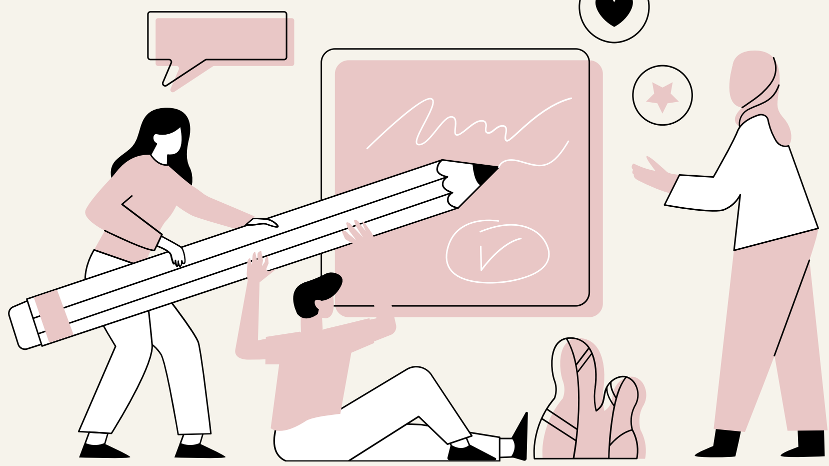
“MILK GRADE-GUARD: ENHANCING FOOD SAFETY THROUGH ML”

Submitted to
KIIT Deemed to be University

Presented By :

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

CONTENTS



INTRODUCTION

Presenting Milk Grade-Guard: an innovative solution leveraging machine learning to revolutionize dairy safety. Traditional methods fall short in ensuring food safety due to limitations and human error. Our approach, powered by comprehensive dataset analysis, promises enhanced contamination detection. Explore how Milk Grade-Guard elevates dairy quality assurance, securing a safer and more reliable supply chain.

BASIC CONCEPTS

Milk Grading Parameters

Milk Grade-Guard, our machine learning system, uses various factors to predict milk quality accurately. Here are the main things we look at:

Fat Content :	An essential marker, deviations may signal cow health issues or water dilution, impacting milk quality.
pH Level :	Fresh milk's slight acidity can indicate proper storage; significant changes suggest bacterial spoilage.
Temperature :	Maintaining optimal storage temperatures prevents microbial growth, preserving milk quality.

Color	:	Natural variations are normal, but severe discoloration may indicate contamination or spoilage.
Taste	:	Unusual flavors could signal contamination or spoilage, though large-scale taste testing may not be feasible.
Odor	:	Like taste, unusual smells detected by specialists may indicate microbial presence or spoilage.
Turbidity	:	Monitoring clarity helps detect contaminants, contributing to more accurate milk grade predictions.

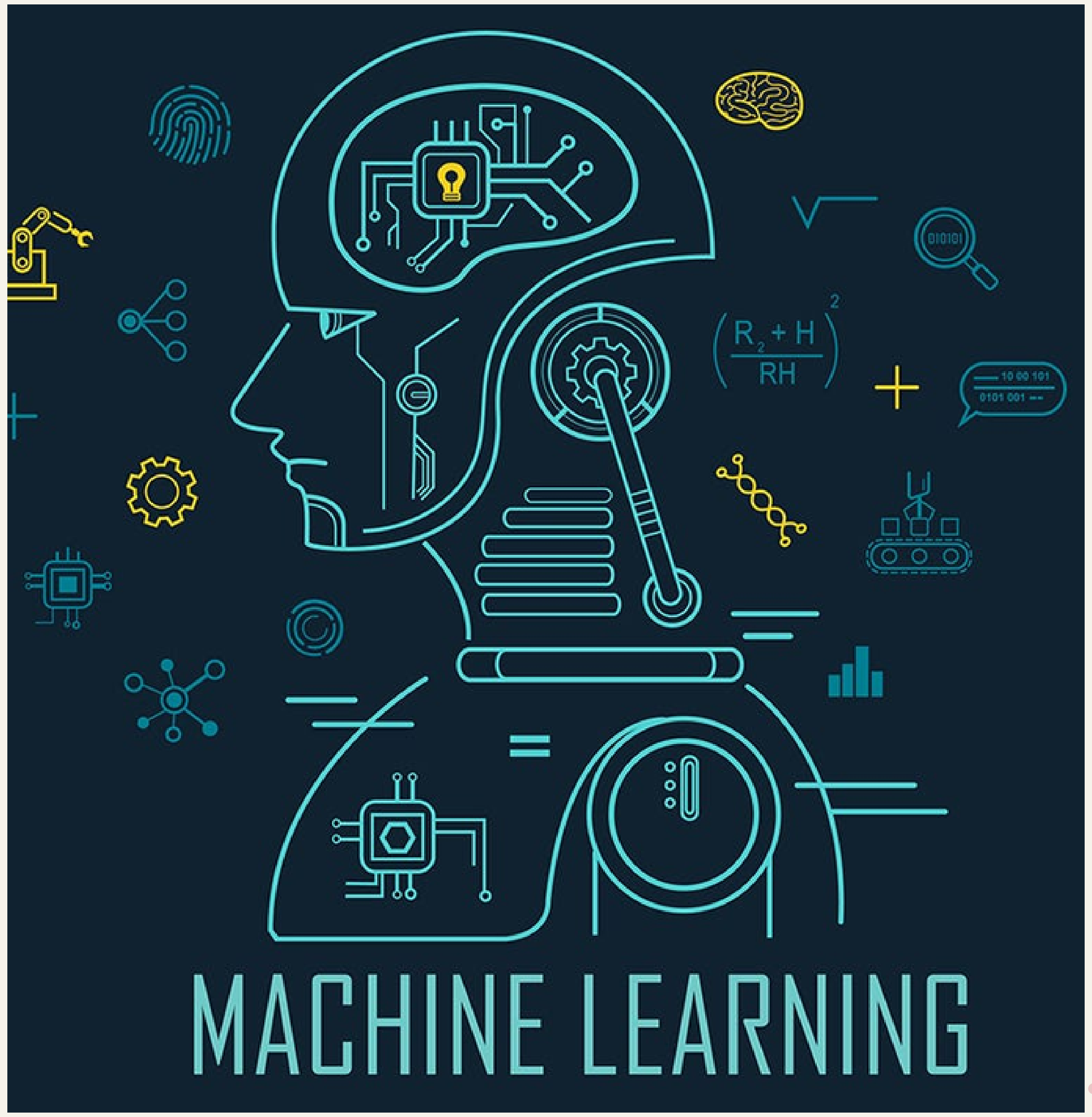
Machine Learning

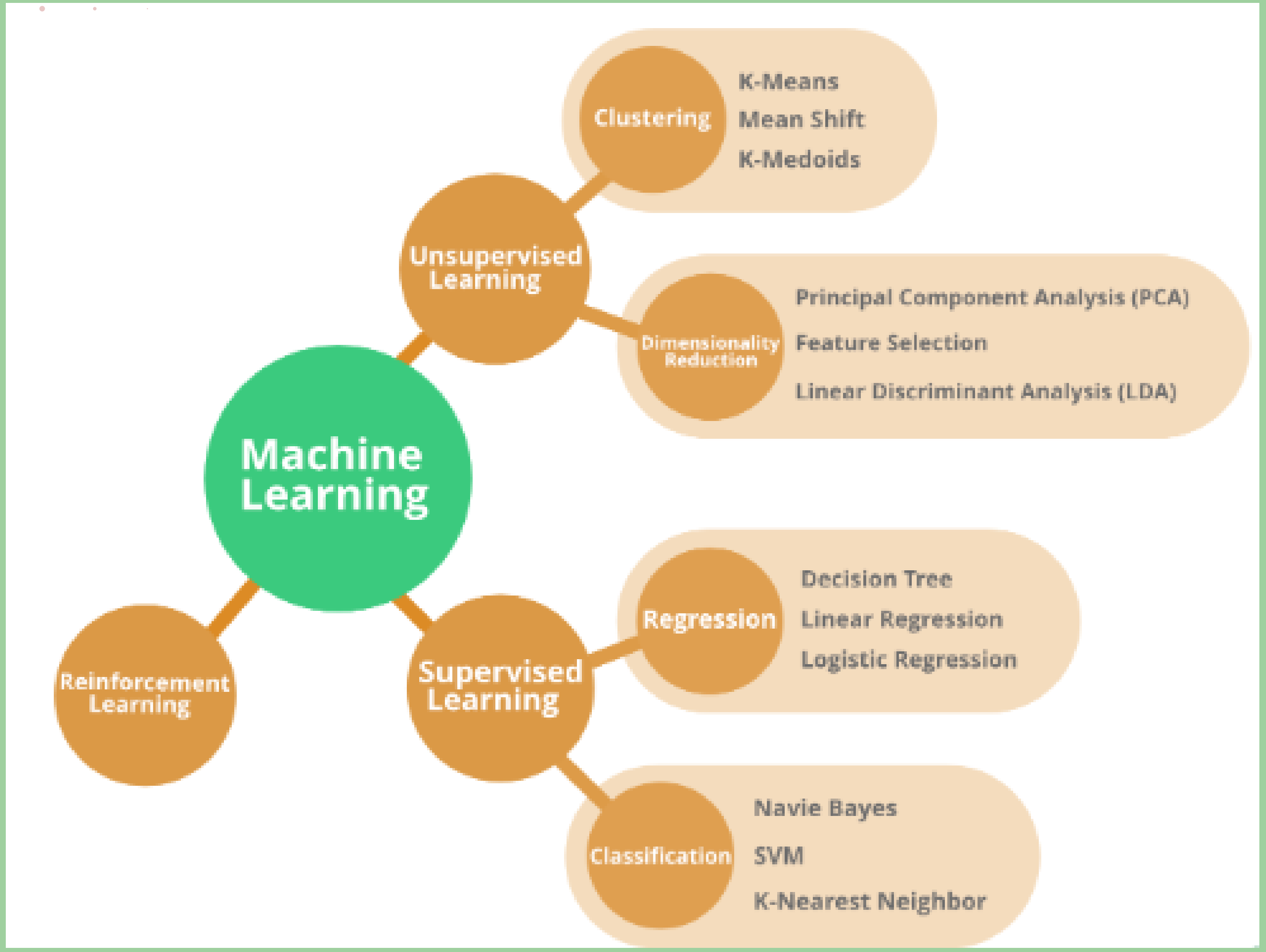
Machine learning is a fast-growing technology that teaches computers to learn from data automatically. It uses different algorithms to make predictions or create models based on historical information. There are three main types:

i.supervised learning

ii.unsupervised learning

iii.reinforcement learning



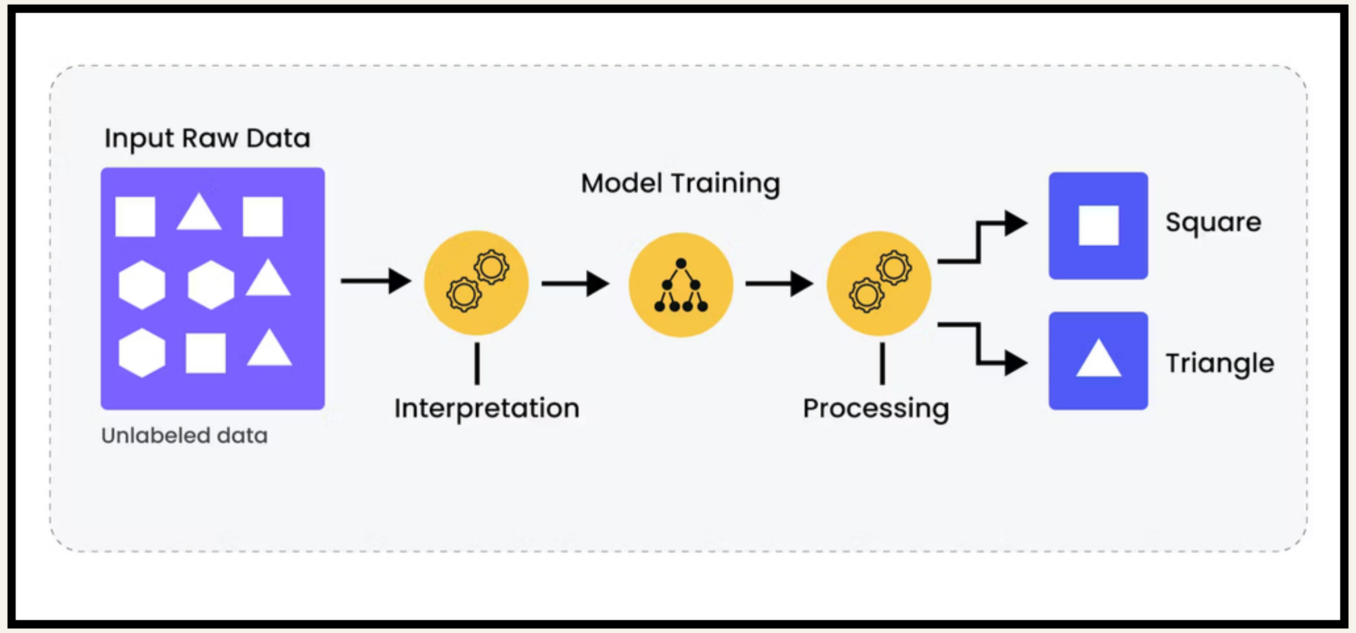


- **Supervised Learning:** Models are trained using labeled data to predict outcomes accurately.
- **Unsupervised Learning:** Algorithms analyze unlabeled data to identify patterns or clusters without predefined categories.
- **Reinforcement Learning:** Models learn decision-making skills by interacting with an environment to maximize cumulative rewards.

Classification Model

Supervised machine learning encompasses two main types of models: Regression and Classification. In this context, we'll focus on classification algorithms, which play a crucial role in categorizing data into predefined classes. Before deploying a model to make predictions on new data, it undergoes rigorous training using existing data and is subsequently evaluated on test data. Several popular classification algorithms include:

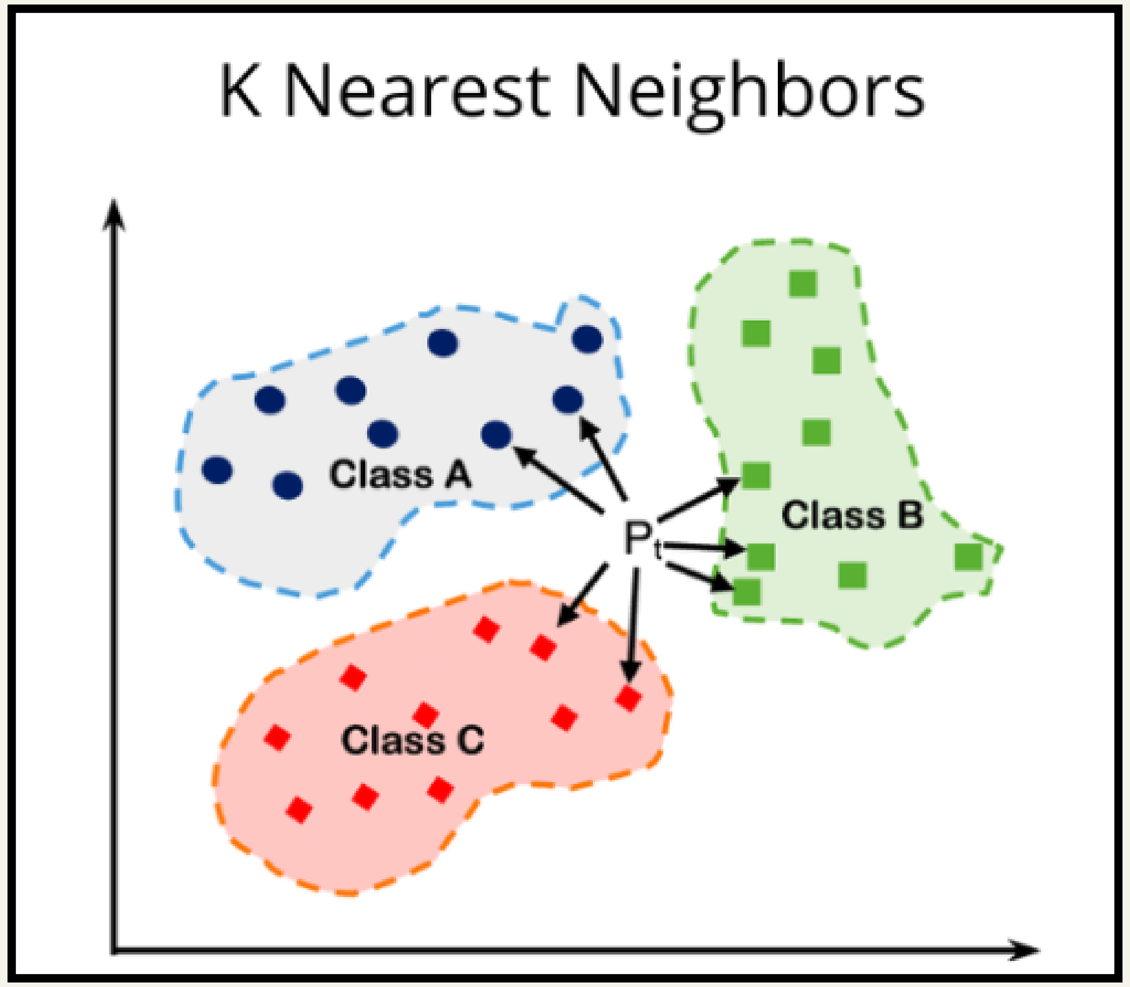
- Logistic Regression
- Naive Bayes Classification
- K-Nearest Neighbors (KNN)
- Decision Trees
- Support Vector Machines (SVM)



K-Nearest Neighbors (KNN)

- Used for classification and regression.
- Considers similar data points (neighbors).
- Determines predictions based on the majority class (classification) or average target values (regression) among the K nearest neighbors.
- The decision boundaries that divide query points into various areas are formed in part by the below distance measurement formula.
- Euclidean Distance (p=2):

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



Naive Bayes Classifier

- This algorithm calculates conditional probabilities.
- It's based on Bayes' Theorem with the assumption of independence among input variables.
- Key applications include spam filtering, text classification, and recommendation systems.
- The formula used for calculation in Naive Bayes classifier is as follows:

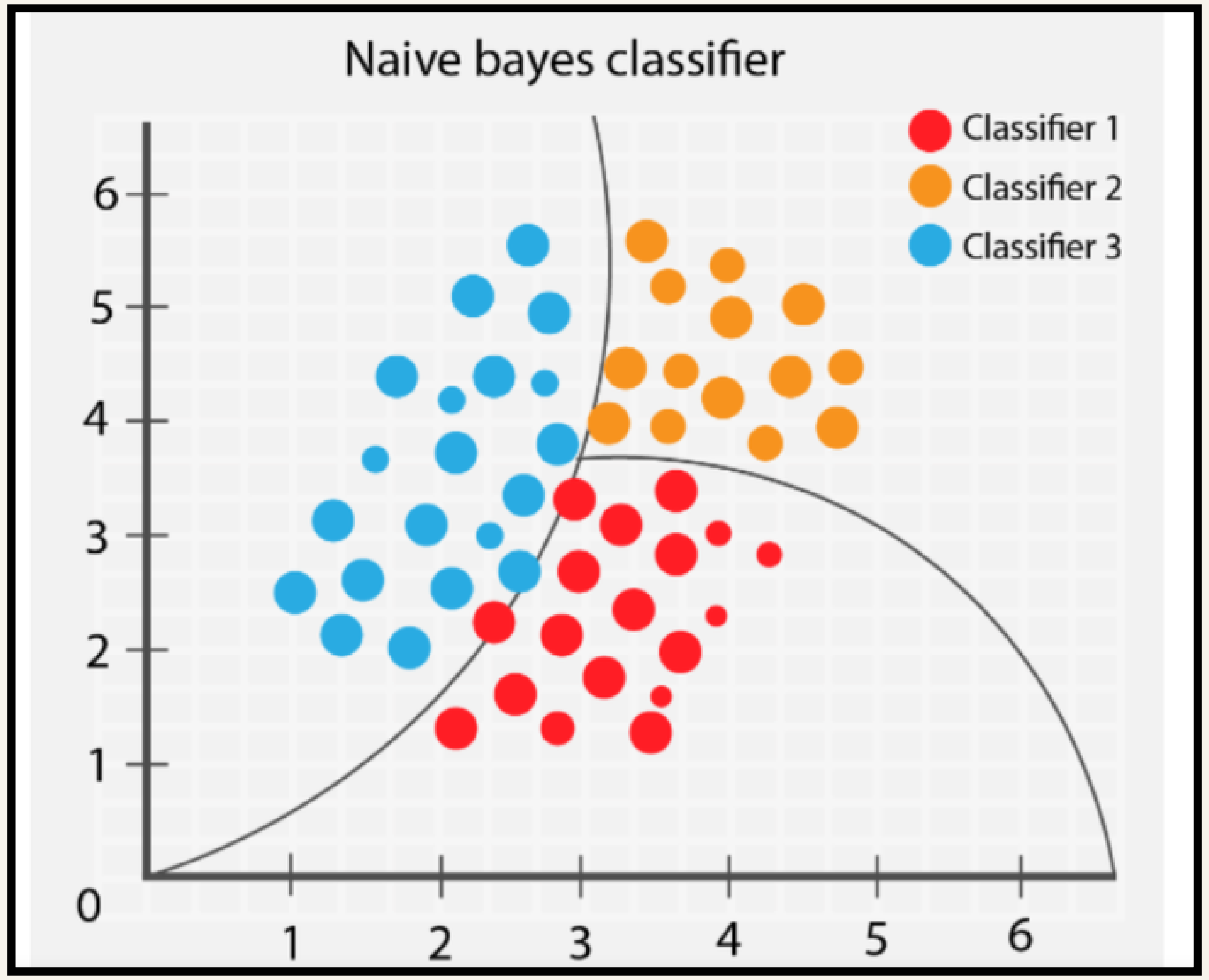
  $P(C|X) = P(X|C)P(C) / P(X)$

  ->(P(C|X)): The posterior probability of class (C) given predictor (X).

  ->(P(X|C)): The likelihood, i.e., the probability of predictor (X) given class (C).

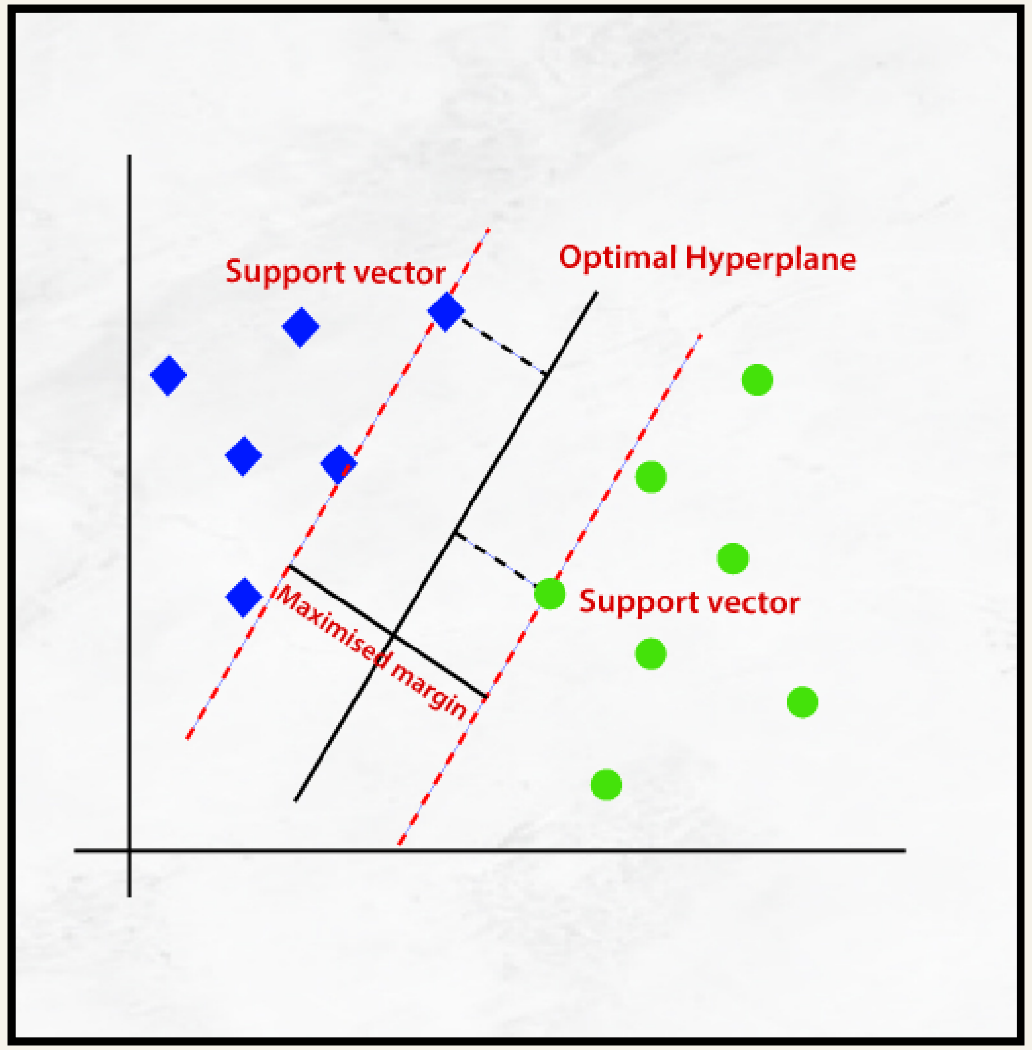
  ->(P(C)): The prior probability of the class.

  ->(P(X)): The prior probability of the predictor.



Support Vector Machine (SVM):

- SVM is a supervised learning algorithm primarily used for classification tasks.
- Its goal is to split data into two or more classes by creating a hyperplane.
- In 2D space, this hyperplane is a line that separates points into different categories.
- To create the decision boundary, we use different SVM kernels (e.g., linear, polynomial, RBF).
- The SVM creates a hyperplane defined by the equation: w.x+b=0
  - (w) is the weight vector.
  - (x) represents the input features.
  - (b) is the bias term.
- The main objective is to maximize the margin between classes, reducing noise and errors.Mathematically, we aim to maximize (2/||w||) subject to y(w.+b) >=1

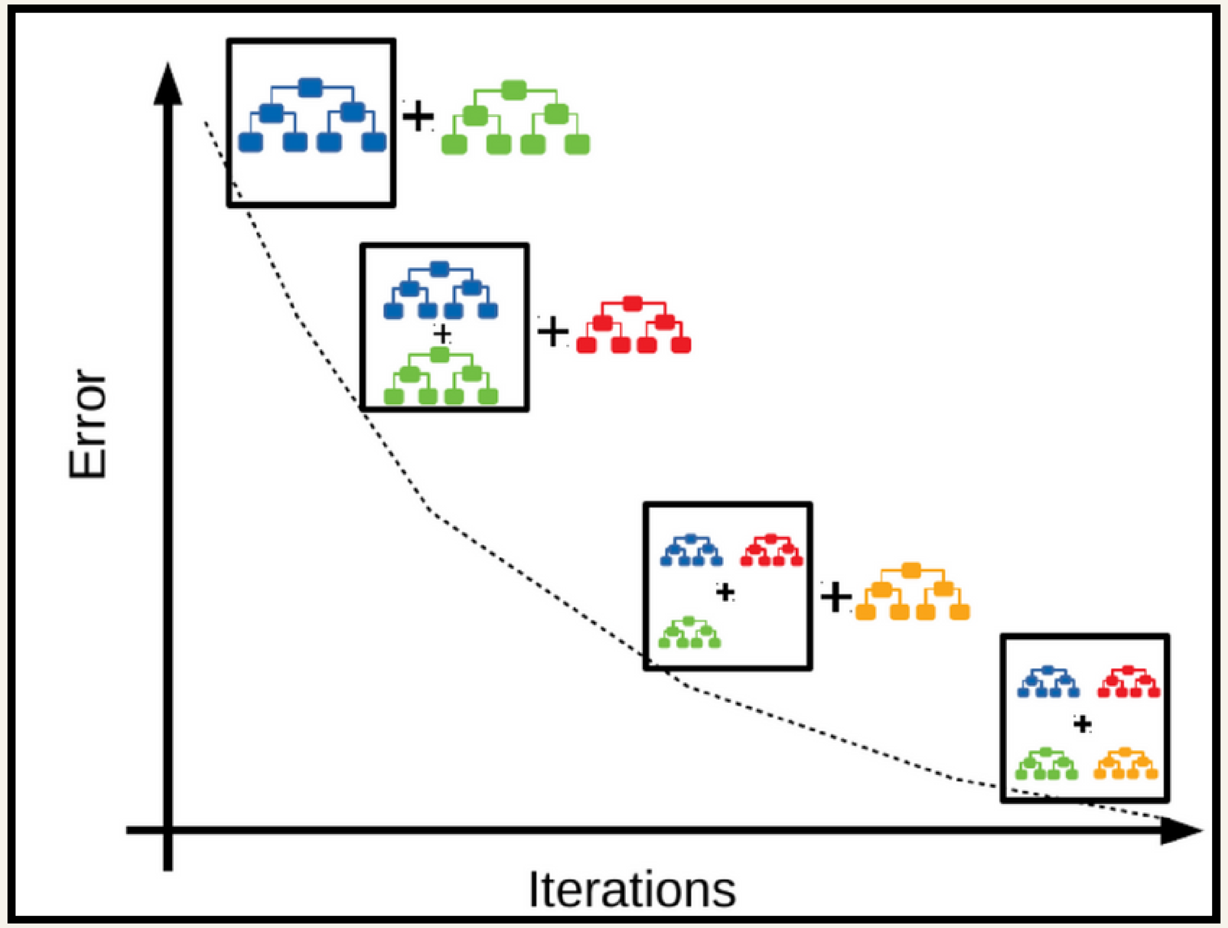


Gradient Boosting:

- Used for classification tasks.
- Builds an ensemble of decision trees sequentially.
- Each tree corrects errors made by the previous one.
- The method trains trees on residual errors from the predecessor.
- Effectively refines the model with each iteration.
- Thus is very efficient in terms of classification models .

  F2(x) = F1(x) +n. H1(x) where ,

    - (F2(x)): Updated model after the 1st iteration.
    - (F1(x)): Model at the 1st iteration.
    - (n): Learning rate.
    - (H1(x)): Weak learner fitted on the gradient of the loss function at the 1st iteration.



PROJECT PLANNING

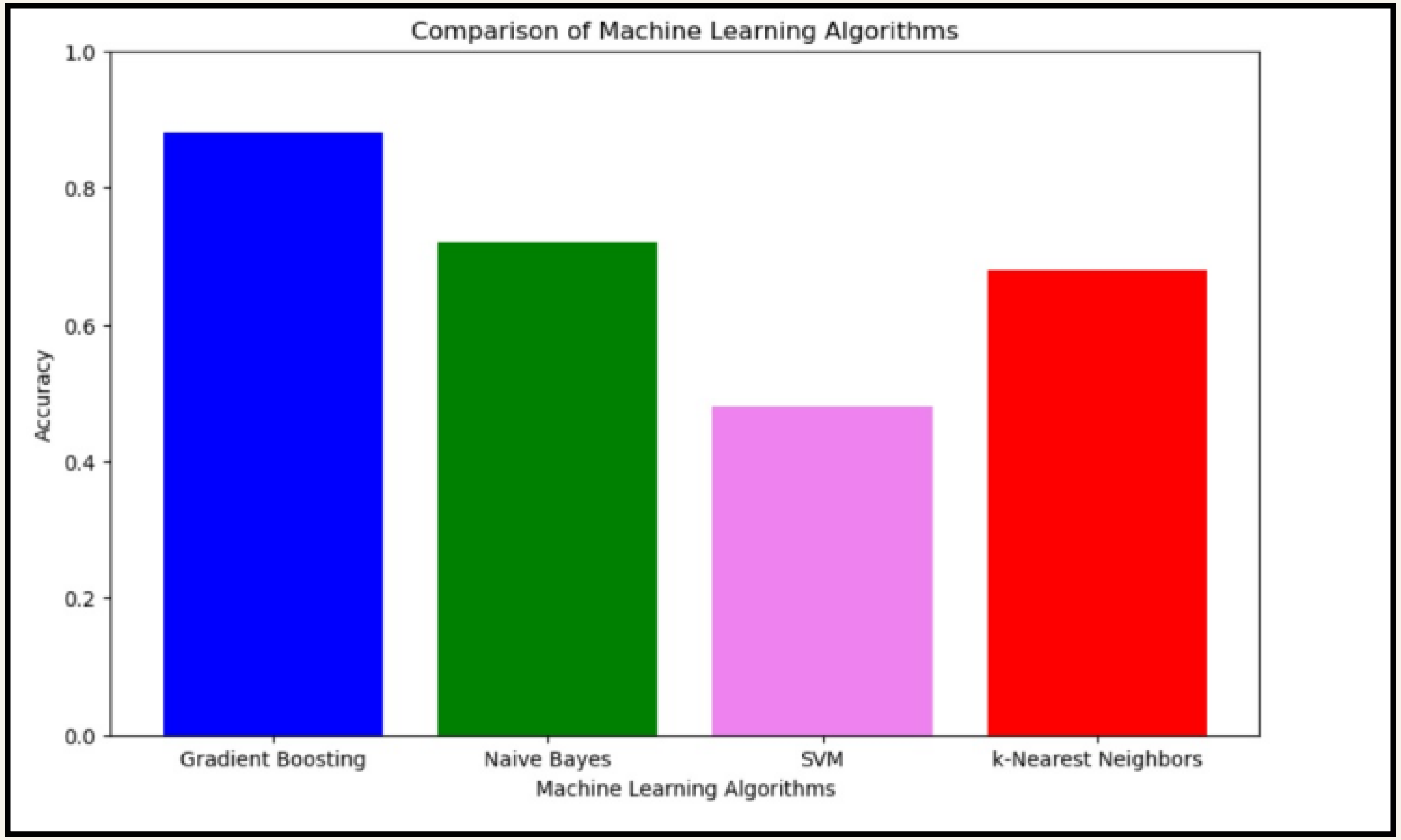
Data Management and Acquisition:

- Data Source: We obtained the dataset from Kaggle, which includes variables like pH, temperature, taste, odor, fat, turbidity, and color. These features help determine the grade of milk.
- Data Preprocessing: We meticulously cleaned the data by removing null values and duplicates to ensure data integrity.
- Splitting Data: Precisely dividing the dataset into training and testing portions is crucial for successful model training.
- Our dataset was partitioned into 90:10 training and testing sets.

Model Development::

- Algorithm Selection: We considered several machine learning algorithms (SVM, Naive Bayes, KNN, and Gradient Boosting) to analyze milk quality.
- Training and Validation: Using labeled data, we trained specific algorithms to identify patterns related to milk quality. Rigorous validation ensured accurate and generalizable models.
- Accuracy: Based on accuracy, we chose Gradient Boosting as our preferred model and evaluated its output.
  ->KNN - 68%
  ->SVM - 55.56%
  ->Naive Bayes - 72%
  ->Gradient Boosting - 88%

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.

SYSTEM DESIGN

Design Constraints

Software:

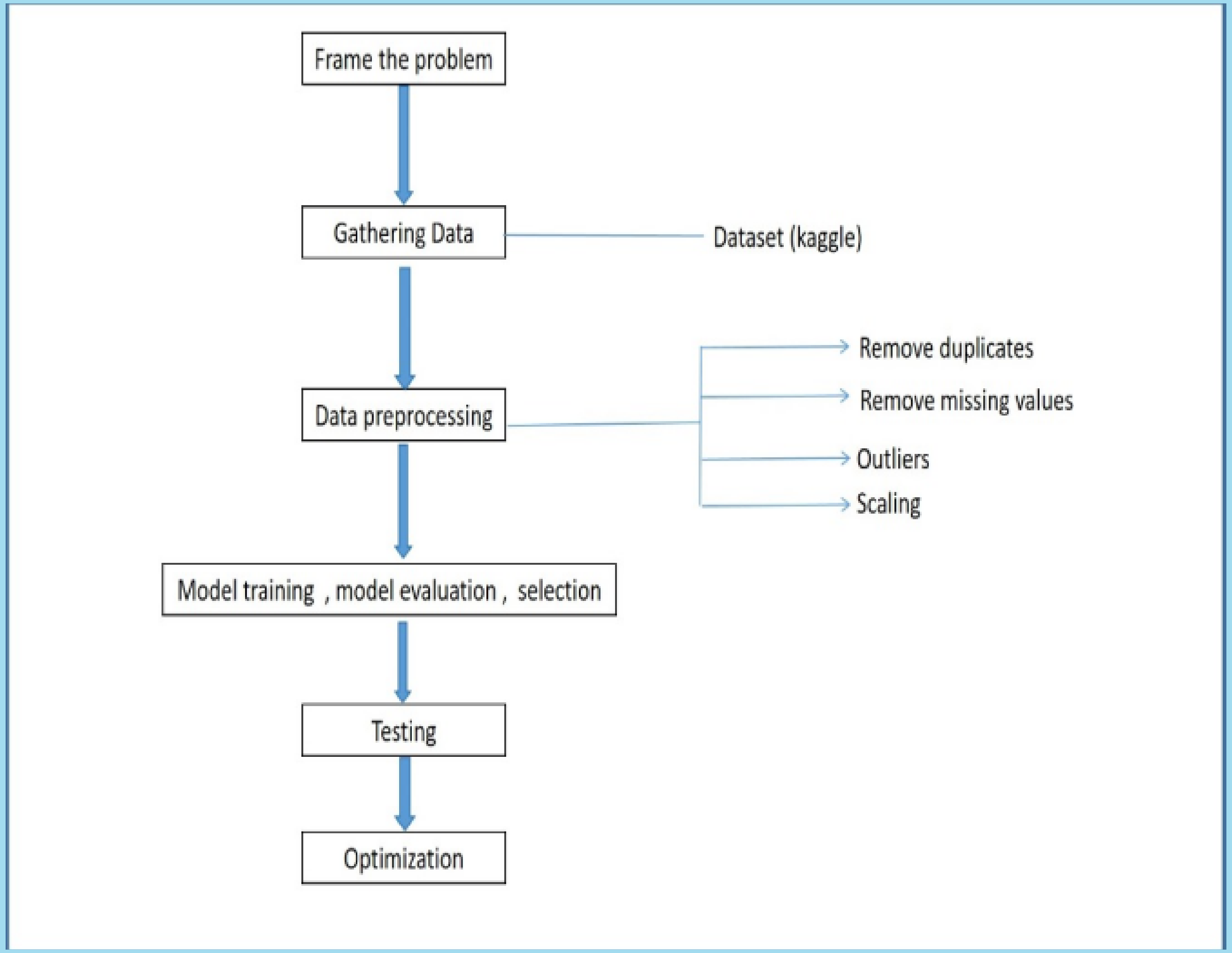
- <u>Python:</u> We use Python for creating machine learning models. Its rich ecosystem of data analysis libraries (such as NumPy, pandas, and scikit-learn) makes it an excellent choice.
- <u>Jupyter Notebook:</u> This interactive programming environment facilitates data analysis, result visualization, and rapid prototyping of ML algorithms.
- <u>Kaggle:</u> We rely on Kaggle to discover and access the required dataset.

Hardware:

<u>Sensor Devices:</u> Internet of Things (IoT) sensors play a crucial role. They collect real-time data on various milk quality parameters, enabling continuous 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

System Architecture OR Block Diagram



Result Analysis

```
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: "))
```

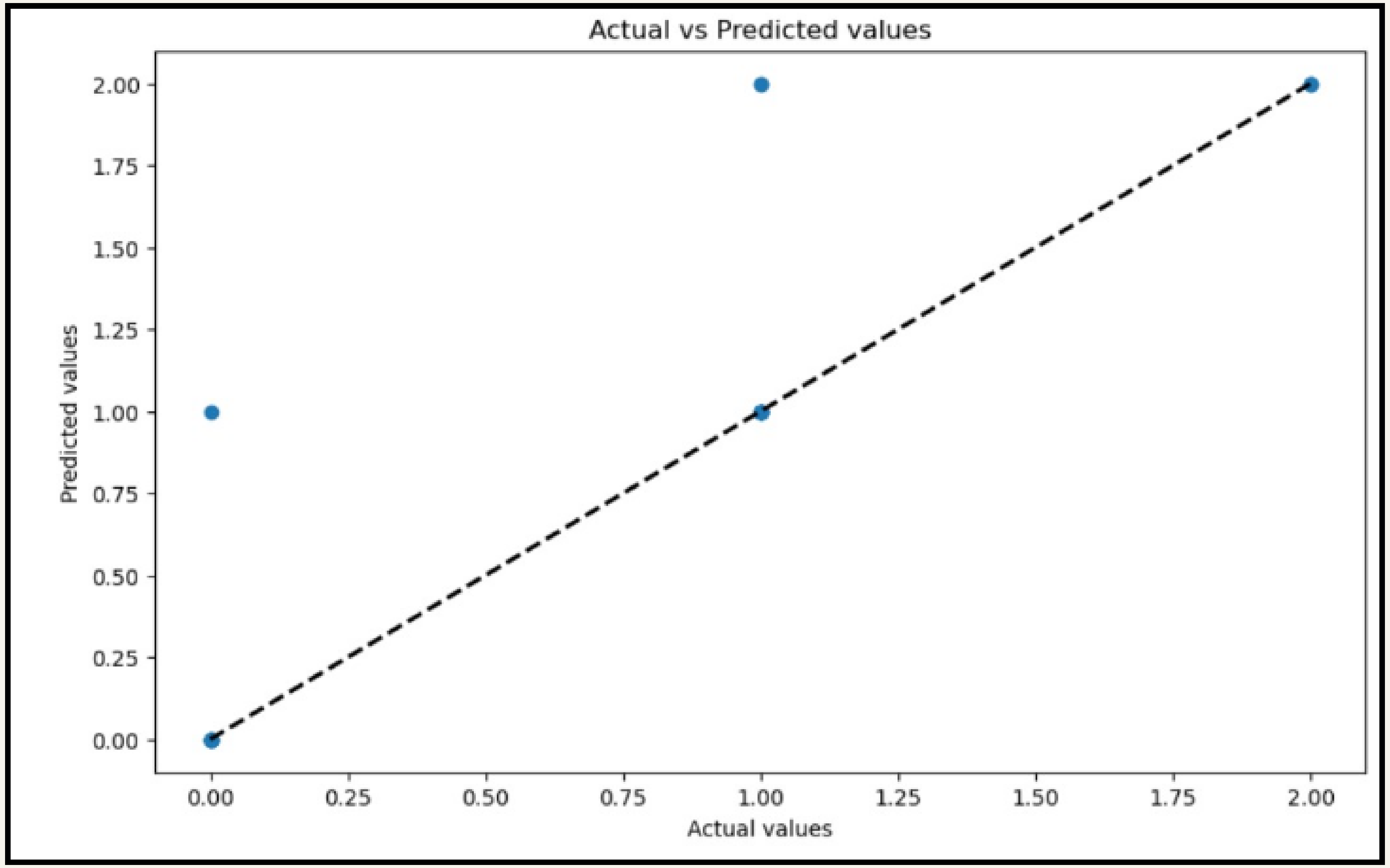
```
Input pH:  6.6
Input the temperature:  50
Input taste:  0
Input odor:  0
Input fat:  0
Input turbidity:  0
Input colour:  255
```

```
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
```

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.

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).

CONCLUSION

Milk Grade Guard (MGG) holds significant promise as a technical advancement toward proactive milk safety. By analyzing various data points and forecasting quality issues, it has the potential to prevent tainted milk from reaching consumers. However, successful deployment requires seamless integration with existing quality control systems. MGG should complement current safety procedures rather than replace them. While further research and testing are necessary, the right technological advancements and system integration could revolutionize the dairy industry, enhancing food safety and productivity.

FUTURE SCOPE

1. Expansion to Other Dairy Products:
    - Extend the algorithm to monitor cheese, cottage cheese (paneer), yogurt, and other dairy products.
2. Real-Time Analysis:
    - Implement a real-time monitoring system for instant feedback and continuous dairy farm monitoring.
3. Consumer Apps:
    - Link the project with consumer apps so users can verify the quality of the milk they consume.
4. Advanced Analysis:
    - Use more sophisticated machine learning algorithms

REFERENCES

- Bhavsar, Drashti & Jobanputra, Yash & Swain, Nirmal & Swain, Drdebabrata. (2023). Milk Quality Prediction Using Machine Learning. EAI Endorsed Transactions on Internet of Things. 10. 10.4108/eetiot.4501.

- J. N. V. R. Swarup Kumar, D. N. V. S. L. S. Indira, K. Srinivas and M. N. Satish Kumar, "Quality Assessment and Grading of Milk using Sensors and Neural Networks," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2022, pp. 1772-1776, doi: 10.1109/ICEARS53579.2022.9752269.

- A. Çelik, “Using Machine Learning Algorithms to Detect Milk Quality”, EJFST, vol. 6, no. 2, pp. 76–87, 2022.

- Habsari, W., F. Udin, and Y. Arkeman. "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.

- Neware, Shubhangi. "Cow Milk Quality Grading using Machine Learning Methods." International Journal of Next-Generation Computing 14.1 (2023).

- www.science.gov

- 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

- www.ijraset.com

THANK YOU