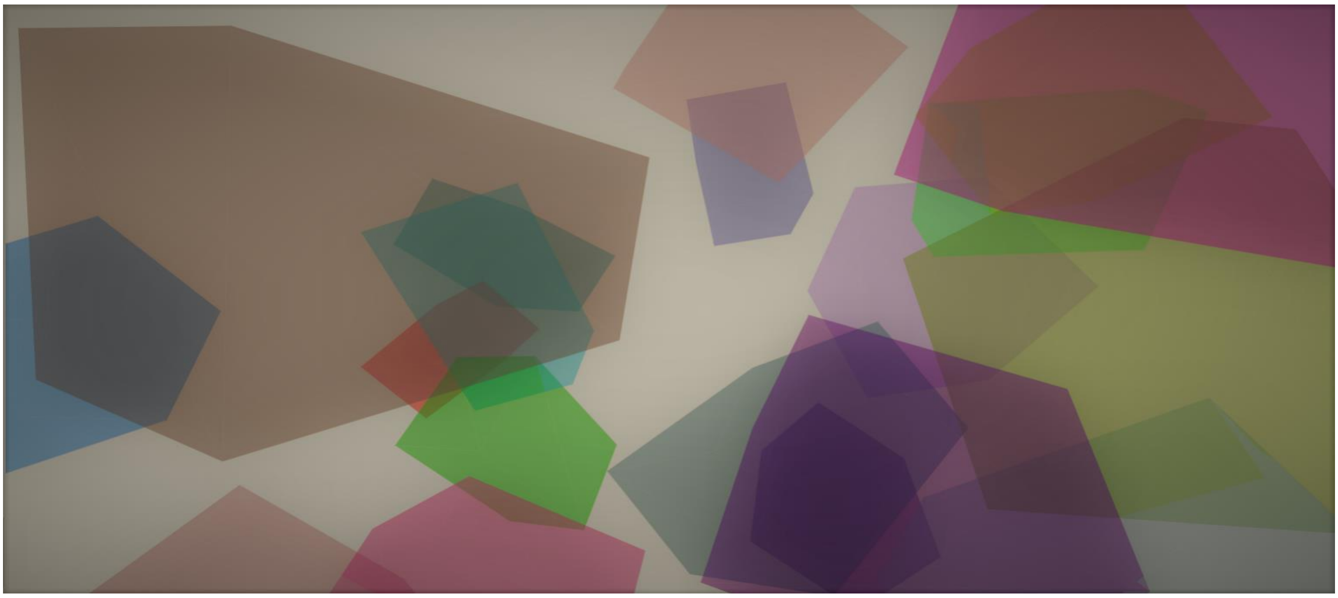

VALIDATION PROCESS FOR SCORING AND RATING MODELS USING NEURAL NETWORKS

Dr. Anatoliy Antonov, Dr. Ventsislav Nikolov
January 9th, 2017

Eurorisk Systems Ltd.
31, General Kiselov str.
9002 Varna, Bulgaria
Phone +359 52 612 367
Fax +359 52 612 371
info@euroriskystems.com
www.euroriskystems.com

Abstract: *This research paper investigates the validation and calibration of models for determination of credit scoring and rating with statistical methods. This is done through a comparison of the results of the model to an alternative model, based on a neural network, and a calculation of different statistical parameters. A prototype of a software system for analysis and evaluations is represented that calculates distance, standard deviation, correlation, cummulative accuracy profile (CAP), as well as accumulation and analysis of historical statistics for default losses.*

Key words: *credit rating, scoring, validation, analysis, calibration, neural networks.*

The validation of models that include selected objective quantitative and subjective qualitative factors, as well as the calibration of the weighted coefficients, saturation degrees, and compulsory conditions (K.O. testings), along with the final classification for scoring and rating levels is an iterative process that is a responsibility of the bank.

Software vendors and implementers can support regulatory authorities during the preparation and implementation of model validations. This is achieved via analysis, preparation of examples and participation in discussions on validation and calibration. The models are implemented via statistical methods and regressive procedures, such as neural networks and CAP (Cumulative Accuracy Profile) calculations. In the literature exist numerous analysis and studies on model validation (e.g. www.google.com: Basel III rating system validation) and can be used to meet the requirements of regulatory authorities.

This research paper introduces specific principles and examples for the validation of models using neural networks. Eurorisk Systems Ltd. has a longstanding experience in using neural networks in various financial areas, such as clustering and prediction of factors, identification of distributions, calculation of implied ratings and multifactor models, which are implemented in different modules within the software systems of financial institutions.

1. The Object of Validation

Given:

Given is a set of borrowers, along with their personal data, such as age, savings, income, real estate, etc. For each borrower, a credit rating/scoring is given that has been determined in the past, using the rating system which is to be investigated.

Goal:

The goal is to create a mathematical alternative model for validation based on existing data about the borrower and its scoring/rating values, so that:

- In case of emergence of new borrowers whose scoring/rating is yet unknown, the model can determine the value of the scoring/rating and can compare it to results of the rating system.

- A comparison between scoring/rating values is possible, using respective rating systems and mathematical models after the calibration, i.e. the training of the mathematical model. The goal implies validation and calculation of statistics as error margin, standard deviation, correlation, CAP, Gini coefficient, etc.

2. Neural Network

The defined mathematical model is based on a multi-layer neural network that operates using borrowers' symbolic and numerical data (factors). This data has been previously converted and processed in a suitable manner. The layers mentioned within the neural network are:

- Input layer: the number of neurons = number of input factors, e.g. 24
- Hidden layer: the number of neurons is determined by rules, e.g. 10
- Output layer: the number of neurons depends on the number of results, e.g. 2 neurons: for scoring und rating.

The network's structure is shown in Fig. 1. The neurons are connected to each other via weighted connections *Wij* and *Wkj* which are automatically adjusted during the training. Thus, the neural network realizes a hierarchical scoring in two levels. Neurons are modeled via a polynomial function, in which the outputs of the preceding neurons participate, and after that an internal non-linear activation function is applied, thus producing the neuron output.

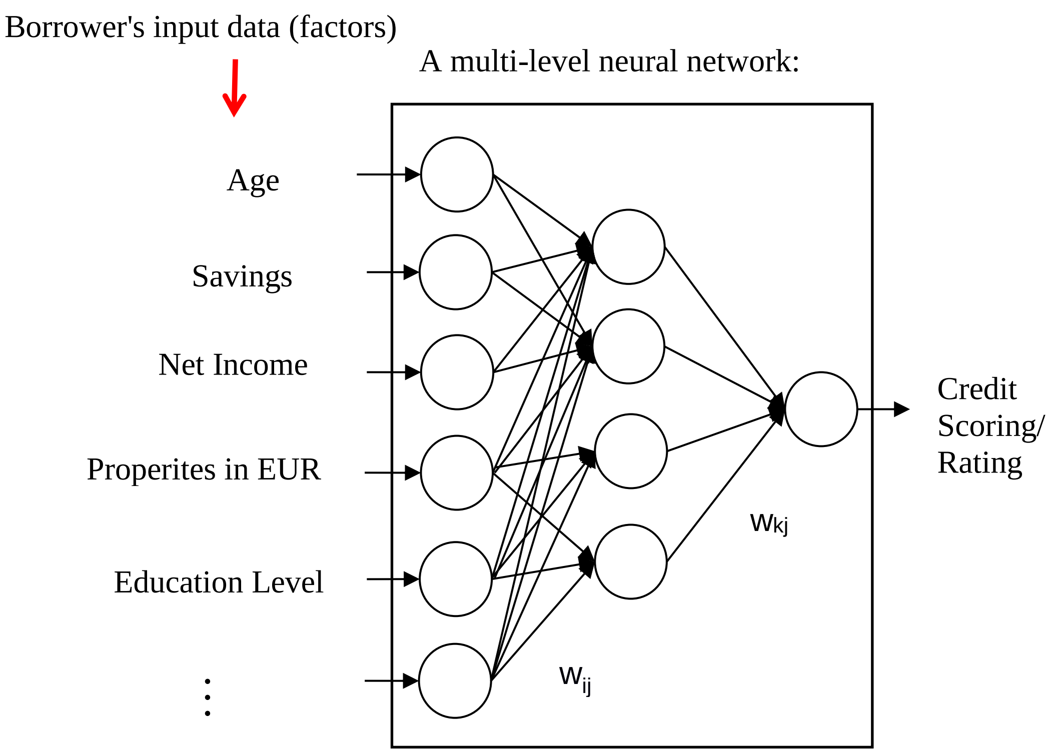

Fig. 1. Structure of a neural network

The neural network works in two stages:

Stage 1: Training the neural network

A set of input data (factors) for each borrower is associated to the scorings/ratings, that are pre-determined by a scoring/rating system, and by them the neural network's internal structure is adjusted in such a way that it has to “learn“ the dependencies between the input factors and the output results by adjustment of the weights of the connections. This stage is shown in Fig. 2. The training takes place automatically in multiple epochs (successive training sessions).

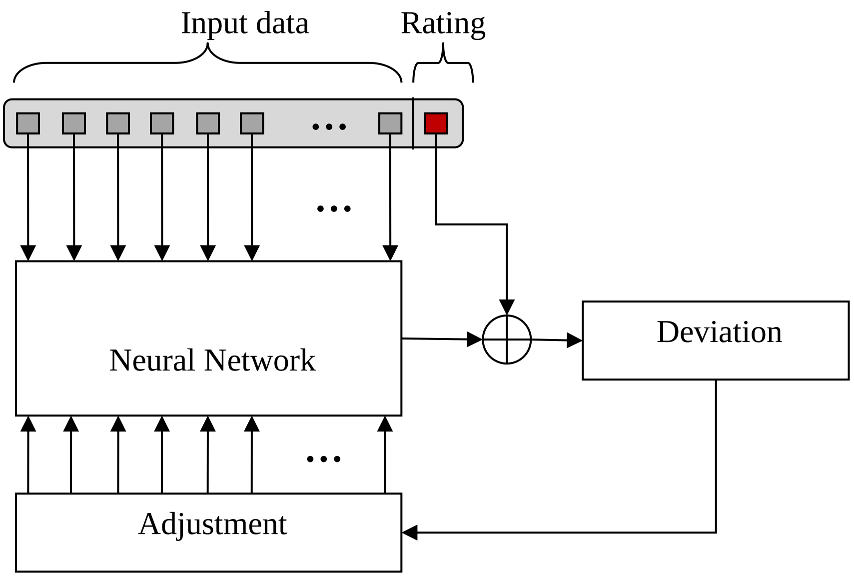

Fig. 2. Training of the neural network

Stage 2: Usage of the neural network

New input data are added to the previously trained neural network, which generates the value of the unknown credit scoring/rating value.

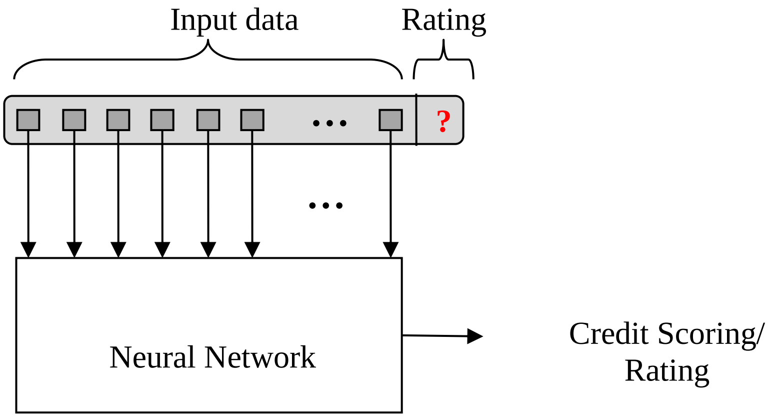

Fig. 3. Definition of the credit scoring/rating via trained neural networks

3. Prototype modul

The work of the neural network is illustrated with a prototype that is based on actual historical data and carries out the validation task (see above: 1. The Object of Validation).

A set of 24 input data (factors) for 821 borrowers (grouped according to their credit type into three groups: mortgage credit, consumer credit, and overdraft) are loaded into the neural network (Fig. 4). First, the credit type is selected. After configuring the network, it is then trained, taking into consideration all the borrowers in the group. Thereupon, individual or batch evaluations are possible.

The validation is based on the comparison between each individual scoring from the rating model and each individual scoring from the trained neural network. The network's accumulated knowledge represents the experience obtained during the training of all presented scores.

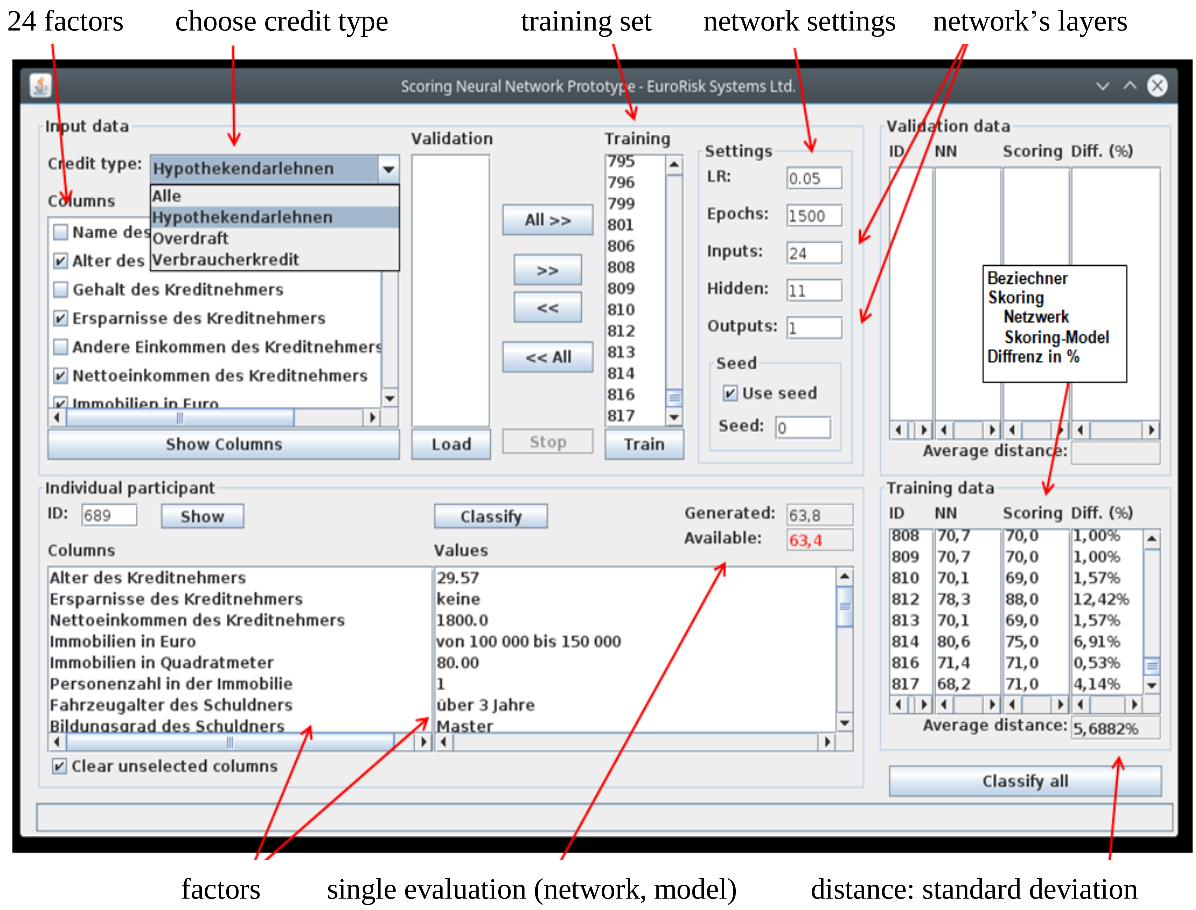

Fig. 4. Prototype of the neural network – training and validation for mortgage loans

These results are statistically summarized and can be used for the evaluation and validation of the scoring/rating models, e.g. as error, standard deviation, correlation, or CAP. The statistics show the extent to which scoring rules within a scoring model are consistent regarding the generalization and whether they meet the general strategy, or if they are based on specific exceptions and individual decisions.

Number	Factor	Example: Mortgage Loan	Example: Overdraft
1	**Age**	51,62	47,38
2	**Savings**	1 000 - 10 000	none
3	**Net Income**	1434	0
4	**Properties (in euros)**	over 150 000	over 150 000
5	**Properties (in squere meters)**	72	200
6	**Number of persons in the household**	3	1
7	**Age of vehicle**	over 3 years	over 3 years
8	**Education level**	Hight School	Master
9	**Marital status**	Married	married
10	**Employer's legal form**	Ltd.	Member companies
11	**Business**	Other	Other
12	**Job title / Employee in**	Local business	Local business

13	**Sources of income**	Indefinite permanent position	Indefinite permanent position
14	**Work experience (in years)**	25	23
15	**Subtype**	Citizen	Citizen
16	**Community**	Berlin	Hamburg
17	**Bank's target group**	No	Target Group I
18	**Credit type**	Mortgage loan	Overdraft
19	**Credit amount**	61360	1500
20	**Interest Rate**	8,15	10,95
21	**Currency**	EUR	BGN
22	**Loan repayment (in months)**	216	24
23	**Repayment type**	Annuity plan	
24	**Income test method**	Verified income	
	Scoring	**72,80**	**58,40**

Fig. 5. Scoring/rating factors, illustrated with two examples

The example in Fig. 5 shows the list of the main factors for two different borrowers. The borrowers' data are processed numerically, prior to its using by the neural network. The last line represents the scoring results that have been calculated using the scoring model. Scoring values lie between 0% and 100%. Data and scoring results of all borrowers from the selected group are used for the training of the network.

With the help of the prototype presented in Fig. 5, a training is executed using only partial data and scoring results of the borrower (training set). The rest (validating set) appears unknown to the neural network. The scorings of the validation set in this case indicate to what extent the network can cope with unknown data of the scoring model, e.g. how powerful the discriminant function of the scoring model is.

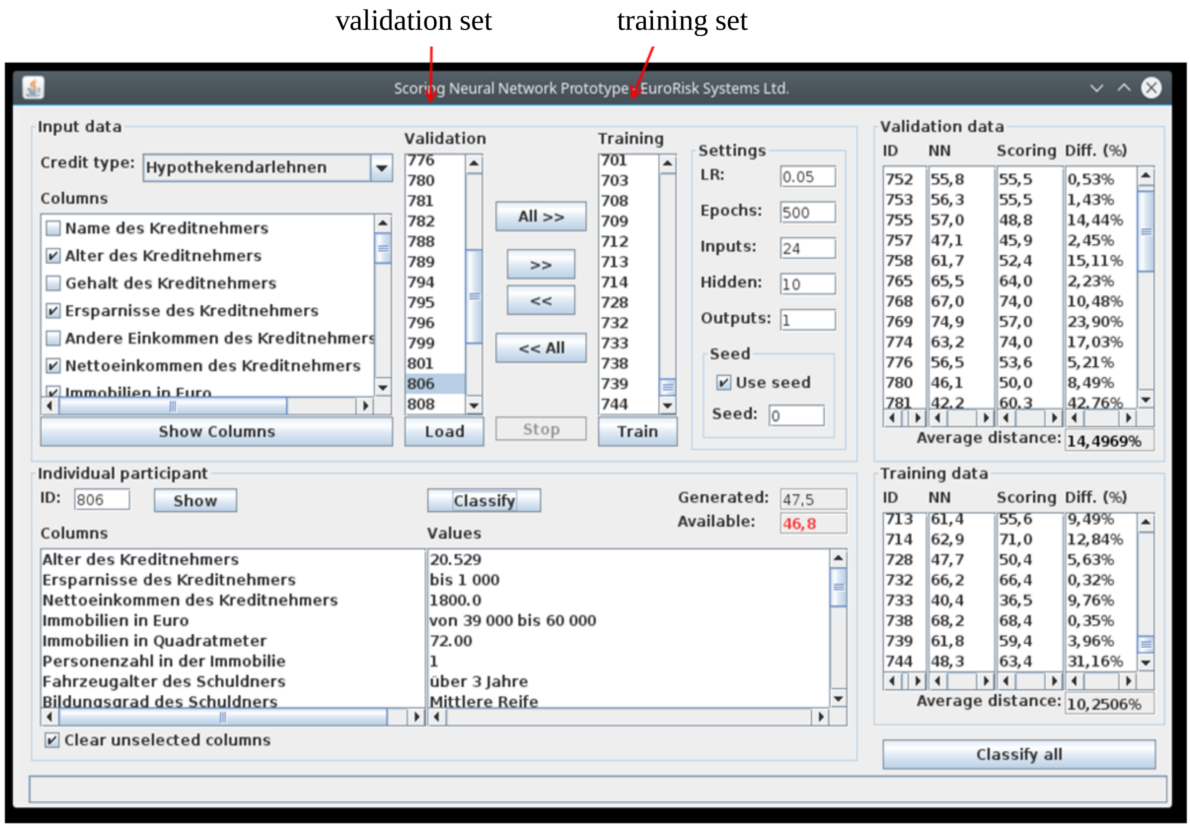

Fig. 6. Prototype of the neural network: training and scoring of new mortgage loans

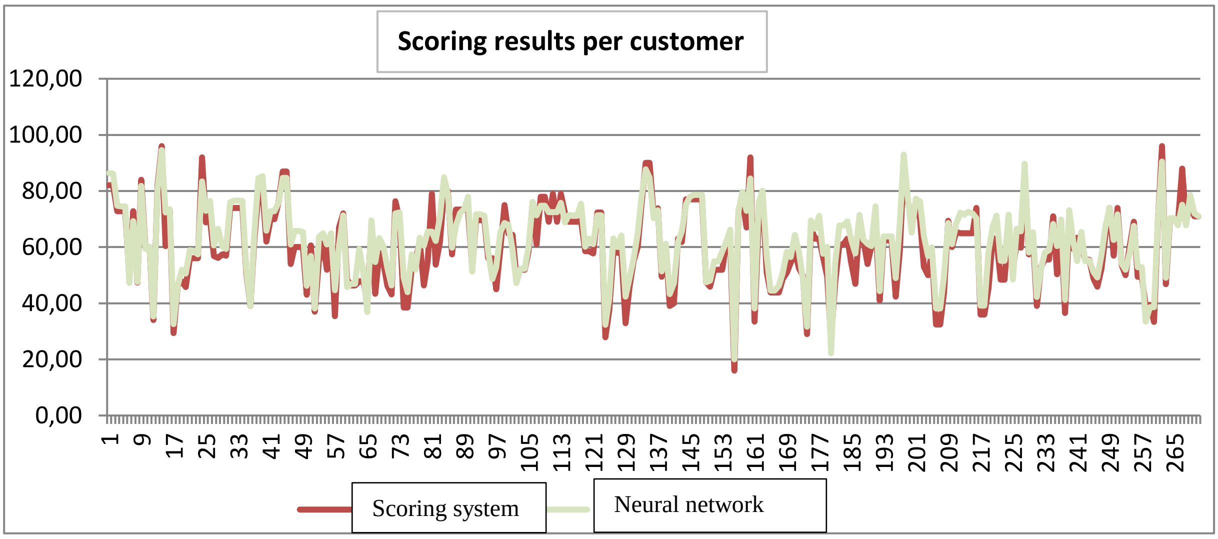

Fig.7. Scoring system versus neural network

Fig. 7 represents a graphical comparison of the scoring values from the scoring model and scoring results from the trained neural network for the mortgage loan group. There is a certain “sobriety and balance” of the network that does not react to minor changes within the scoring model. This behavior is expected, as the scoring of the network is based on integrated “knowledge” gained during the training phase.

These relative changes are illustrated in Fig. 8, where such behavior can be observed as well.

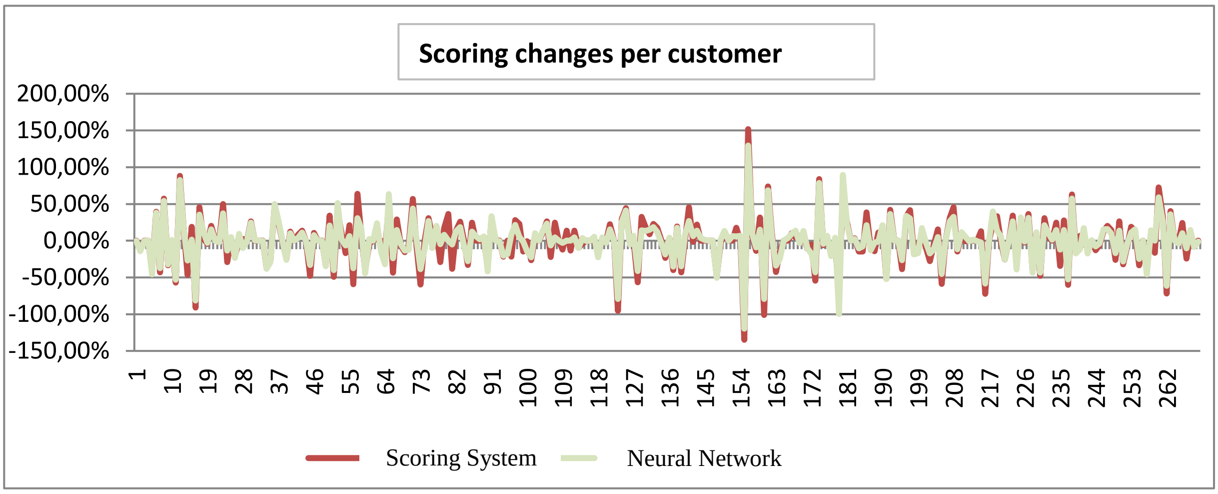

Fig. 8. Scoring system versus neural network

Fig. 9 displays the comparison of scoring results of the first 29 mortgage loan borrowers (a total of 271 borrowers). The statistics show a high positive correlation of 0.9431 between the behavior of the results from the scoring model and those from the neural network. The standard deviation of the difference between both scoring results is 5.27%, which indicates a good conformity and stability of the scoring model. These statistical results ensure stability of the scoring model and in such a manner it is validated.

Remarks:

- The validation of scoring and rating models can also be carried out by means of historical statistics for default losses. This can be achieved assuming the data has been stored, synchronized and accumulated from a larger set of historical data. If this is the case, rating systems can be validated and calibrated using historical data.
- In the methodology mentioned above, the validation is auto-regressive, meaning that a set of scoring results from the scoring models themselves determines the “scoring knowledge”.
- The most significant factors can be identified by excluding individual factors and evaluating the CAP change (CAP = Cumulative Accuracy Profile), that is comparable to the significance of the excluded factor.

Training			Standard Deviation		5,27%
Mean	62,67	60,26	Correlation		0,9431
	Scoring Results			Logarithmic Modifications	
Customer Number	Neural Network	Scoring System	Difference in %	Neural Network	Scoring System
1	86,40	82,00	5,13%	-0,23%	0,00%
2	86,20	82,00	4,85%	-14,32%	-11,90%
3	74,70	72,80	2,52%	-0,27%	0,00%
4	74,50	72,80	2,26%	0,00%	0,00%
5	74,50	72,80	2,26%	-45,43%	-39,59%
12	47,30	49,00	3,64%	38,19%	39,59%
16	69,30	72,80	5,04%	-37,56%	-42,91%
21	47,60	47,40	0,40%	54,14%	57,22%
23	81,80	84,00	2,68%	-32,00%	-33,65%
26	59,40	60,00	1,01%	1,17%	0,00%
27	60,10	60,00	0,25%	-53,21%	-56,80%
32	35,30	34,00	3,61%	82,56%	88,04%
33	80,60	82,00	1,74%	16,02%	15,76%
39	94,60	96,00	1,52%	-26,75%	-46,34%
41	72,40	60,40	16,57%	1,51%	18,95%
46	73,50	73,00	0,73%	-81,60%	-90,95%
48	32,50	29,40	9,64%	35,39%	45,63%
50	46,30	46,40	0,26%	11,61%	2,97%
53	52,00	47,80	8,15%	-2,73%	-4,27%
54	50,60	45,80	9,40%	15,19%	20,11%
57	58,90	56,00	4,86%	-0,34%	0,00%
58	58,70	56,00	4,61%	-2,24%	0,00%
59	57,40	56,00	2,42%	37,36%	49,64%
62	83,40	92,00	10,31%	-13,46%	-29,06%
64	72,90	68,80	5,56%	4,82%	1,73%
68	76,50	70,00	8,54%	-23,30%	-20,72%
73	60,60	56,90	6,14%	9,29%	-1,24%
76	66,50	56,20	15,53%	-9,79%	2,11%
99	60,30	57,40	4,81%	-1,50%	-0,70%
-------	-------	-------	-------	-------	-------
271	Mortgage-	Loans	-------	-------	-------

Fig. 9. Validation of mortgage loans: scoring system versus neural network

The next evaluation is based on a CAP representation (Fig. 10), which compares the scoring model to a perfect model. The ratio is calculated of the area under the curve of the scoring model to the area under the curve of the perfect model, and in this way the Gini coefficient = ap / (ap + ar) is determined. Gini varies between 0% and 100%, whereby higher quality scoring models have a high Gini coefficient.

The decision value for the scoring model and the neural network was set to 48%. Borrowers with a scoring result smaller than 48% are classified into the “default” category. In our example, 36 borrowers in the neural network are assigned to the “default“ category. The scoring model records 54 such borrowers. In both the scoring model and the neural network, 25 borrowers are classified to the “default” category.

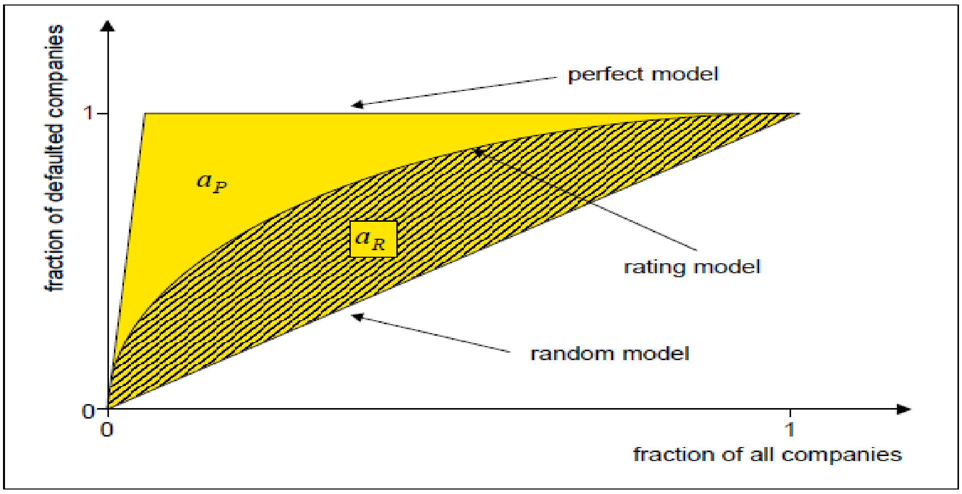

Fig. 10. CAP (Cumulative Accuracy Profile)

Fig. 11 displays the CAP of the scoring data that has been obtained by the scoring model and the neural network. A Gini coefficient of 80% can be observed. It is assumed that the scoring generated by the neural network corresponds to expectations (probabilities) for future defaults.

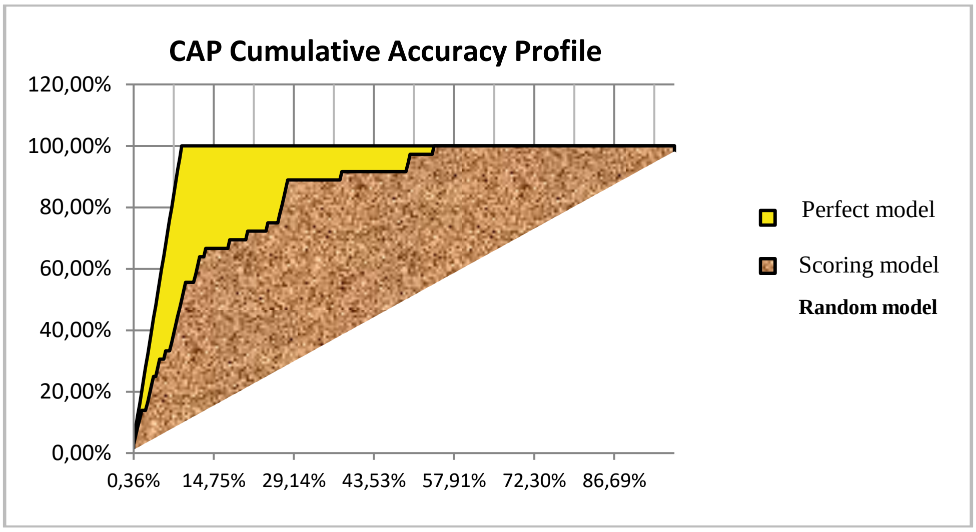

Fig. 11. CAP schema (Cumulative Accuracy Profile) after scoring