UDC 669.331.7:746

A.U.Kalizhanova[1,2]

A.Kh.Kozbakova[1,3]

U.N.Imanbekova[1,2]

A.N.Imanbekova[1,4]

M.M.Kunelbayev[1,4]

COMPARISON OF DIFFERENT CLASSIFICATION ALGORITHMS AND SELECTION OF THE OPTIMAL ONE IN THE TECHNOLOGICAL PROCESS

[1]Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan
[2]Almaty University of Power Engineering and Telecommunications named after G.Daukeyev, Almaty, Kazakhstan
[3]Almaty Technological University, Almaty
[4]Kazakh National University named after Al Farabi

Annotation.
This article examines the main classification methods using the application. The classification of the test data was carried out using different algorithms. We compared the results of the classification algorithms and selected the most appropriate one. Clustering was performed using a decision tree, and the algorithm is tested using examples from each block.

Keywords: classification model, algorithm, optimality, technological process, clustering, cross-validation method.

We chose the Classification Tree classifier, and used Test & Score and Confusion Matrix to analyze the classification quality, since it supports various sampling methods, that is, splitting the input data into training and test samples, and Cross validation splits the data into a user-defined number of blocks [1]. Here (figure -1), the algorithm is tested using examples from each block, while the blocks used for training and prediction are constantly changing (first, the first block is predicted, then the second, and so on, and the remaining blocks are used for training) [2].

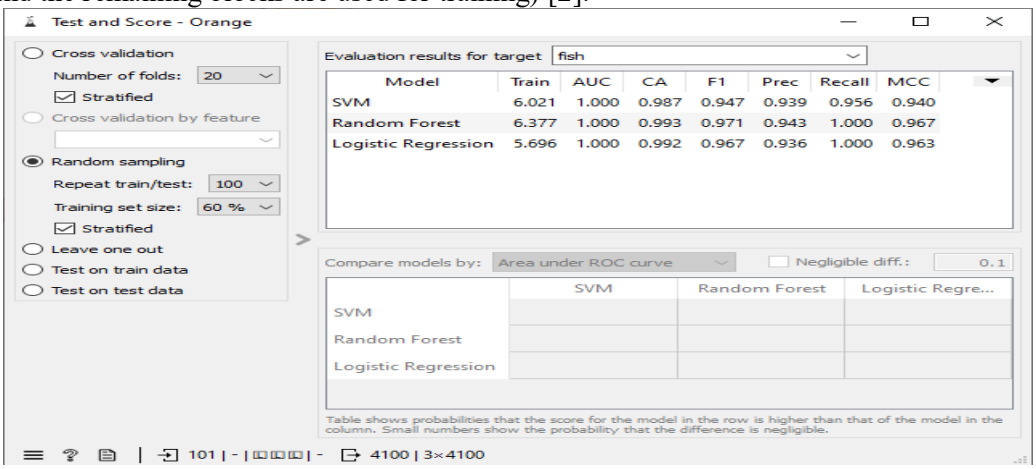

Figure 1 – parameters for testing classification data

In this article, we used the Leave-One-Out (LOO) method. since this method is a cross-validation method that is used to evaluate the performance of machine learning models, because it is a special case of cross-validation, where the number of subsets of data is equal to the number of observations in the training sample. This means that at each iteration, the entire dataset is used for training, with the exception of one instance, which is used for testing the model [3].

Then a random sample was selected. Random Sampling is a method of selecting a subset of data or elements from a wider group (population), in which each element of this group has the same probability of being selected. Here, the main purpose of random sampling is to ensure that the sample is representative so that the results of the analysis can be generalized to the entire population.

Next, we need to use classification functions to build a matrix model. Next, we set the parameters for random hierarchical trees. This is how we set the parameters for the logistic regression in order for our matrix to be correct. And here the final data from three different types of classification comes out. To build our matrix, we also set parameters so that the matrix displays all the information that we received from the data and a matrix model is built based on this data.

Here is a matrix built on mathematical data that displays the correct data from three types of classifications that have gone through training and test data.

Acknowledgment.

Research were carried out within the framework of the Grant Financing Project No. AP19679153 "Research and development of method and technologies for creating composite structures with built-in photonic sensors PSBC (Photonic Smart Bragg Composites)" Institute of Information and Computational Technologies of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

Sources of funding for research presented in a scientific article or scientific article itself.

This work is supported by grant from the Ministry of Science and Higher Education of the Republic of Kazakhstan within the framework of the Grant Financing Project No. AP19679153 "Research and development of a method and technology for creating composite structures with built-in photonic sensors PSBC (Photonic Smart Bragg Composites)", Institute of Information and Computational Technologies CS MSHE RK.

REFERENCES

1. Malfliet A., Lotfian S., Scheunis L., Petkov V., Pandelaers L., Jones P.T., Blanpain B. Degradation mechanisms and use of refractory linings in copper production processes: Journal of the European Ceramic Society. A critical review. 34, 3 (2014), p. 849-876.
2. Kim B.S., Jo S.K., Shin D., Lee J.Ch., Jeong S.B. A physico-chemical separation process for upgrading iron from waste copper slag. International Journal of Mineral Processing. 124 (2013), 124-127.
3. Cheng P., Herreros P., Lalpuria M., Grossmann I. Optimal scheduling of copper concentrate operations under uncertainty. Computers & Chemical Engineering, 140(2020), p.106919. doi.org/10.1016/j.compchemeng.2020.106919

Kalizhanova Aliya Ualiyevna - professor, Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan. Almaty University of Power Engineering and Telecommunications named after G.Daukeyev, Almaty, Kazakhstan. e-mail: kalizhanova.aliya@gmail.com

Kozbakova Ainur Kholdasovna - PhD, associate Professor Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan. Almaty Technological University, Almaty, Kazakhstan; e-mail: ainur79@mail.ru.

Imanbekova Ulzhan Nurkhaliyevna - PhD., associate Professor, Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan. Almaty University of Power Engineering and Telecommunications named after G.Daukeyev, Almaty, Kazakhstan. e-mail: uli.08@mail.ru

Imanbekova Aliya Nurgaliyevna - Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan. Kazakh National University named after Al Farabi. e-mail: aleka.12@mail.ru

Kunelbayev Murat Merkebekovich- Institute of Information and Computational Technologies CS MSHE RK, Almaty, Kazakhstan. Kazakh National University named after Al Farabi. e-mail: *kunelbayev@mail.ru*

ПОРІВНЯННЯ РІЗНИХ АЛГОРИТМІВ КЛАСИФІКАЦІЇ І ВИБІР ОПТИМАЛЬНОГО В ТЕХНОЛОГІЧНОМУ ПРОЦЕСІ

Анотація

У даній статті вивчена основні методи класифікації з використанням програми. Здійснено класифікацію тестових даних, використовуючи різні алгоритми. Порівняли результати роботи алгоритмів класифікації і вибирали найбільш підходящий. Кластеризацію виконали за допомогою дерева рішень, а алгоритм тестується на прикладах з кожного блоку.

Ключові слова: *модель класифікації, алгоритм, оптимальність, технологічний процес, кластеризація, метод крос-валідації.*

Каліжанова Алія Уаліевна*-професор, Інститут інформаційних і обчислювальних технологій Міністерство науки і вищої освіти Республіки Казахстан, Алмати, Казахстан. Алматинський університет енергетики та телекомунікацій ім. Г. Даукеева, Алмати, Казахстан. e-mail: kalizhanova.aliya@gmail.com*

Козакова Айнур Холдасівна*- доцент Інституту інформаційних і обчислювальних технологій Міністерство науки і вищої освіти Республіки Казахстан, Алмати, Казахстан. Алматинський технологічний університет, Алмати, Казахстан; e-mail: ainur79@mail.ru.*

Іманбекова Улжан Нурхаліевна*- доцент, Інститут інформаційних і обчислювальних технологій Міністерство науки і вищої освіти Республіки Казахстан, Алмати, Казахстан. Алматинський університет енергетики та телекомунікацій ім. Г. Даукеева, Алмати, Казахстан. e-mail: uli.08@mail.ru*

Іманбекова Алія Нургаліївна *- Інститут інформаційних та обчислювальних технологій Міністерство науки і вищої освіти Республіки Казахстан, Алмати, Казахстан. Казахський національний університет імені Аль-Фарабі. email: aleka.12@mail.ru*

Кунельбаєв Мурат Меркебекович *- Інститут інформаційних та обчислювальних технологій Міністерство науки і вищої освіти Республіки Казахстан, Алмати, Казахстан. Казахський національний університет імені Аль-Фарабі. email:kunelbayev@mail.ru*