Using Machine Learning Algorithms to Detect Milk Quality

Ahmet ÇELIK

Kütahya Dumlupınar University, Tavşanlı Vocational School, Computer Technologies Department, Kütahya, Türkiye.

**Corresponding author: ahmet.celik@dpu.edu.tr*

ORCID: 0000-0002-6288-3182

Abstract

Machine learning algorithms are used successfully in many sectors. The data formed by the development of digital technology are analyzed with machine learning algorithms and estimation, classification or clustering processes are carried out. Today, the food industry has a very important place and it will be very useful to follow the quality of the products produced and to determine in a short time. Milk is a product that people benefit from raw or processed. Milk is also a perishable product. Each gram milk with poor quality or structure can cause tons of milk to deteriorate, thus causing great financial losses. Millions of bacteria can form in spoiled milk in a very short time. In this way, if people consume milk or dairy products, situations that endanger human health may occur. In this study; A study was conducted in which milk quality was determined by machine learning algorithms. Seven features were used to determine milk quality. In the study, the Milk Quality dataset obtained from the open source Kaggle data repository was used. There are 1059 sample data in the data set. By using 7 attributes of milk samples, low, medium and high quality classification of milk was carried out. In the classification estimation phase, commonly used Neural Network (Neural Network: NN) and Adaptive Boosting (AdaBoost: AB) algorithms were used. Orange platform, which is open source and written in python, was used as the application platform. Orange is a platform with a widely used and see user interface. In the application phase, the results obtained with each algorithm were presented with visual graphics and comparisons were made. In the test phase, 100 milk data samples were used separately for each class in order to achieve a balanced learning. Random samples were selected from the data set for training. According to the results obtained; Classification accuracy (CA) success was achieved by 99.9% with AdaBoost algorithm and 95.4% with Neural Network algorithm. More successful results were obtained with the AdaBoost algorithm than the Neural Network algorithm.

Keywords: Machine learning, Neural network, AdaBoost, Smart decision system, Quality control, Food safety

Research article
Received Date: 27 September 202
Accepted Date: 13 December 2022

INTRODUCTION

With the developing technology, machine learning algorithms is used efficiently while performing quality control in the food industry (Xiao et al., 2019), in the mobile device system (Tahtacı and Canbay, 2020), while detecting malware and attacks on a computer system, server or any network (Singh et al., 2021), while diagnosing diseases in the health sector (Vrindavanam at al., 2021), while making control and error detection in the industry (Kanawaday and Sane, 2017), while making computer vision estimation (Çelik, 2022) while classifying products (Ozkan et al., 2021).

Milk is the most important food source and raw material for human health. Processed milk products (fermented) butter, yogurt, ayran, cheese (white cheese, string cheese, knitted cheese, cottage cheese, kashar etc.) (Küçük and Tapkı, 2020), many products such as kefir and kumiss are produced. With the development of the milk market, more demand comes to food producers. Therefore, improving the quality of fermented milk and reducing the rate of customer complaints has become the focus of food manufacturers.

Determining the milk quality by manual methods can result in high margin of error or loss of time. Since determining the quality of milk with only one feature will not give accurate results, it should be determined by taking into account many features. For this reason, it will be very efficient to analyze the milk data by an intelligent system using many features and to perform quality control.

Xiao et al. (2019) set up a random forest model, LR model, and AdaBoosting model and performed tests to find the most appropriate classification model. In the study, they showed that the color, aroma and taste of milk as attributes have the ability to recognize the quality of milk. According to the results obtained, the success rate was found to be 96.8%.

Sambasivam et al. (2020) estimated the Vitamin D Deficiency (VDD) rate, by using K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Bagging Classifier (BC), ExtraTrees (ET), Stochastic Gradient Descent (SGD), Gradient Boosting (GB), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) machine learning algorithms. According to the results they obtained, 96% correct results were obtained with the Random Forest algorithm.

Titova et al. (2018) performed food quality classification with Multi-layer perceptron (MLP) and AdaBoost algorithm. They used the pattern properties of potatoes as the data set.

Kumar et al. (2019) classified plant species by AdaBoost, k-NN, Decision Tree and Multilayer perceptron algorithms. They showed that leaves, flowers, fruits and seeds can be used while classifying. In the study, a success rate of 95.42% was achieved with machine learning algorithms.

Ricci et al. (2021) used x-rays to detect foreign materials on commercial products. In the study, they carried out a study to determine food safety and protect human health with the Neural Network algorithm, depending on the material density.

Keshavamurthy et al. (2019) proposed a food quality application using the method of identification and review with the OpenCV python library. Convolutional Neural Network (CNN) algorithm has been applied to perform fruit type recognition and quality detection tasks with quantitative data, precisely and reliably.

Vaishnav and Rao (2018) applied machine learning algorithms on the fruit image dataset using the Orange Data Mining Tool. In the study, they performed a comparative study to determine the algorithm with the highest classification accuracy and precision score. Attributes derived from the trained images were used in the decision making phase.

Thange et al. (2021) used the Orange Data Mining Tool to detect the relationship between case symptoms using the COVID-19 dataset shared in the Kaggle data storage and visually showed the results.

In this study, AdaBoost and Neural Network algorithms were applied on the Milk Quality dataset shared in Kaggle data storage and the success of estimating milk quality was compared. In the study, Orange Data Mining Tool was used as the application software interface. According to the results obtained, it was seen that the AdaBoost algorithm made a highly successful quality classification.

MATERIAL and METHOD

In this study, AdaBoost and Neural Network algorithms were applied on the Milk Quality dataset using the Orange Data Mining Tool.

Orange data mining tool

Many tools are available for visualization of data while performing the Machine Learning process. Much of model training adheres to scripting languages. However, with the use of Orange Data Mining Tool, data preprocessing, model training, testing and visualization can be performed with a single software (Vaishnav and Rao, 2018).

The Orange tool was developed by Janez Demsar et al. in 2013 using the Python programming language. Orange offers a hierarchically organized and visualized user-friendly interface of data mining and machine learning algorithm components. Interactive analysis and tests can be performed with its interface and visual components using Orange that contain more than 200 classes. The scripting requirement has been minimized with Orange. Orange has two main packages that are still under active development; scikit Learn (Pedregosa et al., 2011) and mlpy (Albanese et al., 2012) have been developed integrated (Demsar et al., 2013; Orange Data Mining, 2022).

Data management and preprocessing Component: Used for data input and output, data filtering and sampling, feature manipulation (discrete, persist, normalize, scaling, scoring) and feature selection.

Transform Component: It is used for merging, finding pivot point, grouping and preprocessing on data.

Data visualization Component: It is used to draw the scatter graph, temperature map and line graph of the results obtained from the application and the result.

Prediction Component: It is used to apply various supervised machine learning algorithms (Random Forest, AdaBoost, Neural Networks, Naive Bayes, SVM) on the data set.

Classification/Regression Component: It is used to perform ROC Analysis, Complexity matrix, Test and Score, Calibration operations on the data.

Unsupervised learning Component: It includes k-means and hierarchical clustering approaches; it is used in distance matrix, distance map, neighborhood and distance operations.

Figure 1 shows, the model designed on Orange. The application of AdaBoost and Neural Network machine learning algorithms used in the figure is shown. ROC analysis of performance achievements, Complexity matrix and table data can be obtained with Test-Score components.

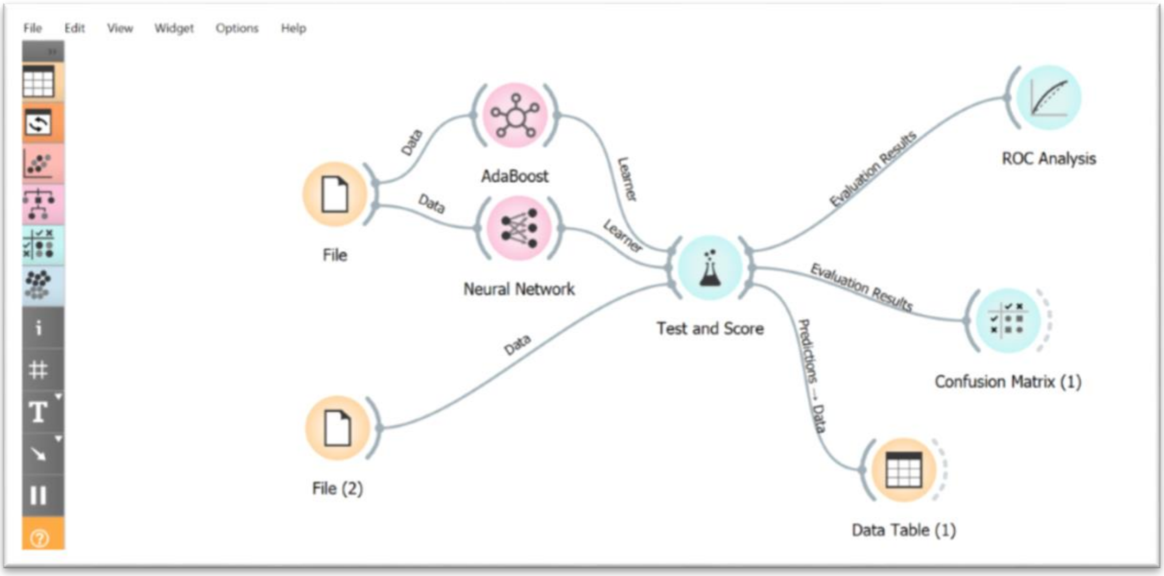


Figure 1. Components of Orange tool

Kaggle data storage platform

Kaggle, backed by Google LLC, is a data storage and sharing platform that provides free online services for data researchers and professionals in software engineering since 2010. Large data sets on Kaggle are stored on cloud systems. Kaggle is a platform for sharing knowledge to find and send datasets to users, research and build models in an electronic data science environment, work with artificial intelligence experts, and address data science challenges (Kaggle, 2022).

Neural network machine learning algorithm

First, in 1943 neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper on how neurons might work. In this article, they modeled a simple neural network using electrical circuits to explain how neurons in the brain work. In 1949, Donald Hebb introduced the concept of Organization of Behavior, pointing out the fact that neural pathways are strengthened each time they are used, and argued that if two nerves fire at the same time, the connection between them becomes stronger (Chung et al., 2009). In the researches of Rosenblat in 1958, Minsky and Papert in 1969, Grossberg in 1976, Hopfield in 1982 and Kohonen in 1984, computers became more advanced and an artificial neural network model was formed by simulating the learning logic of people (Uçan et al., 2006). Later, researchers tested the success of mathematical functions. There are many areas where the neural network model is used. Neural networks are generally examined in 5 basic structures. These are Inputs, weight coefficients, bias (constant), activation function and output value (Balaban and Kartal, 2018). Figure 2 shows the model of the artificial neural network.

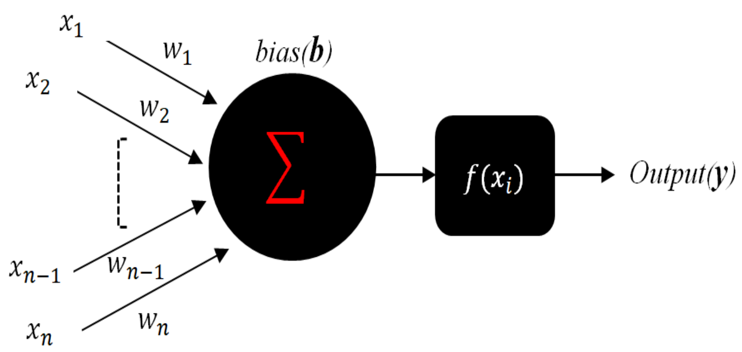


Figure 2. Neural network model

The mathematical expression of the neural network is shown in equation 1. y is the output value, w_i is the weight coefficient, x_i is the input value, and b is the constant coefficient.

$$y = w_i x_i + b \quad (1)$$

AdaBoost machine learning algorithm

By combining many rules, the correct prediction rule creation process is carried out with a machine learning approach. AdaBoost algorithm was first applied in practice by Freund and Schapire in their study in 1996. They proved that this method is a new machine learning algorithm (Freund and Schapire, 1999). It is widely used in many fields with its strong classification ability (Wang et al., 2019).

Adaboost algorithm can be applied to many classifier learning algorithms and strong classification results are obtained. In the AdaBoost algorithm, the "weight" value is calculated by analyzing all the data in the data set (Sun et al., 2006).

With the AdaBoost algorithm, the input value x is divided into two classes as shown in equation 2. The y_ class values are shown as -1 and 1.

$$x_i, x_{i+1}, x_{i+2} \ and \ y_n \in \{-1,1\} \quad (2)$$

Calculation of the weighting coefficient of n pieces of x data is shown in equation 3.

$$w_i = \frac{1}{n} \quad (3)$$

Milk quality dataset

The Milk Quality dataset used in this study was taken from the open source Kaggle data storage area. The dataset data used were obtained from manual observations. Milk samples have 7 attributes: pH, Temperature, Taste Odor, Oil, Turbidity and Color. Generally, the quality of milk is determined by looking at these characteristics. The target is to classify the milk as Low(Poor), Medium(Medium) and High(Good). Taste, Smell, Oil and Turbidity properties take the value 1 or 0. Temperature, pH and Color properties have true color values. Table 1 shows the values of 15 randomly selected milk samples (Kaggle,2022)

Table 1. Milk Quality dataset and classification (Kaggle, 2022)

pH	Temperature	Taste	Odor	Fat	Turbidity	Colour	Grade
6.6	38	1	0	1	0	255	high
6.8	45	1	1	1	1	245	high
6.8	36	0	1	1	0	253	high
6.6	45	0	1	1	1	250	high
6.8	45	1	1	1	1	245	high
6.8	43	1	0	1	0	250	medium
6.8	43	1	0	1	0	250	medium
6.8	43	1	0	1	0	250	medium
6.8	43	1	0	1	0	250	medium
6.8	43	1	0	1	0	250	medium
7.4	65	0	0	0	0	255	low
3	40	1	0	0	0	255	low
9	43	1	1	1	1	248	low
3	40	1	1	1	1	255	low
8.6	55	0	1	1	1	255	low

RESULTS and DISCUSSION

In the study, it was seen that the pH value and temperature value of the milk are an important factor in determining the milk quality. It has been observed that the pH value of high and medium quality milk is between 6-7, and the temperature values are at most 45 degrees. In low quality milk, it was observed that these two conditions did not provide at the same time. The distribution of the results obtained is shown on figure 3.

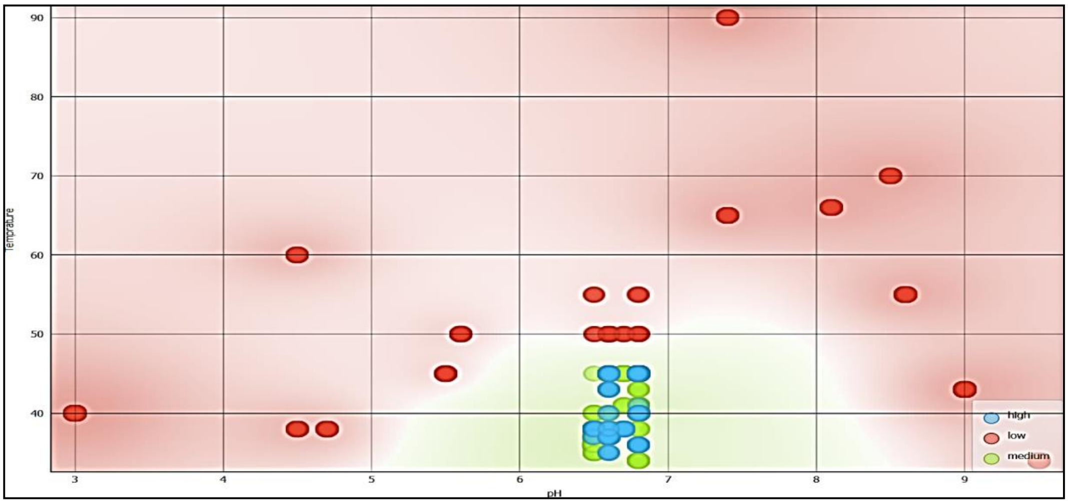


Figure 3. Classification distribution of temperature and pH values

The "high" classification results and attribute data of Milk Quality data with AdaBoost and Neural Network algorithm using Orange Data Mining Tool are shown on Figure 4. The “high” actual values are shown in the Grade field, and the estimated classification prediction results in the AdaBoost and Neural Network field.

With the Neural Network algorithm, in the example with Id number 153, the quality result that should have been "high" was incorrectly estimated as "medium".

Grade	id	AdaBoost	Neural Network	ost	ost	st (n	etwc	etw	work	Fold	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
high	143	high	high	1	2....	2....	0....	0....	0....	5	6.6	45	0	1	1	1	250
high	144	high	high	1	2....	2....	0....	0....	0....	4	6.8	45	0	1	1	1	255
high	145	high	high	1	2....	2....	0....	0....	0....	4	6.6	45	0	1	1	1	250
high	146	high	high	1	2....	2....	0....	0....	0....	4	6.8	45	1	1	1	1	245
high	147	high	high	1	2....	2....	0....	0....	0....	3	6.8	36	0	1	1	0	253
high	148	high	high	1	2....	2....	0....	0....	0....	3	6.6	37	1	0	1	0	255
high	149	high	high	1	2....	2....	0....	0....	6....	4	6.5	38	1	1	1	1	255
high	15	high	high	1	2....	2....	0....	0....	0....	1	6.8	36	0	1	1	0	253
high	150	high	high	1	2....	2....	0....	0....	6....	5	6.8	40	1	1	1	1	255
high	151	high	high	1	2....	2....	0....	0....	0....	1	6.8	45	1	1	1	0	245
high	152	high	high	1	2....	2....	0....	0....	0....	3	6.8	40	1	1	1	1	255
high	153	high	medium	1	2....	2....	0....	0....	0....	4	6.8	45	1	1	1	0	245
high	154	high	high	1	2....	2....	0....	0....	7....	1	6.6	37	1	1	1	1	255
high	155	high	high	1	2....	2....	0....	0....	0....	3	6.7	38	1	0	1	0	255
high	156	high	high	1	2....	2....	0....	0....	0....	4	6.8	43	1	0	1	0	250
high	16	high	high	1	2....	2....	0....	0....	0....	1	6.6	45	0	1	1	1	250
high	17	high	high	1	2....	2....	0....	0....	0....	4	6.8	45	1	1	1	1	245
high	18	high	high	1	2....	2....	0....	0....	0....	4	6.8	36	0	1	1	0	253
high	19	high	high	1	2....	2....	0....	0....	0....	5	6.6	37	1	0	1	0	255
high	2	high	high	1	2....	2....	0....	0....	0....	3	6.8	45	1	1	1	1	245
high	20	high	high	1	2....	2....	0....	0....	0....	3	6.5	38	1	1	1	1	255
high	21	high	high	1	2....	2....	0....	0....	7....	4	6.8	40	1	1	1	1	255
high	22	high	high	1	2....	2....	0....	0....	0....	1	6.8	45	1	1	1	0	245
high	23	high	high	1	2....	2....	0....	0....	7....	1	6.6	37	1	1	1	1	255
high	24	high	high	1	2....	2....	0....	0....	0....	1	6.6	35	0	1	1	1	255
high	25	high	high	1	2....	2....	0....	0....	0....	1	6.8	45	0	1	1	1	255
high	26	high	high	1	2....	2....	0....	0....	6....	4	6.5	38	1	1	1	1	255

Figure 4. "high" classification results and attribute data obtained with AdaBoost and Neural Network algorithms

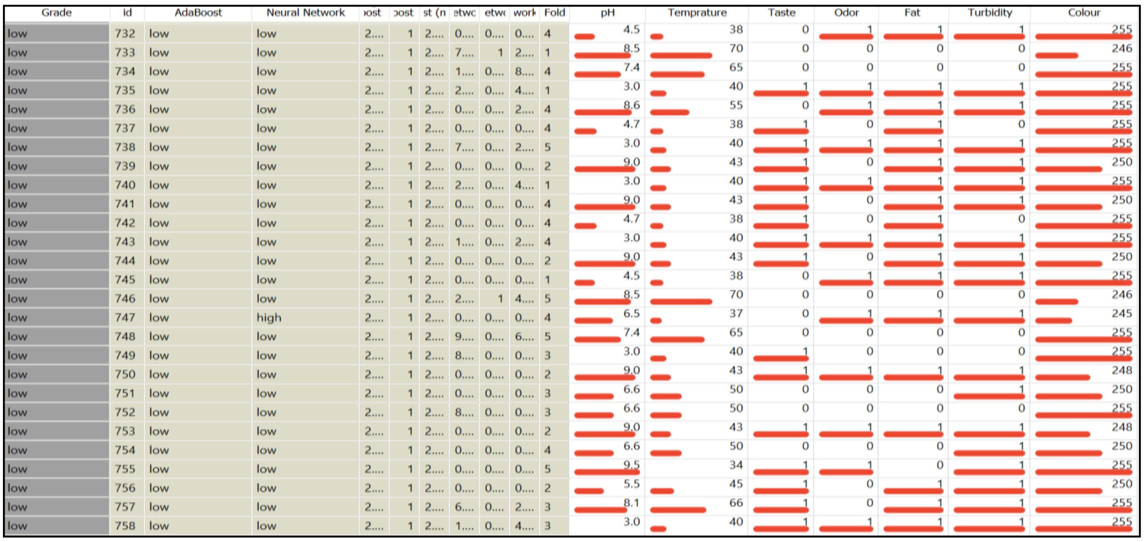
The results of "medium" classification and attribute data of Milk Quality data with AdaBoost and Neural Network algorithm using Orange Data Mining Tool are shown on figure 5. The "medium" real values are shown in the Grade field, and the predicted classification prediction results in the AdaBoost and Neural Network field. In the samples with Id numbers 158,159,168,176 and 178 with the Neural Network algorithm, the quality result that should have been "medium" was incorrectly estimated as "high".

Grade	id	AdaBoost	Neural Network	ost	ost	st (n	etwc	etw	work	Fold	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
medium	158	medium	high	2....	2....	1	0....	0....	0....	2	6.7	45	1	1	1	0	245
medium	159	medium	high	2....	2....	1	0....	0....	0....	4	6.5	38	1	0	1	0	255
medium	160	medium	medium	2....	2....	1	0....	0....	0....	1	6.7	41	1	0	0	0	247
medium	161	medium	medium	2....	2....	1	0....	0....	0....	2	6.8	41	0	0	0	0	255
medium	162	medium	medium	2....	2....	1	7....	0....	0....	5	6.8	38	0	0	0	0	255
medium	163	medium	medium	2....	2....	1	0....	0....	0....	3	6.6	45	0	0	0	1	250
medium	164	medium	medium	2....	2....	1	8....	8....	0....	3	6.5	36	0	0	0	0	247
medium	165	medium	medium	2....	2....	1	0....	0....	0....	4	6.6	38	0	0	0	0	255
medium	166	medium	medium	2....	2....	1	9....	0....	0....	4	6.5	37	0	0	0	0	255
medium	167	medium	medium	2....	2....	1	0....	0....	0....	3	6.7	45	1	1	0	0	247
medium	168	medium	high	2....	2....	1	0....	0....	0....	2	6.7	45	1	1	1	0	245
medium	169	medium	medium	2....	2....	1	0....	0....	0....	4	6.8	45	0	0	1	0	255
medium	170	medium	high	2....	2....	1	0....	0....	0....	2	6.5	38	1	0	1	0	255
medium	171	medium	medium	2....	2....	1	0....	0....	0....	1	6.8	45	0	0	0	1	255
medium	172	medium	medium	2....	2....	1	0....	0....	0....	3	6.5	38	1	0	0	0	255
medium	173	medium	medium	2....	2....	1	0....	0....	0....	5	6.8	40	1	0	1	0	245
medium	174	medium	medium	2....	2....	1	5....	0....	0....	2	6.5	37	0	0	0	0	255
medium	175	medium	medium	2....	2....	1	0....	0....	0....	1	6.7	45	1	1	0	0	247
medium	176	medium	high	2....	2....	1	0....	0....	0....	2	6.7	45	1	1	1	0	245
medium	177	medium	medium	2....	2....	1	0....	0....	0....	1	6.8	45	0	0	1	0	255
medium	178	medium	high	2....	2....	1	0....	0....	0....	5	6.5	38	1	0	1	0	255
medium	179	medium	medium	2....	2....	1	0....	0....	0....	5	6.8	45	0	0	0	1	255
medium	180	medium	medium	2....	2....	1	0....	0....	0....	2	6.5	38	1	0	0	0	255
medium	181	medium	medium	2....	2....	1	0....	0....	0....	3	6.8	40	1	0	1	0	245
medium	182	medium	medium	2....	2....	1	0....	0....	0....	4	6.5	36	0	0	1	0	255
medium	183	medium	medium	2....	2....	1	0....	0....	0....	3	6.5	35	1	0	1	0	246
medium	184	medium	medium	2....	2....	1	4....	0....	0....	2	6.8	34	0	0	0	1	240

Figure 5. "medium" classification results and attribute data obtained with AdaBoost and Neural Network algorithms

The "low" classification results and attribute data of Milk Quality data with AdaBoost and Neural Network algorithm using Orange Data Mining Tool are shown on Figure 6.

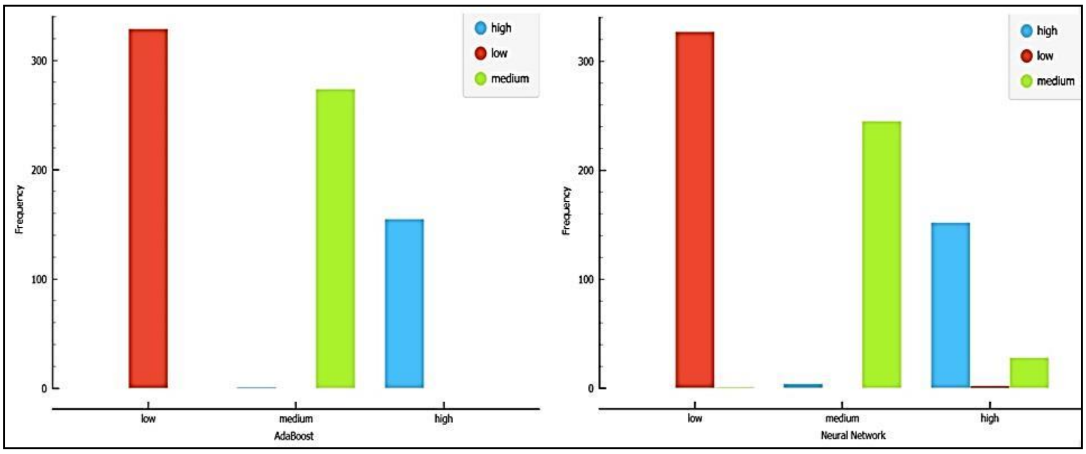
The "low" actual values are shown in the Grade field, and the predicted classification estimation results are shown in the AdaBoost and Neural Network field. In the example with the Id number 747 with the Neural Network algorithm, the quality result that should have been "low" was incorrectly predicted as "high".



Grade	Id	AdaBoost	Neural Network	›ost	›ost	st (n	etwc	etw‹	work	Fold	pH	Temprature	Taste	Odor	Fat	Turbidity	Colour
low	732	low	low	2....	1	2....	0....	0....	0....	4	4.5	38	0	1	1	1	255
low	733	low	low	2....	1	2....	7....	1	2....	1	8.5	70	0	0	0	0	246
low	734	low	low	2....	1	2....	1....	0....	8....	4	7.4	65	0	0	0	0	255
low	735	low	low	2....	1	2....	2....	0....	4....	1	3.0	40	1	1	1	1	255
low	736	low	low	2....	1	2....	0....	0....	2....	4	8.6	55	0	1	1	1	255
low	737	low	low	2....	1	2....	0....	0....	0....	4	4.7	38	1	0	1	0	255
low	738	low	low	2....	1	2....	7....	0....	2....	5	3.0	40	1	1	1	1	255
low	739	low	low	2....	1	2....	0....	0....	0....	2	9.0	43	1	0	1	1	250
low	740	low	low	2....	1	2....	2....	0....	4....	1	3.0	40	1	1	1	1	255
low	741	low	low	2....	1	2....	0....	0....	0....	4	9.0	43	1	0	1	1	250
low	742	low	low	2....	1	2....	0....	0....	0....	4	4.7	38	1	0	1	0	255
low	743	low	low	2....	1	2....	1....	0....	2....	4	3.0	40	1	1	1	1	255
low	744	low	low	2....	1	2....	0....	0....	0....	2	9.0	43	1	0	1	1	250
low	745	low	low	2....	1	2....	0....	0....	0....	1	4.5	38	0	1	1	1	255
low	746	low	low	2....	1	2....	2....	1	4....	5	8.5	70	0	0	0	0	246
low	747	low	high	2....	1	2....	0....	0....	0....	4	6.5	37	0	1	1	1	245
low	748	low	low	2....	1	2....	9....	0....	6....	5	7.4	65	0	0	0	0	255
low	749	low	low	2....	1	2....	8....	0....	0....	3	3.0	40	1	0	0	0	255
low	750	low	low	2....	1	2....	0....	0....	0....	2	9.0	43	1	1	1	1	248
low	751	low	low	2....	1	2....	0....	0....	0....	3	6.6	50	0	0	0	1	250
low	752	low	low	2....	1	2....	8....	0....	0....	3	6.6	50	0	0	0	0	255
low	753	low	low	2....	1	2....	0....	0....	0....	2	9.0	43	1	1	1	1	248
low	754	low	low	2....	1	2....	0....	0....	0....	4	6.6	50	0	0	0	1	250
low	755	low	low	2....	1	2....	0....	0....	0....	5	9.5	34	1	1	0	1	255
low	756	low	low	2....	1	2....	0....	0....	0....	2	5.5	45	1	0	1	1	250
low	757	low	low	2....	1	2....	6....	0....	2....	3	8.1	66	1	0	1	1	255
low	758	low	low	2....	1	2....	1....	0....	4....	3	3.0	40	1	1	1	1	255

Figure 6. "low" classification results and attribute data obtained with AdaBoost Neural Network algorithm

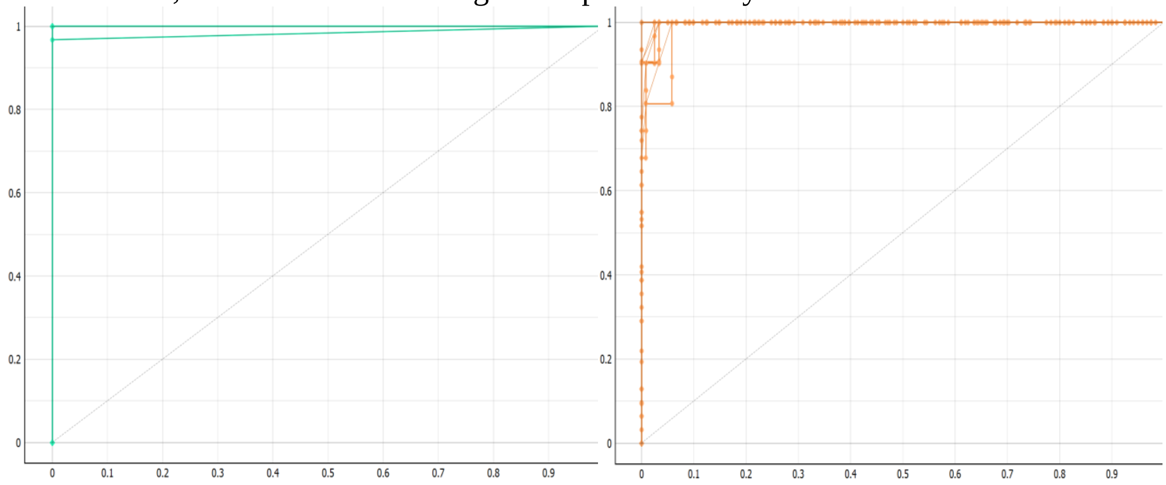
The frequencies of classification distributions of AdaBoost and Neural Network algorithms are shown on figure 7. Blue color indicates "high", Green color indicates "medium" and Red color indicates "low" classification. When the results are compared, it is seen that there is only a very small amount of error in the "medium" classification in the AdaBoost algorithm. However, it is seen that there are incorrect results in all classification predictions with the Neural Network algorithm.



(a) (b)

Figure 7. Classification distributions and frequencies of machine learning algorithms (a) AdaBoost algorithm classification prediction distribution (b) Neural Network algorithm classification prediction distribution

The stability graphs of the algorithms according to the classification estimation results are shown in Figure 8. It is seen that the Adaboost algorithm becomes stable in a very short time. However, the Neural Network algorithm passed steady state slower.



(a) (b)

Figure 8. Classification ROC curves of machine learning algorithms (a) ROC curve of AdaBoost Algorithm (b) ROC curve of Neural Network Algorithm

The complexity matrix value graphs of the algorithms according to the classification estimation results are shown on Figure 9. With the Adaboost algorithm, only one data belonging to the "high" class was incorrectly predicted as the "medium" class. However, with the Neural Network algorithm, 4 data belonging to the “high” class were incorrectly predicted as “medium”, 28 data belonging to the “medium” class were incorrectly predicted as “high” and 2 data belonging to the “low” class were also “high” was incorrectly predicted.

Actual	high	low	medium	∑
high	155	0	1	156
low	0	329	0	329
medium	0	0	274	274
∑	155	329	275	759

(a)

Actual	high	low	medium	∑
high	152	0	4	156
low	2	327	0	329
medium	28	1	245	274
∑	182	328	249	759

(b)

Figure 9. Machine learning algorithms, classification Complexity Matrix values (a) Complexity matrix of AdaBoost Algorithm (b) Complexity matrix of Neural Network Algorithm

The metric values of the results obtained with both algorithms are shown on table 2. Area Under Curve (AUC), Classification accuracy (CA), F1 score, Precision and Recall parameters were used as metric values. With the AdaBoost algorithm, 99.9% accuracy estimation was achieved in all metric values. The highest success rate with the Neural Network algorithm was obtained with the AUC metric value, but the CA parameter was used as the success metric in this study. As a CA parameter, 99.9% success rate was obtained with the AdaBoost algorithm and 95.4% with the Neural Network algorithm.

Table 2. Metric success rates of algorithms

Model Name	AUC	CA	F1	Precision	Recall
AdaBoost	0.999	0.999	0.999	0.999	0.999
Neural Network	0.997	0.954	0.955	0.959	0.954

CONCLUSION

On the basis of machine learning algorithms, artificial intelligence applications are used efficiently in many sectors. Systems developed by using machine learning algorithms facilitate human life and obtain reliable results. The use of these systems in the food industry will be beneficial. The use of machine learning systems will be beneficial, especially for the quality determination of products with sufficient data attributes.

In this study, it has been shown that milk quality can be determined by using machine learning algorithms. AdaBoost and Neural Network algorithms were used as machine learning algorithms. Milk Quality, which was downloaded from Kaggle storage, was used as the data set. There are 1059 milk data samples from this dataset. In the data set, 7 features of each sample were used. As features, pH, Temperature, Taste Odor, Oil, Turbidity and Color data were used. Orange Data Mining Tool was used as the application software interface.

According to the results obtained, in the system trained using 100 samples belonging to each class, 99.9% success rate with AdaBoost algorithm, 95.4% with Neural Network algorithm, classification accuracy were determined. The results are presented visually and compared by giving the ROC curve, complexity matrix and metric values. This study has shown that machine learning algorithms will provide high accuracy success in determining the quality of dairy products.

REFERENCES

Albanese D., Visintainer R., Merler S., Riccadonna S., Jurman G. & Furlanello C. 2012. mlpy: *Machine Learning Python*. arXiv:1202-6548v2

Balaban M.E. & Kartal E. 2018. *Veri madenciliği ve makine öğrenmesi temel algoritmaları ve R Dili ile Uygulamalar*, 2. Basım, Çağlayan Kitap & Yayıncılık & Eğitim, İstanbul, Türkiye, pp. 48-72.

Chung J.R., Kwon J. & Choe Y. 2009. Evolution of recollection and prediction in neural networks, *International Joint Conference on Neural Networks*. Atlanta,14-19 June 2009, pp. 571-577.

Çelik A. 2022. Improving Iris Dataset Classification Prediction Achievement by Using Optimum k Value of kNN Algorithm, *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 3(2), 23-30.

Demsar J., Curk T., Erjavec A., Gorup C., Hocevar T., Milutinovic M., Mozina M., Polajnar M., Toplak M., Staric A., Stajdohar M., Umek L., Zagar L., Zbontar J., Zitnik M. & Zupan B. 2013. Orange: Data Mining Toolbox in Python, *Journal of Machine Learning Research*, 14, 2349−2353.

Kaggle Inc. 2022. Kaggle. https://www.kaggle.com/general (accessed 10 July 2022).

Kanawaday A. & Sane A. 2017. Machine learning for predictive maintenance of industrial machines using IoT sensor data, *8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, 24-26 November 2017, pp. 87-90.

Keshava M., Mariyam S. J., Meghamala M., Meghashree M., Neha R. 2019. Automatized Food Quality Detection and Processing System Using Neural Networks. *4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT),* Bangalore, 17-18 May 2019, pp. 1442-1446.

Kumar M., Gupta S., Gao X.Z. & Singh A. 2019. Plant Species Recognition Using Morphological Features and Adaptive Boosting Methodology, *in IEEE Access*. 7, 163912-163918.

Küçük R. & Tapkı N. 2020. Hatay İlinde Süt ve Süt Ürünleri Üreten İşletmelerin Üretim ve Pazarlama Yapısı, *Hayvan Bilimi ve Ürünleri Dergisi*, 3 (2), 104-119.

University of Ljubljana. 2022. Orange Data Mining. https://orangedatamining.com/, (accessed 12 Aug 2022).

Ozkan I.A., Koklu M. & Saraçoğlu R. 2021. Classification of Pistachio Species Using Improved K-NN Classifier, *Progress in Nutrition*, 23(2), 1-9.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. & Louppe G. 2012. Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2825-2830.

Ricci M., Vasquez J.A.T., Turvani G., Sirena I., Casu M.R. & Vipiana F. 2021. Microwave Sensing for Food Safety: a Neural Network Implementation. *In IEEE Conference on Antenna Measurements & Applications(CAMA)*, Antibes Juan-les-Pins, 15-17 November 2021, pp. 444-447.

Sambasivam G., Amudhavel J. & Sathya G. 2020. A Predictive Performance Analysis of Vitamin D Deficiency Severity Using Machine Learning Methods, *In IEEE Access*. 8: 109492-109507.

Singh P., Kaur S., Sharma S., Sharma G., Vashisht S. & Kumar V. 2021. Malware Detection Using Machine Learning, *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, Tashkent, 10-12 November 2021, pp. 11-14.

Sun Y., Kamel M.S. & Wang Y. 2006. Boosting for Learning Multiple Classes with Imbalanced Class Distribution. *6th International Conference on Data Mining (ICDM'06)*, Hong Kong, 18-22 Dec 2006, pp. 592-602.

Tahtacı B., Canbay B. 2020. Android Malware Detection Using Machine Learning, *Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, 15-17 October 2020, pp. 1-6.

Thange U., Shukla V.K., Punhani R. & Grobbelaar W. 2021. Analyzing COVID-19 Dataset through Data Mining Tool “Orange”, *2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, 9-21 January 2021, pp. 198-203.

Titova T., Nachev V. & Damyanov C. 2018. Collective Neural Classifiers for Food Quality Applications. *ANNA '18; Advances in Neural Networks and Applications*. Bulgaria, 15-17 September 2018, pp. 1-5.

Uçan N. O., Onur O. & Albora M. 2006. *Görüntü İşleme Teknikleri ve Mühendislik Uygulamaları*, 1. Baskı, Nobel Yayın Dağıtım, Ankara, Türkiye, pp.7-25.

Wang F., Li F., He F., Wang R., Yu W. & Nie F. 2019. Feature Learning View Point of Adaboost and a New Algorithm, *In IEEE Access*, 7, 149890-149899.

Xiao L., Xia K. & Tian H. 2019. Research on Classification Model of Fermented Milk Quality Control Based on Data Mining. *International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Shanghai, 21-24 November 2019, pp. 324-327.

Vaishnav D. & Rao B.R. 2018. Comparison of Machine Learning Algorithms and Fruit Classification using Orange Data Mining Tool, *3rd International Conference on Inventive Computation Technologies (ICICT).* Coimbatore, 15-16 November 2018, pp. 603-607.

Vrindavanam J., Srinath R., Shankar H.H. & Nagesh G. 2021. Machine Learning based COVID-19 Cough Classification Models - A Comparative Analysis, *5th International Conference on Computing Methodologies and Communication (ICCMC),* Erode, 08-10 April 2021, pp. 420-426.