A Major Project Stage-1 Report
on

MILK QUALITY MONITORING SYSTEM USING IoT AND ML

Submitted in partial fulfillment of the requirements
for the award of degree of

BACHELOR OF TECHNOLOGY
in
Information Technology
by

G.Harshitha (20WH1A1271)
K.Kavitha (20WH1A1276)
N.manjula (20WH1A1286)

Under the esteemed guidance of
Ch. Sai Lalitha Bala
Assisstant Professor

Department of Information Technology
BVRIT HYDERABAD College of Engineering for Women
Rajiv Gandhi Nagar, Nizampet Road, Bachupally, Hyd–500090
(Affiliated to Jawaharlal Nehru Technological University, Hyderabad)
(NAAC 'A' Grade & NBA Accredited- ECE, EEE, CSE & IT)

December 6, 2023

BVRIT HYDERABAD
College of Engineering for Women
Rajiv Gandhi Nagar, Nizampet Road, Bachupally, Hyderabad – 500090
(Affiliated to Jawaharlal Nehru Technological University Hyderabad)
(NAAC ‘A’ Grade & NBA Accredited- ECE, EEE, CSE & IT)

CERTIFICATE

This is to certify that the Project report on **” Milk Quality Monitoring System Uisng IoT and ML”** is a bonafide work carried out by **G. Harshitha (20WH1A12 71), K. Kavitha (20WH1A1276)** and **N. Manjula (20WH1A1286)** in the partial fulfillment for the award of B.Tech degree in **Information Technology , BVRIT HYDERABAD College of Engineering for Women, Bachupally, Hyderabad** affiliated to Jawaharlal Nehru Technological University, Hyderabad, under my guidance and supervision. The results embodied in the project work have not been submitted to any other university or institute for the award of any degree or diploma.

Internal Guide	**Head of the Department**
Ch. Sai Lalitha Bala	**Dr. Aruna Rao S.L**
Assisstant Professor	**Professor & HoD**
Department of IT	**Department of IT**

External Examiner

DECLARATION

We hereby declare that the work presented in this project entitled **" Milk Quality Monitoring System Uisng IoT and ML"** submitted towards completion of in IV year I sem of B.Tech IT at "BVRIT HYDERABAD College of Engineering for Women", Hyderabad is an authentic record of our original work carried out under the esteemed guidance of **Ch. Sai Lalitha Bala, Assisstant Professor**, Department of Information Technology.

G.Harshitha (20WH1A1271)

K.Kavitha (20WH1A1276)

N.Manjula (20WH1A1286)

ACKNOWLEDGMENTS

We would like to express our profound gratitude and thanks to **Dr K.V.N. Sunitha, Principal, BVRIT HYDERABAD College of Engineering for Women** for providing the working facilities in the college.

Our sincere thanks and gratitude to **Dr. Aruna Rao S.L, Professor & Head, Department of IT, BVRIT HYDERABAD College of Engineering for Women** for all the timely support, constant guidance and valuable suggestions during the period of our project.

We are extremely thankful and indebted to our internal guide, **Ch. Sai Lalitha Bala, Assisstant Professor, Department of IT, BVRIT HYDERABAD College of Engineering for Women** for her constant guidance, encouragement and moral support throughout the project.

Finally, we would also like to thank our Project Coordinators **Dr. J. Kavitha, Associate Professor and Mr. Mukhtar Ahmad Sofi , Assistant Professor**, all the faculty and staff of Department of IT who helped us directly or indirectly, parents and friends for their cooperation in completing the project work.

G.Harshitha (20WH1A1271)

K.Kavitha (20WH1A1276)

N.Manjula(20WH1A1286)

ABSTRACT

The dairy industry relies on the consistent quality of milk products to meet consumer demand and ensure food safety. To address this problem, Milk Quality Monitoring System(MQMS) is used that leverages Internet of Things and Machine Learning technologies. The MQMS utilizes IoT sensors deployed across dairy farms and processing facilities to continuously collect data on key milk quality parameters such as temperature, pH, fat and color. These data streams are transmitted in real-time to a central server for analysis. Most of the existing methods are implemented by using certain parameters like taste, temperature, turbidity which gives inaccurate results. By adding other parameters like fat and color the quality of milk can be tested efficiently. The MQMS ensures consistent milk quality, reducing the risk of substandard products reaching consumers, thus improving food safety and consumer trust by classifying the quality of milk. This system can identify contamination, adulteration of milk thus leading to cost saving for dairy owners.

Keywords: Milk, Quality, MQMS, Parameters, Safety.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Milk Quality Monitoring System, amalgamating Internet of Things (IoT) and Machine Learning (ML), revolutionizes the dairy industry by ensuring the safety and excellence of milk products. Embedded IoT sensors strategically placed across the production and distribution network continuously gather real-time data on crucial parameters like temperature, humidity, and acidity. This comprehensive dataset undergoes rigorous analysis through ML algorithms, empowering the system to discern patterns, detect anomalies, and predict potential quality issues. The system's proactive nature allows for swift interventions, preventing the distribution of substandard or contaminated milk and preserving consumer safety. Beyond ensuring high-quality products, the system offers multifaceted benefits including enhanced efficiency, cost reduction through the prevention of product recalls, and regulatory compliance. Moreover, the incorporation of RFID tags enables robust traceability, providing transparency throughout the supply chain. With a user-friendly interface, the system offers real-time monitoring through web-based dashboards and mobile applications. In essence, the Milk Quality Monitoring System represents a paradigm shift, fostering a more efficient, transparent, and secure dairy supply chain, thereby elevating the standards of both producers and consumers in the dairy market.

1.1 Python Libraries

1.1.1 numpy (np)

NumPy is a Python library facilitating large, multi-dimensional arrays and matrices, coupled with high-level mathematical functions for efficient numerical computations. Integral to scientific computing, it underpins diverse applications like machine learning. NumPy's versatility and performance make it a cornerstone for data manipulation and analysis in Python.

1.1.2 pandas (pd)

Pandas, a robust open-source library, is designed for efficient data analysis and manipulation in Python. Built on NumPy, it introduces powerful data structures, particularly DataFrames, well-suited for handling structured data. Renowned for its speed and flexibility, Pandas simplifies tasks such as cleaning, aggregating, and transforming data, enhancing the capabilities of Python in data-centric workflows.

1.1.3 seaborn (sb)

Seaborn, built on Matplotlib, is a statistical data visualization library for Python. Offering a high-level interface, it simplifies the creation of visually appealing and insightful statistical graphics. Seaborn is particularly useful for effortlessly generating complex visualizations, making it a valuable tool for data analysts and researchers engaged in exploratory data analysis and presentation.

1.1.4 matplotlib.pyplot (plt)

Matplotlib, a Python plotting library, seamlessly integrates with NumPy for numerical mathematics. Its 'pyplot' module offers a user-friendly interface for generating diverse plots. Widely used in scientific and data visualization applications, Matplotlib empowers users to create a range of plots, charts, and graphs to effectively communicate and analyze data in Python programs.

1.1.5 scikit-learn (sklearn)

Scikit-learn is a Python machine learning library, offering straightforward and effective tools for data mining and analysis. Widely utilized for preprocessing data and implementing machine learning models, it simplifies tasks such as classification, regression, clustering, and more. Scikit-learn is a go-to resource for developers and researchers working on machine learning projects in Python.

1.1.6 matplotlib.font manager (fm)

Matplotlib's font - manager module is dedicated to handling font properties in Python. It enables users to manage and customize font-related aspects in Matplotlib plots, providing control over font styles, sizes, and other properties. This module is crucial for fine-tuning the appearance of text elements within plots generated using the Matplotlib library.

1.2 Motivation

The motivation for predicting milk quality using IoT (Internet of Things) and ML (Machine Learning) lies in the pursuit of enhancing dairy industry efficiency, product quality, and overall sustainability. By leveraging IoT devices such as sensors and smart devices installed in milking machines and storage units, real-time data on various parameters such as temperature, pH levels, and bacterial content can be collected. Machine Learning algorithms can then analyze this data to identify patterns, detect anomalies, and predict potential issues in milk quality. This proactive approach allows dairy farmers and producers to take immediate corrective actions, ensuring the production of high-quality milk while minimizing waste and economic losses. Additionally, the integration of IoT and ML enables automated monitoring, reducing the manual workload and increasing the scalability of dairy operations. Ultimately, this convergence of technology aims to optimize milk production processes, guarantee product safety, and contribute to a more sustainable and technologically advanced dairy industry.

1.3 Objective

The primary objective of implementing IoT and ML for the prediction of milk quality is to revolutionize the dairy industry by enhancing operational efficiency, ensuring product quality, and promoting sustainability. By deploying IoT devices, such as sensors, across the milk production and storage chain, real-time data on crucial parameters like temperature, pH, and bacterial content can be collected. The integration of Machine Learning algorithms then allows for comprehensive analysis, pattern recognition, and early detection of potential quality issues. This proactive approach empowers dairy farmers to take immediate corrective actions, minimizing wastage and economic losses, while simultaneously optimizing the production process. Furthermore, the automated monitoring capabilities provided by IoT and ML not only streamline operations but also enable scalability in dairy farming.The motivation for employing these technologies is rooted in the desire to ensure the production of high-quality milk, aligning with consumer expectations and regulatory standards. Additionally, the reduction of manual workload through automation contributes to the overall sustainability of dairy operations. By embracing IoT and ML, the dairy industry can advance towards a more intelligent and technologically sophisticated future, promoting resource efficiency and environmental responsibility. Ultimately, the objective is to create a resilient, high-quality dairy ecosystem that not only meets the demands of the modern market but also addresses global challenges by minimizing the ecological footprint associated with dairy production.

1.4 Problem Definition

The challenge in predicting milk quality through ML and IoT lies in the dairy industry's need for real-time, accurate assessments. The traditional manual methods for evaluating milk quality are time-consuming and susceptible to errors, hindering timely interventions and decision-making. The absence of continuous monitoring exacerbates the difficulty in identifying deviations in parameters like fat content and bacterial load, which are crucial for ensuring product quality. Integrating ML algorithms with IoT sensors is necessary to establish a predictive system capable of monitoring and analyzing key quality indicators in real time. This involves overcoming the limitations of current practices, enhancing precision, and providing dairy farmers with actionable insights for immediate responses to quality fluctuations. The goal is to develop an innovative solution that transforms the dairy industry by ensuring consistent, high-quality milk production, reducing waste, and promoting sustainability through data-driven decision support.

1.5 Aim

"The aim of this dairy-focused project is to introduce an integrated system leveraging advanced machine learning algorithms and IoT technology to enhance the prediction of milk quality. The focus is on surpassing the limitations of traditional methods for quality assessment and optimizing processes with data-driven algorithms. The overarching goal is to revolutionize current practices, ensuring accurate and efficient prediction of milk quality parameters.

The project targets a prediction accuracy exceeding 95% through experimental validation with a diverse dataset. By implementing this integrated system, we seek to significantly improve the overall quality control of dairy products, reduce waste, and contribute to the sustainability and efficiency of the dairy industry. This innovative approach aims to empower dairy farmers and industry stakeholders with real-time insights into milk quality, leading to improved decision-making and enhanced product quality".

1.6 IoT Components

1.6.1 pH Sensor

The pH sensor's integration with IoT and ML revolutionizes milk quality prediction. Continuously measuring acidity, the sensor transmits real-time data, enabling ML algorithms to identify deviations and predict spoilage or contamination. This proactive approach facilitates timely intervention, ensuring the production and distribution of high-quality milk. The system's efficiency and waste reduction benefits are amplified through real-time monitoring, providing actionable insights for stakeholders in the dairy industry. This comprehensive solution optimizes quality control processes and safeguards consumer satisfaction with consistently superior dairy products.

Figure 1.1: Ph Sensor

1.6.2 DS18B20 Temperature Sensor

The DS18B20 temperature sensor is integral to milk quality monitoring systems, precisely measuring temperature in dairy storage. This sensor ensures optimal conditions, preventing bacterial growth and maintaining freshness. It enhances quality control by providing real-time data on temperature fluctuations. Specifically designed for accuracy and reliability, the DS18B20 enables dairy producers to adhere to stringent quality standards, ensuring consumers receive safe, high-quality milk products while optimizing production processes and minimizing waste in the dairy industry.

Figure 1.2: DS18B20 Temperature Sensor

1.6.3 Lactometer

The lactometer, a tool measuring milk density, is crucial for predicting milk quality when integrated with ML and IoT. The lactometer assesses the milk's composition, helping detect potential adulteration or dilution. Combined with ML algorithms and IoT connectivity, real-time data on milk density is collected and analyzed. ML models can then identify patterns indicative of quality deviations, facilitating early intervention. This comprehensive system ensures the production and distribution of genuine, high-quality milk, enhancing efficiency and consumer trust in the dairy industry through advanced predictive capabilities.

Figure 1.3: Lactometer

1.6.4 TCS230 Colour Sensor

The TCS230 color sensor is pivotal in milk quality monitoring systems, detecting variations in milk color that may indicate spoilage or irregularities. By analyzing color changes, it helps identify potential quality issues, ensuring consumers receive fresh and safe dairy products. The TCS230 enhances quality control measures by providing rapid and accurate color readings, enabling dairy producers to take corrective actions promptly. This sensor contributes to maintaining high standards in milk production, promoting consumer satisfaction and minimizing risks associated with sub optimal product quality.

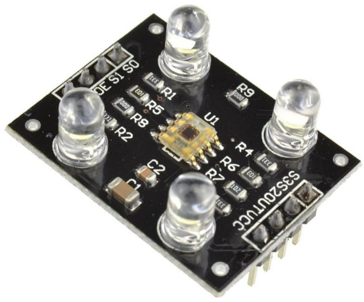

Figure 1.4: Colour Sensor

Chapter 2

Literature Survey

2.1 Research Papers

2.1.1 IoT based detection of adulteration in milk

[1]It revolves around the development of an Internet of Things (IoT)-based system designed for detecting adulteration in milk. This innovative system employs sensors to measure diverse parameters, including air quality, conductivity, pH, and temperature, to comprehensively assess the quality of milk. The execution of this proposed framework is facilitated by the Arduino UNO microcontroller. The sensor readings are then showcased on an LCD screen, and an Android app named Blynk is utilized for real-time monitoring.To enhance communication, the system incorporates a GSM modem interface, enabling the transmission of sensor values via SMS to a predefined mobile number. The study underscores the significance of ensuring the purity of milk and highlights potential health risks associated with the consumption of adulterated milk. Tests conducted using both pure and adulterated milk samples, with adulterants such as detergent. The results were prominently displayed on both the LCD screen and the Blynk application, showcasing the system's efficacy in detecting adulteration.The research concludes by emphasizing the system's potential applications in both small and large-scale milk dairies. Additionally, it explores future prospects for integrating IoT and database management systems to enhance milk quality assessment and streamline payment processing. Furthermore, the study references pertinent literature on milk spoilage detection, particularly using remote-query sensors.IoT-based system emerges as a promising solution for ensuring the quality of milk, with real-time monitoring capabilities and the potential for widespread implementation across various dairy settings. The integration of advanced technologies not only addresses the issue of adulteration but also opens avenues for further research and innovation in the domain of milk quality assurance and management.

2.1.2 Research on dairy products detection based on machine learning algorithm

[2] Approach utilizing an electronic nose model and machine learning algorithms for dairy product detection is presented. The electronic nose, mimicking the olfactory organs of animals, is equipped with a gas sensor array, signal acquisition module, data acquisition module, and signal processing and pattern recognition module. The experiment involved collecting volatile gases from milk samples using the electronic nose, with subsequent data analysis conducted on 1000 standardized data groups, comprising 800 training data and 200 test data. To reduce data dimensionality, principal component analysis (PCA) and linear discrimination analysis (LDA) were employed for cattle farm classification. The machine learning algorithms Logistic Regression, Support Vector Machine, and Random Forest were utilized for classification and regression tasks, demonstrating their effectiveness in distinguishing milk sources, estimating milk fat and protein content, and evaluating milk quality. The study's findings highlight the potential of machine learning algorithms in dairy product detection, offering a low-cost, non-destructive, and accurate method for assessing milk quality and authenticity. This research contributes to the advancement of technology in the dairy industry, with implications for enhancing quality control and ensuring consumer satisfaction.

Method	**Accuracy**
Logistic Regression	91.5%
Support Vector Machine	96%
Random Forest	96%

2.1.2.1: Performance Analysis of used methods

2.1.3 Machine Learning Applied to Milk Sample Classification

[3] The application of machine learning techniques to classify milk samples, with the goal of ensuring the quality and safety of dairy products for human consumption. Emphasizing the significance of providing high-quality dairy products, particularly for vulnerable populations like children, the research evaluates the classification process using machine learning models such as random forest, k-nearest neighbors (KNN), and neural networks. The results reveal that the random forest technique achieved the highest accuracy, ranging from 96 percent to 99 percent , while KNN attained an accuracy level between 77percent and 81 percent. The study also explores the use of support vector machine and linear discriminant

analysis for classification, as well as the potential application of e-nose and brix meter for food quality monitoring. The findings of this research contribute to the advancement of techniques for ensuring the integrity and safety of dairy products, with potential implications for the food industry and public health. The study's focus on leveraging machine learning to enhance the classification of milk samples underscores the importance of technological advancements in ensuring the quality and safety of food products, aligning with global efforts to improve food security and public health.

Method	**Accuracy**
Support Vector Machine	90%
Random Forest	98%

2.1.3.1: Analysis of used methods

2.1.4 Using Machine Learning Algorithms to Detect Milk Quality

[4] "Using Machine Learning Algorithms to Detect Milk Quality" presents a study on the application of machine learning algorithms to determine milk quality. The research utilizes the Milk Quality dataset from Kaggle, comprising seven attributes including pH, temperature, taste, odor, fat, turbidity, and color. By employing the AdaBoost and Neural Network algorithms, the study aims to classify milk quality as low, medium, or high. The results demonstrate high success rates, with the AdaBoost algorithm achieving a 99.9 precent success rate and the Neural Network algorithm achieving 95.4 percent . The study emphasizes the potential of machine learning in enhancing food safety and quality control in the dairy industry. It highlights the importance of analyzing multiple features to accurately assess milk quality, offering promising implications for improving milk quality assessment. The research has the potential to revolutionize milk quality monitoring, benefiting both food producers and consumers. By leveraging intelligent systems, this approach could contribute to the advancement of dairy product quality and have a significant impact on the milk production and consumption landscape. Overall, the study underscores the potential of machine learning algorithms in ensuring the quality and safety of dairy products, with implications for the broader food industry.

Method	Accuracy
AdaBoost	99.9%
Neural Network	95.4%

2.1.4.1: Analysis of used methods

2.1.5 An analysis and design of fresh milk smart grading system based on internet of things

[5] It presents an analysis and design of a fresh milk smart grading system based on the internet of things (IoT). The study aims to improve the quality classification in the dairy industry through the use of machine learning models. The system involves monitoring the temperature of fresh milk in real-time using IoT sensors and collecting quality data to be classified by an artificial neural network (ANN) model. The study also uses Business Process Modelling Notation (BPMN) diagrams to describe the process and sub-process in a smart grading system of fresh milk based on IoT. The result of the requirement analysis showed how smart the grading system involved stakeholders. The machine learning model can help the IoT system classify goods or services. The accuracy value of the classification obtained is 98.74 percent. The attributes used for grouping were temperature and colour. The study provides insights into the potential of IoT and machine learning in improving the quality classification of fresh milk in the dairy industry.

2.1.6 Smart system for Milk quality analysis and billing system

[6] A smart system for milk quality analysis and billing system, which aims to improve the dairy sector in India's economy. The system measures various parameters of milk, including fat content, harmful gas content, pH level, and refrigerated level, using sensors such as LDR and pH sensors. The data is then transmitted to an IoT cloud platform and displayed on LCD monitors, allowing for internet-based data monitoring. The system also includes a Blynk app on an Android phone to aid in billing computations and daily payments. The paper highlights the importance of technology in improving the agricultural way of life and offers a solution that can benefit small-scale farmers in India. Overall, the smart system for milk quality analysis and billing system offers a comprehensive solution to improve the dairy sector's efficiency and productivity.

2.1.7 Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods

[7] It mainly discusses an IoT-based milk monitoring system designed to ensure the quality and quantity of milk, particularly in rural and urban areas. The system aims to address the issue of milk adulteration and the presence of pathogenic organisms in raw milk, which can lead to health risks and a decline in the quality of life. The proposed system utilizes various sensors such as milk level, gas, temperature, viscosity, and salinity sensors, along with an Arduino controller for real-time monitoring and quality checking. Additionally, the system includes power analysis to assess power consumption and battery backup requirements. The implementation results indicate that the system can effectively detect milk adulteration and early microbial activity, contributing to the overall safety and quality of milk products. The use of IoT technology allows for remote monitoring and integration of the physical environment into computer-based systems, leading to improved efficiency, accuracy, and economic benefits. However, challenges related to sensor power consumption and system optimization are also addressed. Overall, the IoT-based milk monitoring system presents a promising approach to ensuring the safety and quality of milk products, with potential benefits for both producers and consumers.

Method	**Accuracy**
Support Vector Machine	75%

2.1.7.1: Performance of listed methods

2.1.8 IoT for Development of Smart Dairy Farming

[8] It explores the potential of Internet of Things (IoT) and data-driven techniques in smart dairy farming (SDF) to improve milk yield and meet the increasing demand for dairy products. The authors propose a framework for enhancing milk production through the adoption of the latest techniques for improving feeding and milking procedures. The framework involves the use of wearable sensors to capture data from cows, which is then transferred to a base station and analyzed in the cloud using IoT-based platforms. The research emphasizes the potential of IoT-enabled smart dairy farming to address traditional farming challenges and increase milk production. Additionally, the article discusses the benefits of using AI techniques, such as smart monitoring, cow observation, feeding, milking, and reproduction management, in dairy farming. The proposed system is expected

to make IoT-based farming more efficient, although it may require heavy initial investment. Overall, the research aims to assist farmers in increasing milk production while balancing the investment and earnings. The findings of the research have implications for both developed and developing countries in the dairy industry, offering potential solutions to improve production and address challenges in dairy farming.

2.1.9 IoT based milk monitoring system for detection of milk adulteration

[9] It presents an IoT-based milk monitoring system that utilizes various sensors to detect adulteration and ensure the quality and safety of milk. The system includes a Milk Level Sensor, Gas Sensor, Temperature Sensor, Viscosity Sensor, and Salinity Sensor, which continuously monitor the milk's composition and detect any early microbial activity or adulteration. The Arduino controller processes the data from these sensors and displays the results on an LCD screen. The system's power consumption is analyzed, and the authors emphasize the importance of real-time monitoring and decision-making to ensure the quality and safety of milk. The proposed system offers a promising solution for addressing milk quality control and safety concerns, while also highlighting the significance of IoT in enhancing efficiency, accuracy, and financial benefits in milk monitoring.

2.1.10 Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot

[10] The development and validation of a near-infrared spectroscopic sensing system for online milk quality assessment in a milking robot. The system allows for the automatic collection of NIR spectra of raw milk, enabling the determination of major milk constituents, somatic cell count, and milk urea nitrogen. The precision and accuracy of the calibration models have been validated, with impressive results. The system has the potential to revolutionize milk quality assessment in the dairy industry, improving production efficiency and reducing costs. It provides detailed information on the system's specifications, including the spectrum sensor, light source, optical fiber, milk chamber surface, spectrometer, and data processing computer. The potential implications of this technology for the dairy industry are significant, as it could lead to more efficient and accurate milk quality assessment, ultimately benefiting both producers and consumers.

Chapter 3

System Design

3.1 Proposed System

The proposed system, referred to as the ”Milk Quality Prediction System,” integrates diverse hardware components and software libraries for the predictive maintenance of milk quality. The system utilizes IoT sensors, microcontrollers such as Arduino Uno, and components including temperature sensor and color sensor, Ph sensor, and communication interfaces. This comprehensive setup aims to ensure optimal milk quality by leveraging IoT and ML to predict and address potential issues. The predictive analytics system can provide timely alerts, enabling proactive interventions to maintain and improve the quality of milk throughout the production and supply chain, thereby ensuring consumer satisfaction and safety.

3.2 System Architecture

The architecture for predicting milk quality through ML and IoT involves a sensor network within the dairy farm, collecting real-time data on various quality parameters such as fat content and bacterial load. This data is transmitted to a centralized cloud platform where ML algorithms analyze and predict milk quality. The results are then accessible to farmers through a user-friendly interface, aiding in timely decision-making. The integration of ML and IoT optimizes quality control processes, providing insights into milk production, minimizing wastage, and enhancing overall dairy efficiency, ultimately contributing to sustainable and high-quality dairy products.

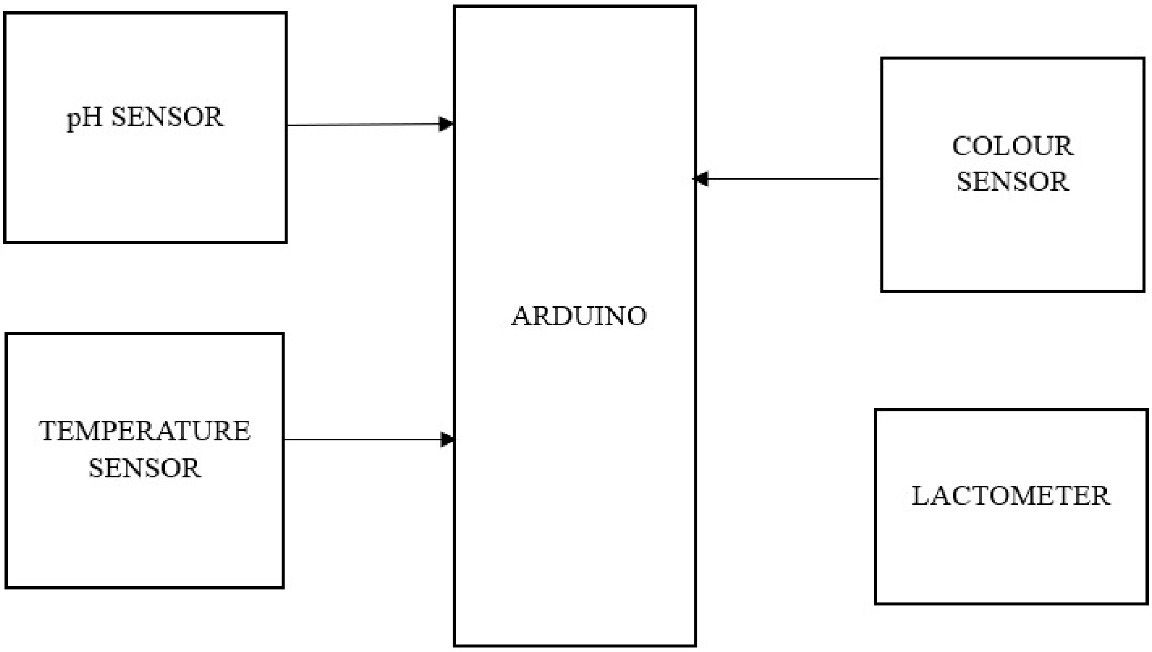

Figure 3.1: Architecture

Chapter 4

Methodology

4.1 Requirement Analysis

- Define the specific functionalities and desired outcomes of the ”Milk Quality Monitoring System” Identify hardware and software requirements, considering components like Arduino Uno, Temperature sensor,Color sensor, pH, Lactometer, Jumper Wires.

4.2 Hardware Configuration

- Assemble and configure the hardware components, ensuring proper connections and compatibility. Verify the functionality of each component, including the Temperature sensor, Color sensor, pH, Lactometer and take the data values which can be used for train the model.

4.3 Software Integration

- Develop the software components using libraries such as numpy(np), pandas(pd),seaborn(sb) , matplotlib.pyplot(plt), scikit-learn(sklearn) ,matplotlib.font manager (fm) and run the ML algorithms by using the train and tesing data for observing the accuracy of each algorithm.

Chapter 5

Implemention

5.1 Importing python Libraries

The code imports essential Python libraries for comprehensive data analysis and machine learning. NumPy is utilized for efficient numerical operations and array handling, while Pandas facilitates data manipulation with its powerful DataFrame structures. Seaborn, built on Matplotlib, enhances statistical data visualization, creating visually appealing plots. Matplotlib's Pyplot and Font Manager modules contribute to customized plot styles and font properties. Scikit-learn, a versatile machine learning library, is employed for data preprocessing and implementing machine learning models like Logistic Regression and Random Forest classifiers. This combination of libraries provides a robust toolkit, enabling seamless data exploration, visualization, and model training within a unified Python environment, catering to diverse aspects of the data science workflow.

Code

```
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
sb.set()
from matplotlib import pyplot as plt, font_manager as fm
```

5.2 Data Splitting

5.2.1 Dataset

This is the source of information for the classification problem.The data source is the file ”milknew.csv”. It contains the data for this example in comma-separated values (CSV) format. The number of columns is 8, and the number of rows is 1060.
The variables are:
pH: This feature defines pH of the milk, which is in the range of 3 to 9.5.
temperature: This feature defines the temperature of the milk, and its range is from 34'C to 90'C.
taste: This feature defines the taste of the milk and takes the possible values: 1 (good) or 0 (bad).
odor: This feature defines the odor of the milk and takes the possible values: 1 (good) or 0 (bad).
fat: This feature defines the fat of the milk and takes the possible values: 1 (good) or 0 (bad).
turbidity: This feature defines the turbidity of the milk and takes the possible values: 1 (good) or 0 (bad).
colour: This feature defines the color of the milk, which is in the range of 240 to 255.
grade: This is the target and takes the values: low quality, medium quality, or high quality.
All variables are inputs, except “grade” it is categorical value. The milk dataset contains 429 instances of low quality, 374 instances of medium quality, and 256 instances of high quality.

The data splitting is conducted using the train-test-split function from scikit-learn. First, the features (X) and the target variable (y), which is the 'Grade' column, are defined. The StandardScaler from scikit-learn is then applied to standardize the feature values. Subsequently, the train-test-split function is employed to divide the dataset into training and testing sets. The test-size parameter is set to 0.3, indicating that 30 percent of the data will be allocated for testing, while the remaining 70 percent will be used for training the machine learning models. Additionally, the random-state parameter is set to 42 for reproducibility. The resulting split consists of X-train and X-test for features, and y-train and y-test for the corresponding target values. This approach ensures that models are trained on a subset of the data and evaluated on an independent subset, facilitating a robust assessment of their performance.

Code

```
# Splitting into train and test
x = milkData.drop(['Grade'], axis=1)
y = milkData['Grade']

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Standardize the features
PredictorScaler = StandardScaler()
PredictorScalerFit = PredictorScaler.fit(x)
x = PredictorScalerFit.transform(x)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(x, y,
test_size=0.3, random_state=42)

# Display the shapes of the resulting sets
print("X Train : ", X_train.shape)
print("X Test  : ", X_test.shape)
print("Y Train : ", y_train.shape)
print("Y Test  : ", y_test.shape)
```

5.3 Training with Different Models

data training with different machine learning models is performed using scikit-learn, a comprehensive machine learning library in Python. Three types of models are utilized: Logistic Regression and Random Forest Classifier. The dataset is split into training and testing sets using the train-test-split function. The features are then standardized using StandardScaler to ensure uniformity in scale. Logistic Regression and Random Forest models are instantiated, fitted to the training data, and subsequently used to predict the target variable on the testing set. Model accuracy is assessed using the metrics.Accuracy-score function, providing insights into how well each model performs on the unseen data. The code exemplifies a standard workflow for training, evaluating, and comparing different machine learning models, demonstrating the flexibility and simplicity offered by scikit-learn for such tasks.

Code

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn import metrics

# Split the data into train and test sets
x = milkData.drop(['Grade'], axis=1)
y = milkData['Grade']

# Standardize the features
PredictorScaler = StandardScaler()
PredictorScalerFit = PredictorScaler.fit(x)
x = PredictorScalerFit.transform(x)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(x, y,
test_size=0.3, random_state=42)
```

```
# Logistic Regression model
model_lr = LogisticRegression()
model_lr.fit(X_train, y_train)
y_pred_lr = model_lr.predict(X_test)
accuracy_lr = metrics.accuracy_score(y_test, y_pred_lr)
print('Accuracy for Logistic Regression (Test Dataset):',
accuracy_lr)
```

Figure 5.1: Logistic Regression Confusion Matrix

20: This value represents the number of True Negatives (TN). These are instances that were actually bad milk (negative) and were correctly predicted by the model as bad milk (negative).

5: This value represents the number of False Positives (FP). These are instances that were actually bad milk (negative) but were incorrectly predicted by the model as good milk (positive).

3: This value represents the number of False Negatives (FN). These are instances that were actually good milk (positive) but were incorrectly predicted by the model as bad milk (negative).

12: This value represents the number of True Positives (TP). These are instances that were actually good milk (positive) and were correctly predicted by the model as good milk (positive).

```
# Random Forest model
model_rfc = RandomForestClassifier(n_estimators=100)
model_rfc.fit(X_train, y_train)
y_pred_rfc = model_rfc.predict(X_test)
accuracy_rfc = metrics.accuracy_score(y_test, y_pred_rfc)
print('Accuracy of the Random Forest model (Test Dataset):',
accuracy_rfc)
```

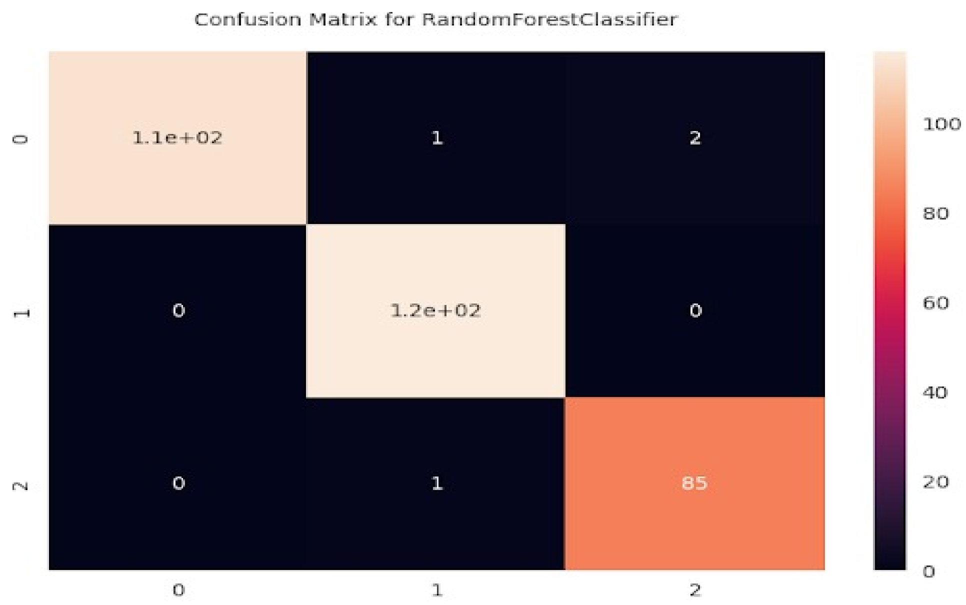

Figure 5.2: Random Forest Confusion Matrix

1.1e+02 and 1.2e+02: These numbers likely represent the number of trees in the forest used by the Random Forest classifier. "e+02" signifies scientific notation, so these translate to 110 and 120 trees, respectively. They indicate the model used multiple decision trees to make predictions and increase its overall accuracy.

85: This number probably represents the number of features used by the classifier. It tells you that the model considered 85 different attributes in your data to make predictions.

Chapter 6

Results and Discussions

The data is trained with different models as shown below and obtained the following accuracies

Method	**Accuracy**
Logistic Regression	85.2%
Random Forest	99.8%

6.0: Analysis of used methods

Classification Report for Logistic Regression (Test Dataset):

	precision	recall	f1-score	support
0	0.72	0.84	0.78	76
1	0.86	0.87	0.87	115
2	0.94	0.83	0.88	127
accuracy			0.85	318
macro avg	0.84	0.85	0.84	318
weighted avg	0.86	0.85	0.85	318

Figure 6.1: Classification Report for Training data and Testing data

Classification Report for Random Forest (Test Dataset):

	precision	recall	f1-score	support
0	0.99	1.00	0.99	76
1	1.00	0.99	1.00	115
2	1.00	1.00	1.00	127
accuracy			1.00	318
macro avg	1.00	1.00	1.00	318
weighted avg	1.00	1.00	1.00	318

Figure 6.2: Classification Report for Training data and Testing data

Chapter 7

Conclusions and future works

7.1 Conclusion

"In this project, a comprehensive solution was developed utilizing advanced technologies for the prediction of milk quality using IoT and ML. The proactive system not only identifies potential issues related to milk quality but also offers timely interventions, showcasing versatile applications across various sectors beyond dairy for heightened quality assurance and safety in critical environments. This innovative approach demonstrates the potential for leveraging IoT and ML to ensure the quality and safety of products in diverse industries".

7.2 Future Scope

"In the future, advancements in the prediction of milk quality through IoT and ML will focus on refining predictive models for enhanced accuracy and early detection of potential issues. Integration with emerging technologies like blockchain could establish transparent and traceable supply chains, assuring consumers of milk quality and authenticity. Additionally, the incorporation of sensor innovations and edge computing may enable real-time data processing at the source, reducing latency and improving the system's responsiveness. Collaborations with industry stakeholders and regulatory bodies will be crucial to establish standardized protocols, ensuring the widespread adoption of these predictive solutions and fostering a more resilient and quality-driven dairy industry."

Bibliography

[1] A. ÇELİK, “Using machine learning algorithms to detect milk quality.,” *urasian Journal of Food Science and Technology*, 2022.

[2] V. V. L. Kilari Jyothi, “Iot based detection of aulteration in milk,” *IEEE*, vol. Volume 10, 2023.

[3] e. a. Zhang, Yang, “Research on dairy products detection based on machine learning algorithm.,” *MATEC Web of Conferences.*, vol. Vol. 355, 2022.

[4] D. O. León, Mia, “Machine learning applied to milk sample classification.,” 2022.

[5] F. U. Habsari, W. 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. Vol. 1063., 2022.

[6] D. S. B. V. B. M Kiran Kumar1, “Smart system for milk quality analysis and billing system,” vol. vo.l.09., 2022.

[7] e. a. Frizzarin, Maria, “Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods.,” *journal of Dairy Science 104.7*, pp. 7438–7447., 2021.

[8] e. a. Akbar, Muhammad Osama, “Iot for development of smart dairy farming,” *Journal of Food Quality 2020*, pp. 1–8, 2020.

[9] e. a. Rajakumar, G., “iot based milk monitoring system for detection of milk adulteration,” *IEEE*, 2017.

[10] S. T. M. M. S. K. M. N. M. Kawasaki, Masataka; Kawamura, “Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot,” 2008.