

Neural Networks Application to Classification of Credit Institutions

Akishina E.P.¹, Ivanov V.V.^{1,2}, Prikazchikova A.S.²

¹ Joint Institute for Nuclear Research, JINR, Dubna

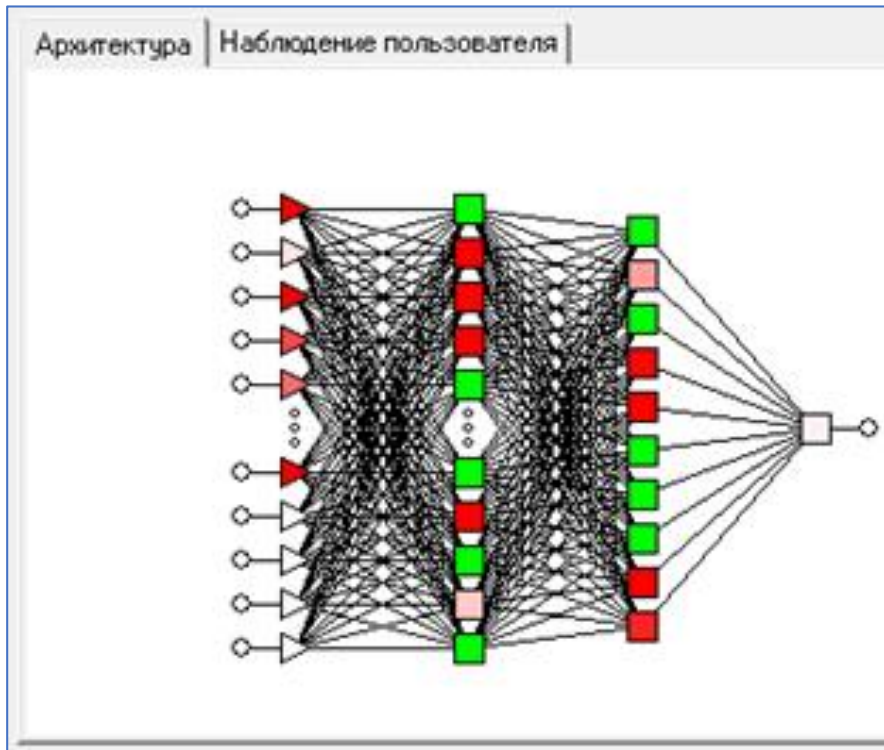
² National Research Nuclear University, MEPHI, Moscow

Neural Networks in the task of classification

As a part of Deep Learning algorithms Neural Networks helps to classify the data. In the classification problem the outcomes are labels that can be applied to the data, for example, in an email to filter spam or notspam, or, in high energy physics data analysis to separate interesting data (signal) from unwanted noise (background) etc.

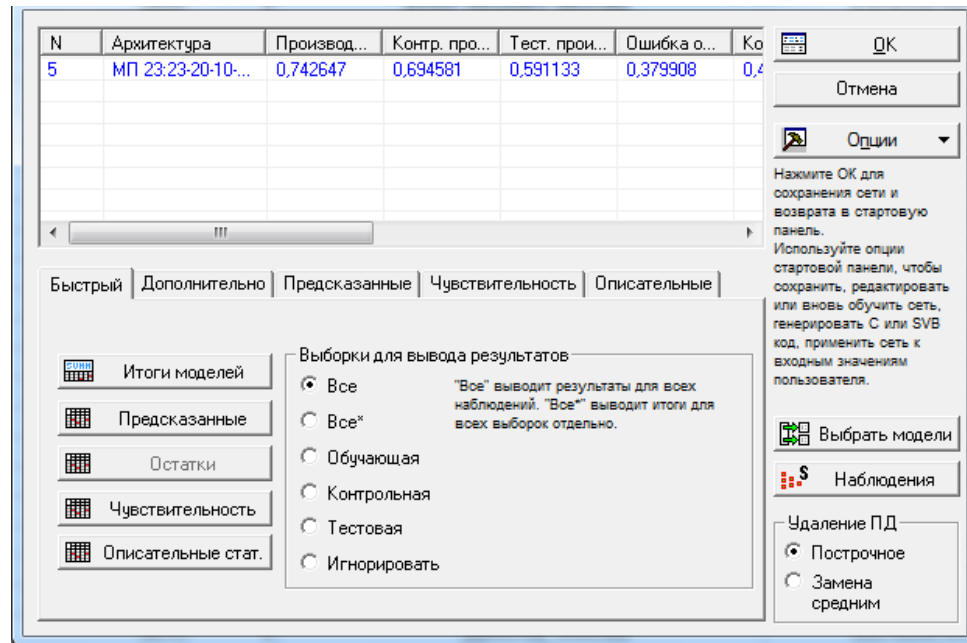
The task presented here is the classification of credit institution for reliable and unreliable (with a high degree of financial risk). To solve it, it seems appropriate to develop a data analysis methodology that involves working with numerous features and characteristic of the system. To conduct this study, the theory of neural networks and the method of principal components of factor analysis were used as mathematical tools.

Architecture of Neural Network



- input: activities data of credit institutions from standard banking statements №101
- training sample with the target indicator "Revocation of license" (two values - Bad/Good). 23 most informative indicators selected
- neural network model - multilayer perceptron
- number of neurons of the input layer is equal to the number of features

Best number of neurons



N	Архитектура	Производ...	Контр. про...	Тест. прои...	Ошибка о...	Ко
5	МП 23:23-20-10...	0.742647	0.694581	0.591133	0.379908	0,4

Быстрый | Дополнительно | Предсказанные | Чувствительность | Описательные

Итоги моделей
Предсказанные
Остатки
Чувствительность
Описательные стат.

Выборки для вывода результатов

- Все
- Все*
- Обучающая
- Контрольная
- Тестовая
- Игнорировать

Нажмите ОК для сохранения сети и возврата в стартовую панель.

Используйте опции стартовой панели, чтобы сохранить, редактировать или вновь обучить сеть, генерировать С или SVB код, применить сеть к входным значениям пользователя.

Выбрать модели
Наблюдения

Удаление ПД

- Построенное
- Замена средним

- Statistica soft as a tool for applying neural network to the sample data
- different number of neurons in hidden layers tested
- best results (highest value of performance) - the model architecture 27-20-10-1

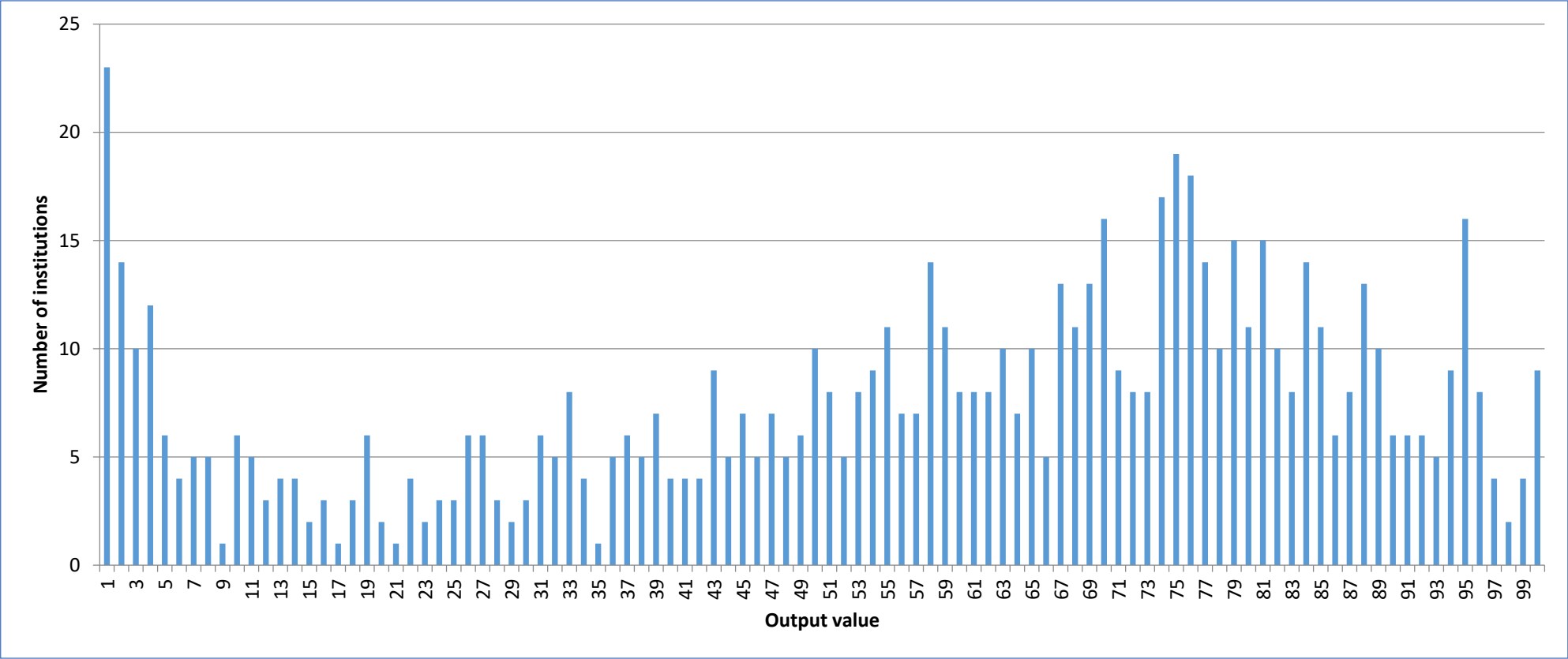
(27 – number of input neurons,
20 – number of neurons at 1st hidden layer,
10 – number of neurons at 2nd hidden layer,
1 – output neuron)

The train-test split evaluation

Net №	Training part, %	Test part, %	Performance, %
1	60	40	72,8
2	70	30	74,3
3	80	20	71,7
4	50	50	74,4

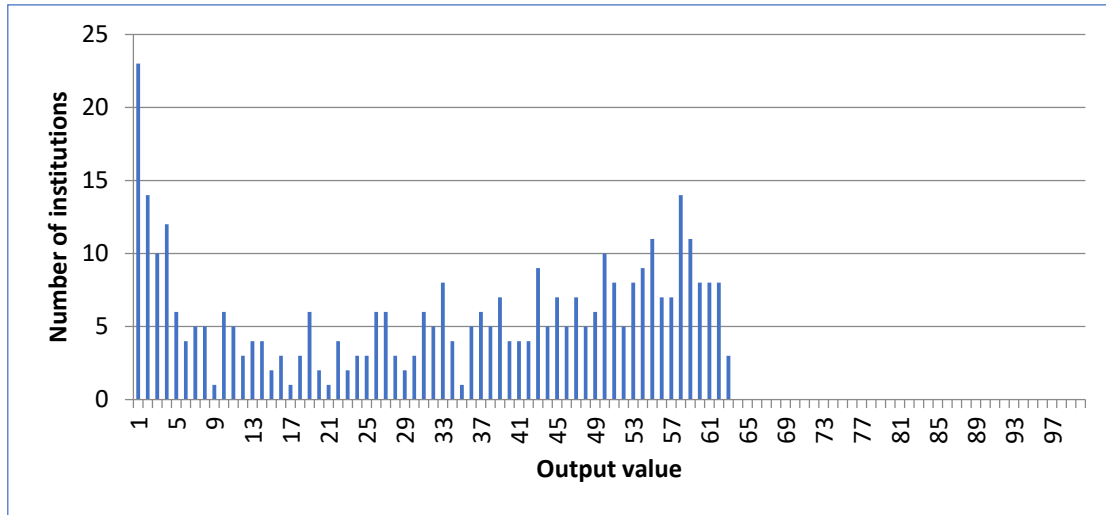
- dataset is divided for 2 subsets: training and test
- best results for our model – split 50% training part, 50 % test part

Neural network output distribution

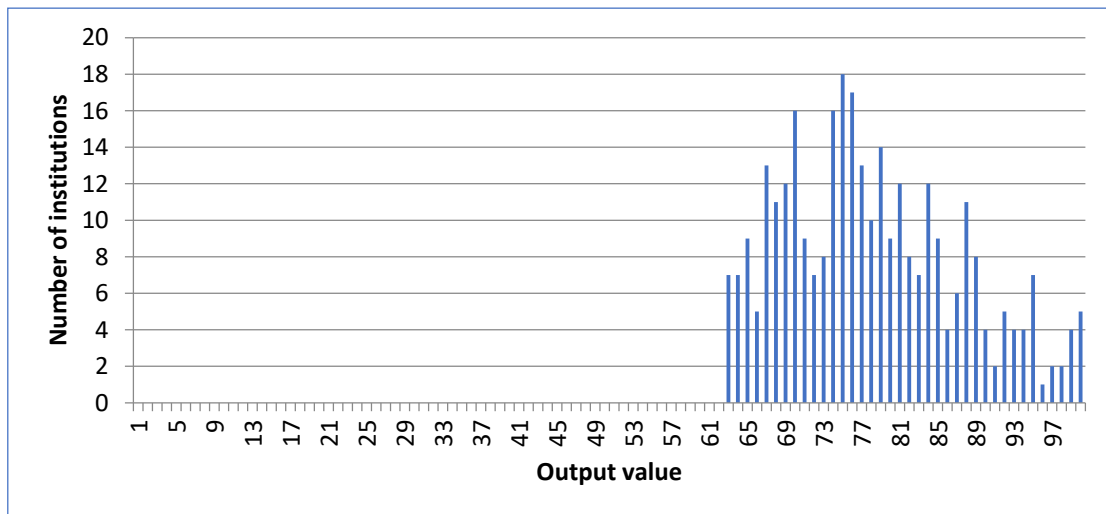


The distribution for both class of institutions: reliable and unreliable

Neural network output distribution

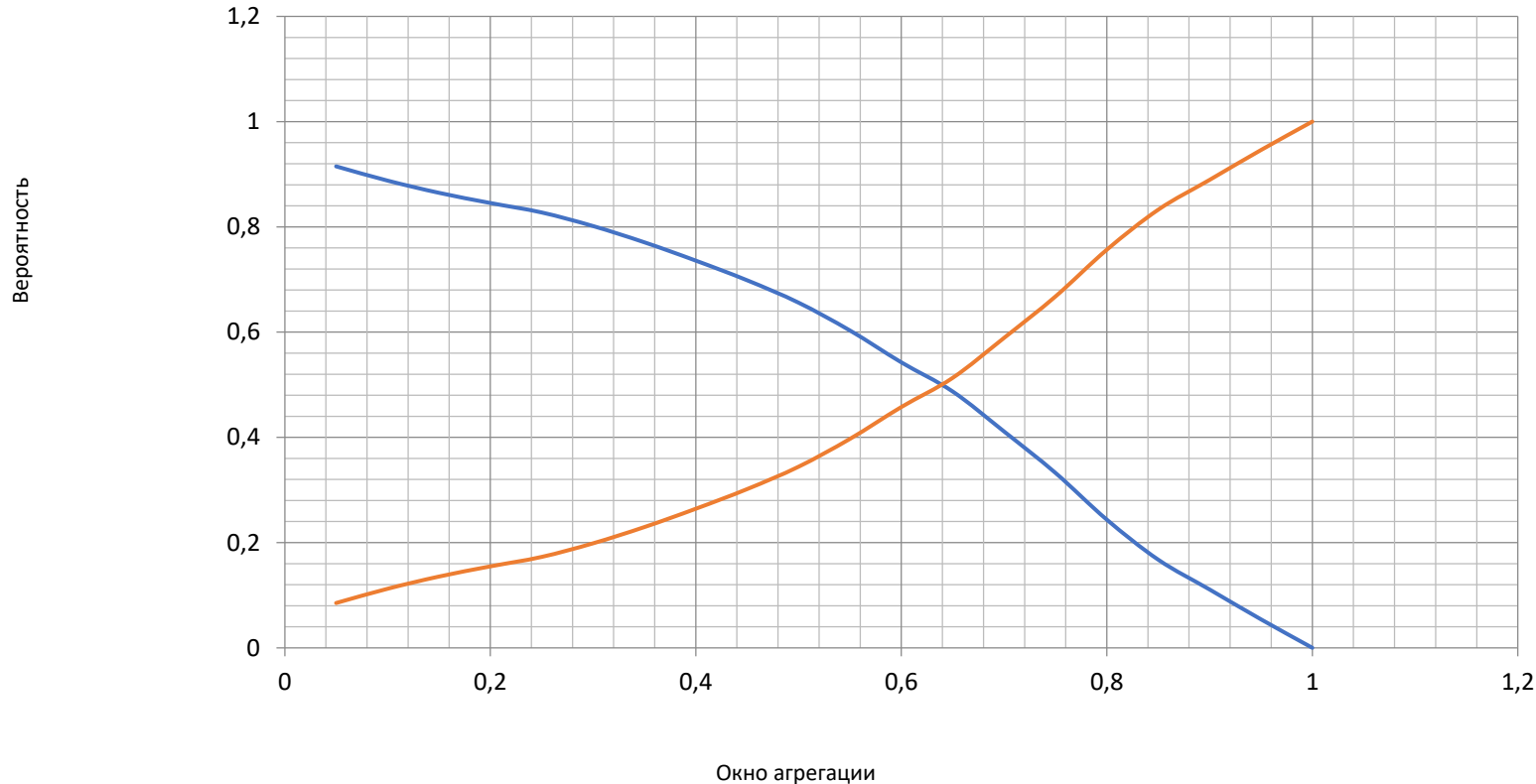


The distribution for class of institutions:
reliable



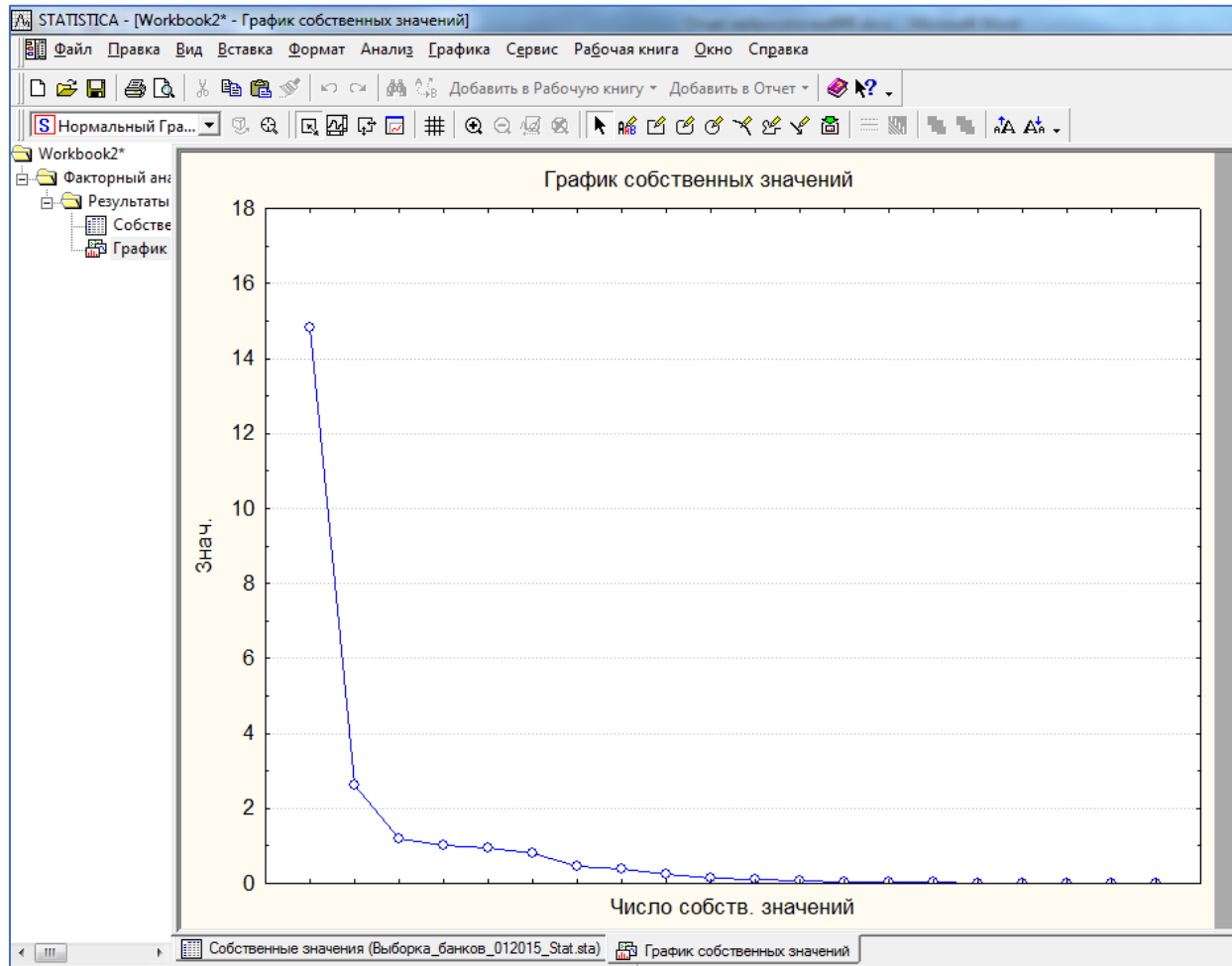
The distribution for class of institutions:
unreliable

Distribution functions behavior



Intersection of the distribution function $f(x)$ for Unreliable institutions (red curve) and the function $1-f(x)$ for Reliable institutions (blue curve) illustrates the separation of the institutions for both types.

Principle Component Analysis (PCA)



- The method was used to try to reduce the features data space to a lower one
- from 23 indicators to 23 new principal components, with the total dispersion of 98%
- major contribution to the system make 12 principal components

Scree plot of eigenvalues of the principal components

. Neural network performance values

Neural network INPUT	Performance, %
23 predictors of initial data	74
23 principal components	72
12 principal components	70

Values of performance for different inputs performance after applying the principal component method and reducing the data space became 70%, which indicates the possibility of reduction without loss of network performance.

Conclusions

- the usage of artificial neural networks for the task of classifying the massive of data and identifying specific data for credit institution is expedient and justified. The performance of the neural network is 74%;
- application of the method of principal components of data analysis allowed to reduce the initial data space from 23 indicators to 12 principal components;
- the performance of the neural network after applying the principal component method and reducing the data space became 70%, which indicates the possibility of reduction without loss of network performance.