The Research on Stability of the Russian Banking System by Machine Learning Methods

Oleg A. Bayuk[*], Dmitry V. Berzin and Bogdan A. Timov

Financial University, Moscow, Russia

Abstract: The problem of stability of the Russian banking system is investigated. To describe the state of a commercial bank, we use a system of indicators, proposed by F.T. Aleskerov and his colleagues. For predicting the development of banking system, the machine learning system implemented in the Azure ML is used. To optimize the work of this software, it is suggested to use integral indicators.

Keywords: Commercial bank, banking system, stability of the banking system, machine learning, decision tree.

INTRODUCTION

The current stage of the Russian banking system is characterized by some stabilization and moderate development after several experienced systemic crises. Commercial banks perform a variety of functions and enter into complex relationships between themselves and other economic entities, carrying out credit, settlement, deposit and other transactions. At the same time, banking activities are subject to numerous risks, underestimation of which can lead to malfunctions and bankruptcy of credit institutions, causing damage to their customers and shareholders.

At this stage, banks are more balanced in evaluating all risks, including the risk of active interbank operations, which include interbank lending, opening and maintaining transactions via NOSTRO accounts ("nostro conto") with other banks, opening deposit accounts with other banks, securities transactions with other banks, etc. On the other hand, bank customers, both legal entities and individuals, have now become more responsible and considerate of the servicing bank. All these factors determine the relevance of development and improvement of methods for analyzing the effectiveness of the bank financial status.

Analysis of the effectiveness of a bank in modern conditions is the basis for making managerial decisions in the bank and establishing trustful and mutually beneficial relations between banks and their clients.

Nowdays, when the economic situation in the Russian Federation has changed (especially the conditions for the commercial banks functioning), the achievement of their goals becomes possible mainly due to changes in the stability of banks. Moreover,

since financial activity is a specialization of banks, it is difficult to overestimate the role of financial analysis in financial sustainability. Commercial banks are the most important link in the market economy. In the process of their activities, most of the money turnover in the state is mediated; the sources of capital for expanded reproduction are formed by redistributing temporarily free cash resources of all participants in the reproduction process – the state, business entities, and the population. At the same time, commercial banks facilitate the transfer of capital from the least efficient sectors and enterprises of the national economy to the most competitive ones.

The relevance of the topic is explained by the fact that commercial banks, mobilizing temporarily free funds in the market of credit resources, with their help meet the need of the national economy for working capital, facilitate the transformation of money into capital, and provide the needs of population in consumer credits. From their clear and competent activity depends both the effectiveness of the functioning of the banking system, and the Russian economy in general. Therefore, the development of an effective mechanism for analyzing their activities, aimed at identifying, at the earliest possible stages of problems in the business of a commercial bank, is vital for the financial and social stability of Russia.

However, the theoretical issues of financial analysis in banks remain until now insufficiently developed, the place and role of financial analysis in the management of a commercial bank has not been clearly identified; therefore, the selected topic is relevant from the point of view of its practical application in the banks activity.

In this paper, we propose to use the methods of machine learning (Bernes, 2015; Brink, Richards, Fetherolf, 2016) to evaluate the stability of banking system (Fetisov, 2011).

*Address correspondence to this author at the Financial University, Moscow, Russia; Tel: +79166078343; Fax: +74992772321; E-mail: OABayuk@fa.ru

In recent years, interesting foreign publications on the subject have appeared. Among them, we should mention (FSB, 2017), which concludes that access to financial services and technological training are enhanced in assessing bank stability. Note also (Addo, Guegan, Hassani, 2018), in which machine learning is used in the task of credit scoring. Special attention should be paid to (Petropoulos, Siakoulis, Stavroulakis, Vlachogiannakis, 2017). In this work, a study of the international banking system is done by machine learning techniques using random forest modeling.

To evaluate the state of commercial banks and ways to predict deterioration of their condition, we use indicators published in open information systems "banki.ru" and "Interfax". In addition, a number of Russian banks were divided into "large", "medium" and "small" classes, as well as classes according to the degree of adherence of banks to business models (Aleskerov, Belousova, Serdyuk, 2008; Alexashin *et al*., 2012). We also investigated the influence of the banks belonging to one of these classes on the degree of their stability.

The subject of the proposed research are the methods of evaluation of commercial banks and the prediction of their future state.

1. DECISION TREES

The decision tree is a data structure, described as follows. In the process of bypassing, in each node, depending on the checking condition, a certain decision is made – moving along one or another branch of the tree from the root to the "leaf" (terminal) vertices. There is the desired value of the attribute of interest in the "leaf" top of the tree. Decision trees can evaluate the values of categorical attributes (a finite number of discrete values), as well as quantitative ones. In the first case, we talk about the classification problem - assigning an object to one of the "classes" defined by an attribute (for example, Yes / No, Good / Satisfactory / Bad, etc.). In the second case, we talk about the problem of regression, that is, about the estimation of a quantitative value.

We consider an algorithm that allows us to construct such a decision tree for estimating and predicting values of a categorical attribute from the analyzed data set based on the values of other attributes (classification problem).

Generally, there are infinitely many ways to build a tree – we can consider attributes in different order,

check different conditions in the tree nodes, stop the process using different criteria. Nevertheless, we are only interested in trees that most accurately estimate the value of the attribute, with a minimum error, and allow us to identify the dependency between attributes and successfully perform predictions of attribute values in new data. Unfortunately, there are no good algorithms, which make it possible to find such an "optimal" tree (within appropriate time). However, there are good enough algorithms that try to build an "almost optimal" tree, performing a certain "local" optimality criterion at each iteration in the hope that the resulting tree will also be "optimal" overall. Such algorithms are called "greedy". We will consider this algorithm.

2. ALGORITHM OF CONSTRUCTING A DECISION TREE

The principle of constructing a tree is as follows. The tree is built "from top to bottom" from the root. The process begins with determining which attribute should be selected for testing in the root of the tree. To do this, each attribute is examined to see how well it classifies the data set, i.e. divides into classes by the target (objective) attribute. When the attribute is selected, a tree branch is created for each of its values, the data set is divided according to the value to each branch, the process is recursively repeated for each branch. Also, it is needed to check the stopping criteria.

The main question is how to choose attributes. In accordance with the idea of the approach, when the target attribute class is located at the end nodes (leaves) of the tree, it is necessary that when the data set is splitted at each node, the resulting data sets would be the most homogeneous in terms of class values. In this case, it is necessary to determine the quantitative criterion in order to evaluate the homogeneity of the splitting.

3. ENTROPY

Consider the set of probabilities p_i describing the probability of the correspondence of the data in our collection (we denote it *X*) to the class *i*. Let's calculate the following value:

$$Entropy(x) = -\sum_{i=1}^{c} p_i \log g_2 p_i$$

This function is called *entropy*. Entropy arose in information theory and describes the amount of information (in bits) that is necessary to encode the

message about the belonging of a randomly selected object (string) from our set *X* to one of the classes and transmit it to a recipient. If the class is only one, the recipient does not need to transmit anything, the entropy is 0. If all classes are equally likely, then it's required $\log_2 c$ bit (*c* is the total number of classes) is required, the maximum of the entropy function.

Next, to select an attribute, for each attribute *A*, the so-called increment of information is calculated:

$$Gain(X, A) = Entropy(X) - \sum_{a \in values(A)} \frac{|X_a|}{|X|} Entropy(X_a)$$

Here *values (A)* are all the accepted values of the attribute A, X_a is the subset of the data set, where $A = a, |X| -$ the number of elements in the set. This value describes the expected decrease in entropy after splitting the data set according to the selected attribute. The second term is the sum of entropies for each subset, taken with its weight. The overall difference describes how entropy decreases, how much we will save bits to encode the class of a random object from the set *X*, if we know the values of the attribute *A* and split the data set into subsets for this attribute.

The algorithm selects an attribute corresponding to the maximum value of the information increment.

When the attribute is selected, the source set is split into subsets according to its values, the original attribute is excluded from the analysis, and the process is recursively repeated.

The process stops when the created subsets become sufficiently homogeneous (one class prevails), namely, when max *(Gain (X, A))* becomes less than some given parameter *Θ* (a value close to 0). Alternatively, one can control the set *X* itself, and when it has become sufficiently small or completely homogeneous (only one class), stop the process.

4. USING MICROSOFT AZURE MACHINE LEARNING TO PREDICT THAT THE STATE OF A BANK IS GETTING WORSE

Machine learning is the "tough nut to crack" of artificial intelligence. The importance of machine learning is great, because this ability is one of the main components of reasonable behavior. For example, an expert system can perform long and time-consuming calculations to solve problems. However, unlike human beings, if to give it the same or similar problem second time, it will not "remember" the decision. It will each time perform the same calculations again – it hardly looks like a reasonable behavior.

Most expert systems are limited by the inflexibility of their decision strategies and the difficulty of modifying large volumes of code. The obvious solution to these problems is forcing programs to learn from experience, analogies, or examples.

Although machine learning is a difficult area, some programs disprove concerns about its inaccessibility. One such program is AM – Automated Mathematician, designed to discover mathematical laws. Starting from the concepts embedded in it and the axioms of set theory, Automated Mathematician succeeded in deriving from them such important mathematical concepts as cardinality, integer arithmetic, and many results of number theory. The Automated Mathematician constructed theorems, modifying his knowledge base, and used heuristic methods to find the best of the set of possible alternative theorems. Of recent results in the field of machine learning, it is possible to note the program of Cotoia, which invents "interesting" integer sequences.

The early works, which influenced the field of machine learning, are the studies of scientists on the derivation of such structural concepts as the construction of "arches" from the "world of blocks" sets. The ID3 algorithm showed the ability to identify common principles from different examples. The Meta-DENDRAL system derives rules for the interpretation of spectrographic data in organic chemistry using examples of information on substances with a known version. The system Teiresias is an intelligent interface for expert systems that converts messages in high-level language into new rules of their knowledge base. The Hacker program in the field of machine learning studies noted that it can build plans for manipulating the "block world" through an iterative process of developing a plan, testing it, and correcting the deficiencies identified.

Speaking about machine learning, we cannot forget about neural networks. After all, thanks to neural networks many different algorithms of machine learning have been developed, each of which has its strong and weak sides.

Note that today many important biological and sociological models of machine learning are known. The success of machine learning programs suggests the existence of universal principles, discovering of

which would allow designing programs that can be trained in real problem areas.

Microsoft Azure Machine Learning

Microsoft Azure Machine Learning (Azure ML) is a cloud-based solution that allows the construction and use of complex machine learning models in a simple and intuitive way.

Why Azure ML? Because Azure Machine Learning is one of the simplest tools for using machine learning, which removes the entry barrier for everyone who decides to use it for their own needs.

The logical process of constructing the algorithm of machine learning can be viewed in the Figure **1**.
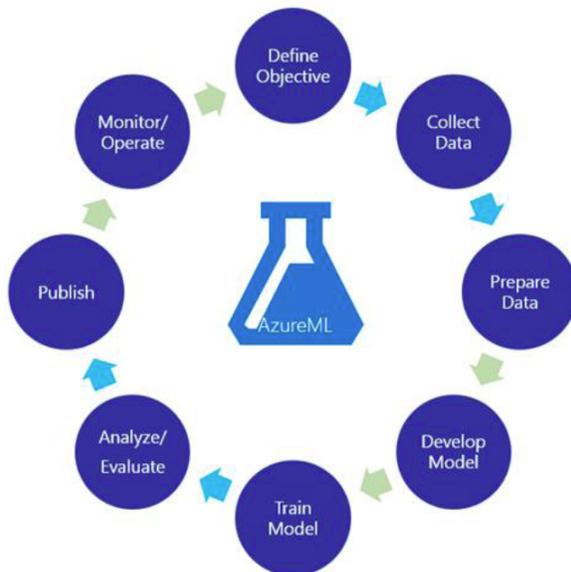
Figure 1: Azure ML flowchart.

Definition of Purpose

All algorithms of machine learning are useless without the clearly defined purpose of the experiment. In this task, the goal is to predict the possibility of revoking a bank license based on a set of characteristics provided by the end user.

Data Collection

During this phase, a sample of data is formed, which is necessary for further training the model. In this case, the data is taken from the Internet site banki.ru

Data Preparation

At this stage, the data are prepared by forming characteristics, removing emissions and splitting the sample into training and test data.

Development of the Model

During the development of the model, one or more data models and corresponding training algorithms are selected, which, in the opinion of the developer, will have to produce the desired result. Often, this process is combined with a parallel study of the effectiveness of several models and a visual analysis of the data in order to find any patterns.

Train the Model

During the training process, the algorithm searches for hidden patterns in the data sample in order to find the prediction method. The model chosen and the training algorithm determine the search process.

Evaluation of the Model

After the model is trained, it is necessary to study its prognostic characteristics. Most often for this, we run it on a test sample and evaluate the resulting error level. Depending on this evaluation and requirements for accuracy, the model can be either adopted as a final or we repeat training after adding new input characteristics or even change the training algorithm.

Using the Model

In case of successful testing of the trained model, the stage of its use begins. And this is the case when Azure ML becomes irreplaceable, giving all the necessary tools for the publication, monitoring and monetization of algorithms.

Carrying Out an Experiment in Azure ML

For an experiment, we create a MS Excel file with a set of bank indicators, such as net assets, net income, equity, loan portfolio, overdue loans, personal deposits, investments in securities, and a change in all these indicators per year, the quality of the loan portfolio, adherence of banks to business models and the size of the bank. All these indicators were chosen experimentally, so that the results of training the model were the best.

Further, after the file downloads in the experiment, select the columns used in the dataset. To do this, you need the module Select Column in Dataset (see Figure **2**). In its settings, you just drag the column headers from one part of the window to the other.

All available data must be split into two parts: the training and control sample. One can do it manually by splitting the source file into your own discretion, or by
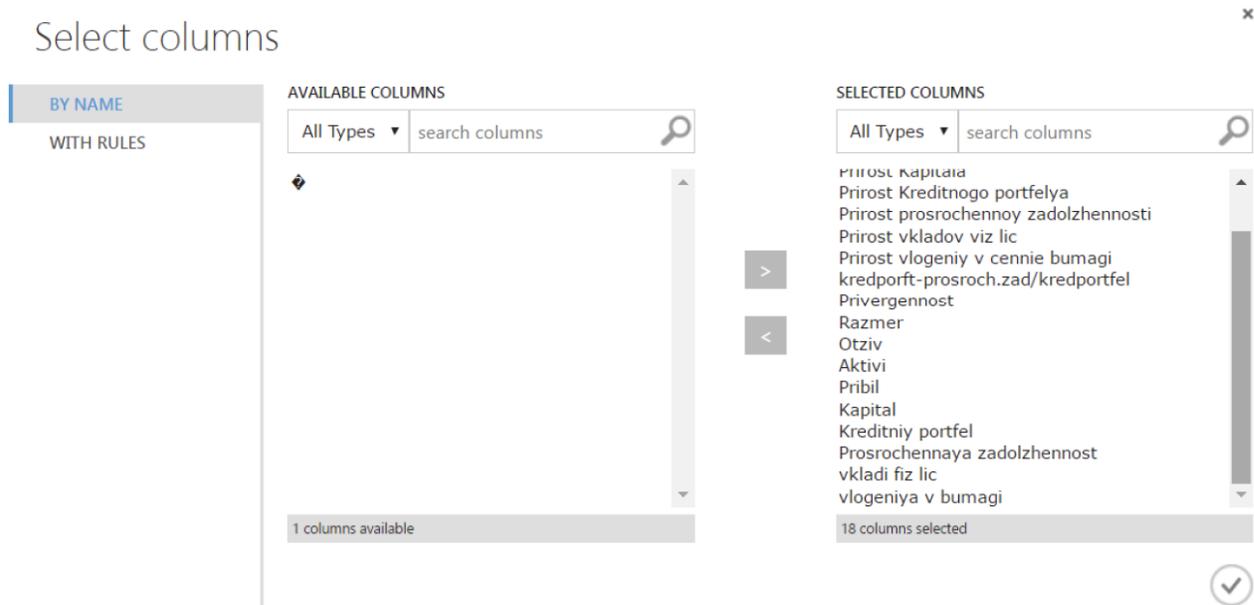
Figure 2: Select appropriate indicators.

using the Split Data module, which allows you to split the data randomly, with the specified parameters (see Figure **3**).
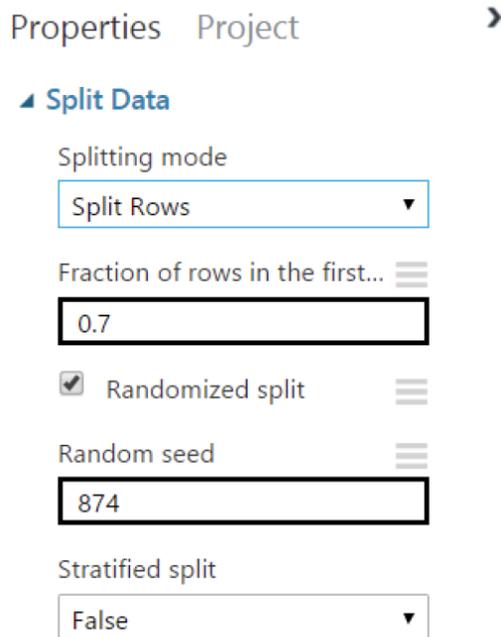
Figure 3: Split the data.

The *Split Rows* value in the *Split* mode parameter means a simple division of the data into two parts. The next line indicates the proportion of data for the "training", respectively 70% of all data will be used to train the model, and 30% to test it. In order for the system to select random lines during the separation, you must enable the Randomized Split option.

As a model, a two-class *Decision Forest* was used. It was chosen experimentally, because with it the most successful outcomes of training the models were obtained. In addition, models such as a two-class neural network (a two-class neural network), a two-class incremental decision tree (Two-Class Boosted Decision Tree) and several others were tested.

In the "Number of decision tree" parameter, we must input as much number of decision trees that can be created as possible. The more decision trees, the more potentially you can get a full coverage of the data, but the training time will increase. For the "Maximum depth of the decision trees" option, one needs to enter a number that will limit the maximum depth of any decision tree. Increasing the depth of the tree can improve accuracy, but with a risk of increased training time. In the "Number of random splits per node", we enter the number of partitions that are used building each node of the tree. Partitioning means that components at each level of the tree (node) are randomly distributed. The last parameter "Minimum number of samples per leaf node" shows the minimum number of cases that are required to create any final node (leaf) in the tree.

Further, after selecting the necessary parameters for the model, the column selected in the *Train Model* module is the result of the prediction, that is, the "revoke" column. It takes the value 0 if the bank license is revoked, and 1 if the bank is OK.

Finally, the constructed experiment looks like in the Figure **4**:
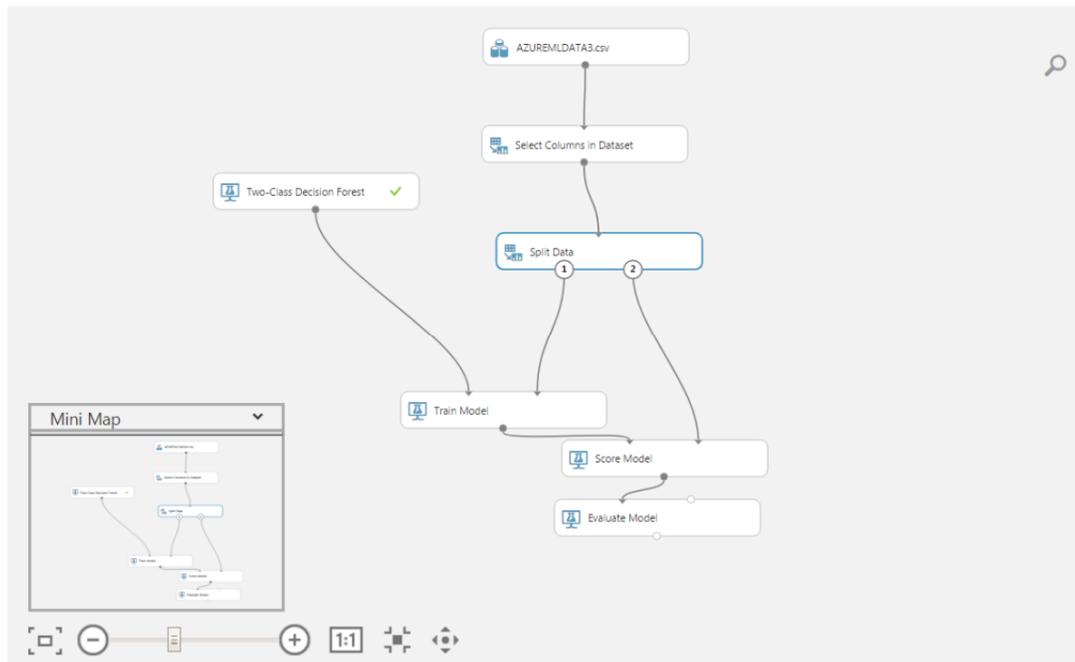
Figure 4: Constructed experiment.

Figure 5: Results of evaluation.

The *Score Model* module is used to create predictions using a trained classification or regression model, and the *Evaluate Model* is used to measure the accuracy of the trained model.

The results for evaluation of the predict model are shown in the Figure **5**:

True Positive means the number of correctly recognized "good" banks;

False Negative is the number of errors of the first kind, that is incorrectly recognized "bad" banks. This means that the bank, which is OK, was identified as unreliable;

False Positive is the number of errors of the second kind, that is, incorrectly recognized "good" banks;

True Negative is the number of correctly recognized "bad" banks.

Accuracy is the proportion of correct answers;

Precision is a measure of accuracy that characterizes how many positive responses resulting from the use of the classification model are correct;

Recall is a measure of completeness that characterizes the ability of the classifier to "guess" as many positive answers as possible from the expected ones (false positive answers do not affect this metric);

F1 Score is the mean harmonic of accuracy and completeness;

ROC (Receiver Operating Characteristic) curve shows the change in the ratio of the proportion of correctly classified "good" clients in their total number to the proportion of "bad" clients incorrectly referred to as "good" when the threshold of the decision rule is varied;

AUC (Area Under ROC Curve) is the area under the ROC curve; the higher the AUC, the better the classifier.
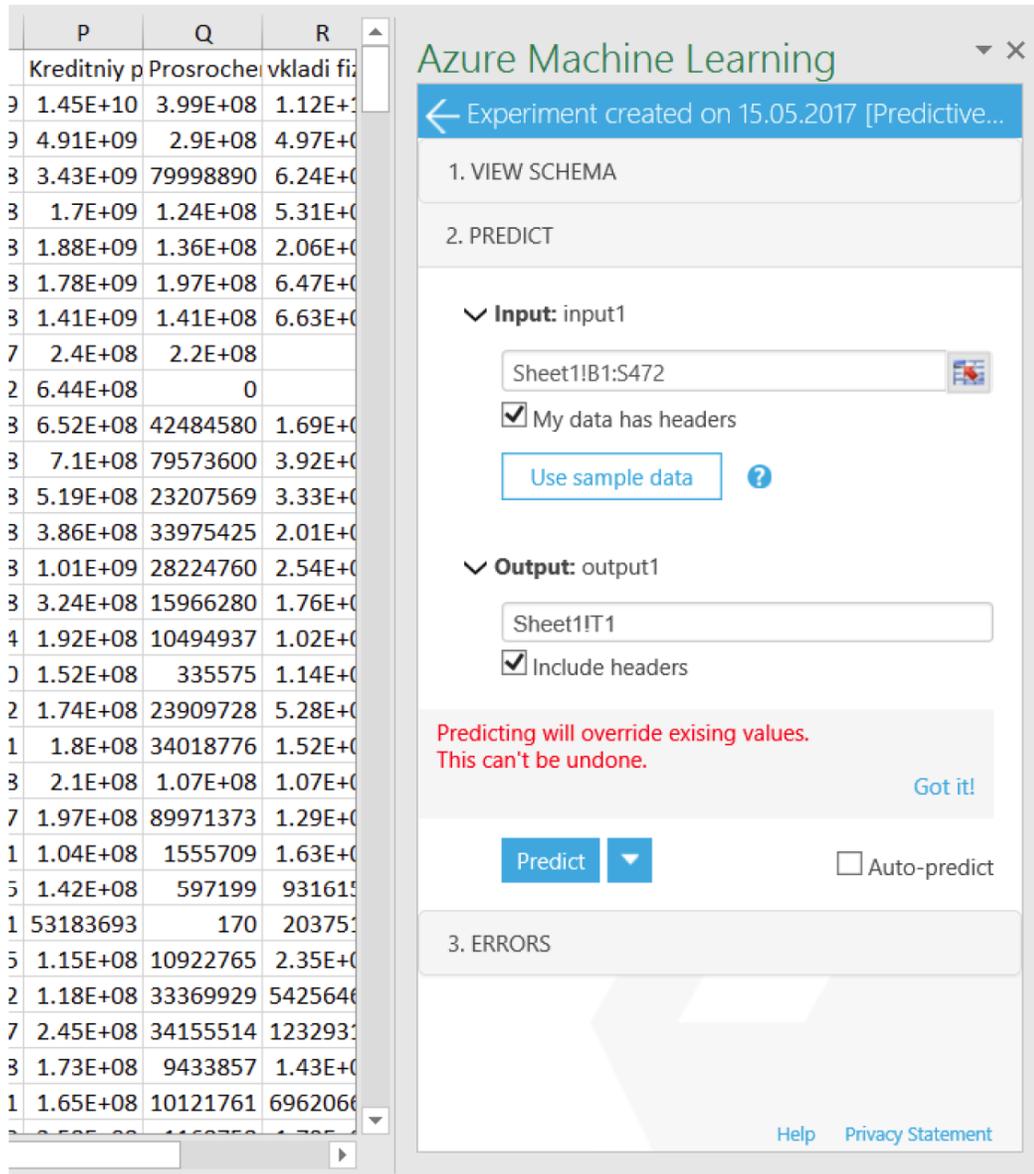
Figure 6: Deal with predicted data.

Next, using the function *Set up web service*, we connect our entire experiment to a file with predicted data (see the Figure **6**):

In the "Input" the range with the data is selected; it is important that the column names correspond to the names from the file with parameters for training the model.

In the "Output" we select the cell from which we want to start inserting the results.

A part of the prediction results are represented below in the Table **1**:

As follows from the represented results, banks, which negative values were observed for a year, fell into the category "bad". Banks that had a growth of profit, capital, etc. were defined as "good". Hence, we can conclude that predicting the probability of revoking a bank license with the help of Azure Machine Learning makes sense, since machine learning allows us to identify quickly bad banking system cells so that, for example, the Central Bank of Russia can further study their condition for a decision concerning revoking a license.

CONCLUSIONS

Thus, during the conducted research the following tasks have been solved:

1.   The set of Russian banks into classes depending on the size of the capital is divided: large,

Table 1: Prediction Results

№	Δ Assets	Δ Profit	Δ Capital	Δ Portfolio	…	Adherence	Size	Prediction
3255	-7367577	150586	-6637689	-6068102	…	1	1	0
3407	-8987946	-29854	-304910	-6861007	…	1	2	0
588	80060800	-35880	178457	1384713	…	1	2	1
3073	8511332	742723	249920	-11644268	…	1	2	1
880	-28467782	19426122	-14904766	-7025570	…	2	1	0
2546	-41537370	4761768	-4635257	7449564	…	3	1	0
467	123483	3843	39268	1778691	…	2	3	1
2655	-3230250	-183739	-217569	-2201765	…	3	3	0
1249	214186	27035	91374	416193	…	1	3	1
2768	-490315	166111	-398871	-823725	…	3	3	0
1411	268701	7086	72102	-745590	…	1	3	1
2763	6906406	-851999	-979951	-3666870	…	3	2	1
732	966799	14393	114910	94416	…	2	3	1
2888	172383069	14000621	141788592	1364187	…	2	1	1
2684	-1911416	-50158	-47495	-505029	…	3	3	0
2156	-173786	74726	99465	417171	…	3	3	1
316	2149832	3493555	1841159	-3256468	…	2	1	1
1810	-10689091	864456	-769703	-12422134	…	2	2	0
2879	6103961	-398730	1397144	-7608026	…	1	2	1

medium, small. The stable belonging of each bank to its class has been revealed.

2. The lack of multi-collinearity of the main indicators characterizing the activity of banks is established.

3. The degree of commitment of all Russian commercial banks in the interval from January 2011 to April 2017 is determined.

4. By means of multiple computational experiments, a system of indicators is developed in order to predict a state of a bank.

5. Integral indicators are included in the system of indicators.

6. Using the system of machine learning Azure ML, we managed to predict the possibility of revoking licenses from banks.

REFERENCES

Addo P., Guegan D., Hassani B., (2018) "Credit Risk Analysis Using Machine and Deep Learning Models", MDPI, Basel, Switzerland

Aleskerov F, Belousova V, Serdyuk M *et al*., (2008) "Stereotypes of the Russian banks behavior", Banking, № 7. pp. 13-15.

Alexashin P., Aleskerov F., Belousova V., Popova E., Solodkov V., (2012) "Dynamic analysis of business models of Russian banks in the period 2006-2009", Higher School of Economics

Bernes J, (2015) "Azure machine learning: Microsoft Azure Essentials", Microsoft Press

Brink H, Richards J, Fetherolf M, (2016) "Real-world machine learning", Manning Publications

Fetisov G, (2011) "Sustainability of a commercial bank and rating systems for its evaluation", Moscow: Finance and Statistics

Petropoulos A., Siakoulis V., Stavroulakis E., Vlachogiannakis N., (2017) "Predicting bank insolvencies using machine learning techniques", 2017 EBA Policy Research Workshop, London

The financial stability board (FSB), (2017) "Artificial intelligence and machine learning in financial services. Market developments and financial stability implications", http://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/