Milk Grade-Guard: Enhancing Food Safety Through ML

Arpita Pal
School of Computer Engineering
KIIT University
Bhubaneswar, India
guddiarpita@gmail.com

G.Gayatri
School of Computer Engineering
KIIT University
Bhubaneswar, India
ggayatri611@gmail.com

Raj Nandani
School of Computer Engineering
KIIT University
Bhubaneswar, India
rnandu1102@gmail.com

Vura Surya Suprathik
School of Computer Engineering
KIIT University
Bhubaneswar, India
suprathikvura12@gmail.com

Soumyam Sharan
School of Computer Engineering
KIIT University
Bhubaneswar, India
soumyam.sharan@gmail.com

Pinaki Sankar Chatterjee
School of Computer Engineering
KIIT University
Bhubaneswar, India
pinaki.sankar.chatterjee@gmail.com

***Abstract*—In the dairy industry, the safety of the milk is crucial. On the other hand, conventional ways of evaluating quality could have drawbacks. Milk Grade-Guard is a cutting-edge strategy that uses machine learning (ML) to improve food safety.The potential of machine learning techniques to evaluate a large dataset of milk quality along the supply chain is examined in this abstract. Beyond conventional techniques, ML can spot minor signs of tainted milk that might go unnoticed otherwise. This feature can greatly enhance the evaluation of milk quality and allow for the early identification of any contamination or spoiling. Milk Grade-Guard provides a multifaceted strategy to improve food safety. The solution can eliminate human error, save costs related to manual testing, and streamline operations by automating quality assessments using machine learning. More significantly, less product waste is produced and dangerous milk is kept from reaching customers thanks to early problem detection enabled by ML-powered prediction. In order to confirm Milk Grade-Guard's function in strengthening food safety protocols within the dairy industry, more investigation and development are necessary.**

***Index Terms*—Food Safety , Milk Grading , Machine Learning , Classification models , Training and Testing**

I. Introduction

Milk, an essential component of a balanced diet, supports millions of people's livelihoods worldwide.Because it is so nutritious, milk is one of the staple meals that people eat all around the world. 3.3% protein, 5% lactose, 87% water, and 3.9% lipids are all found in milk [5]. But protecting it from farm to table is still a top priority. Despite their established nature, traditional techniques of quality assessment can have drawbacks. These restrictions may result in gaps in the complicated dairy supply chain's ability to consistently ensure food safety.The primary product of a dairy farm that provides farmers with a source of money to operate their livestock company is milk. Fresh milk that is suitable for consumption must pass a causative test that adheres to SNI guidelines. Fresh milk has several health advantages, including preventing sickness and improving cognitive function [1]. The majority

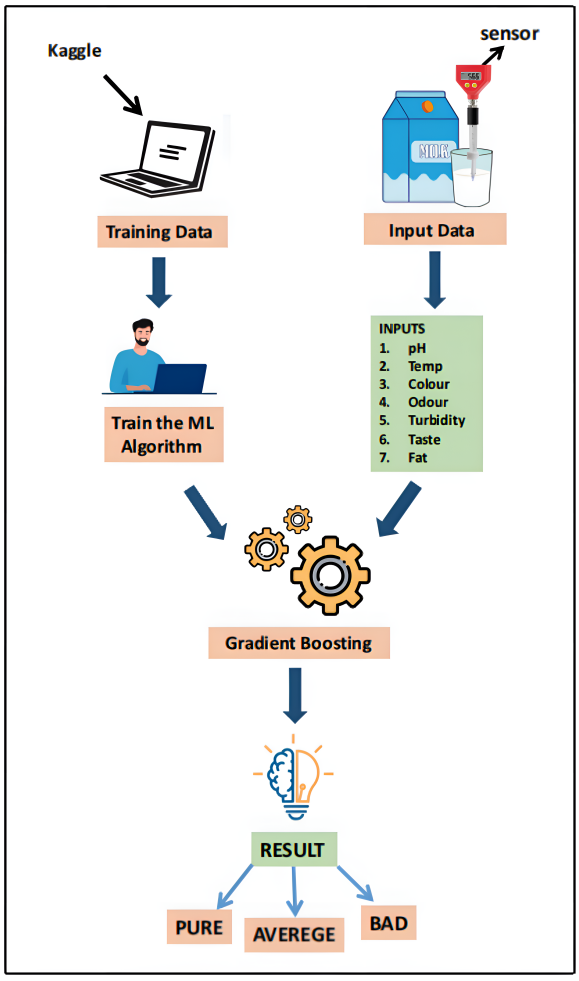

Fig. 1. Diagram Showing Milk-Grade Guard System

of current techniques depend on examining a certain set of factors, such as temperature and bacterial count.Here, sensors are used to determine a number of factors, including as pH, turbidity, and color [10]. Although efficient, this method may overlook minute irregularities that point to tainted milk. Moreover, these techniques are frequently labor-intensive and prone to human error, which causes irregularities in the process of evaluating quality.Water adulteration is a frequent occurrence in milk, which reduces its industrial production and nutritional value while also presenting a serious risk of contamination. Because tarnishing has major repercussions, adulterating milk that people in impoverished nations like India drink can have serious implications [15].A critical reaction to these constraints is Milk Grade-Guard. This project offers a cutting-edge solution that will transform dairy safety by utilizing machine learning (ML). With the use of a thorough data set analysis, Milk Grade-Guard seeks to provide a more advanced and data-driven method of spotting any contamination or spoiling.In this work, we offer a thorough analysis of many machine learning methods used to categorize milk quality [2]. In-depth discussion of Milk Grade-Guard's features and potential to greatly improve food safety in the dairy sector are provided in this paper. The system's technological design, theoretical foundations, and expected advantages for dairy supply chain participants will all be covered in the parts that follow.

This report consists of 6 important sections meant to give a thorough summary of the topic.The first section is the introduction of the project.It discuss about the precise problem or knowledge gap that the research aims to solve is indicated in the next sections. The section II explains about basic concepts required for the project. The third section provides a brief analysis of the project with all the requirements. Subsequently, the fourth section outlines the methodologies of the project, specifying what is and is not included. Section V discuss about the evaluated results obtained. At last in section VI conclusion and future scope , provide readers with a clear understanding of the project and the future scope to develop the same.

II. Basic Concepts

Using machine learning, Milk Grade-Guard re-imagines dairy safety. It compiles a significant dataset on the quality of milk at every stage of the process, from the farm to processing. Robust machine learning algorithms examine this data, gaining knowledge from previous cases of pollution. This enables them to spot minute irregularities in fresh milk samples that could be missed by conventional techniques.Milk Grade-Guard serves as an early warning system by anticipating possible problems early on, allowing for timely intervention and protecting consumers from contaminated milk.

A. Milk Grading Parameters

Our machine learning approach for assessing milk quality, Milk Grade-Guard, depends on a number of factors to reliably predict milk grade. Below is a summary of the primary parameters that will be employed:

- Fat Content: One of the most important markers of milk quality is its fat content. Deviations from the anticipated range could indicate problems with the health of the cows or water dilution.
- pH Level: The pH level of fresh milk is slightly acidic. Significant departures from this range may be a sign of faulty storage or spoiling brought on by bacterial development.
- Temperature: During storage and transit, milk should be kept within a certain range. Deviations may damage quality by hastening the growth of microorganisms.
- Color: Depending on the breed and nutrition, natural milk color can change slightly. Severe discolouration may indicate contamination or spoiling.
- Taste: While taste tests may not be feasible for a large-scale system, the machine learning model may benefit from past data on flavor assessments made by human specialists. Odd tastes may
- be a sign of infection or spoiling.
- Odor: Like taste, odor data (provided by human specialists) can be used to train models. Odd smells can indicate the presence of microorganisms or spoiling [14].

B. Machine Learning

Machine learning is a quickly emerging field of technology which enables computers to automatically learn from the historical data. Machine learning uses various algorithms to create mathematical models and forecasts based on information or historical data. Many industries use machine learning to extract important data. The goal of machine learning is to learn from data [8].Nowadays it is being used for many different things, like image recognition, recommended systems, email filtering, and speech recognition. Machine learning can be broadly classified into three types .These classification is based on the nature of the learning system and the data available . The types are as follows :

- Supervised learning:Labeled data is used to teach models how to predict outcomes.
- Unsupervised Learning: In this process, algorithms sift through unlabeled data in search of patterns or clusters.
- Reinforcement Learning: In this approach, models gain decision-making skills by acting in a way that maximizes a concept of cumulative reward [8].

C. Classification Model

The supervised ML are classified into 2 types namely Regression model and Classification model. The classification algorithms in machine learning are essential for categorizing data into predefined classes. Before using the model to make predictions on newly discovered data, it must first be thoroughly trained on training data and assessed on test data. Some popular classification algorithms include Logistic regression , Naive bayes classification , KNN , Decision trees, SVM .

1) KNN: The k-Nearest Neighbors (kNN) algorithm is a versatile machine learning technique used for classification and regression tasks. It estimates the probability of a data point

belonging to a particular group by examining the class labels of its nearest neighbors. Choosing an appropriate value for k, the number of neighbors considered, is crucial, often favoring odd values [2].

The Euclidean distance metric, commonly used in kNN, calculates the straight-line distance between the query point and other data points, forming decision boundaries that divide areas in the feature space. This distance measure, applicable to real-valued vectors, aids in effectively identifying neighboring points and classifying data.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \tag{1}$$

Overall, kNN is efficient and effective, especially in classification problems, leveraging the collective information of neighboring points to make predictions [2]. The time complexity of kNN is $O(n^2)$ [6].

2) Naive Bayes: This is a classification method based on independence of predictors and Bayes theorem. In short, the Naive Bayes classifier assumes that the presence of a given feature in a class is independent of the presence of any other feature [8].The main field that Naïve Bayes deals with is spam filtering, text classification, and recommendation systems. Depending on the conditional probability, it is mainly used for classification and clustering purposes.

The formula used for calculation in Naive Bayes is as follows:

$$P(\frac{C}{X}) = \frac{P(\frac{X}{C})P(C)}{P(X)} \tag{2}$$

as the above formula assumes the input variables to be independent .We consider,

$$P(\frac{X}{C}) = P(\frac{X1}{C})P(\frac{X2}{C})..P(\frac{Xn}{C}) \tag{3}$$

where, $(P(\frac{C}{X}))$ is the posterior probability of class (C) given predictor (X). $(P(\frac{X}{C}))$ is the likelihood which is the probability of predictor (X) given class (C). (P(C)) is the prior probability of class. (P(X)) is the prior probability of predictor [11].

3) SVM: Support Vector Machine (SVM) is a supervised learning algorithm used for classification tasks. It is used split the data into 2 or more classes by creating a hyperplane .It works by finding the optimal hyperplane that maximizes the margin between different classes. In two-dimensional space, this hyperplane is a line which divides the points into two categories. In higher dimensions, it's a plane .SVM can be linear and non-linear(by kernel trick) [8]. For creating the decision boundary , we can different SVM kernels like linear, polynomial , RBF etc. The SVM creates a hyperplane defined by the equation:

$$W.x + b = 0 \tag{4}$$

Where the , W is the weight vector. x represents the input features. b is the bias term.

The main aim is to maximize the margin to reduce the noise and error between the classes. This can be done by, max($\frac{2}{||W||}$) subject to y(w.x+b) $\geq$ 1 .

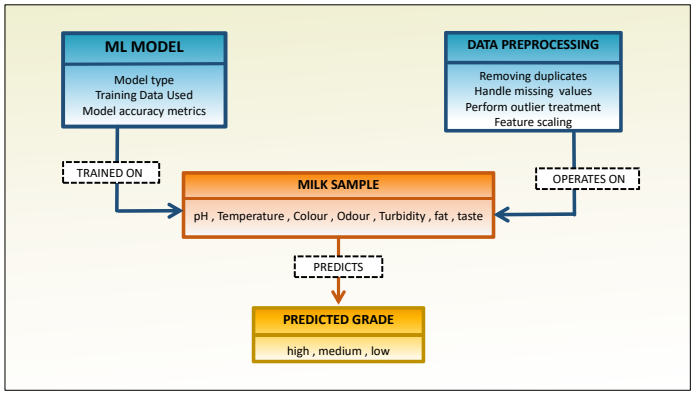

Fig. 2. UML Diagram Showing the process implemented

The memory requirements of our algorithm reach $O(|S|^2)$ when $|S|$ is the size of the support vector set [16].

4) Gradient Boosting: Gradient Boosting is a machine learning technique used for both classification tasks . It builds an ensemble of decision trees, in a sequential manner where each tree tries to correct the errors made by the previous one. The method involves training trees on the residual errors of the predecessor, effectively refining the model with each iteration . Thus is very efficient in terms of classification models .

$$F2(x) = F1(x) + n.H1(x) \tag{5}$$

where , F2(x) is the updated model after the 1st iteration. F1(x) is model at the 1st iteration. n is the learning rate H1(x) is the weak learner fitted on the gradient of the loss function at 1st iteration [3].

III. Project Analysis

For the milk grading calculation, we have undergone the following steps:

- Dataset collection- This is the first step done to determine the grade of milk. The dataset acquired through the online website "Kaggle". The dataset includes a number of variables like pH, temperature, taste, fat, turbidity, colour and odour.
- Data preprocessing-This step is done to clean and process of the data obtained. This involves removing of null values and duplicate values. This ensures the data integrity and consistency.
- Splitting data- For training and testing of the data to be successful, the whole dataset must be split precisely.
- Choose an appropriate machine learning model-Different machine learning classification algorithms like SVM, Naïve bayes, KNN, Gradient Boosting are used to analyse the milk quality.
- Training and validation- Train particular algorithms to find patterns associated with milk quality using labeled data. Thoroughly validate models to ensure they are reliable and applicable to a wider population.
- Accuracy- Select the best-fit model by comparing their accuracy. We selected the Gradient boosting algorithm and evaluated the output with respective algorithm.

TABLE I
PARAMETERS AND SENSORS

Publication	Parameters	Sensors
[13]	pH	Optical pH sensors
[15]	Temperature	DS18B20 sensors
[17]	Taste	Taste-sensing TS-5000Z system
[12]	Odour	E-noses sensors
[15]	Fat	Ultrasonic sensors
[14]	Turbidity	Turbidity sensors
[14]	Colour	TCS 3200

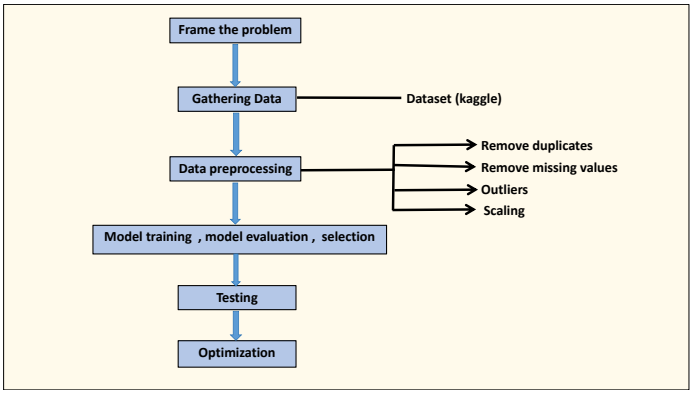

Fig. 3. Block Diagram Showing the process implemented

- Optimizing the model selected.

These steps are shown using the Fig 2.

The software that was used during the whole process are:

- Python: Used for creating machine learning models because of its many data analysis libraries, including NumPy, pandas, and scikit-learn.
- Jupyter Notebook: An interactive environment for data analysis, result visualization, and ML algorithm prototyping.
- Kaggle: to find the required dataset.

The hardware that was used during the whole process are Internet of Things (IoT) sensors are used to gather data in real-time on different parameters in milk grading, allowing for ongoing monitoring and analysis. The Table I shows different sensors with their respective parameters.

IV. METHODOLOGY

The goal of the research was to comprehend the statistical data pertaining to the examination of the milk classification. Develop machine learning models based on this study to ascertain the space's ultimate quality in a secure manner and gain insight into its formation and evolution. This research's technique is descriptive in nature and serves as an explanation. The study was both quantitative and qualitative in nature, utilizing Python programming language machine learning methods for data interpretation along with data analysis [9]. Fig 3 shows all the methods.

A. Data Collection:

1) Source: An organized milk quality dataset from Kaggle was used. The dataset ought to have multiple factors that impact the quality of milk, including pH, temperature, taste, odor, fat content, turbidity, and color. At first it goes through Data Preprocessing:

- Duplicate Removal: To avoid skewing the training process for our machine learning models, we had to remove duplicate entries from the dataset.
- Missing Value Handling: If there are few missing values, we use appropriate methods such as mean/median imputation or deletion to deal with the missing data points (null values).
- Outlier Identification and Management: We locate and address outliers, or extreme values, that may cause the underlying trends in the data to become distorted.
- Scaling: The dataset's features may be measured in several units. Normalization and standardization are scaling strategies that guarantee every feature contributes equally to the training process of the model.

B. Model Training, Evaluation, and Selection:

1) Model Training: Our dataset was partitioned into 70:30 training and testing sets. The model is trained on the training set and evaluated on unseen data using the testing set. To prepare for classification problems, we investigate a variety of machine learning algorithms.The correlations between the milk quality characteristics and the appropriate quality grades (such as "high","medium","low") found in the labeled data will be discovered using these algorithms.

2) Evaluation of the Models: To evaluate our model, we examined four algorithms: Gradient Boosting, Support Vector Machines (SVM), Naive Bayes, and K-Nearest Neighbors (KNN). We determined which technique produced the best results .Accuracy, precision, recall, and F1-score are just a few of the metrics that will be used to thoroughly assess each trained model's performance. Algorithms of the classification models used are as follow:

IMPORT libraries
LOAD dataset
SELECT features and target
SPLIT dataset into training and testing sets
PRINT sizes of the splits
INITIALIZE Classification model
FIT classifier on training data
PREDICT on test data
PRINT classification report

3) Selection: We choose the Gradient Boosting model as it performs best in terms of accuracy and dependability when it comes to determining milk quality based on the evaluation findings with accuracy of 88%.

C. Testing:

The Milk Grade-Guard system's output will be based on this model. To evaluate the chosen model's practicality, we test it on the input dataset from user.

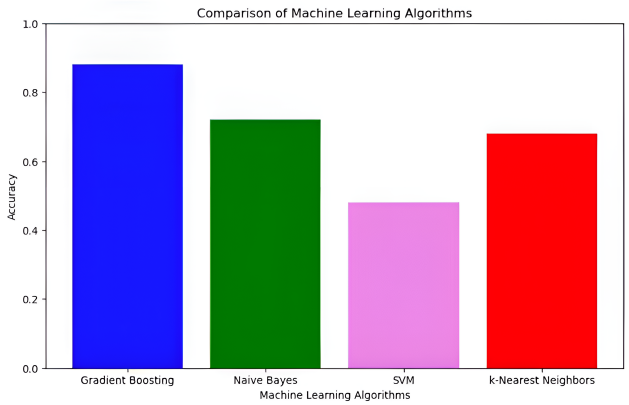

Fig. 4. Accuracy comparisons among Gradient boosting , Naive Bayes , SVM and k-Nearest Neighbours

```
[138]: pH = float(input("Input pH: "))
       temp = float(input("Input the temperature: "))
       taste = float(input("Input taste: "))
       odor = float(input("Input odor: "))
       fat = float(input("Input fat: "))
       turb = float(input("Input turbidity: "))
       color = float(input("Input colour: "))

       Input pH:  6.6
       Input the temperature:  50
       Input taste:  0
       Input odor:  0
       Input fat:  0
       Input turbidity:  0
       Input colour:  255

[139]: input_data = [pH, temp, taste, odor, fat, turb, color]
       prediction = GBC.predict([input_data])
       if prediction[0] == 2:
           print("Milk is Pure")
       elif prediction[0] == 1:
           print("Milk is Average")
       else:
           print("Milk is Bad")

       Milk is Bad
```

Fig. 5. Milk quality prediction

V. Evaluation

The accuracy of all the models are as follows:
KNN - 68%
SVM - 55.56%
Naive Bayes - 72%
Gradient Boosting - 88%

Fig 4 shows the comparison of the accuracy among the models used.This graph clearly concludes that Gradient Boosting(0.88) has the highest accuracy among all. Thus , this is the best fit model.

Fig 5 shows the prediction of quality of milk by taking the set of input from users. Here the model is being tested over a particular set of data.

Fig 6 Graph is used for illustrating the relationship between the actual values (y_test) and the predicted values (y_pred).

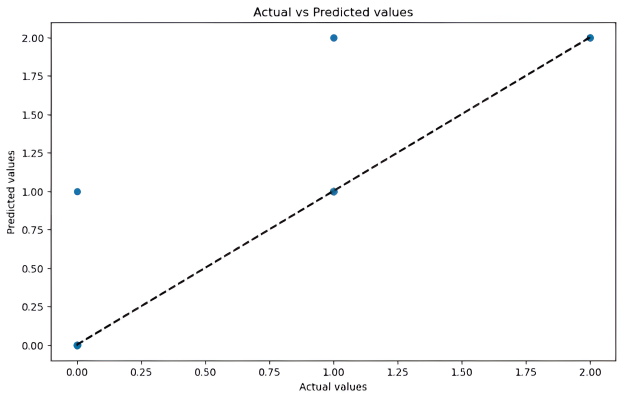

Fig. 6. Graph between the actual and the predicted values of the model

VI. Conclusion and Future Scope

Machine learning-based Milk Grade Guard (MGG) has the potential to improve milk safety. However, further studies and practical testing are required to validate its efficacy.A machine learning system called Milk Grade Guard has the potential to improve food safety in the milk manufacturing process. According to the study, Milk grade guard is capable of efficiently analyzing a variety of data points to forecast any problems with quality or dangers of contamination. This enables prompt action, which may lessen the likelihood that tainted milk will be consumed by customers.In large-scale dairy operations, it is possible to anticipate the quality of milk pretty precisely by comparing it to past samples [14].The caliber and thoroughness of the training data utilized to create the ML models would determine how effective Milk grade guard would be.To confirm grading performance in real-world scenarios across various dairy farms and milk production processes, more study is required.Numerical research identifying various techniques of milk production and fat levels has shown the great efficacy of the proposed strategy [10].A smooth installation of Milk grade guard would need integration with current quality control systems.A notable advancement in proactive milk safety is provided by Milk Grade Guard. Through the examination of several data points, including as sensor readings, farm management techniques, and past quality data, Milk grade guard may be able to forecast variations in milk quality and identify possible contamination hazards at an early stage of the production process. The early warning system has the potential to greatly lower the likelihood of contaminated milk reaching consumers, improving public health outcomes and minimizing financial losses for the dairy sector. Nonetheless, a few crucial factors are still in play. The caliber and comprehensiveness of the training data utilized to create Milk grade guard machine learning models determine how effective the system is. Predictions that aren't correct might be caused by data biases. Furthermore, to guarantee the generalization of Milk grade guard, real-world validation across several dairy farms with different production techniques is essential. Furthermore, the deployment would not be ef-

fective without a smooth interaction with the current quality control systems.A overview of machine learning algorithms is presented in this study. Whether they realize it or not, everyone uses machine learning these days [8]. Milk grade guard should supplement existing safety procedures, not take their place.All things considered, Milk grade guard is a potential technical advancement toward a more proactive approach to milk safety, even though it still needs additional research and testing. With the right technological advancements and system integration, Milk grade guard has the ability to completely transform the dairy business and improve food safety and productivity.

The future developments to this project can be as follows:

- Expansion to other dairy products : By Extending this algorithm , we can help to monitor and improve the quality of other dairy products like cheese , cottage cheese(paneer) , yogurt etc .
- Real-time analysis : by implementing the real-time monitoring system using this algorithm can help to provide us with instant feedback and monitoring of dairy farms .
- Consumer apps : By linking the project and extending it with apps ,so as to help the consumer to verify the milk they are consuming is healthy or not .
- Advanced analysis : By integrating with much more sophisticate machine learning algorithms ,we can also predict the number of days the current "good" quality milk will turn into "bad" milk.
- Smart Dairy Farming: One of the most fascinating and difficult areas of cloud-based data analytics is now smart dairy farming. Since apps are producing more data and rural farms lack internet access, it is presently not possible to transfer raw data from all farms to a central cloud. As a remedy, this can be extended for processing data close to the farm and extracting insights from it through data sharing between on-farm apps and cloud data transfers [7].
- Predictive Intelligence System for Milk Quality: The purpose of this study was to create an intelligent system for pasteurized milk quality prediction and assessment that would aid in the evaluation and prediction of pasteurized milk quality by quality decision makers [4].

These are some of the future advancements that can be done to enhance the "Milk grade- guard" project. This will surely help to become an integral part of dairy industry's push towards innovation and sustainability.

REFERENCES

[1] Habsari.W , Udin.F and Arkeman.Y "An analysis and design of fresh milk smart grading system based on internet of things." IOP Conference Series: Earth and Environmental Science. Vol. 1063. No. 1. IOP Publishing, 2022.

[2] S. Kumari, M. K. Gourisaria, H. Das and D. Banik, "Deep Learning Based Approach for Milk Quality Prediction," 2023 11th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India, 2023, pp. 1-6, doi: 10.1109/ICETET-SIP58143.2023.10151626.

[3] Bentéjac, Candice , Csörgő, Anna and Martínez-Muñoz, Gonzalo. "A Comparative Analysis of XGBoost".(2019)

[4] Septiani, Winnie, and Tatit K. Bunasor. "Intelligent System for Pasteurized Milk Quality Assessment and Prediction." (2007).

[5] Ganesh Kumar, P., A. Alagammai, B. S. Madhumitha, and B. Ishwariya. "IoT based milk monitoring system for the detection of milk adulteration." ESP J Eng Technol Adv 2, no. 2 (2022): 6-9.

[6] Guo, Gongde, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. "KNN model-based approach in classification." In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, pp. 986-996. Springer Berlin Heidelberg, 2003.

[7] Vimalajeewa, Dixon, Chamil Kulatunga, and Donagh P. Berry. "Learning in the compressed data domain: Application to milk quality prediction." Information Sciences 459 (2018): 149-167.

[8] Mahesh, Batta. "Machine learning algorithms-a review." International Journal of Science and Research (IJSR).[Internet] 9, no. 1 (2020): 381-386.

[9] Cunha, Matheus Henrique Lopes, and Hygor Santiago Lara. "Machine learning as a way to have more accuracy when defining milk quality classification." Caderno de ANAIS HOME (2023).

[10] Bhavsar, Drashti, Yash Jobanputra, Nirmal Keshari Swain, and Debabrata Swain. "Milk Quality Prediction Using Machine Learning." EAI Endorsed Transactions on Internet of Things 10 (2024).

[11] Rish, Irina. "An empirical study of the naive Bayes classifier." In IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22, pp. 41-46. 2001.

[12] Yang, Yongheng, and Lijuan Wei. "Application of E-nose technology combined with artificial neural network to predict total bacterial count in milk." Journal of Dairy Science 104, no. 10 (2021): 10558-10565.

[13] Voskoboynikova, Olga, Aleksey Sukhanov, and Axel Duerkop. "Optical pH sensing in milk: A small puzzle of indicator concentrations and the best detection method." Chemosensors 9, no. 7 (2021): 177.

[14] Kumar, JNVR Swarup, D. N. V. S. L. S. Indira, Kalyanapu Srinivas, and MN Satish Kumar. "Quality assessment and grading of milk using sensors and neural networks." In 2022 International Conference on Electronics and Renewable Systems (ICEARS), pp. 1772-1776. IEEE, 2022.

[15] Saravanan, S., M. Kavinkumar, N. S. Kokul, N. S. Krishna, and V. I. Nitheeshkumar. "Smart milk quality analysis and grading using IoT." In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 378-383. IEEE, 2021.

[16] Vishwanathan, S. V. M., and M. Narasimha Murty. "SSVM: a simple SVM algorithm." In Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290), vol. 3, pp. 2393-2398. IEEE, 2002.

[17] Toko, Kiyoshi. "Research and development of taste sensors as a novel analytical tool." Proceedings of the Japan Academy, Series B 99, no. 6 (2023): 173-189.