Studying the features of the use of neural networks and machine learning in the design of food systems

Olga Musina[1*], and *Elena* Nagornykh[1]

[1]Polzunov Altai State Technical University, Barnaul, Russia

Abstract. The features of the use of neural networks and machine learning methods for the design of food systems are studied using the example of processed cheese. A database has been developed that includes 869 cheese recipes, and a program for managing an electronic recipe directory. A neural network has been developed, for the training of which the method "Training with a teacher", the activation function "ReLu" and the author's program written in the Python programming language were used. The neural network consisted of 9 neurons in the input layer, two hidden layers of 65 neurons each, and an output layer consisting of 1 neuron. To determine the linear correlation between columns, a matrix was used showing the relationship between values using the Pearson coefficient. A training sample containing 80% of the total number of recipes, a test sample of 10% of the total number of recipes, and a test sample of 10% of the total number of recipes were selected from the primary data set. As a result of training the neural network, an information-advising system for a food technologist has been developed. The system is designed to predict the quality of food recipes. The information-advising system will speed up the correction of existing recipes and the development of recipes for new products, theoretically predict their quality before launching into production. The information-advising system was tested on a test recipe of a new processed cheese. It has been established that with a certainty of 63.6%, the integral indicator of the quality of the new cheese will be 7.7 conventional units. This predicted value was confirmed during the practical production of cheese according to the designed recipe in laboratory conditions and during approbation in production conditions. The new cheese is really distinguished by high quality, good organoleptic and physico-chemical parameters.

1 Introduction

For the stable operation of a food industry enterprise, it is necessary to continuously regulate a large number of processes, such as: control of logistics chains for the delivery of raw materials; quality control of raw materials entering the enterprise; control of the operation of equipment and lines at all stages of production; control of quality supplies of

---

[*] Corresponding author: musinaolga@gmail.com

spare parts for production lines; monitoring the work of employees at all stages of production; product quality control; control of delivery of products to the customer's warehouse. Along with this, from time to time there is a need both to adjust existing recipes and to develop new recipes for food products, taking into account the requirements of regulatory documents and market demands. For example, the production volumes of processed cheeses with various fillings are increasing every year, which requires constant updating of recipes and updating the product range.

Modern recipe design methods based on the principles of food combinatorics should use the capabilities of digital tools, including neural networks, to solve recipe problems to create food systems with a desired set of characteristics. Therefore, the development of digital tools adapted to the needs of the food industry, including the design of processed cheese recipes, is an urgent scientific task [1].

Artificial intelligence and machine learning programs are gradually being used in the real sector of the economy. However, quite a few information-advising systems have been developed for the food industry, and they are mainly focused on solving production and economic problems, such as calculating the optimal volumes of gross output in physical and value terms [2].

Information-advising systems provide the formation of many alternatives for making decisions on managing an object. Information-advising systems are systems with a sufficiently high degree of intelligence. Nevertheless, the advice-decisions issued by them to the user are recommended for taking into account, and not for immediate execution. In our case, with the help of an information-advising system, variants of recipes for food systems, for example, processed cheeses, are formed. The decision maker (food technologist) chooses a specific option from the proposed alternatives, taking into account his knowledge of the technology and experience with a particular type of food product.

2 Materials and methods

As an auxiliary tool, we have developed the database "Processed cheese recipes" [3] and the web application "Program for managing the electronic directory of processed cheese recipes" [4], which work with a large number of browsers. These tools have been officially registered by the Federal Institute of Industrial Property [5].

To fill the database, the selection of reliable information about the recipes of processed cheeses was carried out from the following sources:
- Officially published collections of technological instructions for the production of processed cheeses.
- Information from articles in publications indexed in international scientific citation databases.
- Materials of dissertations placed in the "Russian State Library".
- Peer-reviewed scientific monographs and study guides.
- Statistically processed results of our own research, obtained by standardized test methods, including arbitration ones.

869 processed cheese recipes were entered into the database.

Next, we developed a neural network, for training which we used the Supervised Learning method and one of the most popular activation functions in machine learning, the ReLu function (Figure 1).

All stages of data analysis and machine learning were carried out using a self-written program in the Python programming language.

A file was prepared for the neural network, including recipes and indicators typical for processed cheeses (mass fraction of protein, mass fraction of moisture, mass fraction of

solids, mass fraction of fat in dry matter, mass fraction of table salt). If the indicator for the recipe was not set, its value was equal to 0.
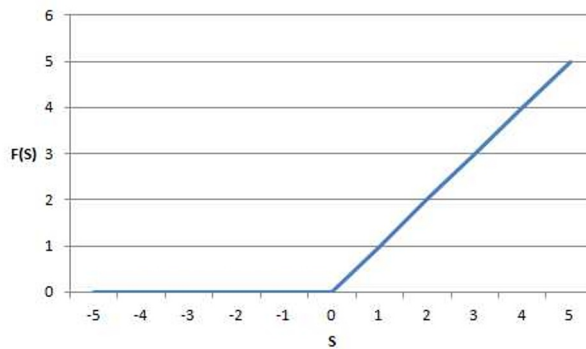

Fig. 1. Graph of the ReLu function.

Also, a new field "QUALITY" was included in the file - a conditional indicator of the integral quality of the recipe that we entered. To train the neural network, the QUALITY field was previously filled in by us by random generation based on the hypothesis that recipes from official sources have a quality score of 6 to 10, and the "training" recipes we modeled have a quality score of 1 to 5.

The set of input training data involved 869 processed cheese recipes, all values are non-zero, have no gaps, all types are correct and are numeric.

Thus, the neural network received input training data containing labels of correct and incorrect answers about recipes and indicators for processed cheese.

The neural network consisted of 9 neurons in the input layer, two hidden layers of 65 neurons each, and an output layer consisting of one neuron.

Column histograms were used to visually analyze the information contained in the set of input training data.

The correlation matrix was used to determine the dependencies between the columns. This matrix shows a linear relationship between column values using Pearson's correlation coefficient. The value of the Pearson coefficient ranges from -1 to 1, where -1 indicates a negative linear correlation between the two columns, +1 indicates a positive linear correlation between the two columns, and 0 indicates no linear correlation between the two columns. The greater the absolute value of the correlation coefficient, the stronger the relationship between the two variables. For convenience of analysis, the matrix is colored in colors depending on the degree of correlation.

Two facts should be noted. First, the diagonal values in the correlation matrix are always 1 because the correlation between the variable and itself is always 1. Second, since the correlation matrix is symmetrical with respect to the diagonal values, the lower triangular part of the matrix should be considered to analyze the values.

3 Results and Discussion

Before training the neural network, a preliminary assessment of the input training data was made, including the analysis of the histogram of the columns (Figure 2) and the correlation matrix (Figure 3).

The histograms of the input training data columns show the correctness of the choice of columns for training the neural network (column values other than 0), the correlation matrix provides information about the dependence of the columns on each other.
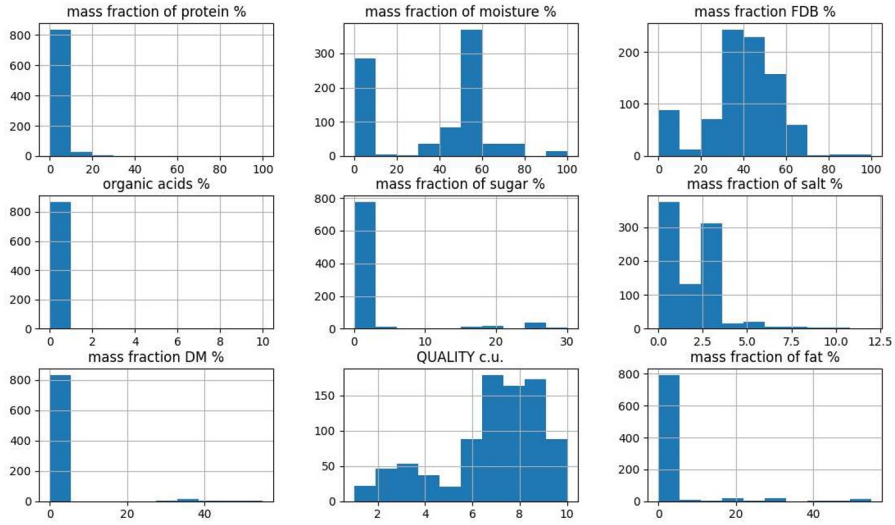
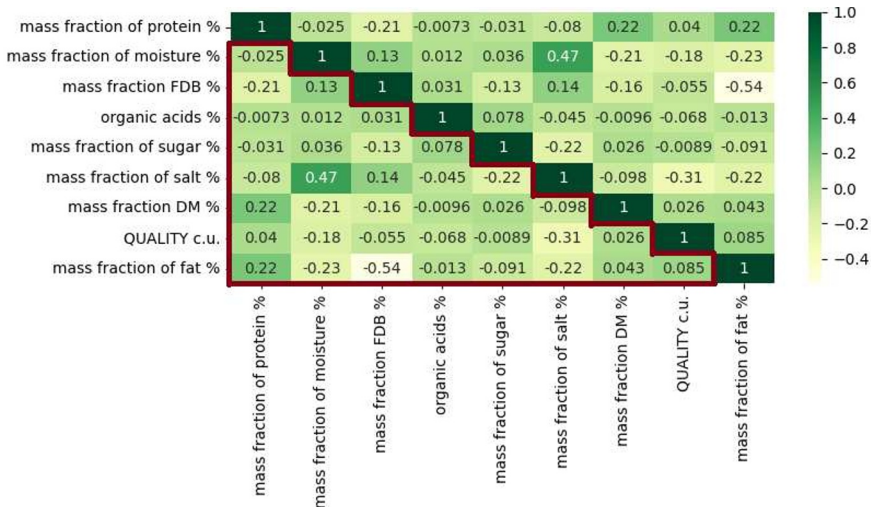**Fig. 2.** Histogram of input training data columns.

Fig. 3. Correlation matrix.

Next comes the actual learning process of the neural network.

From a data set containing 869 processed cheese recipes, we select a training sample containing 695 recipes (80% of the total number of recipes).

We select a test sample, which is 87 recipes (10% of the total number of recipes) and a test sample, which also makes up 87 recipes (10% of the total number of recipes).

On average, small datasets such as ours are subjected to 100 training cycles. Histograms of the dependence of the number of recipes on the quality score of the recipe of the training set are shown in Figure 4, the number of recipes from the quality score of the recipe of the test sample in Figure 5, the number of recipes from the quality score of the recipe of the test sample in Figure 6 and a graph of the average absolute error versus the number of cycles for the training and testing samples - in Figure 7.
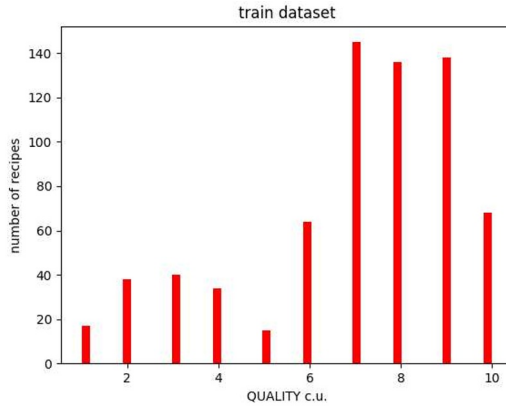
Fig. 4. Histogram of the dependence of the number of recipes on the quality score of the recipe of the training sample.
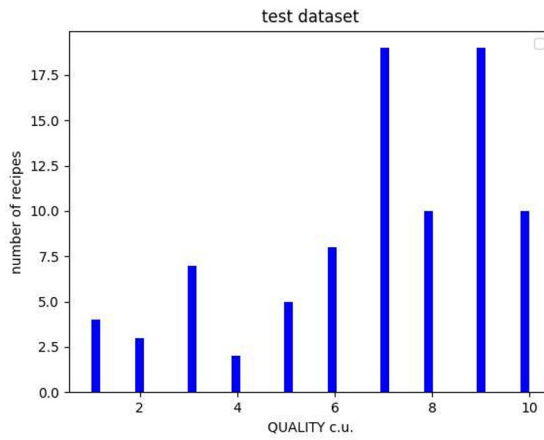

Fig. 5. Histogram of the dependence of the number of recipes on the quality score of the test sample recipe.
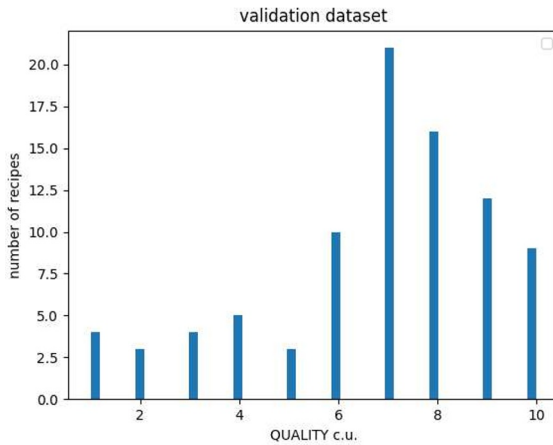

Fig. 6. Histogram of the dependence of the number of recipes on the quality score of the test sample recipe.
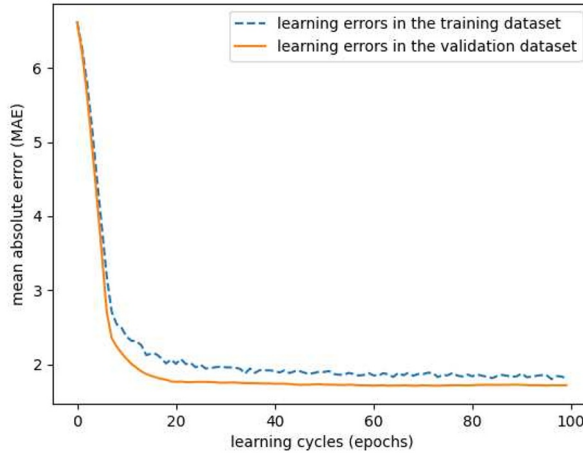
Fig. 7. Graph of the average absolute error versus the number of cycles for the training and testing samples.

The results of training the neural network for the test recipe of the new processed cheese are shown in Figure 8.

```
The result of neural network training:
Accuracy - 63.58576409596202%
Learning completed. Please enter the recipe input parameters. If the recipe parameter is not set, enter 0.
Enter the metric value - mass fraction of fat%
0
Enter the metric value - mass fraction of protein %
0
Enter the metric value - mass fraction of moisture %
52
Enter the metric value - mass fraction FDB %
55
Enter the metric value - organic acids %
6.5
Enter the metric value - mass fraction of sugar %
0
Enter the metric value - mass fraction of salt %
0.2
Enter the metric value - mass fraction DM %
0
[0.0, 0.0, 52.0, 55.0, 6.5, 0.0, 0.2, 0.0]
[[ 0.    0.   52.  55.    6.5  0.    0.2  0. ]]
1/1 [==============================] - 0s 13ms/step
[[7.704522]]
Expected value - [7.704522]
```

Fig. 8. Results of training a neural network for a test recipe for a new processed cheese.

With a certainty of 63.6%, the integral indicator of the quality of the new cheese will be 7.7 conventional units. This predicted value was confirmed during the practical production of cheese according to the designed recipe. Cheese, which received the fantasy name "Syrme", is distinguished by high organoleptic and physico-chemical parameters. The cheese recipe has been industrially tested at the operating enterprise for the production of processed cheeses. Regulatory and technical documentation has been developed for this cheese, which makes it easy to introduce the technology into production.

4 Conclusion

Information-advising systems provide the formation of many alternatives for making decisions on managing an object. The advice-decisions issued by them to the user are recommended for taking into account, and not for immediate execution. Research aimed at developing an information-advising system using neural networks and Machine Learning methods, which facilitates the work of a technologist when designing food systems, is relevant.

A database has been developed that includes 869 cheese recipes, and a program for managing an electronic recipe directory. The developed and trained neural network consisted of 9 neurons in the input layer, two hidden layers of 65 neurons each, and an output layer consisting of 1 neuron. A training sample containing 80% of the total number of recipes, a test sample of 10% of the total number of recipes, and a test sample of 10% of the total number of recipes were selected from the primary data set. As a result of training the neural network, an information-advising system was developed for the food production technologist, which allows more efficient design of recipes. In our case, with the help of an information-advising system, variants of processed cheese recipes are formed. The technologist chooses a specific option from the proposed alternatives, taking into account his knowledge and experience with a particular type of food product.

The system is designed to predict the quality of food recipes. The information-advising system will speed up the correction of existing recipes and the development of recipes for new products, theoretically predict their quality before experimental development. The information-advising system was tested on a test recipe of a new processed cheese. It has been established that with a certainty of 63.6%, the integral indicator of the quality of the new cheese will be 7.7 conventional units. This predicted value was confirmed during the practical production of cheese according to the designed recipe.

References

1. O. Musina, P. Putnik, M. Koubaa, F.J. Barba, R. Greiner, S. Roohinejad, D. Granato, Application of modern computer algebra systems in food formulations and development: a case study, Trends in Food Science & technology, **64**, 48-59 (2017)

2. O.D. Kazakov, A.V. Averchenkov, Development of an information and consulting system for managing production processes, Proceedings of the Voronezh State University of Engineering Technologies, **79**, **2**, 280-284 (2017)

3. O.N. Musina, E.M. Nagornykh, Database "Processed cheese recipes", Certificate of state registration of the database No. 2023620806. Appl. 02/22/2023, publ. 03/06/2023, Bulletin of computer programs, databases, topologies of integrated circuits, **3** (2023)

4. O.N. Musina, E.M. Nagornykh, Computer program "Program for managing an electronic directory of processed cheese recipes", Certificate of state registration of the computer program No. 2023618007. Appl. 04/04/2023, publ. 04/18/2023, Bulletin of computer programs, databases, topologies of integrated circuits, 3 (2023)

5. Programs for computers, databases and TIMS, Federal Institute of Industrial Property, https://www1.fips.ru/iiss/