- Amount of neurons of hidden layer

A way used to calculate amount of neurons in the layer: (number of attributes + number of classes) / 2 + 1

- Training cycles
- Learning rate
- Momentum

For automatically tuning the parameters and finding the best combination of them, we used the statistica program tools. It allows to set some limits for parameters and to train specified number of nets. As a result, we have nets ranked by performance.

## VI. Metric to Estimate the Performance of Models

As was said before, such typical metric as classification accuracy is not a good metric, because if a model correctly classify just instances of majority class, then the model have high accuracy by using the metric of estimation.

In the case, when the minority class is a class which represents fault states of the technical equipment and it is more important to accurately classify the vectors of the class, than vectors of the other class.

That is why, the confusion matrix was chosen as a metric to determine the quality of the model. The matrix allows to estimate the recall of a specific class and getting a clearer representation of the model efficiency.

To calculate the recall of the minority class in the data set (monitoring parameters of technical equipment) it is necessary to split the dataset into 3 parts: training set, validation set, testing set. The first step is to train the classifier with data from the major part, which is more than 50 % of whole data set. The second step is to make some intermediate validation of model efficiency following by tuning the model. The last step is a final one-time testing. It allows to make more accurate estimation of the model performance, because testing on a validation set is some kind of training (changing parameters of a model and retraining the model). The results of the final testing are shown in table 2.

Calculation of classification accuracy:

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn} = \frac{323+2130}{323+2130+132+57} = 0{,}92$$

Therefore, the result is an accuracy equal to 92%, which is a good result for classification. However, the method of estimation does not represent the whole efficiency of the model.

Calculation of the recall of class True:

$$Recall = \frac{tp}{tp+fn} = \frac{323}{323+57} = 0{,}85$$

Comparison of classification results PNN and MLP:

The result is 85 percent, that means class True is recognized with probability of 85 percent. So with propability of 15 percent a failure or failts will not be recognized, that's not enough for effective operation of a system.

Training the models (MLP and PNN) on unbalanced data/ unrepresentative data showed, performance/accuracy of the models are equal, however it is not enough for effective performing of the current task of technical state of technological equipment determination.

TABLE II
Confusion Matrix

| | | **Real values / classification** | |
|---|---|---|---|
| | | True | False |
| Predicted values / classification | True | tp = 323 | fp = 132 |
| | False | fn = 57 | tn = 2130 |

## VII. Conclusion

In the paper it was explained what unbalanced data are and what influence it has on the performance of a system to classify different states of a technical equipment. Principal approaches of handling unbalanced data are discussed. Making a representative training set by using k-means clustering is shown. Several machine-learning algorithms such as probabilistic neural net and multilayer perceptron have been chosen as models to deal with the task of determination of different states of the technical equipment.

The traditional methods of diagnosis and control of parameters of the technological equipment are time-consuming and it doesn't present a possibility for express analysis. The results of the approaches presented in the article can be useful in some difficult situations as additional information to make a good decision.

Future research will be concentrated on applying all of that approaches and methods of handling with unbalanced data. The main issue is to figure out how much does this affect the final performance of the model. It is important to know which type of neural net fits to the case most.

## References


[1] Basmanov M., Menshikov S., Morozov I., Strebkov A.," System of parameter's diagnostic of GTU: modern approach" Delovaya Rossia N7, 2011.-42-43 p. (In Russian)

[2] Icamaan B. Viegas da Silva; Paulo J. L. Adeodato, "PCA and Gaussian noise in MLP neural network training improve generalization in problems with small and unbalanced data sets" Proceedings of International Joint Conference on Neural Networks, San Jose,

[3] Robert E. Schapire, "The Strength of Weak Learnability" Machine Learning, 5(2):197–227, 1990

[4] Statsoft program "Statistica". [Online]. Available: http://www.statsoft.ru/

[5] Dolina O.N. and Kyzmin A.K. Particularities of developing expert systems based on neural net modeling" Journal of Saratov technic state university, 2009.-266-272 p. (In Russian)

[6] S. Sathyanarayana, "A gentle introduction to backpropagation", July 22, 2014

[7] Specht, D. F. "Probabilistic neural networks". Neural Networks 3:1990.- 109–118 p.

[8] A.I. Gavrilov and P.V. Evdokimov, "Neural Network optimum parameters determining under industrial process mathematical model construction" Vestnik of ISPU N4, 2007(In Russian)