# Training a Classifier for Evaluation of Credit Ratings

Kamila Yangirova[1,2], Mikhail Alexandrov[2,3], Oleksiy Koshulko[4]

[1]*Moscow Institute of Physics and Technology, Moscow, Russia*
[2]*Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia*
[3]*University of Barcelona, Barcelona, Spain*
[4]*Glushkov Institute of Cybernetics, NASU, Kyiv, Ukraine*

kamila68@mail.ru, malexandrov@mail.ru, koshulko@gmail.com

**Abstract.** *Credit rating is one of the most important indicators reflecting risk of investments. The problem of rating calculation is the problem of classification on* 12 *the well-known categories. In the paper we show how to construct such a classifier using the technique of GMDH. The source information for our experiments are: (a)* 80 *large companies of corporate sector with their ratings given by* 3 *the most famous rating agencies (S&P, Fitch, Moody's), and (b)* 23 *parameters describing business and financial activities of these companies taken from their financial reports. The results proved to be very promising: with neuro-like algorithm from the package GMDH Shell we reached Accuracy = 100%, F-measure=100% for 2 classes, and Accuracy=88%, F-measure=83% for 4 classes. These classes cover the best* 6 *categories. Such results mean that the proposed technology can be used for self-evaluation of companies before their application to rating agencies.*

## Keywords

Risk of investments, rating, classification, GMDH

## 1 Problem setting

Investments need preliminary evaluation of company creditworthiness. There are several respectable rating agencies, which give such assessments. These assessments are distributed on the 12 categories presented in the Table 1. In our research we considered only the best categories (1-6) because of the absence of necessary experimental material. These categories were distributed on 2 and 4 classes presented in the Table 1 as well.

**Tab. 1**. Categories of creditworthiness

| *Rating* | *Level of creditworthiness* | *Classes* | *Classes* |
|---|---|---|---|
| AAA | The best one | 1 | 1 |
| AA | Very high | 1 | 1 |
| A | High | 1 | 2 |
| BBB | Good | 2 | 3 |
| BB | Speculative | 2 | 4 |
| B | Very speculative | 2 | 4 |
| CCC, CC, C | Essential Risk | - | - |
| DDD, DD, D | Default | - | - |

Our goal is to construct a classifier of credit ratings using the GMDH technique [1]. For this we use the package GMDH Shell [2] and test its algorithms. The paper reflects the main results of master thesis [3].

# 2 Experimental material

Experimental material is: (a) 80 large companies of corporate sector with their ratings given by 3 the most famous rating agencies (S&P, Fitch, Moody's), and (b) 23 parameters describing business and financial activities of these companies. Table 2 shows the part of the data and Table 3 gives the set of parameters the agencies use to evaluate the creditworthiness of companies.

**Tab. 2**. Companies and their ratings

| *Company* | *S&P* | *Fitch* | *Moody's* | *Compose* |
|---|---|---|---|---|
| Microsoft | AAA | AAA | Aaa | AAA |
| Procter&Gamble | AA | AA | Aa3 | AA |
| Google | AA | AA | Aa2 | AA |
| Lukoil | BBB | BBB | Ba1 | BBB |
| Megafone | BB | BB | Ba1 | BB |
| ..... | ..... | ..... | ..... | ..... |

The concordance of agencies assessments are equal 91.7% for 2 classes and 82.5% for 4 classes. These values can be considered as the desired level of *Accuracy* or *F-measure*, when we evaluate the quality of model to be constructed.

**Tab. 3**. Parameters of business and financial activities

| *Parameter* | *Notation* | *Description* |
|---|---|---|
| $P_1$ | LR | Absolute liquidity |
| $P_2$ | CR | Current liquidity |
| ..... | ..... | ..... |
| $P_{23}$ | DY | Norm of dividend income |

The parameters from the Table 3 are taken from financial documents and analytical reviews related to risk of investments. These 23 parameters reflect 5 integral characteristics of company activity: liquidity, financial stability, profitability, business activity and market value. But nobody knows concrete parameters and their combination that are used in a given agency, which is know-how.

# 3 Modeling (constructing classifier)

For modeling we use tool GMDH Shell (GS) in the mode of classification. In this mode GS suggests the following 5 algorithms: the combinatorial, neuro-similar, step-by-step with addition, mixed step-by-step and also the random forest. GS itself is able to try all algorithms and to suggest the best one with typical values of algorithm parameters.

## 3.1 Modeling with 2 classes

GS selected the combinatorial algorithm, which provided the following indicators of the classifier's quality: *Accuracy*=93.8% and *F-measure*=93.7%. It is easy to see that these values exceed the desirable level of precision for 2 classes (~91.7%). The model is presented in the following form:

$$Y = -\,0.533 + (P_4)*0.627 + (P_9^{1/3})*0.143 + (P_{19}^{1/3})*0.178$$

where: $Y$ is class, $P_4$ is carrying circulating value, $P_9$ is coefficient of percent fill rate, $P_{19}$ is market capitalization.

Then we repeated the experiment with neuro-similar algorithm and manual tuning. Here we could reach the best values: *Accuracy* = **100%** and *F*-measure =**100%**.

### 3.2 Modeling with 4 classes

GS selected the neuro-like algorithm, which provided *Accuracy*=75.0% and *F-measure*=74.7%. These values couldn't reach the desired level of precision for 4 classes (~82.5%). The model is presented in the following form:

$$Y = -\,0.008 + P_{20}*N_2*0.009 + N_2*0.708,$$

$$N2 = -\,0.008 - P_{17}*0.0004 + P_{17}*N_3*0.0180 + N_3*0.678,$$

$$N3 = 0.243 - P_7*0.109 + P_7*N_4*0.240 + N_4*0.485,$$

$$N4 = -\,0.051 + P_{18}*0.021 + N_5*0.971,$$

$$N5 = 0.346 - P_2*0.239 + P_2*N_6*0.630,$$

$$N6 = -\,0.085 + P_9*0.003 + P_9*N_7*0.004 + N_7*0.963,$$

$$N7 = 0.121 - P_4*0.410 + P_4*P_{21}*0.085,$$

where: $Y$ is класс, $N_i$ are neurons (partial models of *i*-th layer), $P_2$ is current liquidity, $P_4$ is carrying circulating value, $P_7$ is coefficient of financial dependence, $P_9$ is coefficient of percent fill rate, $P_{17}$ is book cost of capital per share, $P_{18}$ is ratio of market capitalization to book cost of capital, $P_{21}$ is ratio of market capitalization to clear profit.

We repeated this experiment with the same algorithm but using careful manual tuning. In this case we could essentially improve the result: *Accuracy*=87.5% and *F-measure*=82.5%. Now these values exceed and are equal the desired precision (~82.5%).

## 4 Conclusions

The main results are:

- In the research we used accessible data in financial reports of large private companies. It proves they contain enough information to be used in models of creditworthiness.
- GMDH allows constructing classifiers which give results being close to opinion of the leading rating agencies.

The planned future works may be:

- We completed our research using only the best ratings of the categories A and B. In future we suppose to repeat this research with data reflecting all ratings i.e. A, B, C, D.
- We intend to test noise immunity of constructed models that is to test sensibility of models to variation of source data.

## References

[1] Stepashko, V.: *Ideas of Academician O. Ivakhnenko in the Inductive Modeling field from historical perspective*, In: Proc. of 4th Intern. Conf. on Induct. Modeling (ICIM-2013), Sept. 2013, Kyiv, Ukraine, pp. 31-37.

[2] GMDH Shell platform: http:// www.gmdhshell.com.

[3] Yangirova, K: *Constructing multicriteria classification on the example of credit ratings*, Master thesis, RPANEPA, 2015. (In Russian)