Milk Grade-Guard: Enhancing Food Safety Through ML

by VURA SURYA SUPRATHIK

Submission date: 10-Apr-2024 10:54AM (UTC+0530)
Submission ID: 2345341198
File name: Report-format_6.doc (691.02K)
Word count: 4160
Character count: 22904

A PROJECT REPORT

on

"Milk Grade-Guard: Enhancing Food Safety Through ML"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

Vura Surya Suprathik	2105340
Arpita Pal	2105359
G. Gayatri	2105368
Raj Nandani	2105397
Soumyam Sharan	2105411

Under the guidance of
PINAKI SANKAR CHATTERJEE

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

April 2024

A PROJECT REPORT

on

"Milk Grade-Guard: Enhancing Food Safety Through ML"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

Vura Surya Suprathik	2105340
Arpita Pal	2105359
G.Gayatri	2105368
Raj Nandani	2105397
Soumyam Sharan	2105411

UNDER THE GUIDANCE OF
PINAKI SANKAR CHATTERJEE

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA -751024

April 2024

KIIT Deemed to be University
School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

“Milk Grade-Guard: Enhancing Food Safety Through ML”

submitted by

Vura Surya Suprathik	2105340
Arpita Pal	2105359
G. Gayatri	2105368
Raj Nandani	2105397
Soumyam Sharan	2105411

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under our guidance.

Date: 09/04/2024

Pinaki Sankar Chatterjee

Acknowledgements

We are profoundly grateful to **Pinaki Sankar Chatterjee** of **Affiliation** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. His timely counsel, constructive criticism, and encouragement have been invaluable in helping us mold our project and steer it in the direction of a fruitful completion. His breadth of experience and knowledge has been a great asset to our project. Furthermore, we would like to thank Dr. Pinaki Sankar Chatterjee for his efforts in giving us the resources and assistance for the project. Without his direction, our project would not have been possible to complete. We are grateful for the opportunity to collaborate with Dr. Pinaki Sankar Chatterjee and his willingness to share his knowledge and expertise with us. He has been an incredible teacher and we have learned so much from him. We would like to thank Dr. Pinaki Sankar Chatterjee again for his invaluable counsel and support during this research.

Vura Surya Suprathik
Arpita Pal
G.Gayatri
Raj Nandani
Soumyam Sharan

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.

Keywords: Food Safety , Milk Grading , Machine Learning , Classification models , Training and Testing

Contents

List of Figures

Chapter 1

Introduction

Milk, an essential component of a balanced diet, supports millions of people's livelihoods worldwide. 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 majority of current techniques depend on examining a certain set of factors, such as temperature and bacterial count. 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.

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 dataset analysis, Milk Grade-Guard seeks to provide a more advanced and data-driven method of spotting any contamination or spoiling. 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.

Chapter 2

Basic Concepts/ Literature Review

Using machine learning, Milk Grade-Guard reimagines 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.

2.1 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.
Turbidity: Clarity-impairing contaminants might alter the quality of milk. Turbidity readings are used by Milk Grade-Guard into its machine learning model to help it forecast milk grade more accurately.

2.2 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. Nowadays it is being used for many different things, like image recognition, recommender 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, unsupervised learning, and reinforcement learning

- 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.

2.3 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 .

2.4 KNN

The KNN algorithm is commonly employed as a classification algorithm, based on the premise that comparable points can be located next to one another, while it can be used for regression or classification problems as well.

The distance between a query point and the other data points must be computed in order to ascertain which data points are closest to a particular query point. The decision boundaries that divide query points into various areas are formed in part by the below distance measurement formula.

Distance Euclidean (p=2):This distance metric is restricted to real-valued vectors and is the most widely used one. It measures a straight line between the query location and the other point being measured using the formula below.

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

2.5 Naive Bayes

Naive Bayes Classifier algorithm calculates conditional probabilities .This algorithm is based on Bayes' Theorem with the assumption of independence among input variables . It is known for high efficiency in handling large datasets . It is used widely in applications like spam filtering, text classification, and recommendation systems.

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

P(C|X) = P(X|C)P(C) / P(X)

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

P(X|C)=P(X1|C) P(X2|C)...P(Xn|C)

(P(C|X)) is the posterior probability of class (C) given predictor (X).

(P(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.

2.6 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 . For creating the decision boundary , we can different SVM kernels like linear, polynomial , RBF etc.

The SVM creates a hyperplane defined by the equation:

$w.x+b=0$

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, $\max(2/\|w\|)$ subject to $y(w.+b) >= 1$

2.7 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 .

$F2(x) = F1(x) + n. H1(x)$ 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.

Chapter 3

Problem Statement / Requirement Specifications

For the sake of consumer confidence and public health, milk product quality and safety must always be guaranteed. However, human examination and subjective assessment are frequently used in conventional techniques of evaluating milk quality, which leaves opportunity for error and inconsistent results. In addition, the complexity of the supply chain and the rising demand for dairy products make it difficult to uphold strict quality control standards.

The problem statement in this instance centers on the requirement for a dependable and efficient system to improve food safety in the dairy sector through the application of machine learning (ML) technologies. Creating a Milk Grade-Guard system that can reliably and effectively evaluate milk quality at every stage of production and distribution is the main problem.

3.1 Project Planning

The essential phases in project planning are broken down as follows:

3.1.1. Data Management and Acquisition:

A. Data Source: The dataset was acquired via Kaggle, includes a number of variables including pH, temperature, taste, odor, fat, turbidity, and color that are used to determine the grade of milk.

B. Data Preprocessing: Strict preprocessing and cleaning methods, such as removing null and duplicate values from the dataset, were used to guarantee data integrity and consistency.

C. Splitting Data: Precise splitting of the dataset is necessary for training and testing of the data to be successful.

3.1.2. Model Development:

A. Choose an appropriate machine learning algorithm: To analyze the milk quality , ML algorithms such as SVM, Naive baye's , KNN and Gradient boosting were used.

B. Training and Validation: Using labeled data, train specific algorithms to identify patterns related to milk quality. Make sure models are accurate and generalizable by thoroughly validating them.

C. Accuracy: On the basis of accuracy, we selected the preferred model which is Gradient boosting and evaluated the output with respective algorithm.

3.2 Project Analysis

After evaluating the accuracy of all the ML algorithms used , we found that Gradient Boosting is the best-fit model for the project.

3.3 System Design

3.3.1 Design Constraints

Software :

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

Hardware:

Sensor Devices: 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.

PARAMETERS	SENSORS
1)pH	Optical pH sensors
2)temperature	DS18B20 sensor
3)Taste	taste-sensing TS-5000Z system
4)Odour	E-noses sensors
5)Fat	Ultrasonic Sensor
6)Turbidity	TS/300B sensor
7)color	TCS 3200

3.3.2 System Architecture OR Block Diagram

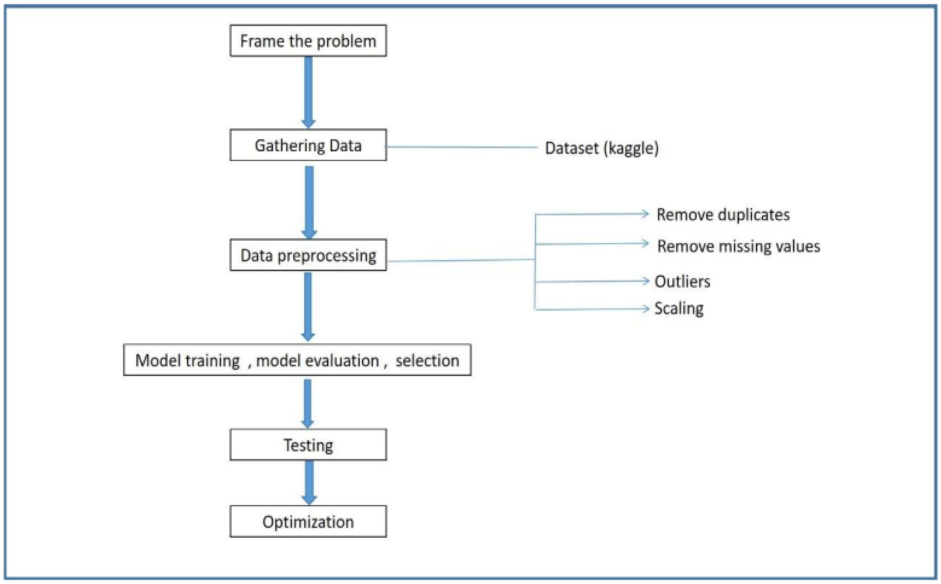

Fig1 : Block Diagram Showing the process implemented

Chapter 4

Implementation

4.1 Methodology OR Proposal

4.1.1. Data Collection:

Source:
A 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.

Data Preprocessing:

i.Duplicate Removal: To avoid skewing the training process for our machine learning models, we had remove duplicate entries from the dataset.

ii. 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).

iii. Outlier Identification and Management: We locate and address outliers, or extreme values, that may cause the underlying trends in the data to become distorted.

iv. 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.

4.1.2. Model Training, Evaluation, and Selection:

Model Training:
Our dataset was partitioned into 90:10 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 and 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.

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:

SVM

\IMPORT libraries

LOAD dataset
SELECT features and target

SPLIT dataset into training and testing sets

INITIALIZE SVM with RBF kernel
SET parameters gamma and C
FIT SVM on training data

PREDICT using SVM on test data

CALCULATE accuracy and F1 score
PRINT performance metrics

NAIVE BAYES

IMPORT libraries
SUPPRESS warnings

STANDARDIZE features

SPLIT dataset into training and testing sets
PRINT shapes of the splits

INITIALIZE Gaussian Naive Bayes classifier
FIT classifier on training data

PREDICT on test data
CALCULATE accuracy

PRINT accuracy and classification report

KNN

IMPORT libraries

LOAD dataset
SELECT features and target

SPLIT dataset into training and testing sets

INITIALIZE KNN classifier with specified number of neighbors
FIT KNN on training data

PREDICT using KNN on test data

CALCULATE confusion matrix and classification report
PRINT evaluation metrics

GRADIENT BOOSTING

IMPORT libraries

LOAD dataset
SELECT features and target

SPLIT dataset into training and testing sets
PRINT sizes of the splits

INITIALIZE Gradient Boosting Classifier
FIT classifier on training data

PREDICT on test data

PRINT classification report

Model 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. The Milk Grade-Guard system's output will be based on this model. The accuracy of all the models are as follows:
KNN - 68%
SVM - 55.56%
Naive Bayes - 72%
Gradient Boosting - 88%

4.1.3. 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.

4.2 Result Analysis OR Screenshot

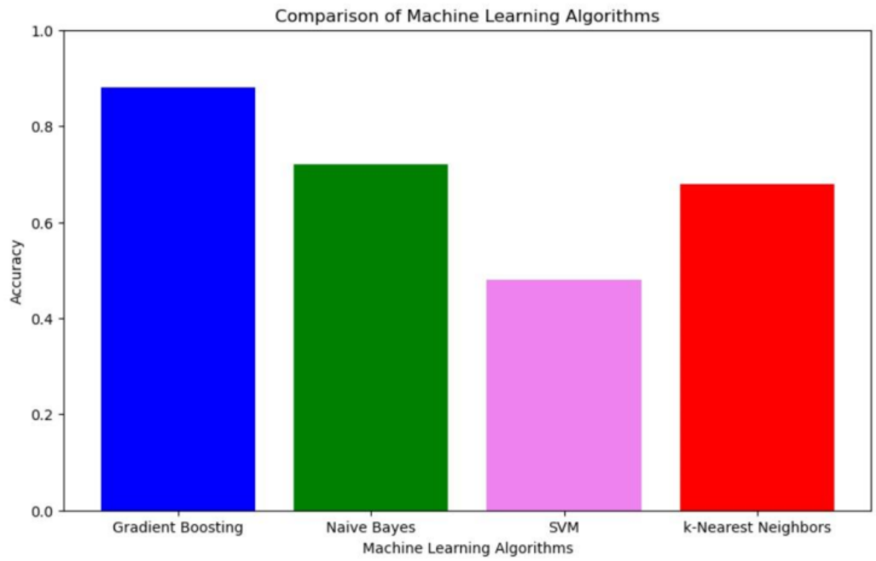

Fig 2: Accuracy comparisons among Gradient boosting , Naive Bayes , SVM and k-Nearest Neighbours

This 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.

```
[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 3:Milk quality prediction

This 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.

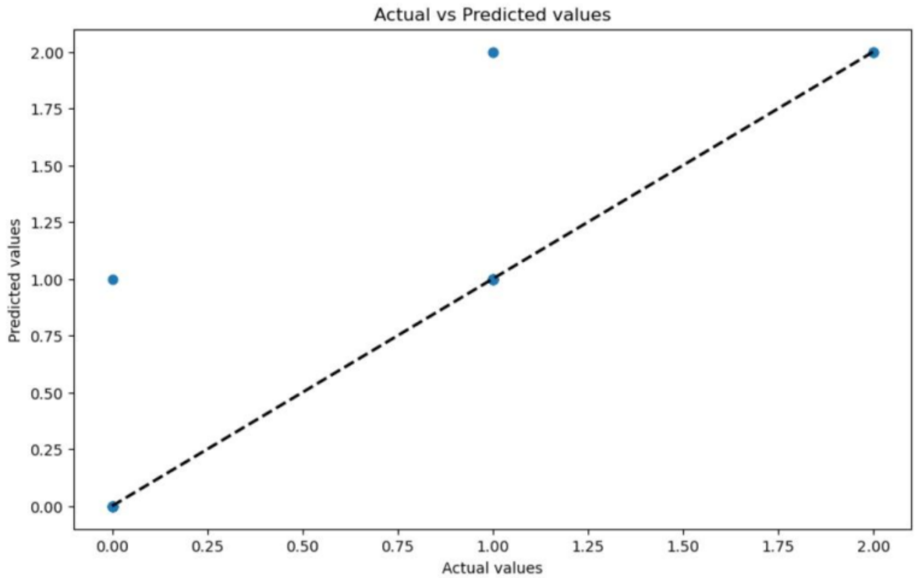

Fig 4: Graph between the actual and the predicted values of the model

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

Chapter 5

Standards Adopted

5.1 Design Standards

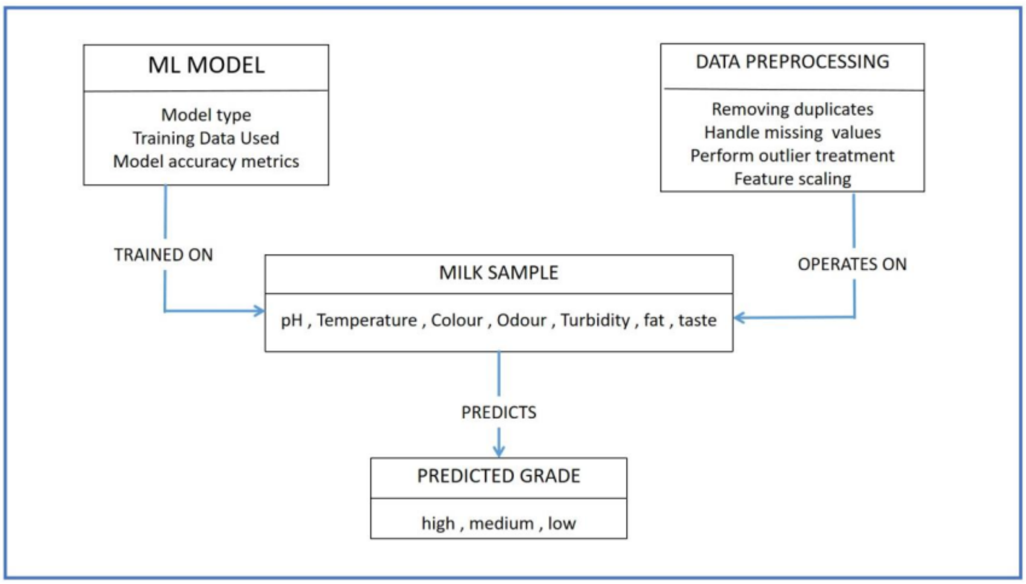

Fig 5: UML diagram showing the project

5.2 Coding Standards

26

Naming Conventions:
Use descriptive names for variables, functions, and modules in accordance with naming conventions.
Choose names that are meaningful and convey the variable or function's purpose.

Comments:
Comments help other developers understand the code better.
Use comments to add context to the code or to clarify complex logic.

Formatting:
Use consistent indentation.
Maintain consistent and clear formatting across the codebase.

Version Control:
Track code changes via a version control system, such as Git.

Evaluation Metrics:
Choose relevant evaluation measures for your problem.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

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.
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.
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 generalizability of Milk grade guard, real-world validation across several dairy farms with different production techniques is essential. Furthermore, the deployment would not be effective without a smooth interaction with the current quality control systems. 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.

6.2 Future Scope

As we all know , food safety is a very critical aspect of public health . It is very necessary for protecting consumers from hazards like microbiological contamination , chemicals and other toxins.
Through this project , we contributed to the betterment of public health by creating an algorithm that can test the given sample of milk and catagorize it into good , bad , average milk.
This project has a very promising future scope . It can potentially revolutionize in the dairy sector and ensure high standard milk are served to consumers.

The future developments to this project can be as follows:

a)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 .

b) 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 .

c)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 .

d)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.

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.

Milk Grade-Guard: Enhancing Food Safety Through ML

ORIGINALITY REPORT

19%	15%	5%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to KIIT University Student Paper	4%
2	Submitted to Banaras Hindu University Student Paper	3%
3	www.coursehero.com Internet Source	2%
4	Submitted to University of Hull Student Paper	1%
5	Submitted to Liverpool John Moores University Student Paper	1%
6	Submitted to University of KwaZulu-Natal Student Paper	1%
7	fastercapital.com Internet Source	1%
8	Ton Duc Thang University Publication	1%
9	Saumya Kumari, Mahendra Kumar Gourisaria, Himansu Das, Debajyoty Banik. "Deep	1%

Learning Based Approach for Milk Quality Prediction", 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), 2023
Publication

10	Submitted to University of Hertfordshire Student Paper	1 %
11	journal.uod.ac Internet Source	1 %
12	assets.researchsquare.com Internet Source	1 %
13	cdn-cms.f-static.net Internet Source	<1 %
14	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1 %
15	insideaiml.com Internet Source	<1 %
16	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
17	www2.mdpi.com Internet Source	<1 %
18	www.frontiersin.org Internet Source	<1 %

19	www.ijraset.com Internet Source	<1%
20	Submitted to Jacobs University, Bremen Student Paper	<1%
21	Submitted to National Institute of Technology Warangal Student Paper	<1%
22	Pablo Ferri Borredà. "Deeep Continual Multimodal Multitask Modells for Out-of-Hospital Emergency Medical Call Incidents Triage Support in the Presence of Dataset Shifts", Universitat Politecnica de Valencia, 2024 Publication	<1%
23	www.science.gov Internet Source	<1%
24	dokumen.pub Internet Source	<1%
25	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
26	truelearn.readthedocs.io Internet Source	<1%
27	www.diva-portal.org Internet Source	<1%