	0.034	0.0569
5 <= DOE	6		49.0	795.0	844.0	0.058	-0.4937

Table 8: Debt on equity grouping

Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	2		0.0	0.0	0.0	0.0	0.0
ROC < -0.78	1	-0.78	60.0	783.0	843.0	0.071	-0.71143
-0.78 <= RO...	2	0.26	25.0	828.0	853.0	0.029	0.21991
0.26 <= ROC...	3	1.0	21.0	818.0	839.0	0.025	0.38212
1 <= ROC < ...	4	2.29	18.0	821.0	839.0	0.021	0.53993
ROC >= 2.29	5		29.0	817.0	846.0	0.034	0.05812

Table 9: Return on costs grouping

Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	3		0.0	0.0	0.0	0.0	0.0
CR < 0.74	1	0.74	51.0	789.0	840.0	0.061	-0.54128
0.74 <= CR ...	2	1.1	43.0	790.0	833.0	0.052	-0.36939
1.1 <= CR < ...	3	1.53	28.0	826.0	854.0	0.033	0.10417
1.53 <= CR ...	4	2.67	12.0	835.0	847.0	0.014	0.9623
CR >= 2.67	5		19.0	827.0	846.0	0.022	0.49314

Table 10: Current ratio grouping

Value	Group	Cutoff	Event Count	Non Event Count	Total	Event Rate	WOE
MISSING	4		0.0	0.0	0.0	0.0	0.0
PT < 1.14	1	1.14	43.0	801.0	844.0	0.051	-0.35556
1.14 <= PT <...	2	2.72	42.0	801.0	843.0	0.05	-0.33203
2.72 <= PT <...	3	4.48	36.0	808.0	844.0	0.043	-0.16918
4.48 <= PT <...	4	7.63	22.0	823.0	845.0	0.026	0.34169
PT >= 7.63	5		10.0	834.0	844.0	0.012	1.14343

Table 11: Payables turnover grouping

The WoE values (which are in the last columns of tables above) are entering the transformation process (page 53) to get the values in the interval 0 – 1.

7.2 Score Calculation

Now everything is ready to perform calculations regarding score. As mentioned in Introduction section, the company score will be the area of polygon (in this case pentagon).

The final polygon area calculation is described by equation (65), where polygon is divided into N triangles and the area is the sum of triangles' areas. N will be equal to five for pentagon area calculation.

$$S = \frac{1}{2}\sum_{i=0}^{N-1} a_i \times a_{(i+1)\,\mathrm{mod}\,N} \times \sin\alpha_i \times w_i \times w_{(i+1)\,\mathrm{mod}\,N} \tag{65}$$

N – vertex count (equal to 5 in this calculation), α_i – angle between consecutive axes
a_i – value on i-th axis, $a_{(i+1)\,\mathrm{mod}\,N}$ – value on axis next to i-th axis,
w_i – weight of the value on i-th axis , $w_{(i+1)\,\mathrm{mod}\,N}$ – weight of the value on axis next to i-th axis

The calculation according to formula (65) is depicted on figure 28. The area of yellow triangle is the value on ROC axis multiplied by altitude perpendicular on ROC (red line) axis and divided by two. Altitude is calculated as sine of angle α multiplied by the value on D/E axis. This is repeated five times.

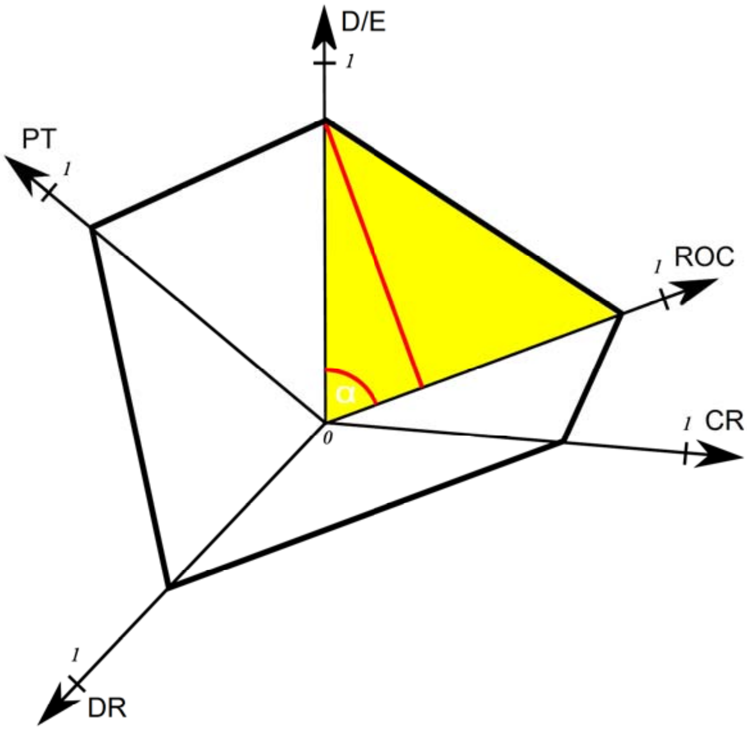


Figure 28: Score calculation for one triangle

The final score (area) depends also on parameters order for the number of parameters greater than three. This is not a parameter in equation but it is implicitly assumed and considered during the calculation process.

7.3 Parameters Determination

The core idea is to maximize Gini and Kolmogorov-Smirnov indexes values by changing angles, weights and parameters' order. This is the main difference between classifiers and chosen approach, which is actually not a classifier.

$$Gini(\boldsymbol{\alpha}, \boldsymbol{w}, \boldsymbol{o}) \rightarrow MAX \tag{66}$$
$$KS(\boldsymbol{\alpha}, \boldsymbol{w}, \boldsymbol{o}) \rightarrow MAX \tag{67}$$

α – vector of angles, w – vector of weights, o – order vector

Mathemathica software was used to perform maximization process. Numerical methods had to be used due the fact that analytical solution of maximization expressed by equations (66) and (67) is practically not possible. Gini and KS indexes computation algorithms are time consuming and require a lot of memory. Each calculation of Gini index requires ordering and calculations over Cartesian product and it leads asymptotically to quadratic computational and memory complexity. When any of parameters changes, recalculation has to be performed.

There are 5! = 120 permutations for axes order. Circular permutations [78] can be used to perform optimisation. The reason is that the rotation of pentagon does not affect the score (pentagon area). This property can reduce original 120 various orders to 12. A number of circular permutation can be calculated using equation (68).

$$P = \frac{1}{2}(n-1)! \tag{68}$$

Circular permutation used in calculations are in table 12.

1	2	3	4	5
1	2	3	5	4
1	2	4	3	5
1	2	4	5	3
1	2	5	3	4
1	2	5	4	3
1	3	2	4	5
1	3	2	5	4
1	3	4	2	5
1	3	5	2	4
1	4	2	3	5
1	4	3	2	5

Table 12: Circular permutations that represent all cases of axes order

The simplest calculation scenario configuration was with angles equal to 360° / 5 = 72°, weights were equal to one and the order of parameters was changing. The maximization was done for Gini and Kolmogorov-Smirnov indexes separately. It means that there are two result sets for each calculation scenario. Another scenarios calculation are analogic, but there are more changing variables. These scenarios were already described in the Introduction section.

Calculation of one permutation took between tenths of minutes and several hours when weights and angles are changing. Whole calculation took about one day. This was when five parameters were used. In case of six parameters, there is 60 circular permutations and the calculation time can grow five times. It means that the whole calculation procedure can take around one week. This is not optimal and it is one shortcoming of used maximisation methods.

8 Results and Discussion

The table 13 shows the results for Gini index optimization in scenario A (constant weights and angles). It is obvious that the order of parameters does not have the significant impact on the Gini index value. The maximum and minimum values are highlighted. The difference between maximum and minimum values are 1.3 % for Gini index. The best result for scenario A is Gini index value equal to 29.18 % on validation data. Order of parameter in the best case is [1, 4, 2, 3, 5]. This is comparable to the best resulting alternative methods. The majority of the Results and Discussion section is based on [20].

Order	Gini Train	Gini Validation
1 2 3 4 5	0.3984	0.2804
1 2 3 5 4	0.4016	0.2888
1 2 4 3 5	0.3992	**0.2788**
1 2 4 5 3	0.4003	0.2893
1 2 5 3 4	0.4021	0.2862
1 2 5 4 3	0.3997	0.2806
1 3 2 4 5	0.3994	0.2878
1 3 2 5 4	0.4011	0.2801
1 3 4 2 5	0.3986	0.2791
1 3 5 2 4	0.4010	0.2833
1 4 2 3 5	0.4009	**0.2918**
1 4 3 2 5	0.4004	0.2845

Table 13: Results for Gini index of scenario A. The maximum and minimum values are highlighted.

The table 14 contains results for Kolmogorov-Smirnov index optimization in A scenario. Impact of parameters' order is slightly higher than in Gini index optimization. The difference between maximum and minimum values is 2.7 %. The scenario with constant weights and angles was chosen because there is no impact of changes in these parameters on results. The best result for this scenario is KS index value 27.58 % for parameters' order [1, 2, 4, 5, 3].

Order	KS Train	KS Validation
1 2 3 4 5	0.3268	0.2708
1 2 3 5 4	0.3268	0.2735
1 2 4 3 5	0.3170	0.2658
1 2 4 5 3	0.3333	**0.2758**
1 2 5 3 4	0.3203	**0.2489**
1 2 5 4 3	0.3268	0.2509
1 3 2 4 5	0.3301	0.2766
1 3 2 5 4	0.3268	0.2585
1 3 4 2 5	0.3301	0.2490
1 3 5 2 4	0.3235	0.2570
1 4 2 3 5	0.3268	0.2716
1 4 3 2 5	0.3268	0.2509

Table 14: Results for KS index of scenario A. The maximum and minimum values are highlighted.

ID	Sc.	Order	Weights					Angles				
1	D	1 2 5 3 4	3.7	2.1	3.9	5.5	4.2	40.0	29.4	127.6	64.6	98.4
2	B	1 2 5 3 4	1	1	1	1	1	18.7	8.6	120.8	95.7	116.2
3	C	1 2 5 3 4	6.5	1.8	4.3	9.1	5.3	72	72	72	72	72
4	D	1 4 3 2 5	10.0	9.0	8.0	6.7	0.3	54.3	1.0	69.7	180.0	55.0
5	B	1 3 2 5 4	1	1	1	1	1	106.0	14.7	92.0	116.6	30.7
6	C	1 2 3 4 5	9.1	7.9	7.5	9.1	0.6	72	72	72	72	72

Table 15: The best performing settings of parameters for scenarios B, C, D.

ID	Sc.	Gini Train	Gini Valid.	KS Train	KS Valid.
1	D	0.4148	0.3133		0.2979
2	B	0.4163	**0.3155**	N/A	0.3006
3	C	0.4146	0.3135		0.2841
4	D		0.2938	0.3529	0.2998
5	B	N/A	0.2884	0.3464	0.3028
6	C		0.2863	0.3497	0.3080

Table 16: The first part is the result of Gini maximization and the second part is the result of Kolmogorov-Smirnov index maximization for parameters from table 16.

The results (validation part in the table 16) are close to each other so it is not necessary to search for the values of all three groups of variables (order, weights, angles) and one group can stay constant with the equal values. This is very important for calculation time, which was about several hours up to one day in case of determination of all parameters. The maximization of Gini index leads also to very good[19] results for KS index. This is obvious from the table 16 where KS values are the same in case of Gini maximization and KS maximization – the highest Gini index value on validation data is 31.55 % and KS value is equal to 30.06 % for the same configuration. On the other hand, the maximization of KS index does not lead to the highest possible values of Gini index. The second part of the table 16 shows that Gini index values are 2 – 3 % lower than values obtained by Gini maximization. The highest KS index value from KS index maximization is 30.80 % but corresponding Gini index value is only 28.63 %.

The standard error of Gini index is 0.075 in case of validation dataset and 0.052 in case of training dataset. The calculation using equation (24) is not very sensitive on Gini index value, but it is sensitive on number of defaulted and non-defaulted companies in actual sample. Changing Gini index values by 0.1 results in change of standard error on third decimal place and therefore standard error is not stated separately for each calculation and in the comparison with other methods.

Finally, the configuration with parameters order [1, 2, 5, 3, 4], angles values (18.7°, 8.6°, 120.8°, 95.7°, 116.2°) with all weights equal to one is the best and leads to the result 31.6 % for Gini index value and 30.1 % for KS index value on validation data. The aim was to choose

[19] "Very good" is a term of wide comprehension. Gini and KS indexes values around 30 % on validation dataset are the best values achieved on given data even when using alternative methods from "Comparison with Other Methods" section. Values similar to these ones are considered as very good.

one result for proposed method. The configuration with the highest Gini index value was chosen because also KS index value is very good (not the best, but higher than 30 %). On the other hand, if the configuration with highest KS index value was chosen, corresponding Gini index value would be lower by 3 % from the best case which is not optimal. This was a kind of multi-criterial decision where both criterions were assumed and the aim was to have the maximum value of one indicator and reasonable value of another indicator. Aim was fulfilled and the results were presented.

Equation (69) shows the final score calculation. Variables entering the formula are the transformed weight of evidence values, not the original financial ratios. For proper score usage, the original values have to be transformed onto weight of evidence values and consequently transformed by formula (62). The constant in each term represents sine of corresponding angle. This is due the fact, that weights are equal to one. Otherwise, also the weights should be included into the calculation.

$$Score = \frac{1}{2} \times (0.321 \times PT \times DOE + 0.15 \times DOE \times DR + 0.897 \times DR \times ROC + 0.859 \times ROC \times CR + 0.995 \times CR \times PT) \quad (69)$$

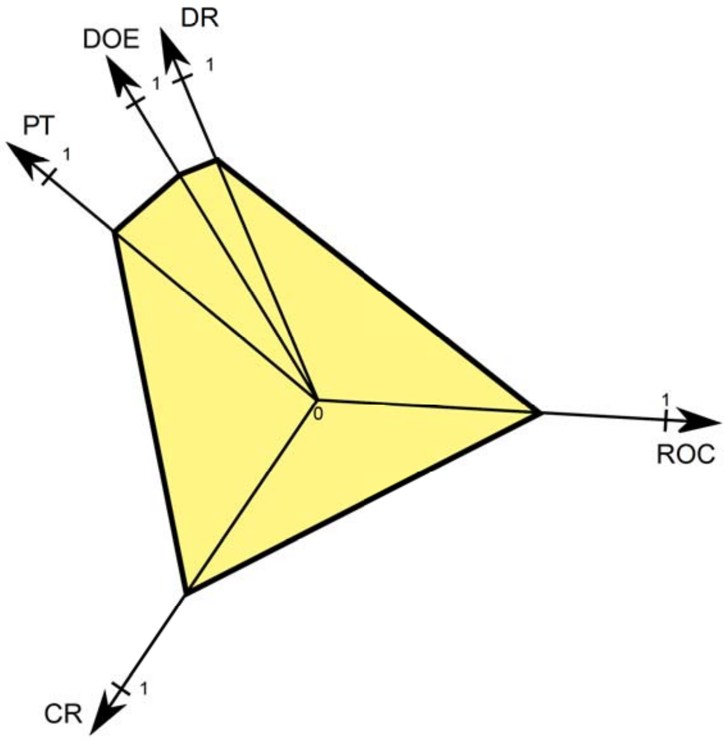


Figure 29: The pentagon diagram of the best rated company (score = 0.82).

Figures 29 and 30 show the polygon (pentagon) diagram for the best and the worst rated company from validation data set. One can notice only a small angle between DOE and DR axes. These two indicators are from the same group of indicators and therefore reduction of one of them should be considered. This can be also a way how maximization procedure can reduce unnecessary indicators.

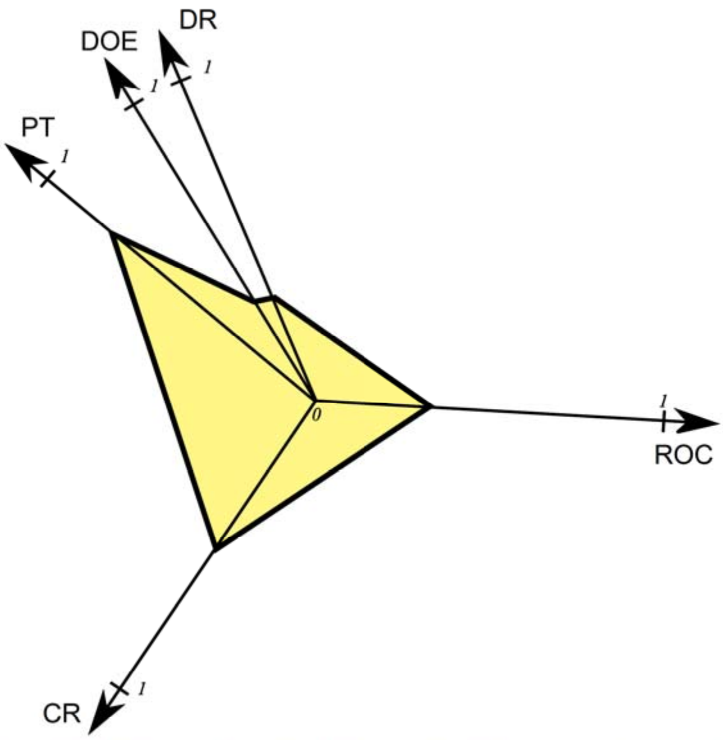


Figure 30: Pentagon diagram of the worst rated company (score = 0.2)

Based on newly discovered information the DOE indicator was removed the angle between PT and DR indicator was set to the sum of angles between PT and DOE and DR axes. The new configuration order is [1, 4, 2, 3], angles values (27.3°, 120.8°, 95.7°, 116.2°) with all weights equal to one. The Gini and Kolmogorov-Smirnov indexes are equal to 31.25 % and 28.53 % respectively on validation data set. The values for quadrangle are practically the same for Gini index value and slightly lower (by 1.5 %) for KS index value. These results are still fine and competitive to alternative methods with the best results. The big advantage is that performance is preserved even with reduction of one parameter.

Equation (70) shows simplified score calculation where DOE parameter is omitted. The other calculation logic is the same as for formula (69).

$$Score = \frac{1}{2} \times (0.459 \times PT \times DR + 0.897 \times DR \times ROC \times 0.897 + 0.859 \times ROC \times CR + 0.995 \times CR \times PT) \tag{70}$$

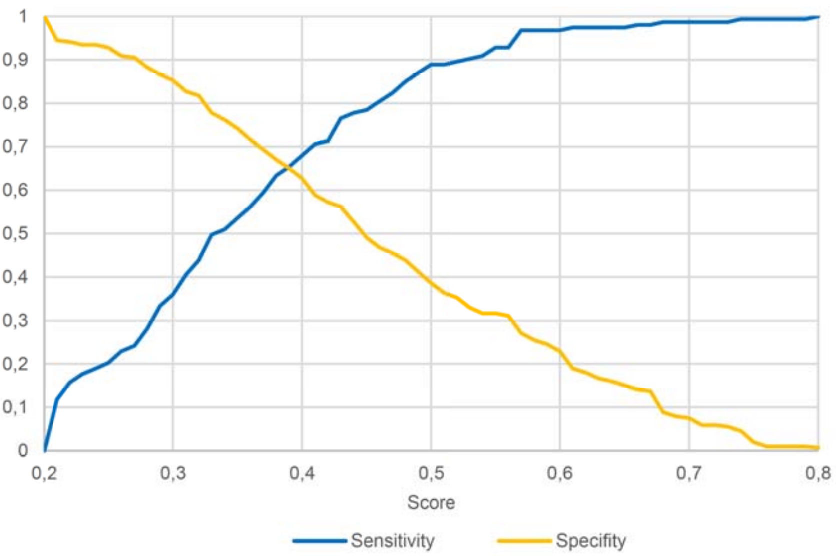


Figure 31: Sensitivity and specificity on training set dependent on threshold score value

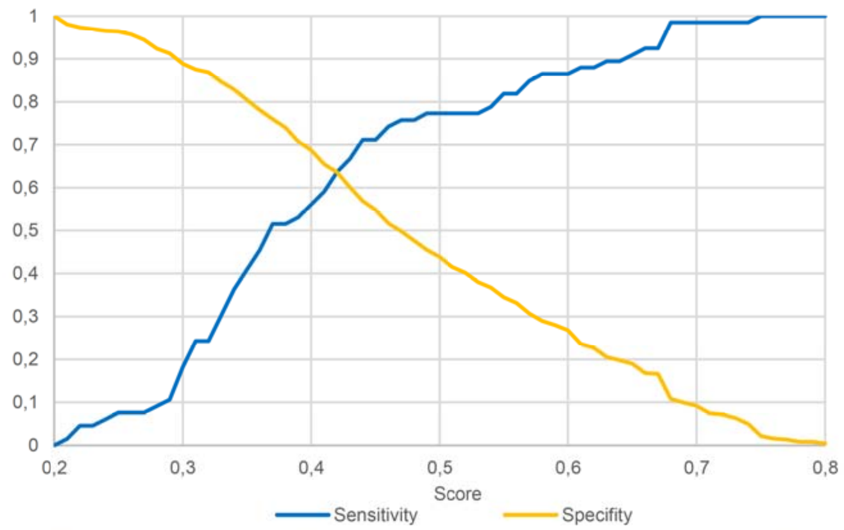


Figure 32: Sensitivity and specificity on testing set dependent on threshold score value

The figures 31 and 32 show sensitivity and specifity based on threshold level. Threshold is the score value that distinguishes between defaulted and non-defaulted companies. This threshold level can be set according to requirements on true positive rate (sensitivity) and true negative rate (specifity). This level does not affect Gini or KS indexes. Increasing sensitivity means decreasing specifity and vice versa. Set up of threshold value can be used also in other binary classification tasks. The problem can be in interpretation where score is represented by default probability (e.g. in logistic regression). In this case the logical threshold value should be 50 % and companies with default probability higher than 50 % should have be interpreted as default companies and probability lower than 50 % non-default companies. Despite of this it is not necessary to be conform to mentioned fact and threshold value can be moved to the level where sensitivity and specifity is acceptable for given task.

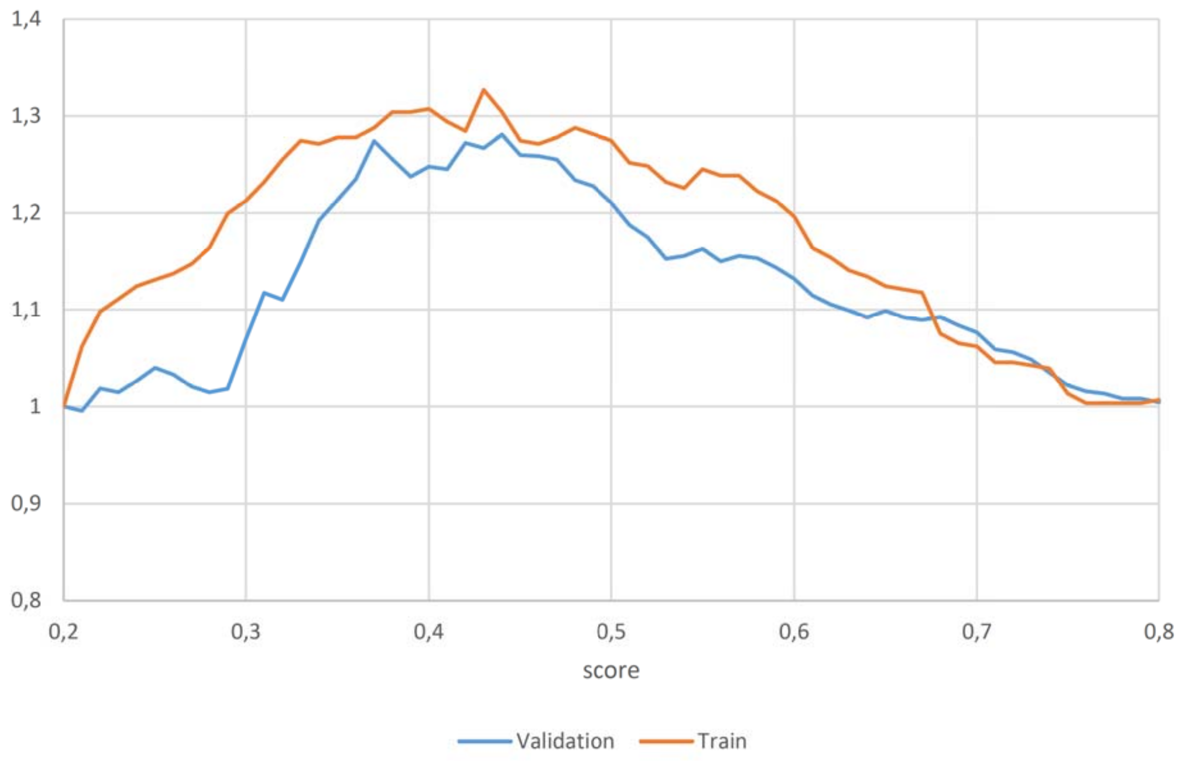


Figure 33: Positive assessment – dependency of sensitivity + specifity on threshold value

Since sensitivity is true positive rate and specificity is true negative rate, both of these indicators represent the correctness (or accuracy) of assessing by the model. One of the possible way of choosing the threshold value is to choose the score value where the sum of the specificity and sensitivity values is the highest. The figure 33 shows the dependence of the sum of sensitivity and specificity on the score. The maximum value is for the score equal to 0.43 in the training sample. The value of sensitivity is 76.47 % and specificity is 56.21 %. The maximum in validation sample is at score value 0.44, which is near to the maximum on training sample, and therefore training value 0.43 perfectly represents the purpose of choosing this value. The sensitivity and specificity values for threshold value 0.43 are 66.66 % and 60.04 % respectively. Another technique how to set the threshold value is using false positive rate, which represents type I error or false negative rate (type II error). This can be easily done by ordering of all observations by the score and calculate false positive rate (or false negative rate) corresponding to specific score. There is no single correct technique that must be used. Model developer can choose the one, which is suitable based on problem requirements.

8.1 Comparison with Other Methods

Following tables show comparison between ordering performance of the new proposed method with the methods where calculation was performed by SAS software. Radial basis neural network with three hidden neurons performs best on training dataset in both Gini and KS indexes calculations. Despite of these excellent training results, the performance on validation dataset dramatically falls below the performance even of methods that had lower training score.

Method	Gini Train	Gini Validation
Proposed (Scenario B)	0.416	0.312
Altman	N/A	0.207
Logistic Regression	0.420	0.301
Neural Network (Back Propagation 3H)	0.454	0.291
Neural Network (Radial Basis 3H)	0.508	0.265
Memory Based Reasoning (K-Nearest Neighbours)	0.422	0.199

Table 17: Comparison of proposed method with other methods – Gini index

Method	KS Train	KS Validation
Proposed (Scenario B)	N/A[20]	0.301
Altman	N/A	0.248
Logistic Regression	0.317	0.299
Neural Network (Back Propagation 3H)	0.337	0.292
Neural Network (Radial Basis 3H)	0.412	0.225
Memory Based Reasoning (K-Nearest Neighbours)	0.340	0.216

Table 18: Comparison of proposed method with other methods – KS index

The big difference between training and training stage can be caused by overfitting where model is well adopted to training data, but performance on validation data is poor. From the result tables 17 – 21 it is obvious that the best (highest) values of Gini and KS indexes are around 30 % on validation dataset. Logistic regression, neural networks are used in the most cases for company failure prediction and the results show that this is legitimate because performance indicators for these methods are among the best results. From the other classification algorithms used in machine learning the linear Bayes classifier and grammatical evolution belong also to the group of the best performing methods.

[20] The results of proposed method for the best scenario B (table 16). Train value is not relevant because optimization was performed on Gini index and the KS value was calculated on validation dataset.

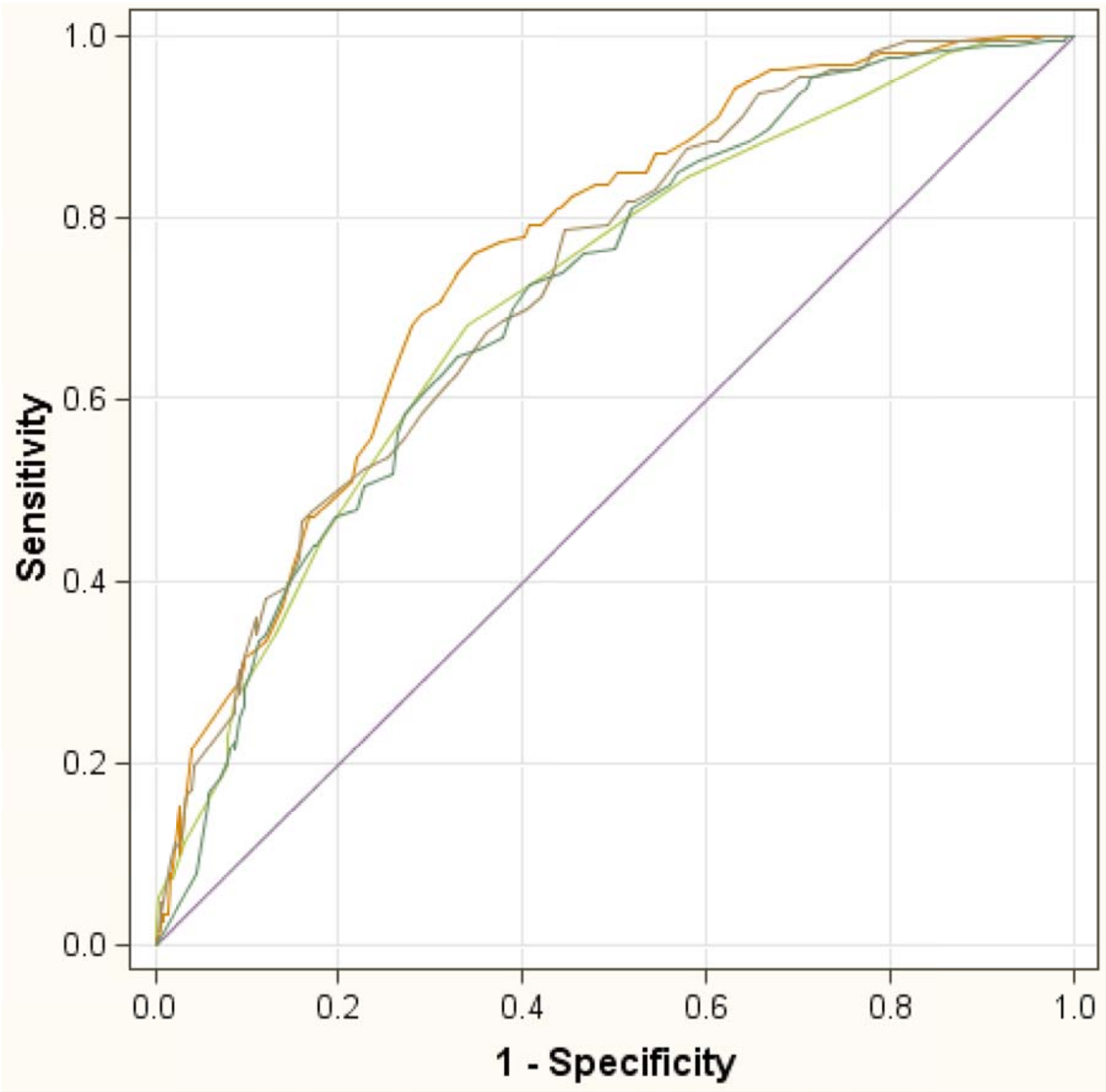


Figure 34 (screenshot from SAS): ROC curve on training data for specific alternative methods.

Green	**Logistic regression**
Red	**Backpropagation neural network with 3 nodes in hidden layer**
Brown	**Radial basis neural network with 3 nodes in hidden layer**
Yellow	**Memory based reasoning**
Violet	**Random model**

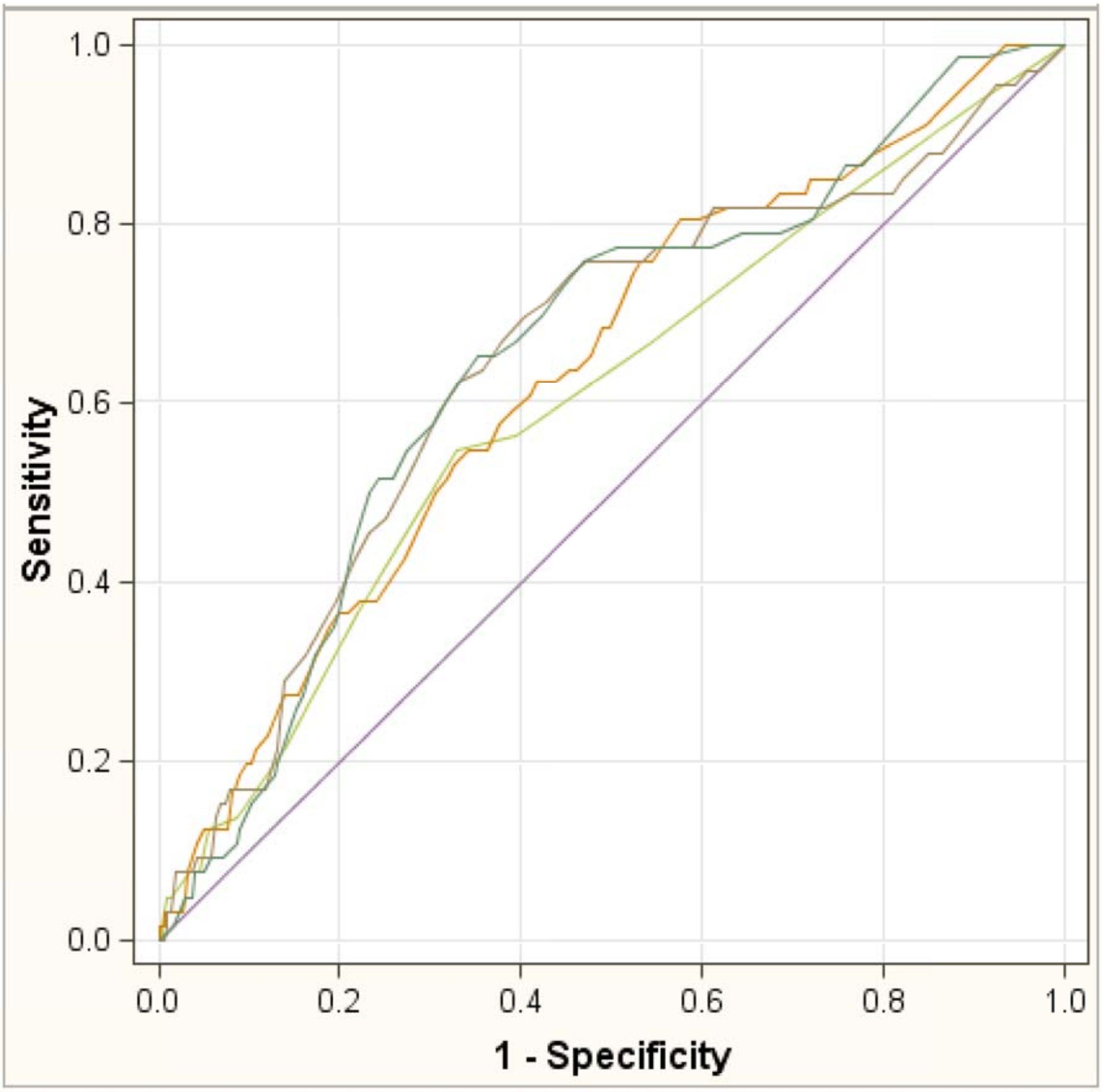


Figure 35 (screenshot from SAS): ROC curve on validation data for specific alternative methods

Green	**Logistic regression**
Red	**Backpropagation neural network with 3 nodes in hidden layer**
Brown	**Radial basis neural network with 3 nodes in hidden layer**
Yellow	**Memory based reasoning**
Violet	**Random model**

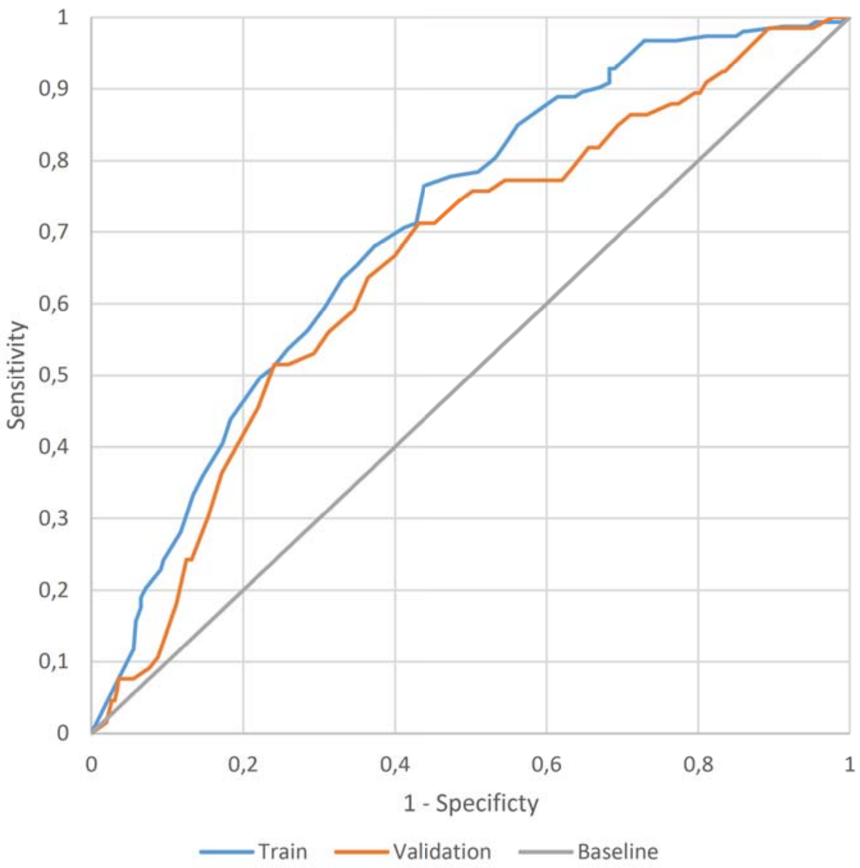


Figure 36: ROC curve for proposed method on training and validation data sets

Figures 34 - 36 show comparison of ROC curves for several methods and proposed methods. The shape of curves is in conformance with the Gini and KS indexes results. Low performance of memory based reasoning classifier and significant drop in performance of radial basis neural network on validation dataset is clear from the figure 35. The performance drop between values 0.6 – 0.8 is also obvious. Due the fact that this drop is visible for all plotted methods, the reason is data quality with highest probability.

Method	Gini Train	Gini Validation
Linear perceptron	0.2507	0.1621
Neural Network (Back Propagation 2H)	0.3907	0.3076
Neural Network (Radial Basis 2H)	0.7552	0.2032
Neural Net (Levenberg–Marquardt)	0.4129	0.3017
Support Vector Machine	0.3566	0.2594
Naive Bayes	0.4185	0.2869
Linear Bayes	0.4192	0.3043
Adaptive Boosting	0.3989	0.2681
Quadratic Bayes	0.4330	0.2702
Mixture of Gaussian and Bayes	0.4625	0.2316
C4.5 Decision Tree	0.2712	0.2036
Random Forest	0.3043	0.2020
Scaled Nearest Mean	0.4008	0.2853

Table 19: Comparison with other classification methods – Gini index

Method	KS Train	KS Validation
Linear perceptron	0.0065	0.0314
Neural Network (Back Propagation 2H)	0.3072	0.2644
Neural Network (Radial Basis 2H)	0.5392	0.2043
Neural Net (Levenberg–Marquardt)	0.3203	0.3084
Support Vector Machine	0.3203	0.2579
Naive Bayes	0.3268	0.2706
Linear Bayes	0.3333	0.3069
Adaptive Boosting	0.3301	0.2979
Quadratic Bayes	0.3366	0.2835
Mixture of Gaussian and Bayes	0.3464	0.2855
C4.5 Decision Tree	0.2810	0.2537
Random Forest	0.2908	0.1919
Scaled Nearest Mean	0.3333	0.2658

Table 20: Comparison with other classification methods – KS index

A brief comparison of other results achieved by Matlab software is shown in tables 19 and 20. It is noticeable that radial basis neural network with two hidden neurons has perfect results on training data but performance on validation dataset is much worse in comparison to training dataset and also to other methods. The methods based on Bayesian classifier are well performing with exception of mixed Gaussian and Bayes classifier where dramatic drop between training and validation dataset is present. The nearest mean algorithm, which belongs to memory based reasoning algorithms, is also well performing.

The best results for grammatical evolution were achieved by maximizing Gini index. The KS index maximization resulted lower performance for both Gini and KS values similarly to proposed method where this behaviour was also present. The median values from all runs are presented in the following table.

Method	Gini Train	Gini Validation	KS Train	KS Validation
Grammatical evolution	0.4035	0.3023	0.3148	0.2869

Table 21: The results for grammatical evolution

From this comparison, it is obvious that proposed method performance is competitive to other methods. The decision trees methods are least performing, their Gini index values are around 20 % and these methods do not look suitable for company failure prediction based on given data. Other methods are usable. There is at least one method within each group with very good results. For example, backpropagation neural networks are the best performing from the neural networks group, linear Bayes classifier is the best performing from Bayesian models group, scaled nearest mean has also good results and belongs to memory based reasoning algorithms. Grammatical evolution and logistic regression have also excellent results.

8.2 Power Distribution Companies Assessment

As mentioned in the Introduction section, the proposed method can be also used for the relative comparison of companies' performance because the method gives a score and not the only Boolean information whether company will default or not. The score is calculated based on financial statements data and information about company failure in the meaning of ability to fulfil obligations. The simple assumption is that companies that are able to fulfil their obligations are well performing and companies that are not able to fulfil their obligations are bad performing. In other words high score does not only mean that a company is less probable to default, but also means that a company's core business is healthy and a company performs well in its business area. This idea can be used for the comparison of several companies, in this case electricity distribution companies.

Distribution company	Abbreviation	Country
AEW Energie AG	AEW	Switzerland
SP Distribution Limited	SPD	United Kingdom
RARIK ohf	RARIK	Island
Northern PowerGrid PLC	NPG	United Kingdom
CEZ Razpredelenje Bulgaria AD	CEZ BG	Bulgaria
Adger Energi Nett AS	ADGER	Norway
E.ON Distribuce, a.s.	EON	Czech Republic
SRD Réseaux de Distribution	SRD	France
Electricity North West Limited	ENW	United Kingdom
Sadales Tikls As	SADALES	Latvia
HEP – Operator distribucijskoj sustava d.o.o.	HEP	Croatia
Eesti Energia Jaotusvork OÜ	EESTI	Estonia
ČEZ Distribuce, a.s.	CEZ CZ	Czech Republic
Elektro Maribor d.d.	EM	Slovenia

Table 22: The list of electricity distributor companies used in the comparison

Several distribution companies were chosen for comparison. It is subset of companies used in doctoral thesis [6]. Subset was chosen due the data quality:

- Financial statement data for some distributors were not publicly available.
- Available data were not detailed enough.
- Available financial data were consolidated and required items did not correspond to electricity distribution only.
- Unbundling[21] in energy sector was not done (compulsory only of EEA[22] countries) and data were available for whole company only, not divided from distribution business only.

Distributor	Score	Score4	Y	CRS	VRS	MBM
AEW	0,703116	0,702787	142,6221	100	100	89,61897
SPD	0,44404	0,437988	86,45842	100	100	89,49859
RARIK	0,427207	0,403424	67,92594	46,3537	89,97662	71,77769
NPG	0,356241	0,350338	61,89781	100	100	80,68371
CEZ BG	0,53047	0,513525	51,30338	50,64468	100	67,61712
ADGER	0,339062	0,345712	45,72081	29,01074	33,57188	73,73312
EON	0,475791	0,458846	42,53348	21,75375	35,61574	66,24917
SRD	0,50767	0,506248	38,48301	66,44753	77,1767	71,88257
ENW	0,422321	0,415192	35,29487	30,92568	100	73,36799
SADALES	0,39534	0,387233	9,435276	7,36649	40,51413	61,04242
HEP	0,327594	0,291801	16,63032	8,07047	37,40485	57,77276
EESTI	0,352303	0,336868	20,28964	7,66583	32,51902	62,98078
CEZ CZ	0,572528	0,572199	29,65544	13,42613	34,28563	66,01035
EM	0,603368	0,584208	20,71334	12,14659	37,7721	65,07278

Table 23: Scores of various methods used for comparison

The table 23 shows the result of application of proposed method and benchmarking methods from doctoral thesis [6] on selected distribution companies. The column "Score" is the result of proposed method and "Score4" is the result of modified proposed method where only four input parameters were used. The difference in results between five and four used parameters is small and therefore the modification with four parameters can be used without a significant impact on final score. The "Y" column represents "Y" performance index, CRS represent constant return to scale (CCR) model, VRS is the variable return to scale model (BCC) and MBM is represents referential point method. The description of these methods (models) is in Benchmarking Methods section.

Figures 37 and 38 are showing the best and the worst performing company according to proposed method. This assessment is conform also with other benchmarking models as it is shown in the table 24. *AEW Energie AG* belongs between the first and the third place in all

[21] The process of separation of networks from activities of generation and supply, which is the part of European Union's Third Energy Package. Directive 2009/72/EC concerning common rules for the internal market in electricity [79].

[22] European Economic Area (EU + Switzerland, Norway, Island and Lichtenstein)

methods used. On the other hand *HEP – Operator distribucijskoj sustava d.o.o.* is between the 10th and the 14th place from the 14 companies.

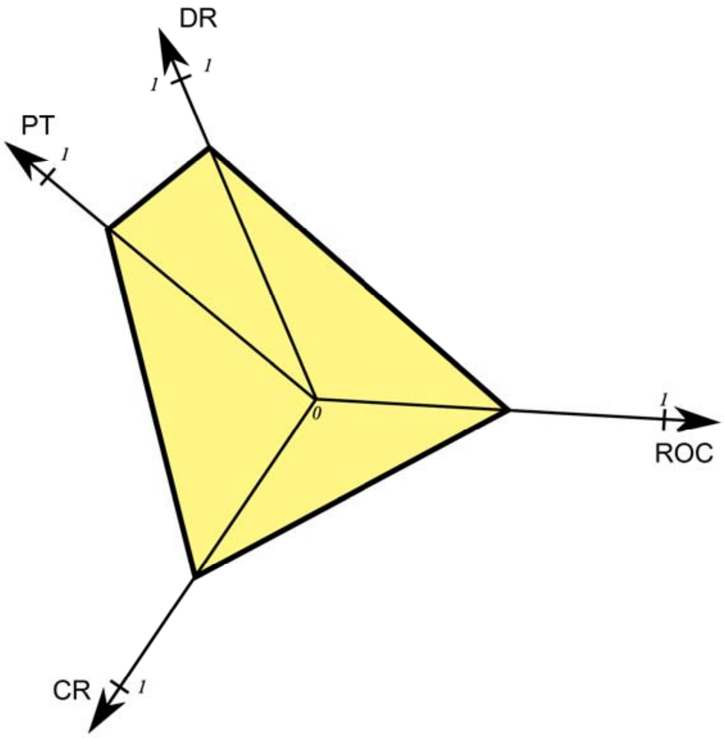


Figure 37: The quadrangle of the best performing distribution company – AEW Energie AG (score = 0.70)

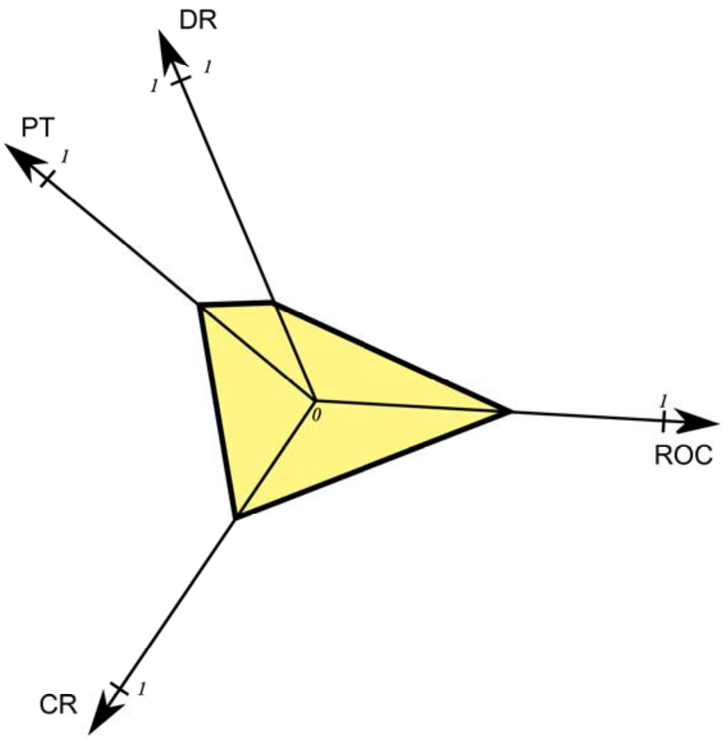


Figure 38: The quadrangle of the worst performing company – HEP – Operator distribucijskoj sustava d.o.o. (score = 0.29)

The biggest difference in assessment between proposed method and *Y performance index* is for the distribution company *Elektro Maribor d.d.* (figure 39). Proposed method puts this company on the second place and *Y performance index* on the 11th place.

The big Y score difference between AEW and EM is caused mainly by the last two terms in calculation (30). Terms x_5[23] and x_6[24] have significantly higher values in case of AEW and therefore the results are very different. On the other hand, these two terms do not enter the calculation of proposed method. Y index profitability ratio EBT/Assets is substituted by EAT/Costs in proposed method and costs on employer are not taken into account at all. In addition, other technical indicators are not considered in proposed method. This is the reason why there is in some cases a significant difference in assessment by benchmarking methods and proposed method.

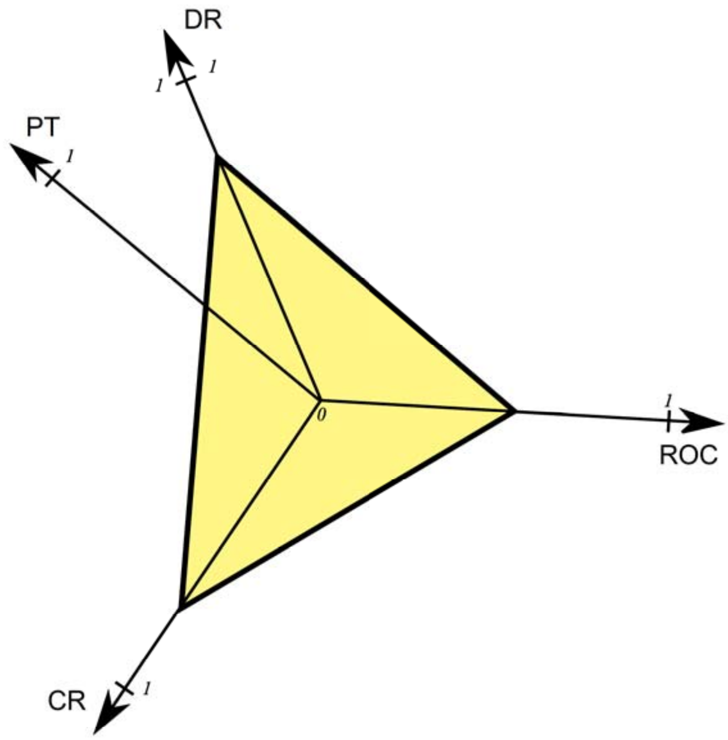


Figure 39: Elektro Maribor d.d.

The quadrangle diagrams for other assessed companies are in the appendix (Appendix Figure 7 – Appendix Figure 17).

[23] Earnings Before Tax / Total Assets
[24] Personal Costs / Number of Employers

Distributor	Score4	Y	CRS	VRS	MBM
AEW	1	1	2	3	1
EM	2	11	11	9	11
CEZ CZ	3	10	10	12	10
CEZ BG	4	5	5	3	8
SRD	5	8	4	7	6
EON	6	7	9	11	9
SPD	7	2	2	3	2
ENW	8	9	7	3	5
RARIK	9	3	6	6	7
SADALES	10	14	14	8	13
NPG	11	4	2	3	3
ADGER	12	6	8	13	4
EESTI	13	12	13	14	12
HEP	14	13	12	10	14

Table 24: Orders of distribution companies

Relative performance order for various method used is shown in the table 24. Distribution companies are ordered according to the result of proposed method. The table 25 represents the differences between proposed method and benchmarking methods from [6].

Distributor	Y	CRS	VRS	MBM
AEW	1	1	1	1
EM	3	3	3	3
CEZ CZ	3	3	3	3
CEZ BG	1	1	1	2
SRD	1	1	1	1
EON	1	1	2	1
SPD	2	2	2	2
ENW	1	1	2	1
RARIK	2	1	1	1
SADALES	2	2	1	1
NPG	3	3	3	3
ADGER	2	2	1	3
EESTI	1	1	1	1
HEP	1	1	2	1

Table 25: Differences between proposed method and benchmarking methods

The differences were divided into three groups that represent the difference in order of proposed method and method in column heading.

- Group 1 (green): the order difference is between 0 and 3.
- Group 2 (orange): the order difference is between 4 and 6.
- Group 3 (red): the order difference is between 7 and 9.

It is noticeable that the proposed method assesses distribution companies very similarly in a half of cases. Seven or eight companies belong to the first group, which means that the difference in assessment between proposed method and benchmarking models is low. Approximately quarter of cases belongs to the second group, which means the considerable difference between proposed and benchmarking models and finally in the quarter of cases the difference is big.

Differences between proposed method and methods from [6] are caused by different input parameters and different data on which methods were trained on. One cannot say which method is better in general, but methods developed on data from power distribution companies should have higher performance for power distribution companies' assessment. On the other hand, their usage is limited to power distributors and cannot be used for any company as the proposed method can. Nevertheless, proposed method performed well for power distribution companies that can be considered as very specific kind of data.

9 Conclusions

The aim of this thesis was to propose a new scoring method for companies' default prediction. The proposed method is based on *magic square* concept that is used in macroeconomic evaluation of individual countries. Concept of *magic square* was extended to general polygon with many possible settings. Input variables represented by companies' financial ratios were transformed to achieve desired properties (better variable value is located further on axis than worse value). Therefore, area of polygon can represent score for particular company.

The core of proposed method development was in changing of parameters (weights, angles, axes order) in order to maximize the value of Gini and Kolmogorov-Smirnov indexes. Consequently comprehensive performance comparison was performed and method was used also for benchmarking purposes.

Proposed method is not a standard classifier but it can be used in the very similar way as the classification algorithms. The performance of the proposed method is comparable to logistic regression, various neural networks models, Bayes classifiers, evolutionary algorithms etc. Both Gini and KS indexes values are around 30 % on validation data set for the methods with the highest performance. This number is not high and it is lower than expected in real scoring models. This is caused by the fact that data quality is not ideal and deep data pre-processing was not the aim of this paper. The important is that the new proposed method is as good as competitive methods applied for required classification task. Indisputable advantage of proposed method is in visualisation that was inspired by the magic square, which is often used in comparison between economies of different countries. The result of company scoring is not the only number but also a diagram, which shows strength of individual factors influencing the final score and overall performance.

Proposed method output score can be also used for relative companies' performance comparison. This comparison was performed for 14 electricity distribution companies. The results were compared to the benchmarking methods results, especially to the new benchmarking method designed in doctoral thesis, which was dealing with the problematics of benchmarking models. The comparison of proposed method and new benchmarking method was done as the result of research cooperation at the Czech Technical University in Prague. Results show that proposed method can be also used for benchmarking under limited conditions. Results are good with respect to that fact that proposed method was developed on the companies from all economic sectors and the benchmarking was done on a specific industrial sector (electricity distribution companies).

The advantage of proposed method is that it is universal. It can be used for default prediction, credit scoring and for benchmarking purposes. Another advantage is in simple visualisation, which is indisputable benefit for users.

9.1 Answers on Research Questions

This section formulates answers on research questions. The deeper discussion is presented in the section Results and Discussion on page 58.

Does axes order have significant impact on Gini and Kolmogorov-Smirnov index values when the weights and angles are constant and equal?

Parameters order does not have significant impact on Gini index in case of constant weights and angles. Kolmogorov-Smirnov index is more sensitive on axes order than Gini index and its value is affected by parameters order on the above-mentioned assumptions.

The situation can change (impact can be significant) in case of extreme angles values and weight values. Significance falls with growing number of input parameters due the fact that polygon area is divided into more triangles and relative impact of one parameter is smaller than in case of low number of input parameters.

Is the performance of proposed scoring method competitive with other methods such as logistic regression or neural networks?

Yes, the performance of proposed method is competitive with other methods used for company scoring. Results show that the performance of proposed method belongs to the group of methods with the highest performance.

Is it necessary to change all groups of parameters or some of them (weights, angles) can remain constant?

No, the results show that the highest Gini and Kolmogorov-Smirnov indexes can be achieved when angles or weights are constant. The result with constant weights was selected as the best one.

Models with constant weights along with constant angles have lower performance than models with constant only one group of these parameters.

Is the relative order of electricity distribution companies obtained by proposed method different from assessing by the Y performance index used in dissertation thesis [6] ?

The relative order of electricity distribution companies by proposed method differs from the Y performance index results. There is significant group of distribution companies that were assessed similarly by the both methods, but assessment of the whole sample was not the same and therefore this research question cannot be answered positively.

9.2 Future Research

The future research can be focused on the construction of better performing input variables even though that their interpretation can be more difficult. Indicators used in this paper are standard financial ratios, which are well known but their information value can be lower than complex indicators. On the other hand, complex indicators will not be as easily understandable as known and widely used financial ratios.

A polygon can be constructed for various groups of companies and this can lead to the better performing models. Identification of patterns in polygons or their shape can be useful for finding of common characteristics, which can lead to better prediction power.

Proposed method has several ways how to eliminate unnecessary parameters, for example assigning very small angle leads to low parameter significance because area of corresponding triangle will be low. Future research could try to answer whether this behaviour is stable or not.

Including non-financial industry specific variables can be also very useful for improving performance. This can be essential also for benchmarking purposes where for example reliability of electricity supplies can be used in construction of benchmarking model used in energy industry.

Appendix

		Condition (as determined by "Gold standard")			
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive	False discovery rate (FDR) = Σ False positive / Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative	Negative predictive value (NPV) = Σ True negative / Σ Test outcome negative
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = Σ True positive / Σ Condition positive	False positive rate (FPR, Fall-out) = Σ False positive / Σ Condition negative	Accuracy (ACC) = Σ True positive + Σ True negative / Σ Total population	
	Negative likelihood ratio (**LR−**) = FNR/TNR	False negative rate (FNR) = Σ False negative / Σ Condition positive	True negative rate (TNR, Specificity, SPC) = Σ True negative / Σ Condition negative		
	Diagnostic odds ratio (**DOR**) = LR+/LR−				

Appendix Figure 1: Measurement of performance for binary classification models.
Figure source http://en.wikipedia.org/wiki/Sensitivity_and_specificity

	P_0	P_1	P_2	P_3	P_4	P_5	P_10	P_90	P_95	P_96	P_97	P_98	P_99	P_100
ROA	-5.19887	-0.91235	-0.48945	-0.33418	-0.25366	-0.21097	-0.08572	0.164773	0.259696	0.288048	0.320881	0.39272	0.500468	2.481648
ROE	-4029.5	-4.45133	-2.31096	-1.51043	-1.08943	-0.78652	-0.24572	0.777639	1.178692	1.370025	1.898551	2.715266	5.869565	2810
ROCE	-1365	-1.84967	-1.05553	-0.68944	-0.45595	-0.33781	-0.10865	0.495803	0.803544	0.937038	1.116573	1.336411	2.52381	212.8113
ROS	-25450	-25450	-1031	-7.12733	-2.2381	-0.9802	-0.22561	0.174043	0.315409	0.368794	0.457133	0.578995	2.03179	101
ROC	-3130	-87	-29.9913	-16.5	-10.3846	-7.99547	-2.93933	3.562711	6.496779	8.081395	11.32836	20.98904	59.29915	66630
CR	-2448.02	0.01664	0.051896	0.089418	0.129547	0.175511	0.417765	4.820554	9.691943	12.0101	15.81998	29.36842	171	1602
QR	-2250	0.007187	0.040489	0.080581	0.110855	0.14343	0.28974	5.364074	11.33797	13.91216	20.96296	42.72881	1043	1602
CASHR	-140	-0.42051	-0.17857	-0.07402	0	0.000673	0.008717	2.601297	6.232218	8.25	11.37378	26.72596	1934	162788
DR	-1.55789	0.028012	0.073129	0.108261	0.132074	0.159026	0.252449	1.028686	1.200179	1.289258	1.413129	1.626839	2.362265	12.02995
DOE	-4412	-97.0706	-40.1006	-25.5801	-17.7653	-14.1749	-3.13949	11.77716	24.42066	31.47208	45.54118	71.40151	160.9943	323366
IC	-322	-6.27094	-3.31894	-1.80263	-1.0347	-0.73495	-0.20499	0.740486	1.175024	1.468553	1.875	2.908149	6.631851	100
CAP	-2205	-91.8947	-8.01278	-3.42609	-2.11167	-1.04005	-0.0657	14.11363	123.9296	4420	4420	4420	4420	4420
AT	-0.22594	0	0	0.001484	0.010445	0.02686	0.094848	3.669919	4.838846	5.247429	5.89095	7.327918	9.917061	41.98738
FAT	-2372.88	0	3.36E-05	0.004883	0.019236	0.039227	0.1185	141.0604	24811	24811	24811	24811	24811	24811
RT	-56.2182	0	0.000445	0.056069	0.221912	0.413905	1.251694	21.86693	43.97388	57.17325	98.81081	329.3391	16077	16077
PT	-22.5243	0	0	0.003834	0.033432	0.063751	0.304201	11.62998	17.3011	19.65157	24.33763	32.88365	63.63636	8770
CFLTD	-41558	-93.9421	-40.0062	-18.3714	-10.7567	-10	-10	10	10	11.12636	20.35043	51.63202	182.3333	44829.8
CFL	-10392.6	-21.7633	-9.00895	-4.18305	-2.71802	-1.78748	-0.61017	0.584089	1.105818	1.535337	2.426807	4.436548	14.24138	3193.579
CFINT	-3387981	-2657.2	-730.42	-387.165	-214.194	-150.073	-34.9138	35.71517	139.5808	198.3824	343.3003	729	3138	729659
T1_WCTA	-11.3008	-1.31601	-0.94538	-0.79608	-0.67261	-0.57967	-0.31563	0.549117	0.687731	0.726113	0.782088	0.83363	0.921511	7.386276
T2_RETA	-29.493	-1.31342	-0.81005	-0.58594	-0.46607	-0.33817	-0.13839	0.467847	0.586058	0.635469	0.677556	0.739384	0.827284	6.867868
T3_EBITTA	-5.19137	-0.54644	-0.32161	-0.20665	-0.14746	-0.11651	-0.03417	0.160826	0.236068	0.269524	0.31043	0.364789	0.487129	3.074792
T4_ETL	-33.8356	-0.61212	-0.40052	-0.30373	-0.237	-0.17524	-0.03562	2.886847	4.894866	5.873762	7.611439	11.11878	23.47645	162809.2
T5_STA	-0.22594	0	0	0.000837	0.007573	0.017808	0.087367	3.638505	4.794793	5.184591	5.834923	7.322702	9.851713	41.98738

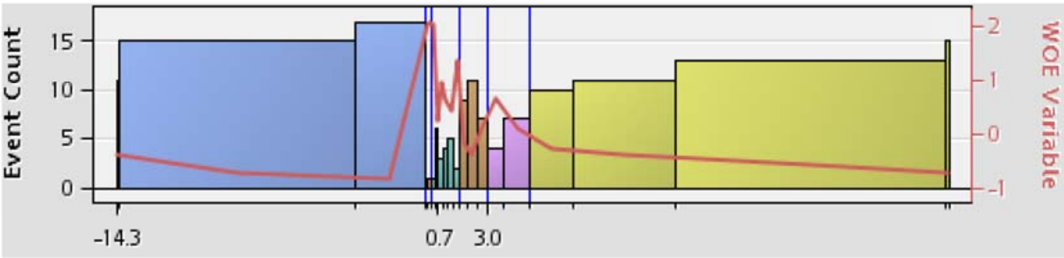
Appendix Table 1: Percentiles for value of observed indicators

```
if ROA < -0.912350598 then ROA = -0.912350598;
     else if ROA > 0.5004679656 then ROA = 0.5004679656;
if ROE < -4.451327434 then ROE = -4.451327434;
     else if ROE > 5.9158699809 then ROE = 5.9158699809;
if ROCE < -1.853917663 then ROCE = -1.853917663;
     else if ROCE > 2.5238095238 then ROCE = 2.5238095238;
if ROS < -6.730769231 then ROS = -6.730769231;
     else if ROS > 2.2034267066 then ROS = 2.2034267066;
if CR < 0.0159854504 then CR = 0.0159854504;
     else if CR > 29.377655172 then CR = 29.377655172;
if QR < 0.0062594431 then QR = 0.0062594431;
     else if QR > 45.505415162 then QR = 45.505415162;
if CASHR < -0.4419778 then CASHR = -0.4419778;
     else if CASHR > 26.725955204 then CASHR = 26.725955204;
if DR < 0.0272547004 then DR = 0.0272547004;
     else if DR > 2.3622654155 then DR = 2.3622654155;
if DOE < -40.14290552 then DOE = -40.14290552;
     else if DOE > 71.401509952 then DOE = 71.401509952;
if IC < -6.270935961 then IC = -6.270935961;
     else if IC > 6.6318508692 then IC = 6.6318508692;
if CAP < -62.45894737 then CAP = -62.45894737;
     else if CAP > 128.85915493 then CAP = 128.85915493;
if AT < 0 then AT = 0;
     else if AT > 9.9170607494 then AT = 9.9170607494;
if FAT < 0 then FAT = 0;
     else if FAT > 140 then FAT = 140;
if RT < 0 then RT = 0;
     else if RT > 100 then RT = 100;
if PT < 0 then PT = 0;
     else if PT > 32 then PT = 32;
if CFLTD < -38.44779582 then CFLTD = -38.44779582;
     else if CFLTD > 54.525944907 then CFLTD = 54.525944907;
if CFL < -21.76326173 then CFL = -21.76326173;
     else if CFL > 14.580474934 then CFL = 14.580474934;
if CFINT < -115.0380117 then CFINT = -115.0380117;
     else if CFINT > 74.719425823 then CFINT = 74.719425823;
if ROC < -93.96 then ROC = -93.96;
     else if ROC > 66.5 then ROC = 66.5;
if T1_WCTA < -1 then T1_WCTA = -1;
     else if T1_WCTA > 1 then T1_WCTA = 1;
if T2_RETA < -1 then T2_RETA = -1;
     else if T2_RETA > 1 then T2_RETA = 1;
if T3_EBITTA < -0.5 then T3_EBITTA = -0.5;
     else if T3_EBITTA > 0.5 then T3_EBITTA = 0.5;
if T4_ETL < -0.5 then T4_ETL = -0.5;
     else if T4_ETL > 10 then T4_ETL = 10;
if T5_STA < 0 then T5_STA = 0;
     else if T5_STA > 10 then T5_STA = 10;
```

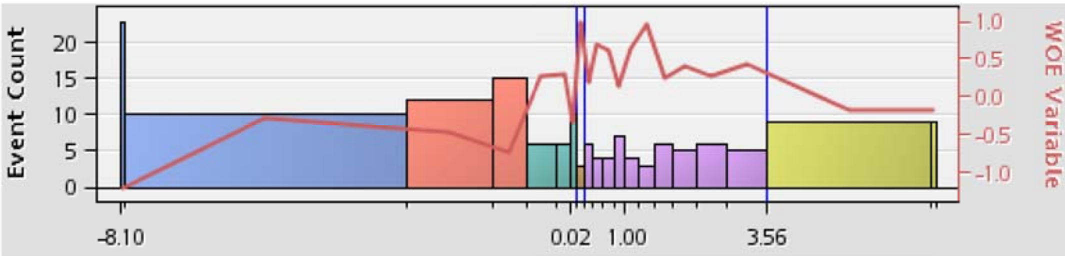
Appendix Figure 2: Values of cut-off points.
Variables beginning with “T” letter are those entering Altman score calculation

	ROA	ROE	ROCE	ROS	ROC	CR	QR	CASHR	DR	DOE	IC	CAP	AT	FAT	RT	PT	CFLTD	CFL	CFINT	T1_WCTA	T2_RETA	T3_EBITTA	T4_ETL	T5_STA
ROA	1	0,055833	0,200116	0,22823	0,444176	0,096318	0,071024	0,082994	-0,48221	0,020856	0,018448	0,146739	0,036517	0,031185	0,027486	0,2345	0,147857	0,066992	0,17983	0,382318	0,164854	0,676902	0,201861	0,037006
ROE	0,055833	1	0,338859	-0,03398	0,001686	0,003602	-0,0025	0,0116	0,089003	-0,12019	-0,02685	-0,00714	0,048241	0,041737	0,038582	-0,00424	0,025403	-0,00117	0,014269	-0,03237	-0,04615	0,09044	-0,04048	0,049826
ROCE	0,200116	0,338859	1	0,037667	0,094969	-0,02425	-0,02895	-0,02635	0,015154	0,050329	0,005595	0,039462	0,136273	0,123942	0,015419	0,06419	0,089972	0,032828	0,067391	0,00132	-0,02511	0,201734	-0,00133	0,135107
ROS	0,22823	-0,03398	0,037667	1	0,289253	-0,10113	-0,12805	-0,09413	-0,15574	0,054558	0,113129	0,019116	0,178826	0,054941	-0,03151	0,148763	0,054479	0,030353	0,048456	0,160357	0,086063	0,210259	0,043122	0,175299
ROC	0,444176	0,001686	0,094969	0,289253	1	0,035701	0,028347	0,037834	-0,19172	0,020875	0,009632	0,076682	0,062109	0,016531	0,00115	0,119714	0,084198	0,024304	0,102124	0,158826	0,069569	0,246724	0,03798	0,062057
CR	0,096318	0,003602	-0,02425	-0,10113	0,035701	1	0,837182	0,68714	-0,23356	-0,05178	-0,00504	0,151834	-0,10812	0,053156	0,088383	0,276151	0,013532	-0,02808	-0,01846	0,422612	0,144592	0,066403	0,381654	-0,10837
QR	0,071024	-0,0025	-0,02895	-0,12805	0,028347	0,837182	1	0,797264	-0,16417	-0,04101	-0,00873	0,13382	-0,11635	0,043459	0,077494	0,265262	0,001299	-0,03617	-0,00285	0,293307	0,087317	0,042313	0,257514	-0,11758
CASHR	0,082994	0,0116	-0,02635	-0,09413	0,037834	0,68714	0,797264	1	-0,15206	-0,0333	-0,00476	0,100583	-0,13254	0,003396	0,177672	0,193384	0,027596	-0,00656	0,037469	0,193157	0,05844	0,045566	0,244989	-0,13375
DR	-0,48221	0,089003	0,015154	-0,15574	-0,19172	-0,23356	-0,16417	-0,15206	1	0,061868	-0,00806	-0,22006	0,114925	0,057054	0,030666	-0,39638	-0,04943	0,008147	-0,02845	-0,63118	-0,55966	-0,31216	-0,66541	0,114585
DOE	0,020856	-0,12019	0,050329	0,054558	0,020875	-0,05178	-0,04101	-0,0333	0,061868	1	0,061709	0,068451	-0,01716	0,025313	-0,04098	-0,05841	0,003289	0,01262	0,020561	-0,02004	-0,03714	-0,00063	-0,09469	-0,01838
IC	0,018448	-0,02685	0,005595	0,113129	0,009632	-0,00504	-0,00873	-0,00476	-0,00806	0,061709	1	0,011537	-0,03122	-0,00989	0,041964	0,005008	-0,00583	0,001904	0,01006	-0,01346	-0,01408	0,025472	-0,01828	-0,03059
CAP	0,146739	-0,00714	0,039462	0,019116	0,076682	0,151834	0,13382	0,100583	-0,22006	0,068451	0,011537	1	0,111685	0,629517	0,039383	0,184002	-0,00244	-0,0493	-0,02184	0,354977	0,121198	0,060442	0,143767	0,111904
AT	0,036517	0,048241	0,136273	0,178826	0,062109	-0,10812	-0,11635	-0,13254	0,114925	-0,01716	-0,03122	0,111685	1	0,449108	0,183502	0,465015	0,0394	-0,00722	-0,00475	0,047599	-0,01204	0,118531	-0,08078	0,996445
FAT	0,031185	0,041737	0,123942	0,054941	0,016531	0,053156	0,043459	0,003396	0,057054	0,025313	-0,00989	0,629517	0,449108	1	0,04992	0,209504	-0,00075	-0,04345	-0,02761	0,270496	0,002763	0,062922	-0,01903	0,449509
RT	0,027486	0,038582	0,015419	-0,03151	0,00115	0,088383	0,077494	0,177672	0,030666	-0,04098	0,041964	0,039383	0,183502	0,04992	1	0,192184	0,007155	0,008272	0,021384	-0,08111	-0,07119	0,013785	0,003094	0,184553
PT	0,2345	-0,00424	0,06419	0,148763	0,119714	0,276151	0,265262	0,193384	-0,39638	-0,05841	0,005008	0,184002	0,465015	0,209504	0,192184	1	0,036653	-0,0185	0,008064	0,359172	0,270507	0,231375	0,420933	0,461569
CFLTD	0,147857	0,025403	0,089972	0,054479	0,084198	0,013532	0,001299	0,027596	-0,04943	0,003289	-0,00583	-0,00244	0,0394	-0,00075	0,007155	0,036653	1	0,508597	0,541258	0,034958	0,005396	0,132363	0,055051	0,041248
CFL	0,066992	-0,00117	0,032828	0,030353	0,024304	-0,02808	-0,03617	-0,00656	0,008147	0,01262	0,001904	-0,0493	-0,00722	-0,04345	0,008272	-0,0185	0,508597	1	0,438644	-0,02581	-0,02255	0,050131	-0,00044	-0,00745
CFINT	0,17983	0,014269	0,067391	0,048456	0,102124	-0,01846	-0,00285	0,037469	-0,02845	0,020561	0,01006	-0,02184	-0,00475	-0,02761	0,021384	0,008064	0,541258	0,438644	1	-0,03094	-0,01325	0,149093	0,009168	-0,00099
T1_WCTA	0,382318	-0,03237	0,00132	0,160357	0,158826	0,422612	0,293307	0,193157	-0,63118	-0,02004	-0,01346	0,354977	0,047599	0,270496	-0,08111	0,359172	0,034958	-0,02581	-0,03094	1	0,419873	0,279163	0,432065	0,046346
T2_RETA	0,164854	-0,04615	-0,02511	0,086063	0,069569	0,144592	0,087317	0,05844	-0,55966	-0,03714	-0,01408	0,121198	-0,01204	0,002763	-0,07119	0,270507	0,005396	-0,02255	-0,01325	0,419873	1	0,173846	0,385221	-0,01111
T3_EBITTA	0,676902	0,09044	0,201734	0,210259	0,246724	0,066403	0,042313	0,045566	-0,31216	-0,00063	0,025472	0,060442	0,118531	0,062922	0,013785	0,231375	0,132363	0,050131	0,149093	0,279163	0,173846	1	0,166761	0,115034
T4_ETL	0,201861	-0,04048	-0,00133	0,043122	0,03798	0,381654	0,257514	0,244989	-0,66541	-0,09469	-0,01828	0,143767	-0,08078	-0,01903	0,003094	0,420933	0,055051	-0,00044	0,009168	0,432065	0,385221	0,166761	1	-0,07989
T5_STA	0,037006	0,049826	0,135107	0,175299	0,062057	-0,10837	-0,11758	-0,13375	0,114585	-0,01838	-0,03059	0,111904	0,996445	0,449509	0,184553	0,461569	0,041248	-0,00745	-0,00099	0,046346	-0,01111	0,115034	-0,07989	1

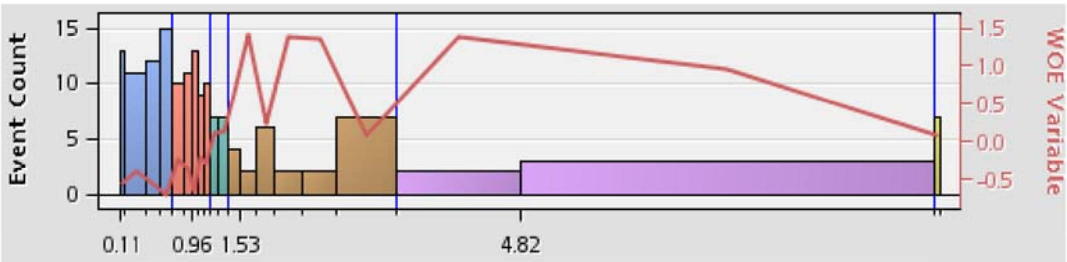
Appendix Table 2: Correlation matrix for all observed variables. The yellow squares represent variables from the same group. Positive correlation is indicated by the green strip and negative correlation by the red strip.



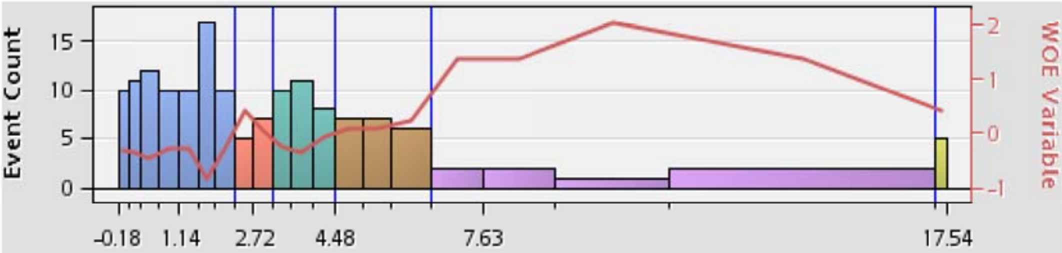
Appendix Figure 3: Default ratio values based on DOE values



Appendix Figure 4: Default ratio values based on ROC values



Appendix Figure 5: Default ratio values based on CR values



Appendix Figure 6: Default ratio values based on PT values

o1	o2	o3	o4	o5	w1	w2	w3	w4	w5	a1	a2	a3	a4	a5	GINI T	GINI V	KS T	KS V
1	2	3	4	5	5,1	7,8	5,0	9,7	7,6	79,1	20,0	80,6	0,7	179,5	0,4181	0,2966		0,3130
1	2	3	5	4	4,2	4,9	6,9	4,5	6,2	31,3	86,6	74,6	80,4	87,1	0,4067	0,2949		0,2850
1	2	4	3	5	3,5	1,8	6,0	4,0	4,2	36,9	76,4	106,1	106,6	34,1	0,4102	0,3122		0,2881
1	2	4	5	3	5,3	2,5	4,7	3,5	5,0	32,2	111,7	33,9	88,2	94,0	0,4094	0,2919		0,2809
1	2	5	3	4	3,7	2,1	3,9	5,5	4,2	40,0	29,4	127,6	64,6	98,4	0,4148	0,3133		0,2979
1	2	5	4	3	4,2	3,0	2,2	6,1	4,5	32,6	16,8	124,3	69,1	117,2	0,4153	0,3006		0,2695
1	3	2	4	5	5,0	7,6	4,0	10,0	5,5	94,1	44,1	41,7	0,3	179,8	0,4155	0,2892		0,2645
1	3	2	5	4	2,8	6,0	5,7	0,5	6,2	132,9	110,5	15,5	43,7	57,4	0,4180	0,2453		0,2455
1	3	4	2	5	4,7	5,4	4,7	8,9	5,1	56,8	107,8	25,4	0,5	169,6	0,4164	0,2759		0,2348
1	3	5	2	4	8,1	10,0	10,0	6,2	10,0	43,9	174,0	1,3	136,3	4,4	0,4133	0,2750		0,3053
1	4	2	3	5	8,3	6,2	1,8	9,3	5,1	53,9	82,1	97,3	125,8	0,9	0,4120	0,2884		0,2900
1	4	3	2	5	3,9	7,7	5,9	9,2	6,8	78,3	86,5	51,8	1,0	142,4	0,4203	0,2618		0,2460
1	2	3	4	5	1	1	1	1	1	15,4	50,9	136,2	4,1	153,4	0,4128	0,3012		0,2956
1	2	3	5	4	1	1	1	1	1	1,0	92,8	138,2	0,3	127,7	0,4137	0,2970		0,2779
1	2	4	3	5	1	1	1	1	1	15,3	32,8	124,0	29,6	158,2	0,4097	0,2957		0,2653
1	2	4	5	3	1	1	1	1	1	10,9	32,1	77,8	134,3	104,9	0,4068	0,3048		0,2872
1	2	5	3	4	1	1	1	1	1	18,7	8,6	120,8	95,7	116,2	0,4163	0,3155		0,3006
1	2	5	4	3	1	1	1	1	1	8,9	0,5	111,9	116,2	122,5	0,4108	0,3044		0,2912
1	3	2	4	5	1	1	1	1	1	41,9	157,0	0,2	158,3	2,6	0,4123	0,2961		0,2927
1	3	2	5	4	1	1	1	1	1	71,7	104,4	6,2	153,3	24,4	0,4143	0,2724		0,2519
1	3	5	2	4	1	1	1	1	1	134,4	178,8	0,8	40,5	5,4	0,4128	0,2787		0,2966
1	4	2	3	5	1	1	1	1	1	42,2	50,8	87,4	70,6	109,0	0,4025	0,2905		0,2655
1	4	3	2	5	1	1	1	1	1	126,6	154,6	76,4	2,1	0,3	0,4225	0,2799		0,2911
1	3	4	2	5	1	1	1	1	1	173,2	175,8	8,4	1,1	1,6	0,4132	0,3011		0,2993
1	2	3	4	5	5,9	7,4	8,8	9,7	0,0	72	72	72	72	72	0,4179	0,2997		0,3096
1	2	3	5	4	0,4	0,5	7,5	0,0	7,0	72	72	72	72	72	0,4179	0,2975		0,2967
1	2	4	3	5	5,0	3,6	4,4	10,0	0,6	72	72	72	72	72	0,4130	0,2980		0,2803
1	2	4	5	3	7,2	1,5	6,4	4,3	9,1	72	72	72	72	72	0,4081	0,2979		0,2947
1	2	5	3	4	6,5	1,8	4,3	9,1	5,3	72	72	72	72	72	0,4146	0,3135		0,2841
1	2	5	4	3	8,3	2,7	3,6	6,1	8,6	72	72	72	72	72	0,4139	0,2838		0,2459
1	3	2	4	5	5,7	9,7	4,5	10,0	0,0	72	72	72	72	72	0,4155	0,2909		0,2604
1	3	2	5	4	3,3	3,5	9,1	0,1	7,2	72	72	72	72	72	0,4181	0,2613		0,2781
1	3	4	2	5	7,0	8,2	8,3	6,6	1,3	72	72	72	72	72	0,4168	0,2966		0,2952
1	3	5	2	4	5,6	5,6	5,2	1,3	3,6	72	72	72	72	72	0,4090	0,2712		0,2820
1	4	2	3	5	5,9	2,5	2,7	10,0	0,0	72	72	72	72	72	0,4148	0,2783		0,2917
1	4	3	2	5	8,6	7,1	6,3	10,0	0,0	72	72	72	72	72	0,4208	0,3085		0,2948

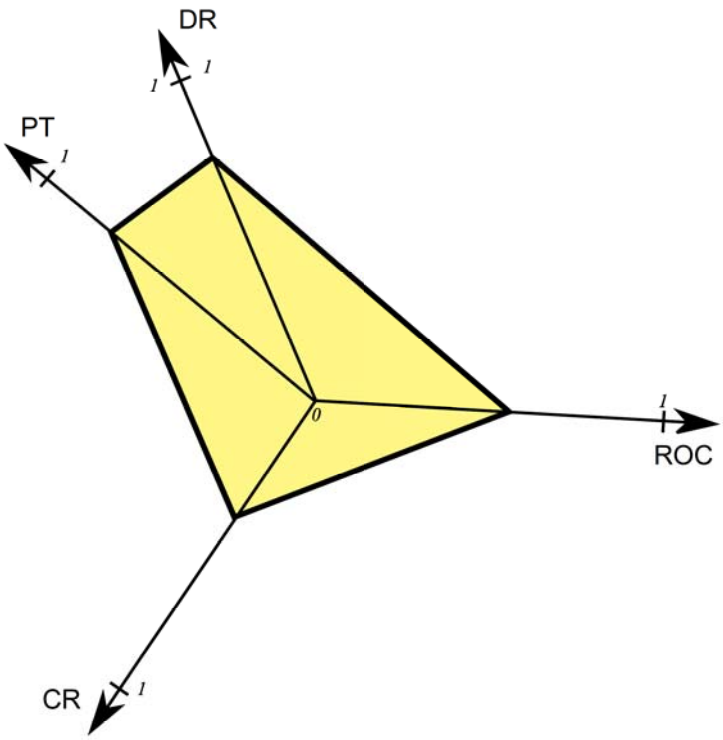
Appendix Table 3: The results of Gini index optimization.
“Gini T” column is the value for training set and “Gini V” is value for Validation set.
Kolmogorov-Smirnov index value for validation set is also supplied.

o1	o2	o3	o4	o5	w1	w2	w3	w4	w5	a1	a2	a3	a4	a5	GINI T	GINI V	KS T	KS V
1	2	3	4	5	5,7	5,5	9,8	3,1	6,0	39,4	7,8	159,2	116,1	37,5		0,2737	0,3431	0,2597
1	2	3	5	4	10,0	5,5	2,2	0,6	7,7	7,7	73,8	19,5	124,5	134,5		0,2825	0,3464	0,2957
1	2	4	3	5	3,8	7,9	6,4	8,4	7,3	9,9	26,5	163,4	147,0	13,2		0,3061	0,3464	0,2756
1	2	4	5	3	4,6	5,6	4,5	6,4	6,6	91,9	46,6	73,8	72,8	75,0		0,2861	0,3464	0,2669
1	2	5	3	4	4,1	6,3	1,5	6,2	6,8	160,7	29,8	3,5	75,0	91,0		0,2692	0,3464	0,2419
1	2	5	4	3	8,8	10,0	1,6	7,0	8,1	86,8	14,2	70,1	91,6	97,3		0,2639	0,3497	0,2407
1	3	2	4	5	10,0	10,0	9,2	9,7	3,0	117,5	0,0	61,6	158,0	22,8		0,2624	0,3464	0,2930
1	3	2	5	4	9,2	8,5	8,0	0,7	7,4	161,8	39,3	12,8	87,5	58,6		0,2562	0,3562	0,2547
1	3	4	2	5	8,7	10,0	7,2	5,4	10,0	147,6	29,9	78,3	96,6	7,5		0,2808	0,3431	0,2383
1	3	5	2	4	6,2	7,9	0,5	3,8	5,6	38,9	154,0	15,2	88,0	63,8		0,2590	0,3562	0,2278
1	4	2	3	5	10,0	9,0	8,0	6,7	0,3	54,3	1,0	69,7	180,0	55,0		0,2717	0,3464	0,2925
1	4	3	2	5	10,0	9,0	8,0	6,7	0,3	54,3	1,0	69,7	180,0	55,0		0,2938	0,3529	0,2998
1	2	3	4	5	1	1	1	1	1	12,9	32,1	94,8	98,9	121,2		0,3052	0,3464	0,2652
1	2	3	5	4	1	1	1	1	1	12,9	32,1	94,8	98,9	121,2		0,3077	0,3399	0,2710
1	2	4	3	5	1	1	1	1	1	12,9	32,1	94,8	98,9	121,2		0,3047	0,3431	0,2706
1	2	4	5	3	1	1	1	1	1	45,0	0,7	162,2	86,1	66,1		0,2906	0,3464	0,2658
1	2	5	3	4	1	1	1	1	1	45,0	0,7	162,2	86,1	66,1		0,2974	0,3497	0,2936
1	2	5	4	3	1	1	1	1	1	12,9	32,1	94,8	98,9	121,2		0,3044	0,3431	0,2818
1	3	2	4	5	1	1	1	1	1	109,6	14,9	130,5	54,1	50,9		0,2895	0,3431	0,2804
1	3	2	5	4	1	1	1	1	1	106,0	14,7	92,0	116,6	30,7		0,2884	0,3464	0,3028
1	3	4	2	5	1	1	1	1	1	109,6	14,9	130,5	54,1	50,9		0,2935	0,3464	0,2849
1	3	5	2	4	1	1	1	1	1	12,9	32,1	94,8	98,9	121,2		0,2835	0,3529	0,2399
1	4	2	3	5	1	1	1	1	1	94,2	0,2	43,3	151,2	71,1		0,2844	0,3366	0,2952
1	4	3	2	5	1	1	1	1	1	173,6	2,2	29,6	149,2	5,4		0,3133	0,3464	0,3028
1	2	3	4	5	9,1	7,9	7,5	9,1	0,6	72	72	72	72	72		0,2863	0,3497	0,3080
1	2	3	5	4	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2829	0,3464	0,2940
1	2	4	3	5	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2941	0,3431	0,2741
1	2	4	5	3	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2798	0,3431	0,2714
1	2	5	3	4	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2895	0,3464	0,2877
1	2	5	4	3	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2919	0,3529	0,2882
1	3	2	4	5	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2864	0,3464	0,2708
1	3	2	5	4	9,2	5,7	6,0	9,7	3,0	72	72	72	72	72		0,2965	0,3497	0,2920
1	3	4	2	5	9,1	7,9	7,5	9,1	0,6	72	72	72	72	72		0,3080	0,3431	0,2926
1	3	5	2	4	9,1	7,9	7,5	9,1	0,6	72	72	72	72	72		0,2920	0,3595	0,3128
1	4	2	3	5	9,1	4,8	7,6	7,3	0,0	72	72	72	72	72		0,2969	0,3497	0,2876
1	4	3	2	5	3,4	7,4	8,1	8,5	2,1	72	72	72	72	72		0,3129	0,3497	0,2924

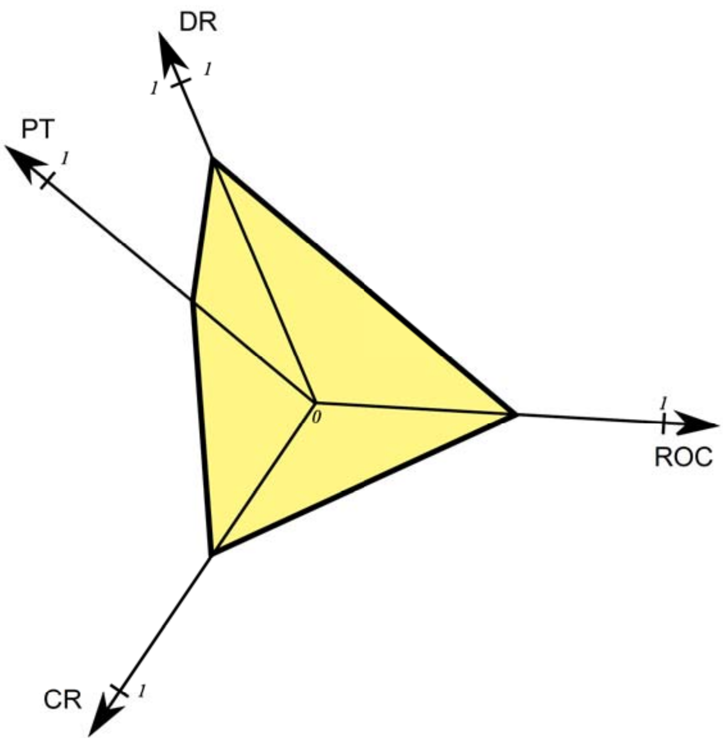
Appendix Table 4: The results of Kolmogorov-Smirnov index optimization. "KS T" column is the value for training set and "KS V" is value for Validation set. Gini index value for validation set is also supplied.

o1	o2	o3	o4	o5	w1	w2	w3	w4	w5	a1	a2	a3	a4	a5	GINI T	GINI V	KS T	KS V
1	2	3	4	5	1	1	1	1	1	72	72	72	72	72	0,3984	0,2804	0,3268	0,2708
1	2	3	5	4	1	1	1	1	1	72	72	72	72	72	0,4016	0,2888	0,3268	0,2735
1	2	4	3	5	1	1	1	1	1	72	72	72	72	72	0,3992	0,2788	0,317	0,2658
1	2	4	5	3	1	1	1	1	1	72	72	72	72	72	0,4003	0,2893	0,3333	0,2758
1	2	5	3	4	1	1	1	1	1	72	72	72	72	72	0,4021	0,2862	0,3203	0,2489
1	2	5	4	3	1	1	1	1	1	72	72	72	72	72	0,3997	0,2806	0,3268	0,2509
1	3	2	4	5	1	1	1	1	1	72	72	72	72	72	0,3994	0,2878	0,3301	0,2766
1	3	2	5	4	1	1	1	1	1	72	72	72	72	72	0,4011	0,2801	0,3268	0,2585
1	3	4	2	5	1	1	1	1	1	72	72	72	72	72	0,3986	0,2791	0,3301	0,249
1	3	5	2	4	1	1	1	1	1	72	72	72	72	72	0,401	0,2833	0,3235	0,257
1	4	2	3	5	1	1	1	1	1	72	72	72	72	72	0,4009	0,2918	0,3268	0,2716
1	4	3	2	5	1	1	1	1	1	72	72	72	72	72	0,4004	0,2845	0,3268	0,2509

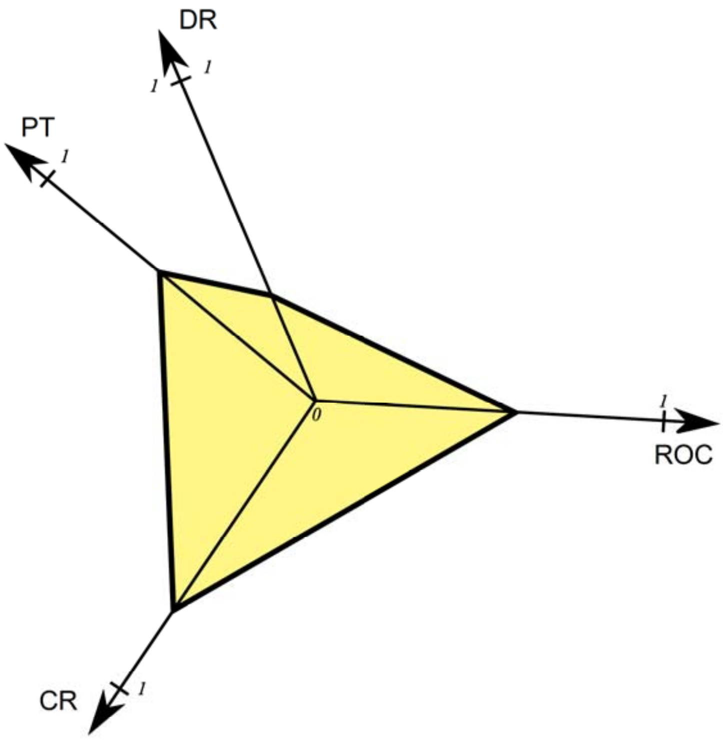
Appendix Table 5: The result of Gini and KS indexes optimization for scenario where weights and angles are constant and only parameters order is changing



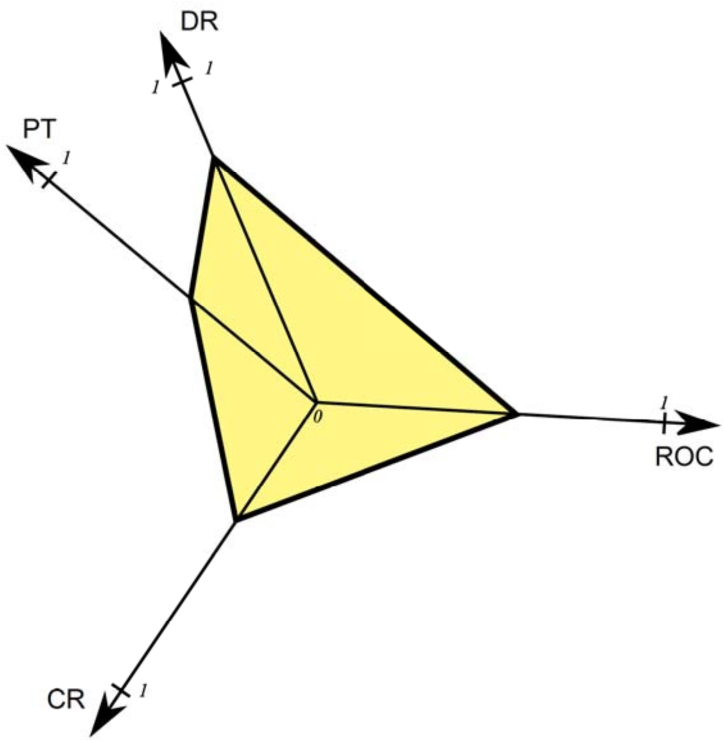
Appendix Figure 7: ČEZ Distribuce, a.s.



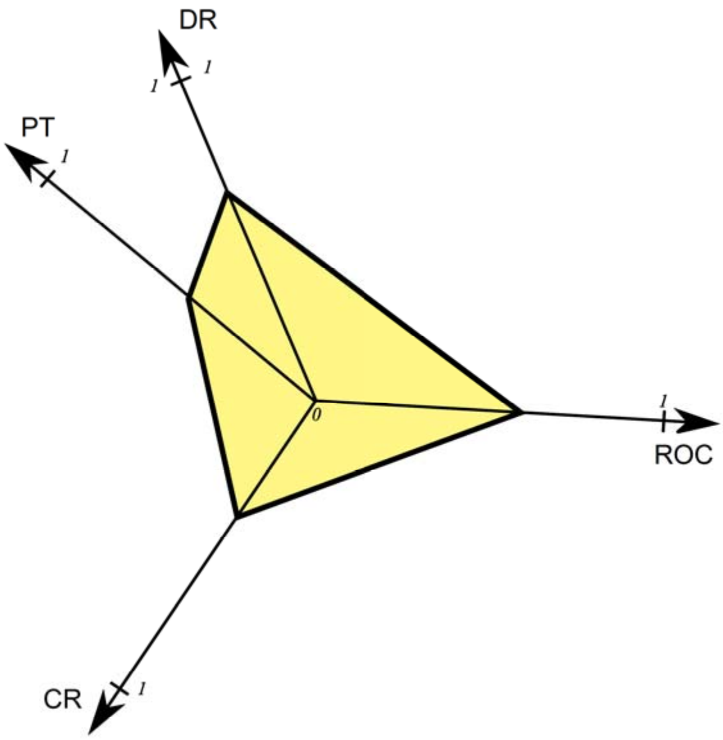
Appendix Figure 8: CEZ Razpredelenje Bulgaria AD



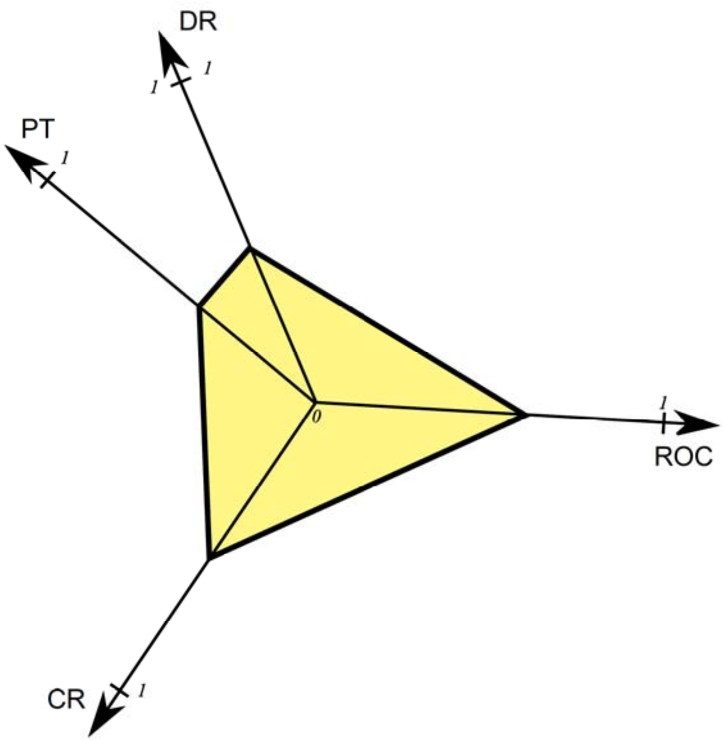
Appendix Figure 9: SRD Réseaux de Distribution



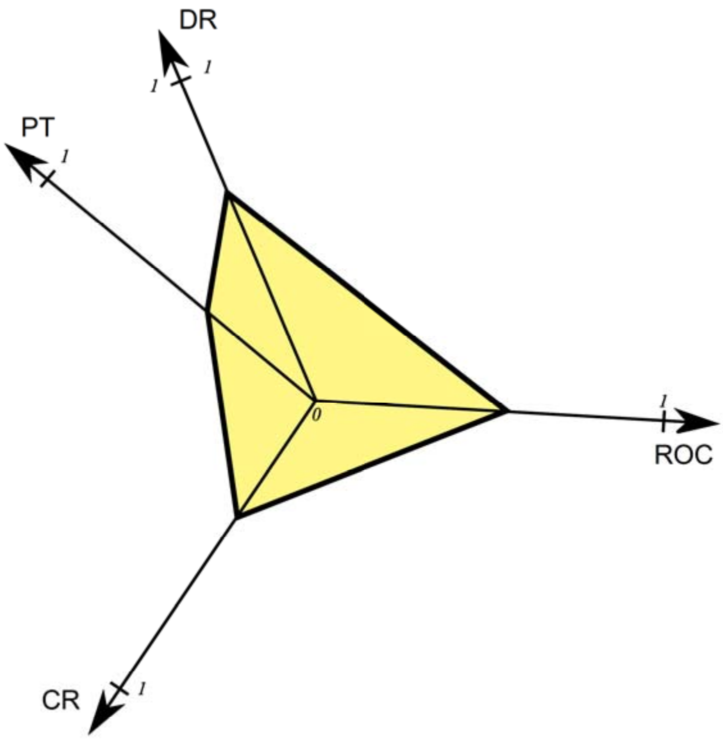
Appendix Figure 10: E.ON Distribuce, a.s.



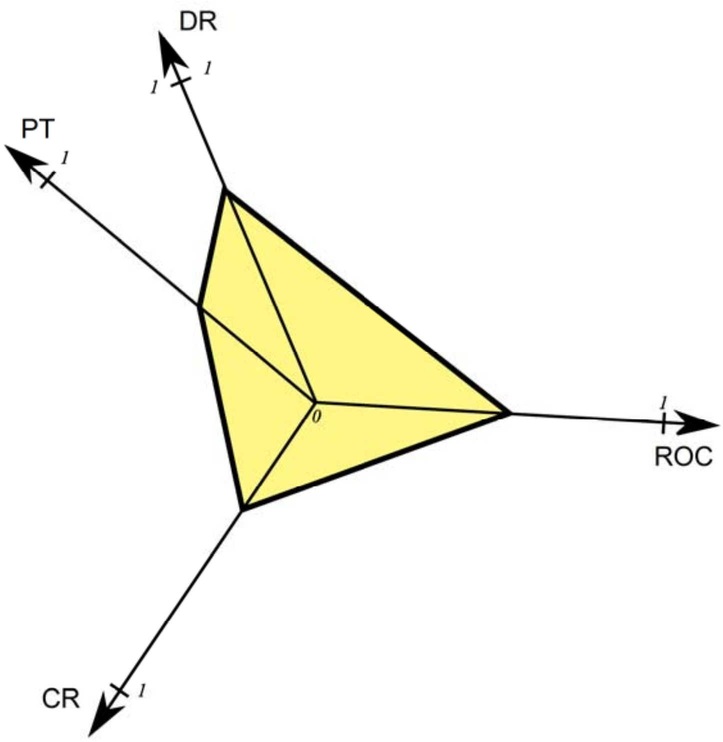
Appendix Figure 11: SP Distribution Limited



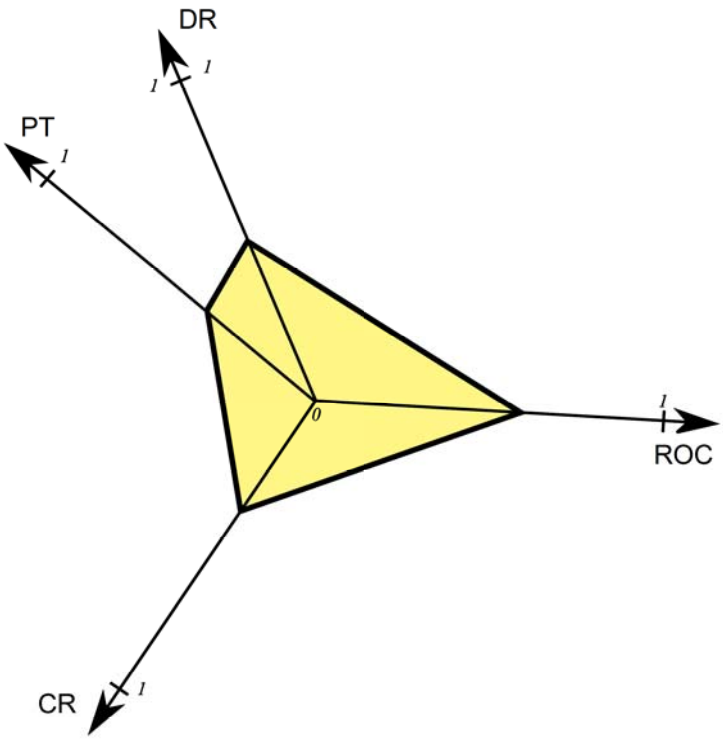
Appendix Figure 12: Electricity North West Limited



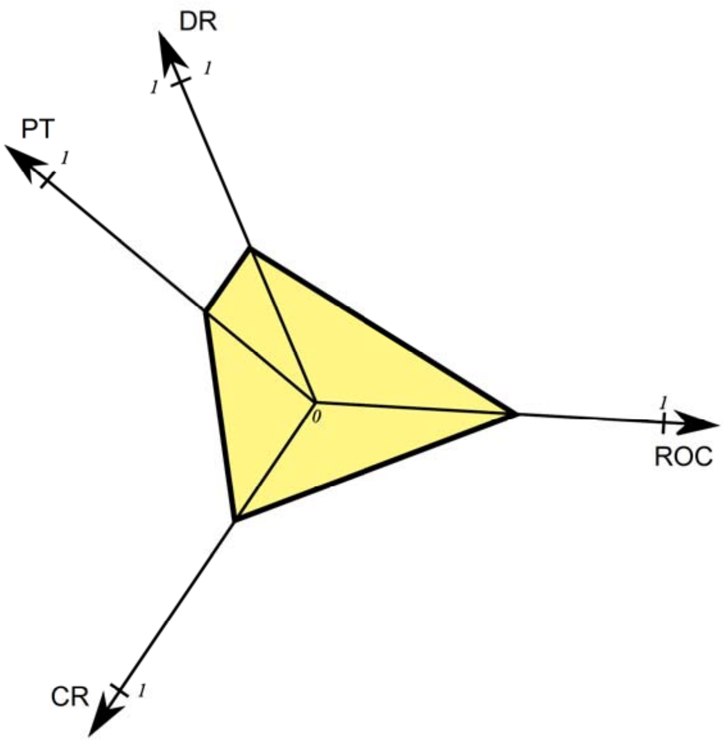
Appendix Figure 13: RARIK ohf



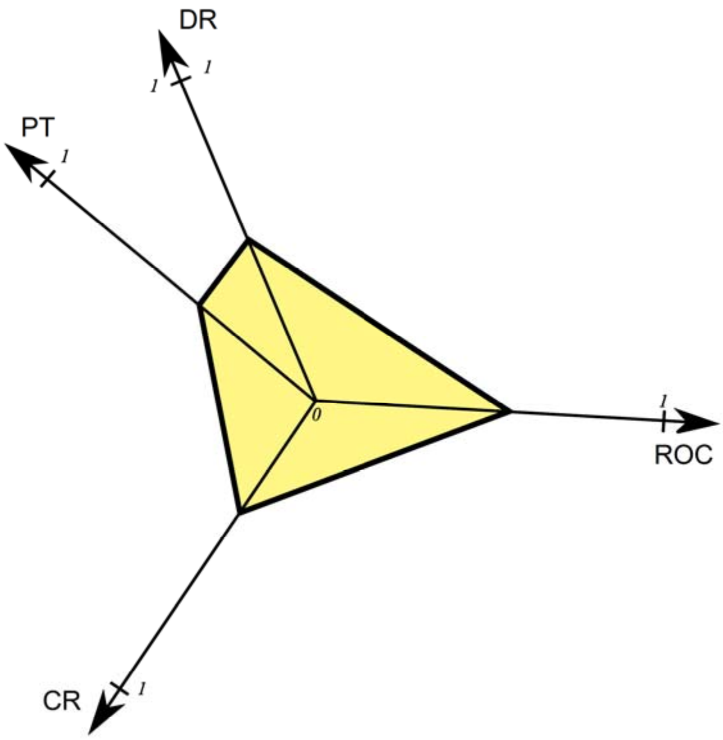
Appendix Figure 14: Sadales Tikls AS



Appendix Figure 15: Northern PowerGrid PLC



Appendix Figure 16: Adger Energi Nett AS



Appendix Figure 17: Eesti Energia Jaotusvork OÜ

References

[1] R. Medrano-B, Teixeira, J., "A KALDORIAN MACROECONOMIC INDEX OF ECONOMIC WELFARE," *Revista de História Econômica & Economia Regional Aplicada,* vol. 8, 2013.

[2] H. Fialová, *Úvod do ekonomiky - Základy makroekonomiky*: Czech Technical University in Prague, 2006.

[3] N. Kaldor, "Conflicts in National Economic Objectives," *The Economic Journal,* vol. 81, pp. 1-16, 1971.

[4] D. Pontiggia, "Optimal long-run inflation and the New Keynesian model," *Journal of Macroeconomics,* vol. 34, pp. 1077-1094, 2012.

[5] E. Altman, "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance,* vol. 23, pp. 589-609, 1968.

[6] M. Karajica, "Metody hodnocení výkonnosti elektro-distribučních společností," Doctoral thesis, Czech Technical University in Prague, 2014.

[7] L. Medema, R. H. Koning, and R. Lensink, "A practical approach to validating a PD model," *Journal of Banking & Finance,* vol. 33, pp. 701-708, 4// 2009.

[8] W. Xu, Xiao, Z., Dang, X., Yang, D., Yang, X., "Financial ratio selection for business failure prediction using soft set theory," *Knowledge-Based Systems,* vol. 63, pp. 59-67, 2014.

[9] D. Hosmer, Jr., Lemeshow, S., Sturdivant, R., *Applied Logistic Regression*, 3 ed. USA: John Wiley & Sons, 2013.

[10] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics,* vol. 35, pp. 352-359.

[11] C. Bolton, "Logistic regression and its application in credit scoring," Mathematics and Applied Mathematics, University of Pretoria, Pretoria, South Africa, 2009.

[12] O. A. Abdalla, M. N. Zakaria, S. Sulaiman, and W. F. W. Ahmad, "A comparison of feed-forward back-propagation and radial basis artificial neural networks: A Monte Carlo study," in *Information Technology (ITSim), 2010 International Symposium in*, 2010, pp. 994-998.

[13] X. Tiantian, Y. Hao, and B. Wilamowski, "Comparison between traditional neural networks and radial basis function networks," in *Industrial Electronics (ISIE), 2011 IEEE International Symposium on*, 2011, pp. 1194-1199.

[14] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*: MIT Press, 1995.

[15] C. M. Bishop, *Neural Networks for Pattern Recognition*: Oxford University Press, Inc., 1995.

[16] B. D. Ripley and N. L. Hjort, *Pattern Recognition and Neural Networks*: Cambridge University Press, 1995.

[17] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *Neural Networks, IEEE Transactions on,* vol. 5, pp. 989-993, 1994.

[18] Y. Freund and R. Schapire, "{A} {S}hort {I}ntroduction to {B}oosting," *Journal of Japanese Society for Artificial Intelligence,* vol. 14, pp. 771-780, // 1999.

[19] R. Irwin, Irwin, T., "Appraising Credit Ratings: Does the CAP Fit Better than the ROC?," *IMF Working Paper,* 2012.

[20] J. Bemš, M. Macaš, P. Pošík, J. Žegklitz, and O. Starý, "Innovative Default Prediction Approach," *(submitted to Expert Systems With Applications @11/2014),* 2014.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*: Wiley-Interscience, 2000.
[22] J. R. Quinlan, "Improved use of continuous attributes in C4.5," *J. Artif. Int. Res.,* vol. 4, pp. 77-90, 1996.
[23] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-32, 2001/10/01 2001.
[24] M. Marinaki, Y. Marinakis, M. Doumpos, N. Matsatsinis, and C. Zopounidis, "A comparison of several nearest neighbor classifier metrics using Tabu Search algorithm for the feature selection problem," *Optimization Letters,* vol. 2, pp. 299-308, 2008/06/01 2008.
[25] B. V. Dasarathy, "Nearest neighbor ({NN}) norms:{NN} pattern classification techniques," 1991.
[26] J. H. Holland, *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan Press, 1975.
[27] J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*: MIT Press, 1992.
[28] C. Ryan, J. J. Collins, and M. Neill, "Grammatical evolution: Evolving programs for an arbitrary language," in *Genetic Programming*. vol. 1391, W. Banzhaf, R. Poli, M. Schoenauer, and T. Fogarty, Eds., ed: Springer Berlin Heidelberg, 1998, pp. 83-96.
[29] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision-making units," *European Journal of Operational Research,* vol. 3, p. 339, 7// 1979.
[30] M. D. Pettypool and M. D. Troutt, "Recent improvements to data envelopment analysis," *Mathematical and Computer Modelling,* vol. 11, pp. 1104-1106, // 1988.
[31] J. Dudorkin, "Systémové inženýrství a rozhodování," ed: ČVUT, 2003.
[32] S. Satchell and W. Xia, "Analytic Models of the ROC Curve: Applications to Credit Rating Model Validation," ed, 2007.
[33] M. V. D. Burgt, "Calibrating Low-Default Portfolios, using the Cumulative Accuracy Profile," 2007.
[34] Corinna and Mehryar, "Confidence Intervals for the Area Under the ROC Curve," in *Advances in Neural Information Processing Systems 17*, L. Saul, Y. Weiss, and Bottou, Eds., ed: MIT Press, 2005, pp. 305-312.
[35] B. Engelmann, E. Hayden, and D. Tasche, "Measuring the Discriminative Power of Rating Systems," ed: DEUTSCHE BUNDESBANK, 2003.
[36] A. Bradley, "ROC curve equivalence using the Kolmogorov–Smirnov test," *Pattern Recognition Letters,* vol. 34, pp. 470-475, 2013.
[37] H. Dalton, "The Measurement of the Inequality of Incomes," *The Economic Journal,* vol. 30, pp. 348-361, 1920.
[38] W. Orth, "The predictive accuracy of credit ratings: Measurement and statistical inference," ed: University of Cologne Statistics and Econometrics, 2011.
[39] C. Gini, *Variability and Mutability*: Tipogr. di P. Cuppini, 1912.
[40] W. Greene, *Econometric Analysis*: Prentice Hall, 2003.
[41] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *Journal of Clinical Epidemiology,* vol. 49, pp. 1373-1379, 12// 1996.
[42] I. J. Myung, "Tutorial on maximum likelihood estimation," *J. Math. Psychol.,* vol. 47, pp. 90-100, 2003.

[43] (November 10). *Maximum Likelihood Estimation*. Available: https://onlinecourses.science.psu.edu/stat414/node/191

[44] (November 20). *Biological Neurons* Available: http://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/Biology/index.html

[45] L. Mastin. (2010, November 20). *NEURONS & SYNAPSES*. Available: http://www.human-memory.net/brain_neurons.html

[46] (2014, November 20). *Axon*. Available: http://en.wikipedia.org/wiki/Axon#mediaviewer/File:Neuron_Hand-tuned.svg

[47] Y. Mochizuki and S. Shinomoto, "Analog and digital codes in the brain," *Physical Review E,* vol. 89, p. 022705, 02/04/ 2014.

[48] (2014, November 20). *Biological Neural Network*. Available: http://en.wikibooks.org/wiki/File:Basic_Synapse.svg

[49] (2000, November 20). *Artificial Neuron*. Available: http://lcn.epfl.ch/tutorial/english/aneuron/html/components.html

[50] (2014, November 25). *Artificial Neural Networks/Activation Functions*. Available: http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Activation_Functions#mediaviewer/File:ArtificialNeuronModel_english.png

[51] S. I. Inc. (2000, November 25). *The NEURAL Procedure*. Available: http://support.sas.com/documentation/onlinedoc/miner/em43/neural.pdf

[52] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*: Springer-Verlag New York, Inc., 2006.

[53] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*: MIT Press, 1998.

[54] (2014, November 25). *Artificial Neural Networks/Neural Network Basics*. Available: http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Neural_Network_Basics#mediaviewer/File:Neural_network.svg

[55] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature,* vol. 323, pp. 533-536, 10/09/print 1986.

[56] A. J. P. D. S. Kim. (November 25). *Backpropagation*. Available: http://www.cse.unsw.edu.au/~cs9417ml/MLP2/

[57] S. Roweis, "Levenberg-Marquardt Optimization," ed: New York University.

[58] H. Yu and B. M. Wiamowski, "Levenberg–Marquardt Training," ed: Auburn University, 2010.

[59] DTREG. (2014, December 1). *Support Vector Machine*. Available: http://www.dtreg.com/solution/view/20

[60] F. v. d. Heijden, *Classification, parameter estimation, and state estimation : an engineering approach using MATLAB*. Chichester: John Wiley & Sons, 2004.

[61] scikit-learn. (2014, January 8). *Underfitting vs. Overfitting*. Available: http://scikit-learn.org/stable/auto_examples/plot_underfitting_overfitting.html

[62] Gringer. (2007, January 8). *Overfitting*. Available: http://commons.wikimedia.org/wiki/File:Overfitting_svg.svg

[63] D. Delen, Kuzey, C., Uyar, A., "Measuring firm performance using financial ratios: A decision tree approach," *Expert System with Applications,* vol. 40, pp. 3970-3983, August 2013 2013.

[64] V. Easton and J. McColl. (1997, November 10). *Hypothesis testing*. Available: http://www.stats.gla.ac.uk/steps/glossary/hypothesis_testing.html

[65] W. W. Daniel and J. C. Terrell, *Business Statistics: Basic Concepts and Methodology*: Houghton Mifflin, 1975.

[66] C. H. Brase and C. P. Brase, *Understandable Statistics: Concepts and Methods, 10th Edition*: Richard Stratton, 2012.

[67] StatTrek. (2014, November 10). *What is Hypothesis Testing?* Available: http://stattrek.com/hypothesis-test/hypothesis-testing.aspx?tutorial=ap

[68] A. K. Akobeng, "Understanding diagnostic tests 1: sensitivity, specificity and predictive values," *Acta Pædiatrica,* vol. 96, pp. 338-341, 2007.

[69] A. K. Akobeng, "Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice," *Acta Pædiatrica,* vol. 96, pp. 487-491, 2007.

[70] A. K. Akobeng, "Understanding diagnostic tests 3: receiver operating characteristic curves," *Acta Pædiatrica,* vol. 96, pp. 644-647, 2007.

[71] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Trans. on Knowl. and Data Eng.,* vol. 21, pp. 1263-1284, 2009.

[72] FSC, "Guidance Note V3, Capital Requirements Directive, The IRB Approach," ed. Gibraltar: Financial Services Commission, 2013.

[73] B. C. o. B. Supervision, "International Convergence of Capital Measurement and Capital Standards ", ed. Basel, Switzerland Bank for International Settlements 2004.

[74] (2014, December 20). *Accounts Receivable Turnover Ratio*. Available: http://www.myaccountingcourse.com/financial-ratios/accounts-receivable-turnover-ratio

[75] A. Z. Lin, "Variable Reduction in SAS by Using Weight of Evidence and Information Value," presented at the SAS Global Forum 2013, San Francisco, 2013.

[76] B. Lund and D. Brotherton, "Information Value Statistic," presented at the MWSUG 2013 Conference, Omaha, Nebraska, 2013.

[77] StatSoft. (January 02). *Weight of Evidence (WoE) Introductory Overview*. Available: http://documentation.statsoft.com/STATISTICAHelp.aspx?path=WeightofEvidence/WeightofEvidenceWoEIntroductoryOverview

[78] E. Weisstein. (January 10). *Circular Permutation*. Available: http://mathworld.wolfram.com/CircularPermutation.html

[79] *DIRECTIVE 2009/72/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning common rules for the internal market in electricity and repealing Directive 2003/54/EC*, 2009.