**MILK QUALITY PREDICTION BY USING SUPERVISED MACHINE LEARNING AND IOT**

### A PROJECT REPORT

***Submitted by***

|  |  |
| --- | --- |
| **ANITHA S** | **[REGISTER NO: 211419104012]** |
| **GAYATHRI M** | **[REGISTER NO: 211419104076]** |
| **HINDUJA N** | **[REGISTER NO: 211419104104]** |

***in partial fulfillment for the award of the degree of***

## BACHELOR OF ENGINEERING

**IN**

**COMPUTER SCIENCE AND ENGINEERING**



## PANIMALAR ENGINEERING COLLEGE

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

### APRIL 2023

**PANIMALAR ENGINEERING COLLEGE**

**(An Autonomous Institution, Affiliated to Anna University, Chennai)**

## BONAFIDE CERTIFICATE

Certified that this project report “**MILK QUALITY PREDICTION BY USING SUPERVISED MACHINE LEARNING AND IOT**” is the bonafide work of **ANITHA S [211419104012], GAYATHRI M [211419104076] , HINDUJA N**

**[211419104104]** who carried out the project work under my supervision.

### SIGNATURE SIGNATURE

**Dr. L. JABASHEELA, M.E., Ph.D., Dr.T.JACKULIN M.E.,Ph.D., HEAD OF THE DEPARTMENT SUPERVISOR**

### PROFESSOR

DEPARTMENT OF CSE, DEPARTMENT OF CSE,

PANIMALAR ENGINEERING COLLEGE, PANIMALAR ENGINEERING COLLEGE, NASARATHPETTAI, NASARATHPETTAI,

POONAMALLEE, POONAMALLEE,

CHENNAI-600 123. CHENNAI-600 123.

Certified that the above mentioned students were examined in the End Semester Project Viva-Voce Examination held on...........................

**INTERNAL EXAMINER EXTERNAL EXAMINER**

## DECLARATION

We… **ANITHA S [211419104012], GAYATHRI M [211419104076],**

**HINDUJA N [211419104104]**…. hereby declare that this project report titled “**MILK QUALITY PREDICTION BY USING SUPERVISED MACHINE**

**LEARNING AND IOT**” , under the guidance of **Dr. T. JACKULIN M.E, Ph.D.,**

is the original work done by us and we have not plagiarized or submitted to any other degree in any university by us.

## ANITHA S GAYATHRI M HINDUJA N

**ACKNOWLEDGEMENT**

We would like to express our deep gratitude to our respected Secretary and Correspondent **Dr. P. CHINNADURAI, M.A., Ph.D.,** for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We express our sincere thanks to our beloved Directors **Tmt. C. VIJAYARAJESWARI**, **Dr. C. SAKTHI KUMAR, M.E., Ph.D.,** and **Dr. SARANYASREE SAKTHI KUMAR B.E., M.B.A., Ph.D.,** for providing us

withthe necessary facilities to undertake this project.

We also express our gratitude to our Principal **Dr. K. MANI, M.E., Ph.D.,** who facilitated us in completing the project.

We thank the Head of the CSE Department, **Dr. L. JABASHEELA , M.E., Ph.D.,** for the support extended throughout the project.

We would like to thank our guide **Dr. T. JACKULIN M.E, Ph.D.,**and all the faculty members of the Department of CSE for their advice and encouragement for the successful completion of the project.

## ANITHA S GAYATHRI M HINDUJA N

**ABSTRACT**

For milk quality analysis, several contaminants like water, whey, and urea are utilized, along with some other dangerous equipment. Although these instruments are precise, they are difficult to use in the field, expensive, and need time-consuming expert processes to obtain a reading. The literature includes reports on portable sensors-based instruments that study direct/indirect transduction events. For detecting milk contaminants, 1 conductive sensor with selective sensing films are most frequently utilized, however these sensors need calibration and exhibit drift over time. A simple functional detecting system that is quick, accurate, sensitive, and cost-effective is required. In this paper, an accurate and straightforward functioning prototype model of an across-conductance sensor for milk is presented together with its theory. In order to determine the portability of milk, machine learning is implemented by using algorithm.

**KEYWORDS**: AC Cross Conductance Sensor, Adulteration, Milk, Response characteristics, Algorithms.

### LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| **TABLE NO** | **NAME OF THE TABLE** | **PAGENO** |
| **A1** | Sample Dataset | **41** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **NAME OF THE FIGURE** | **PAGE NO.** |
| 4.1 | Use Case Diagram | 16 |
| 4.2 | Sequence Diagram | 17 |
| 4.3 | State Chart Diagram | 18 |
| 4.4 | Data Flow Diagram – level 0 | 19 |
| 4.5 | Data Flow Diagram – level 1 | 19 |
| 5.1 | System Architecture | 22 |
| 5.2 | Work flow Diagram | 23 |
| 5.3 | Block Diagram | 24 |
| 7.1 | Screenshot of User Login Web Page | 34 |
| 7.2 | Screenshot of Input Values taken from IOT module | 34 |
| 7.3 | Screenshot of Input Values taken from IOT module | 35 |
| 7.4 | Screenshot of Grade Predictions | 35 |
| 7.5 | Screenshot of Admin / Officer login | 36 |
| 7.6 | Screenshot of Output Report and Admin Feedback | 36 |
| 7.7 | Screenshot of Milk Quality Feedback Received By User | 36 |
| 7.8 | Screenshot of Accuracy of Random Forest Algorithm | 37 |
| 7.9 | Screenshot of Accuracy of Logistic Regression Algorithm | 37 |

VII

**FIGURE NO. NAME OF THE FIGURE PAGE.NO**

|  |  |  |
| --- | --- | --- |
| 7.10 | Screenshot of Accuracy of XG Boost Classifier Algorithm | 38 |
| 7.11 | Screenshot of Accuracy of MLP Classifier Algorithm | 38 |
| B2.1 | Data Preprocessing Coding | 53 |
| B2.2 | Data Visualization Coding | 53 |
| B2.3 | Random Forest Algorithm Coding | 54 |
| B2.4 | Logistic Regression Algorithm Coding | 54 |
| B2.5 | XG Boost Classifier Algorithm Coding | 54 |
| B2.6 | MLP Classifier Algorithm Coding | 54 |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** |  |
|  | **LIST OF FIGURES** | vii |
| **1** | **INTRODUCTION** | 1 |
|  | 1.1 Overview | 2 |
|  | 1.2 Problem Definition | 2 |
| **2** | **LITERATURE SURVEY** | 3 |
|  | 2.1 Introduction | 4 |
|  | 2.2 Literature survey | 4 |
| **3** | **SYSTEM ANALYSIS** | 10 |
|  | 2.1 Existing System | 11 |
|  | 2.2 Proposed System | 11 |
|  | 2.3 Feasibility Study | 12 |
|  | 2.4 Hardware Requirements | 14 |
|  | 2.5 Software Requirements | 14 |
| **4** | **SYSTEM DESIGN** | 15 |
|  | 4.1 UML Diagrams | 16 |
|  | 4.1.1 Use Case Diagram | 16 |
|  | 4.1.2 Sequence Diagram | 17 |
|  | 4.1.3 State Chart Diagram | 18 |
|  | 4.2 Data Flow Diagram | 19 |
| **5** | **SYSTEM ARCHITECTURE** | 20 |
|  | * 1. System Architecture   2. Work Flow Diagram   3. Block Diagram | 21  22  23 |

|  |  |  |
| --- | --- | --- |
| **CHAPTER.NO** | **TITLE** | **PAGE.NO** |
| **6** | **SYSTEM IMPLEMENTATION** | 24 |
|  | 6.1 Module Description | 25 |
|  | 6.1.1 Data Preprocessing | 25 |
|  | 6.1.2 Data visualization | 26 |
|  | 6.1.3 ML Model Development | 27 |
|  | 6.1.3.1 Random Forest Algorithm | 28 |
|  | 6.1.3.2 Logistic Regression | 28 |
|  | 6.1.3.3 XG Boost Classifier Algorithm | 29 |
|  | 6.1.3.4 MLP Classifier Algorithm | 30 |
|  | 6.1.3.5 Deployment Using Flask | 31 |
| **7** | **PERFORMANCE EVALUATION** | 32 |
|  | 7.1 Results and Discussions | 33 |
|  | 7.2 Comparative Analysis | 36 |
| **8** | **CONCLUSION** | 38 |
|  | 8.1 Conclusion and Future Enhancement | 39 |
|  | **APPENDICES** | 40 |
|  | A.1 Sample Dataset | 40 |
|  | A.2 Sample Coding | 42 |
|  | 6.1 Module 1 | 42 |
|  | 6.2 Module 2 | 43 |
|  | 6.3 Module 3 | 44 |
|  | 6.4 Module 4 | 46 |
|  | 6.5 Module 5 | 47 |
|  | 6.6 Module 6 | 48 |
|  | 6.7 Module 7 | 49 |
|  | **REFERENCES** | 54 |

# CHAPTER 1 INTRODUCTION

### CHAPTER 1 INTRODUCTION

* 1. **OVERVIEW**

Milk adulteration by various impurities, such as water, whey, and urea, and some other harmful instruments are used for milk quality analysis. These instruments are accurate but these are complex, expensive, inconvenient for field use, and require laborious skilled procedures to get the reading. Portable sensors- based instruments working on direct/indirect transduction phenomena are also reported in the literature. Most commonly, the conductive sensors with selective sensing films are used for detecting the milk impurities but such sensors require calibration and show drift due to aging. There is a need to develop a rapid, accurate, sensitive, and cost effective simple working detection system. This article presents the theory, design, fabrication, and test results of an accurate and simple working prototype model of an across-conductance sensor for milk adulteration detection.

### PROBLEM DEFINITION

Milk Quality is one of the major important factors in healthcare domain. There are lot of people who are actively having milk. So it is necessary to know the quality of the milk. Most commonly, the conductive sensors with selective sensing films are used for detecting the milk impurities but such sensors require calibration and show drift due to aging. The goal is to develop a machine learning model for Milk Quality Prediction, to potentially replace the updatable supervised Machine Learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm. So the project can easily find out the milk quality.

# CHAPTER 2 LITERATURE SURVEY

### CHAPTER 2 LITERATURE SURVEY

* 1. **INTRODUCTION**

A literature review is a body of text that aims to review the critical points of current knowledge on and/or methodological approaches to a particular topic. It is secondary sources and discuss published information in a particular subject area and sometimes information in a particular subject area within a certain time period. Its ultimate goal is to bring the reader up to date with current literature on a topic and forms the basis for another goal, such as future research that may be needed in the area and precedes a research proposal and may be just a simple summary of sources.

### LITERATURE SURVEY

* + 1. **A Direct AC Cross Conductive Sensor For Milk Quality Measurement Author(s):** Tarikul Islam, Arshi Salamat, Sandeep K.Singh, Mahfoozur Rehman **Year:** 2022

**Description:** This article presents the theory,design,fabrication, and test results of an accurate and simple working prototype model of an ac cross conductance sensor for milk adulteration detection.The sensor was fabricated and tested in the laboratory environment to analyze the quality of milk having different amount of impurities, such as water,whey and urea.The sensor can differentiate between pure and adulterated milk with significant resolution and precision.

### Milk Adulterant Detection

**Author(s):** Rupak Nagraik,Avinash Sharma,Deepak Kumar

**Year:** 2021

**Description:** Milk adulteration is a social problem. It exists both in the backward and advanced countries. Consumption of adulterated milk causes serious health problems and a great concern to the food industry.This article represents the conventional biosensor based approaches,Biosensor is an analytical device which is composed of three components :a bio-receptor,a transducer and a signal reading device.It was developed for label free detection of melamine in raw milk.

* + 1. **Simultaneous Detection of Fat Content and Adulteration Author(s):** Moupali Chakraborty,Karabi Biswas

**Year:** 2020

**Description:** This article reports an integrated milk-testing instrument for the simultaneous detection of fat content and adulterants in milk.The prototype consists of a small temperature control system and an integrated signal conditioning circuit for fat and adulteration detection.The proposed instrument is also compared with the commercially available milk-testing instruments to understand the market challenges. In addition to this,a study is carried out to understand the chemical kinetics behavior of the sensors in milk for the selection of operating frequency. Power consumed by the modules is measured to improve the design of the instrument for fieldwork.

### Use of a Smartphone for visual detection of melamine in milk based on Au @Carbon quantum dots nanocomposites

**Author(s):** Muhammad Arslan,Xueato Hu

**Year:** 2019

**Description:** Gold nanoparticles@Carbon quantum dots nanocomposites (Au@CQDs) were designed for analyzing melamine in milk visually. Fluorescent emission of Au@CQDs enhanced with the increase of melamine concentration. A calibration curve (R2 = 0.9856) and fluorescence standard array were established for detection of melamine in the range of 1 μM–10 μM. Limit of quantification and limit of detection were 12 nM and 3.6 nM, respectively. The approximate concentration of melamine adulterated in milk samples were detected visually by the fluorescence standard array and smartphone.

### Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis

**Author(s):** Francesca Trimboli,Nicola Costanzo

**Year:** 2019

**Description:** The addition of cow milk during the production of buffalo mozzarella is a common fraud in dairy industries because of the lower price and greater availability of cow milk throughout the year. The aim of this study was to develop a new, rapid, and robust [capillary electrophoresis](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/capillary-electrophoresis) method for detecting and quantifying cow milk in buffalo milk by exploiting cow α-lactalbumin as a marker of adulteration. In particular, a linear calibration curve was generated, using a training set of calibrators consisting of 7 series of 17 buffalo/bovine whey mixtures, obtained after casein precipitation, with increasing percentages of cow whey. Moreover, the minimum amount of detectable fraudulent cow milk was 1%, and the limit of quantification was 3.1%.

* + 1. **A Thin Film Porous Alumina Based Cross Capacitive Humidity Sensor Author(s):** Zubair Hassan Zargar,Tarikul Islam

**Year:** 2019

**Description:** In this paper, the principle of the cross-capacitance proposed by Thompson-Lampard theorem has been utilized for the first time to fabricate a precise and accurate relative humidity (%RH) sensor. A thin film of porous alumina (γ -Al2O3) was deposited on the inner wall of the quartz tube for adsorbing humidity. The four symmetrical silver electrodes of the cross-capacitor are formed around the quartz tube by the screen printing technique. The main advantage of the cross-capacitor is the single-dimensional accuracy. The sensitivity of the sensor for humidity without any hydrophilic film was found to be less.

### In-Line Technologies for the analysis of important Milk parameters during the Milking Process

**Author(s):** Radim Kunes,Petr Bartos,Pavel Kriz

**Year:** 2021

**Description:** Milk quality can be affected by external conditions, such as temperature and contamination in the feedstock; by management practices, such as hygiene, milking frequency, treatment, and feedstuff quality; and by diseases, genetics, or age. Somatic cell count, electric conductivity, and contents of urea, fat, protein, and lactose were reviewed as likely parameters of milk representing its quality with respect to feedback for consumers and breeders. Methods for evaluating milk constituents and parameters are still being developed to provide in-line information. These methods allow the avoidance of enormous economic losses every year caused by milk discard, health treatments, or cow replacements.

### Genetic Relationships of lactose and freezing point with minerals and coagulation traits predicted from milk mid-infrared spectra in Holstein cows

**Author(s):** A.Costa,G.Visentin

**Year:** 2019

**Description:** The aim of the present study was to assess the relationships of lactose percentage (LP), lactose yield (LY), and freezing point (FRP) with minerals and coagulation properties predicted from mid-infrared spectra in [bovine milk](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/bovine-milk). To achieve this purpose, we analyzed 54,263 test-day records of 4,297 [Holstein](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/holstein) cows to compute (co)variance components with a linear repeatability animal model. Parity, stage of lactation, season of calving, and herd- test-date were included as fixed effects in the model, and additive [genetic animal](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/animal-genetics), within- and across-lactation permanent environment, and residual were included as random effects. Being strongly associated with milk yield, LY did not favorably correlate with coagulation properties, likely because of the negative correlation of this trait with protein and casein percentages.

* + 1. **A new ultrasonic method to detect chemical additives in branded milk Author(s):** S Mohanan,PG Thomas Panicker

**Year:** 2022

**Description:** A new ultrasonic method – thermoacoustic analysis- is reported for the detection of the added chemical preservatives in branded milk. The nature of variation and shift in the thermal response of the acoustic parameters specific acoustic impedance, adiabatic compressibility and Rao’s specific sound velocity for different samples of branded milk as compared to the chemical added pure milk are explained as due to the presence of chemicals in these branded samples.

### Validation of a fast real-time PCR method to detect fraud and mislabeling in milk and dairy products

**Author(s):** M.Di Domencio,C.Camma

**Year:** 2017

**Description:** The method was validated by means of laboratory-prepared samples mixing different species. Moreover, 18 commercial dairy samples were analyzed by both real-time PCR and [isoelectric focusing](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/isoelectric-focusing), the official European Union reference method. The 4 TaqMan assays were confirmed to be a useful tool for milk and dairy product authentication. An internal amplification control was used to detect possible false negatives.

### Rapid and Automated Measurement of Milk Adulteration using a 3D printed Optofluidic Micro Viscometer

**Author(s):** Abhishek Sharma ,Santhosh,Dubey

**Year:** 2016

**Description:** The working principle is the viscosity-dependent capture of the microchannel width by the fluids flowing inside the microchannel under the laminar flow based on the pressure gradient between the inlets and the outlet. In this paper, around 60 milk samples with various adulteration ratios of various adulterants ranging from 1% to 10% have been tested. A best fit curve for every adulterant was defined, and the device was found to be accurate enough to measure the entire range of adulteration ratios with a high accuracy of 0.95.

# CHAPTER 3 SYSTEM ANALYSIS

### CHAPTER 3 SYSTEM ANAYSIS

* 1. **EXISTING SYSTEM**

Most commonly, the conductive sensors with selective sensing films are used for detecting the milk impurities but such sensors require calibration and show drift due to aging. There is a need to develop a rapid, accurate, sensitive, and cost-effective simple working detection system. This article presents the theory, design, fabrication, and test results of an accurate and simple working prototype model of an across-conductance sensor for milk adulteration detection. As the measured conductance is related to the conductivity by a single length measurement, a precise measurement of conductivity is made possible. For the adulteration of water and milk, the sensor can discriminate among different kinds of adulterants. The fabrication of the sensor is simple and is portable for field use and suitable for household applications at a low cost.

### DISADVANTAGES OF EXISTING SYSTEM

* Machine learning is not implemented.
* Prediction cost will be expensive.
* Deployment has not done**.**

### PROPOSED SYSTEM

The proposed method is to build a machine learning model for the classification of milk quality. The process carries from data collection where past data related to milk quality are collected. Data mining is a commonly used technique for processing enormous data. Machine learning is now applied where it reduces manual effort and a better model makes error less which leads to preventing problems. The data analysis is done on the dataset, proper variable identification is done that is both the dependent variables and independent

variables are found. The proper machine learning algorithms are applied to the dataset where the pattern of data is learned. After applying different algorithms, abetter algorithm is used for the prediction of the outcome.

### ADVANTAGES OF PROPOSED SYSTEM

* + - Machine learning is implemented.
    - More algorithms are compared.
    - Prediction cost will be low.
    - Deployment done.

### FEASIBILITY STUDY DATA WRANGLING

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

### DATA COLLECTION

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms, Support vector classifier (SVC), Multilayer Perceptron are applied on the Training set and based on the test result accuracy, Test set prediction is done.

### PREPROCESSING

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done.

### Building the classification model

The prediction of stellar classification, a high accuracy prediction model is effective because of the following reasons: It provides better results in classification problem.

* It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.
* It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

### Construction of a Predictive Model

Machine learning needs data gathering have lot of past data. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can’t be used directly. It’s used to pre-process then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

### PROJECT REQUIREMENTS FUNCTIONAL REQUIREMENTS

The software requirements specification is a technical specification of requirements forthe software product. It is the first step in the requirements analysis process. It lists requirements of a particular software system. The following details to follow the special libraries like sk-learn, pandas, numpy, matplotlib and seaborn.

### NON-FUNCTIONAL REQUIREMENTS

Process of functional steps,

1. Problem define
2. Preparing data
3. Evaluating algorithms
4. Improving results
5. Prediction the result

### ENVIRONMENTAL REQUIREMENTS

1. **HARDWARE ENVIRONMENT**

Processor : Intel i3

Hard disk : minimum 80 GB

RAM : minimum 2 GB

Microcontroller :Arduino Atmega 328p

Sensors : Ph,Temperature, Turbidity,colour

Electronic Display :LCD

### SOFTWARE ENVIRONMENT

Operating System : Windows 10 or later

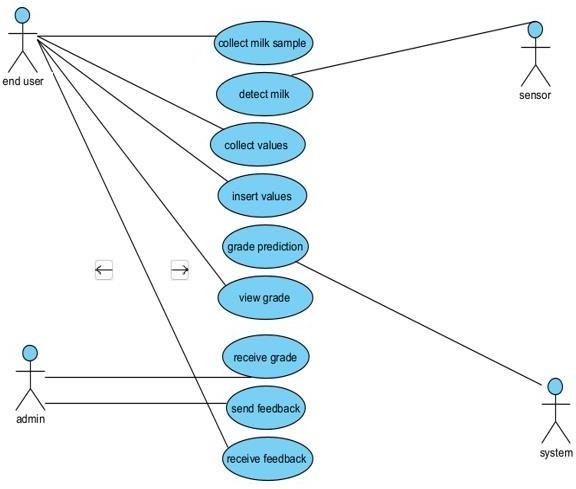
Tool(s) : Anaconda with Jupyter Notebook

# CHAPTER 4 SYSTEM DESIGN

### CHAPTER 4 SYSTEM DESIGN

* 1. **UML DIAGRAMS**

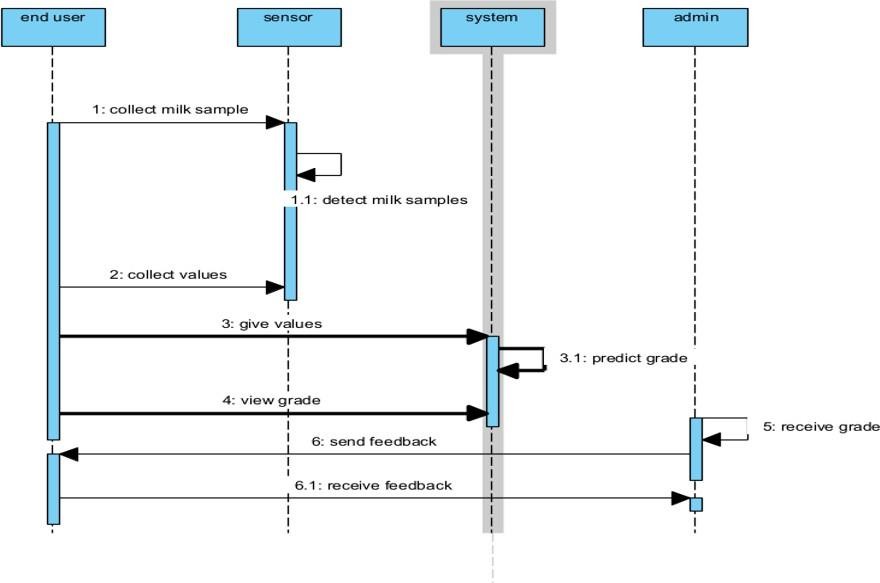
### USE CASE DIAGRAM



**Figure 4.1** Use Case Diagram

The above Use case Diagram Contains end user like people who collects the milk samples,sensor that detects the milk,end user collects the detection value by using sensor and insert to the system and grade prediction can be done by the system based on the trained model. Admin like people who can view the grade of milk and shares the feedback to the user.

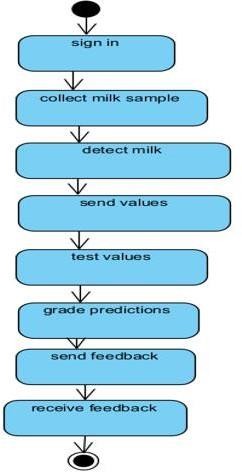
### SEQUENCE DIAGRAM



**Figure 4.2** Sequence Diagram

The above diagram depicts all the sequence of activities in detecting the grade from the input given by the user as well as the IOT module.Initially user collects the milk sample and it can be detected by the sensor.Those detected value can be fed to the system for grade prediction.The predicted grade report can be sent to the admin.The admin can able to view the predicted grade result and can send feedback to the user.

### STATE CHART DIAGRAM

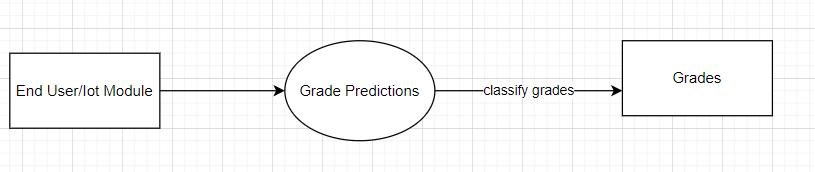


**Figure 4.3** State Chart Diagram

The above diagram depicts the different states in detecting the milk.Initially user and admin needs to sign in the webpage.Milk samples are collected by the user and it can be detected by the sensor.Those detected values are send to the system in order to know the levels of grade.Admin can avle to view the grade and feedback can be sent.

### DATA FLOW DIAGRAM

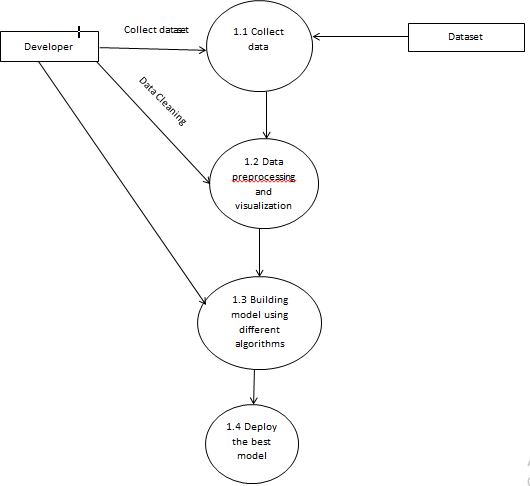
* + 1. **LEVEL 0**



**Figure 4.4** Level 0 Dataflow Diagram

The Level 0 DFD diagram shows that the input will be given as text by the end user and IOT module and the grades are predicted/classified as high,low,medium.

### LEVEL 1



**Figure 4.5** Level 1 Data Flow Diagram

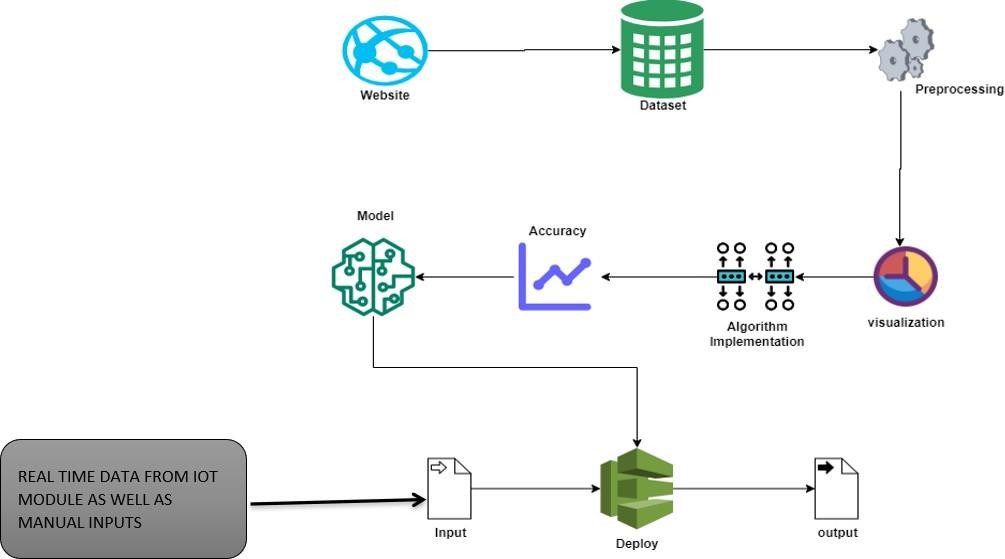
The below diagram shows the flow of data in this project starting from collection of data to deployment.

# CHAPTER 5 SYSTEM ARCHITECTURE

### CHAPTER 5 SYSTEM ARCHITECTURE

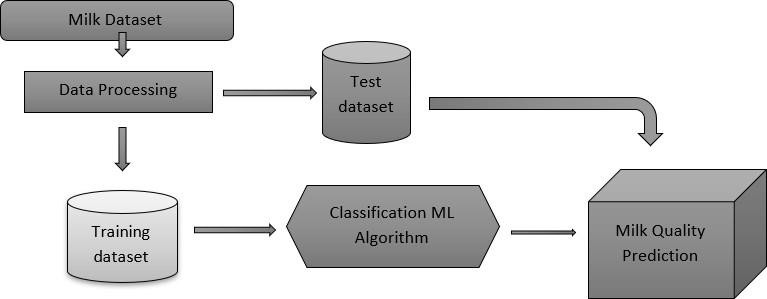
* 1. **Architecture Overview**

The Milk Quality Prediction includes six different stages in order to predict the grade of milk. Generally, the first step will be collecting input from the sensor who wants to know what grade it is. Parallely, the network will have several collection of features stored in dataset for prediction. The data will be preprocessed by removing all the cleaning the raw data and make it as a valuable dataset. Then, the next stage will be data visualization which is done to have a better glance of vast amount of data. It can be either in graphs or in charts. Now the dataset provided by the user will be compared with the four machine learning algorithms such as Logistic Regression, Random Forest, XG Boost and MLP Classifier. After the comparison is done, the system will predict the best algorithm which gives efficient accuracy and builds the model. Then the deployment will be carried out.



**Fig 5.1** System Architecture

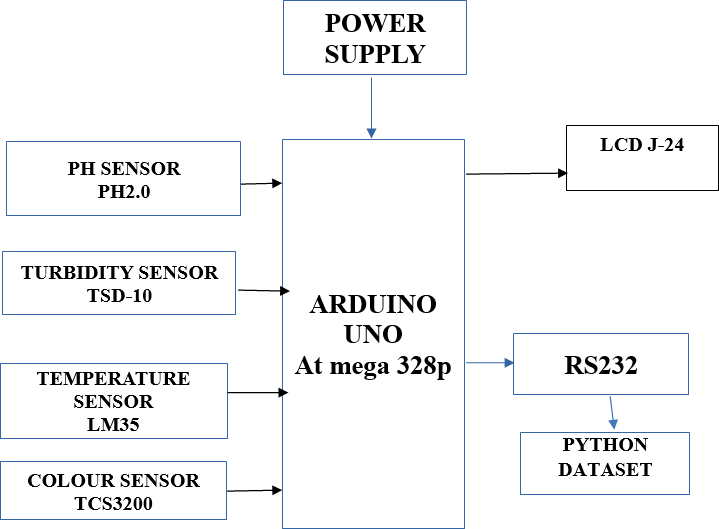
### WORK FLOW DIAGRAM



**Figure 5.2** Work Flow Diagram for Proposed System

The workflow diagram shows the flow of proposed work. That is it starts from dataset collection, then processing the data and cleaning. After splitting it into test and train data, the algorithm is implemented to train the model on grade prediction and then deploying it as webpage by finding the best mode

### BLOCK DIAGRAM



**Fig 5.3** Block Diagram

In this module we get the values from the milk to check the quality with python dataset prediction. We get the parameters for the prediction with sensors like ph. turbidity, colour, and temperature sensors. The PH sensor is used to check the PH level of the milk. The colour sensor is used to check the colour of the milk. Temperature for checking the temperature of the milk and the turbidity for checking the milk is dirty or not, and we send the data’s from the sensors to the python dataset for milk quality prediction. We send the datas with the help of RS232 protocol.

# CHAPTER 6

**SYSTEM IMPLEMENTATION**

### CHAPTER 6

**SYSTEM IMPLEMENTATION**

### MODULE DESCRIPTION LIST OF MODULES

* + - Data Pre-processing
    - Data Analysis of Visualization
    - Implementing Random Forest Algorithm
    - Implementing Logistic Regression Algorithm
    - Implementing XG Boost Classifier Algorithm
    - Implementing MLP Classifier Algorithm
    - Deployment Using Flask

### DATA PRE-PROCESSING

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.Here are some typical reason why data is missing,

* User forgot to fill in a field.
* Data was lost while transferring manually from a legacy database.
  + - * There was a programming error.
      * Users chose not to fill out a field tied to their beliefs about how the results would be used or interpreted.

### DATA CLEANING

A number of different data cleaning tasks using Python’s [Panda library](https://pandas.pydata.org/) and specifically, it focus on probably the biggest data cleaning task, missing values and it able to more [quickly clean data](https://www.dataoptimal.com/data-cleaning-with-python-2018/). The time for cleaning data should me less, and more time should be spent for exploring and modeling. It is the process of cleaning the imbalanced data, unwanted data, duplicate data, missing values.It involves

* + - * + Tokenising the data
        + Removing stopwords
        + Eliminating punctuation
        + Encoding the data

### DATA VISUALIZATION

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. Thiscan be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end. In this project, we use “matplotlib” and “seaborn” library for data visualization. Word clouds are a type

of data visualization that can be used to visualize the most commonly used words in a text.

### ML MODEL DEVELOPMENT

After preprocessing and visualization of data, algorithm implementation takes place. It starts with importing necessary libraries such as pandas, regular expressions, stopwords and WordNetLemmatizer from the Natural Language Toolkit (nltk) for text preprocessing.The preprocessing method is the same for all the four algorithms. Then it loads the dataset using Pandas and preprocesses the text data using the function preprocess\_text(), which converts the text to lowercase, removes punctuation and digits, removes stopwords and lemmatizes the words. After preprocessing, the text data is converted to numerical vectors using TfidfVectorizer, which is a method for converting text data into a numerical form that can be used by machine learning algorithms. Then, the data is split into training and testing sets using train\_test\_split from scikit-learn, and data is trained on the by using the corresponding libraries and functions. For example, Decision Tree Classifier uses “from sklearn.tree import DecisionTreeClassifier”. Next, the trained model is used to predict the emotions on the testing data and evaluate the performance using metrics like accuracy, precision, recall, and f1-score.

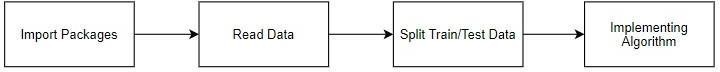
The below four algorithms compared

* + - * Random Forest Algorithm
      * Logistic Regression Algorithm
      * XG Boost Classifier Algorithm
      * MLP Classifier Algorithm

### RANDOM FOREST ALGORITHM

The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample. It can perform both regression and classification tasks. A random forest produces good predictions that can be understood easily. It can handle large datasets efficiently. The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

Random forests is great with high dimensional data since we are working with subsets of data. It is faster to train than decision trees because we are working only on a subset of features in this model, so we can easily work with hundreds of features.

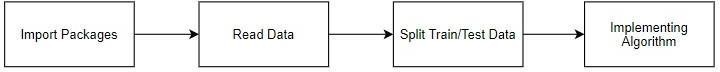


### LOGISTIC REGRESSION

It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

In other words, the logistic regression model predicts P(Y=1) as a function of X. Logistic regression Assumptions:

* + - * + Binary logistic regression requires the dependent variable to be binary.
        + For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
        + Only the meaningful variables should be included.
        + The independent variables should be independent of each other. That is, the model should have little.
        + The independent variables are linearly related to the log odds.
        + Logistic regression requires quite large sample sizes.

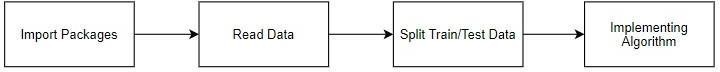


### XG BOOST CLASSIFIER ALGORITHM

[XG Boost](https://xgboost.ai/), which stands for Extreme Gradient Boosting, is a scalable, distributed [gradient-boosted](https://en.wikipedia.org/wiki/Gradient_boosting) decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

It’s vital to an understanding of XG Boost to first grasp the machine learning concepts and algorithms that XG Boost builds upon: supervised machine learning, decision trees, ensemble learning, and [gradient boosting](https://developer.nvidia.com/blog/gradient-boosting-decision-trees-xgboost-cuda/).

Supervised machine learning uses algorithms to train a model to find patterns in a dataset with labels and features and then uses the trained model to predict the labels on a new dataset’s features.

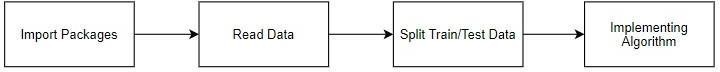


### MLP CLASSIFIER ALGORITHM

Multilayer [perceptron](https://www.sciencedirect.com/topics/engineering/perceptron) (MLP) is a supplement of feed forward [neural](https://www.sciencedirect.com/topics/chemical-engineering/neural-network) [network](https://www.sciencedirect.com/topics/chemical-engineering/neural-network). It consists of three types of layers—the input layer, output layer and hidden layer. The input layer receives the input signal to be processed. The required task such as prediction and classification is performed by the output layer.

An arbitrary number of hidden layers that are placed in between the input and output layer are the true computational engine of the [MLP](https://www.sciencedirect.com/topics/engineering/perceptron). Similar to a [feed](https://www.sciencedirect.com/topics/computer-science/feedforward-network) [forward network](https://www.sciencedirect.com/topics/computer-science/feedforward-network) in a MLP the data flows in the forward direction from input to output layer.

The neurons in the MLP are trained with the [back propagation](https://www.sciencedirect.com/topics/computer-science/backpropagation) learning algorithm. MLPs are designed to approximate any continuous function and can solve problems which are not linearly separable. The major use cases of MLP are pattern classification, recognition, prediction and approximation.



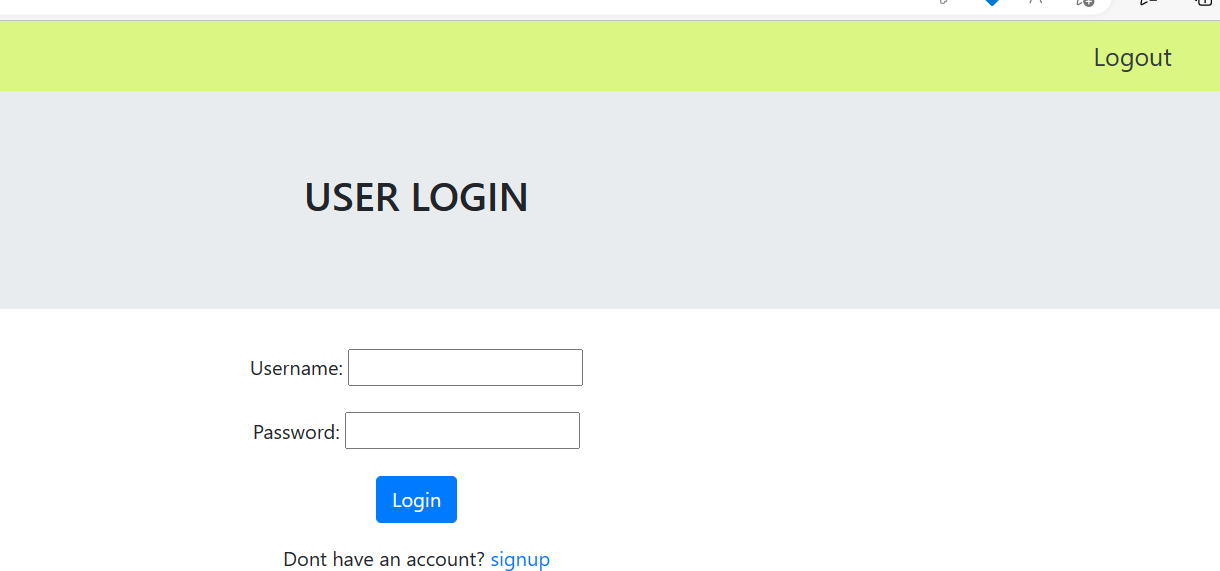
### DEPLOYMENT USING FLASK

Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It is free and open source, has a thriving and active community, great documentation, and many options for free and paid-for support.

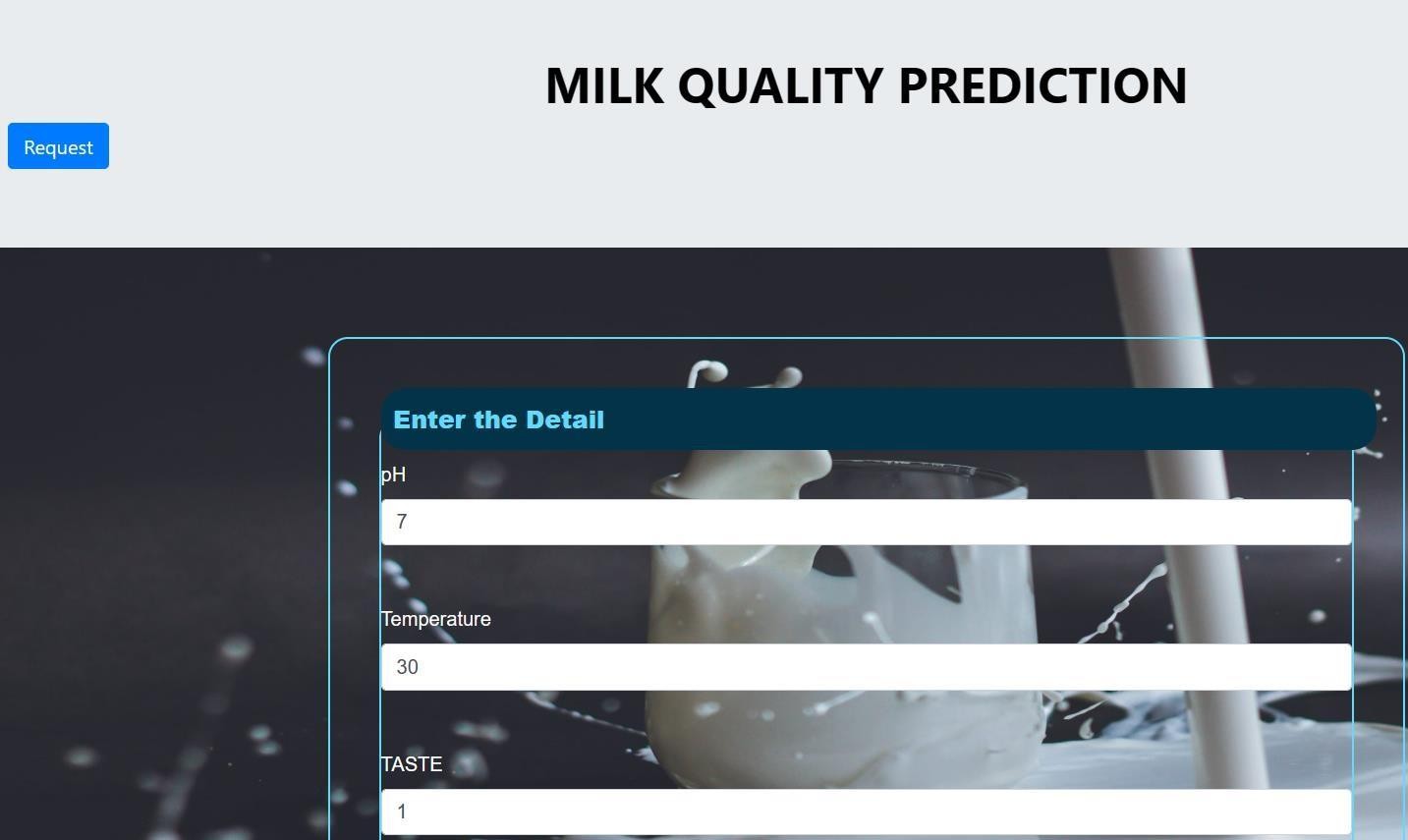
# CHAPTER 7 PERFORMANCE EVALUATION

### CHAPTER-7 PERFORMANCE EVALUATION

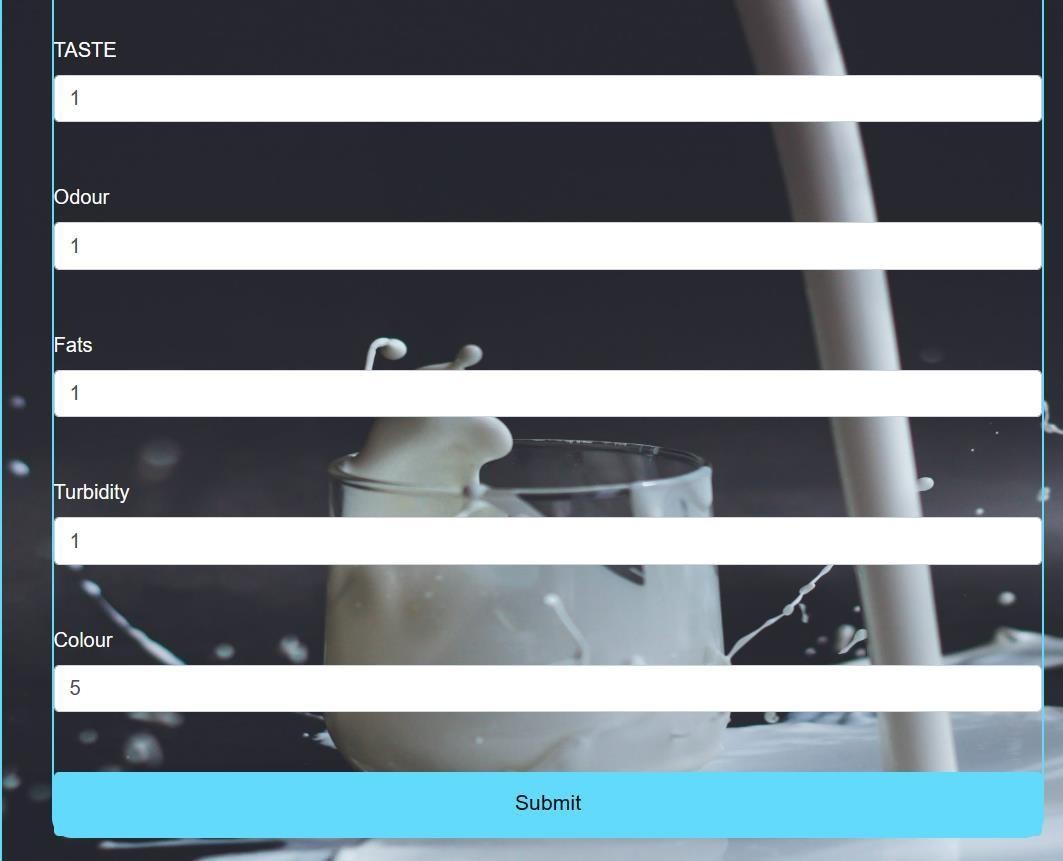
* 1. **RESULTS AND DISCUSSIONS**



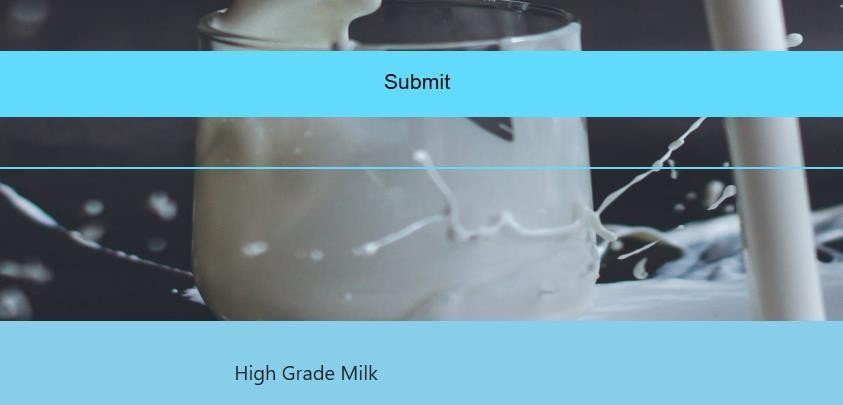
**Fig 7.1** Screenshot of User Login Web Page



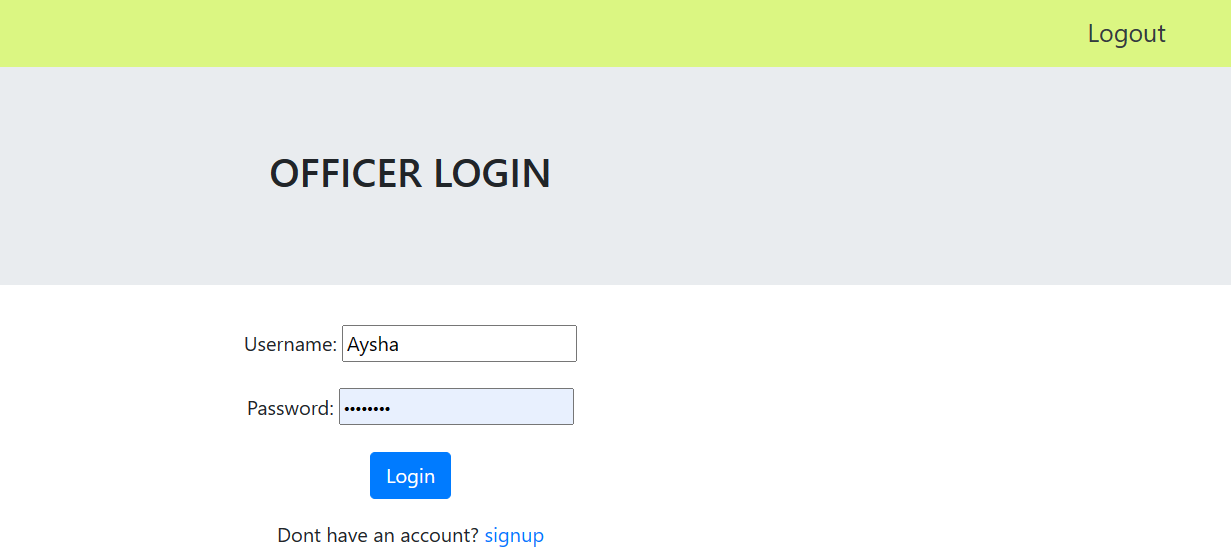
**Fig 7.2** Screenshot of Input Values taken from IOT module



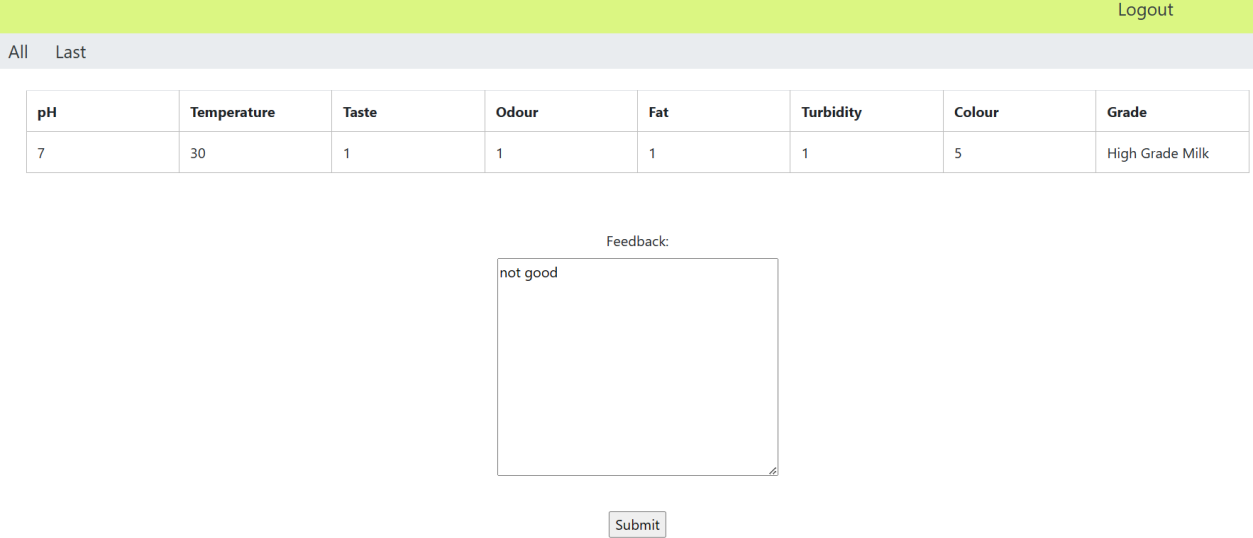
**Fig 7.3** Screenshot of Input Vlaues taken from IOT module



**Fig 7.4** Sreenshot of Grade Predictions



**Fig 7.5** Screenshot of Admin/Officer Login

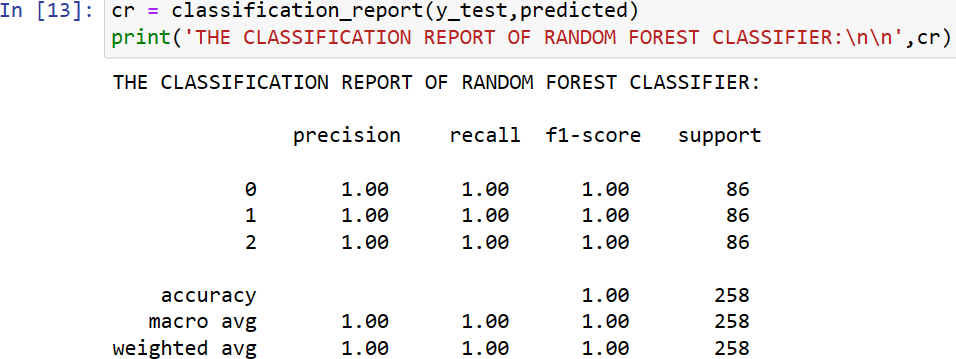


**Fig 7.6** Screenshot of Output Report and Admin Feedback



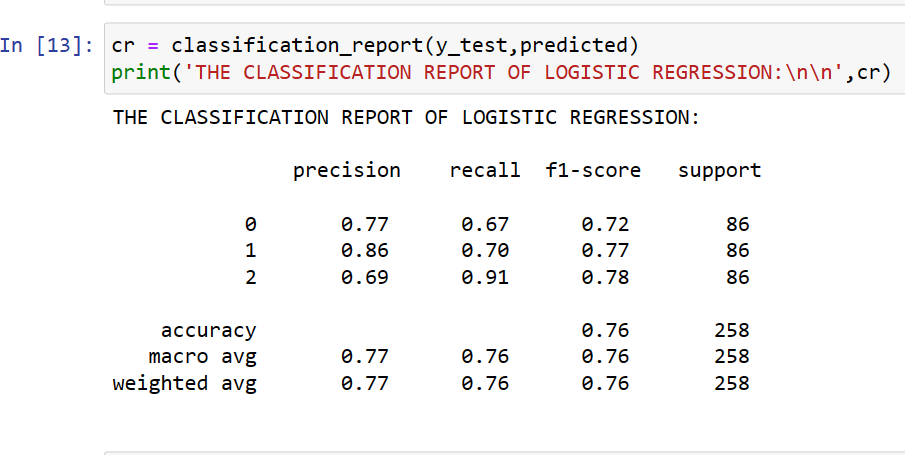
**Fig 7.7** Screenshot of Milk Quality Feedback Received By User

### COMPARTIVE ANALYSIS



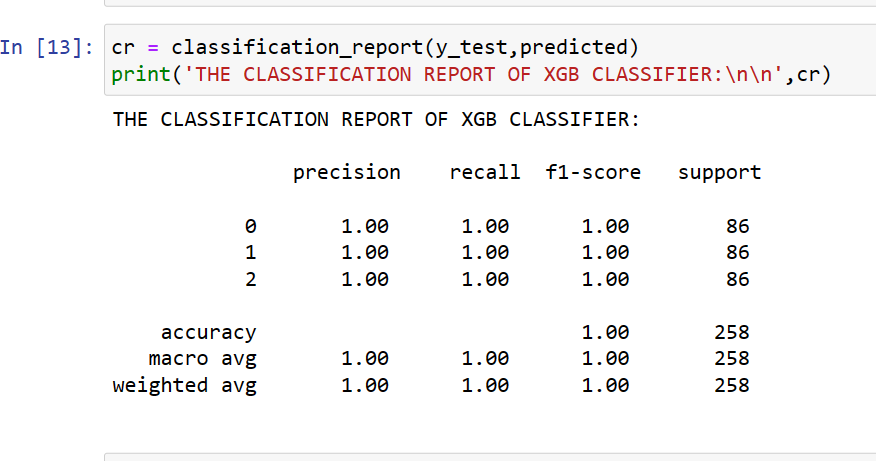
**Fig 7.8** Screenshot of Accuarcy of Random Forest Algorithm

The above figure shows the accuracy,precision,recall,f1-score obtained by training themodel using Random Forest Algorithm. The accuracy obtained is 100%.



**Fig 7.9** Screenshot of Accuarcy of Logistic Regression Algorithm

The above figure shows the accuracy,precision,recall,f1-score obtained by training model using Logistic Regression Algorithm.The accuracy obtained is 75%.



**Fig 7.10** Screenshot of Accuarcy of XG Boost Classifier Algorithm

The above figure shows the accuracy ,precision,recall,f1-score obtained by trained model using XG Boost Classifier Algorithm.The accuracy obtained is 100%.

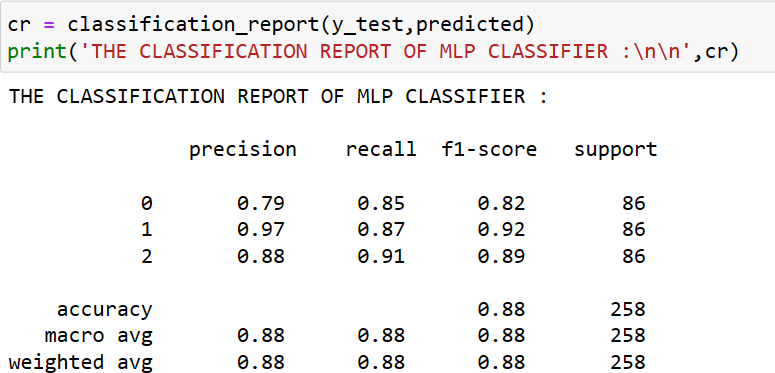


Fig 7.11 Screenshot of Accuarcy of MLP Classifier Algorithm

The above figure shows the accuracy,precision,recall,f1-score obtained by trained model using MLP Classifier Algorithm.The accuracy obtained is 87%.

# CHAPTER 8 CONCLUSION

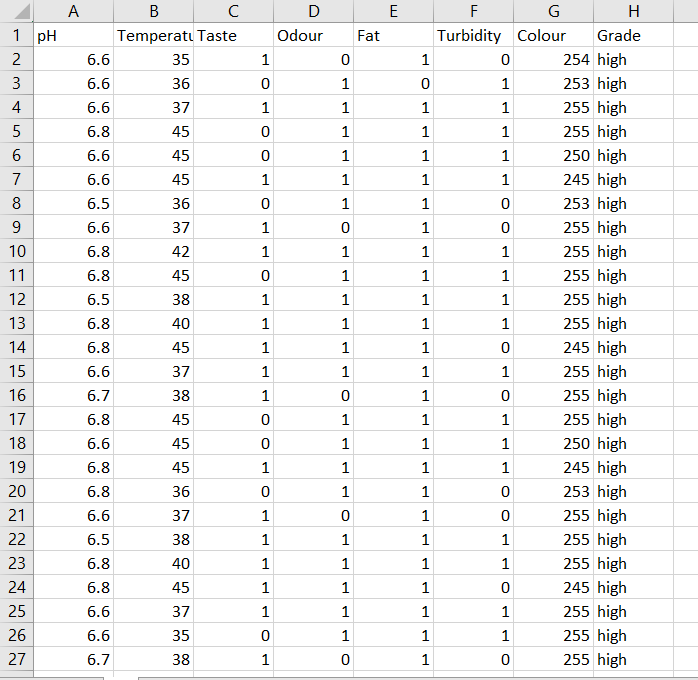
### CHAPTER 8 CONCLUSION

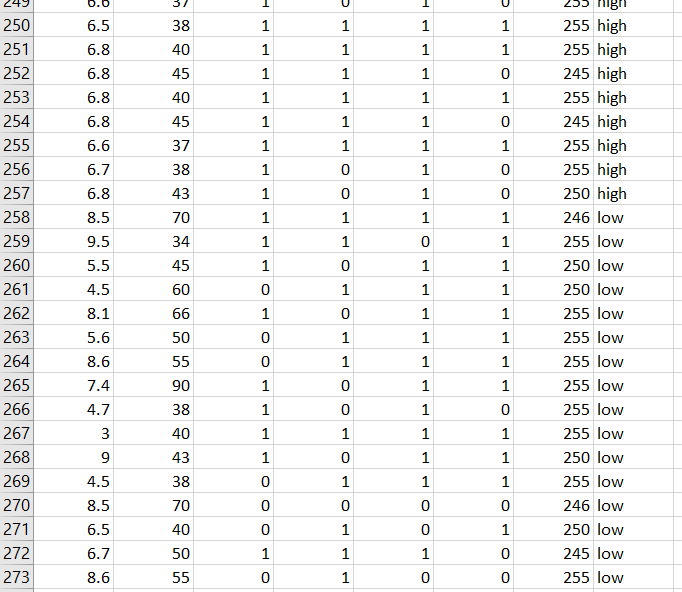
**8.1 CONCLUSION AND FUTURE ENHANCEMENT**

In this work, we proposed an analysis for attributing milk samples based on their behavioral characteristics/features following their activities. The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score will be find out. This application can help to find the Prediction of Milk Quality. This way can help the organisations to know more about the quality of milk and also will be able to identify the grades of milk.The future works can be implemented on Crypto Milk Quality prediction which can be connected with the cloud model. The optimization of this work can also be better if the system and the prediction model is connected with the embedded system.

### APPENDICES

**A.SAMPLE DATASET**





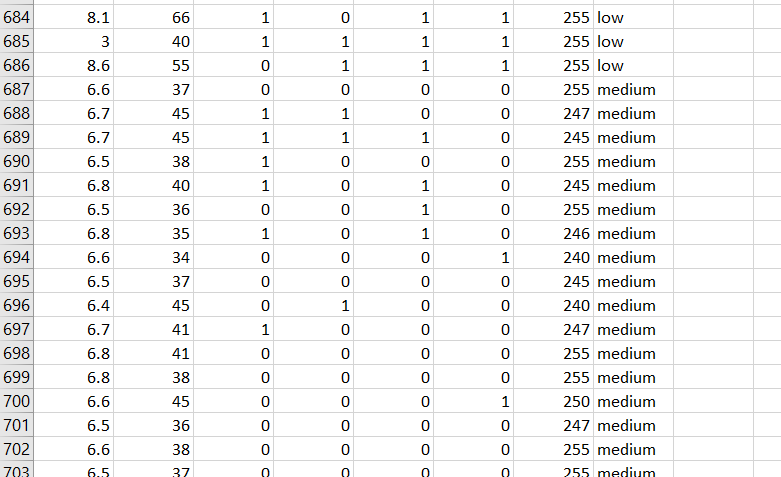


TABLE A1 SAMPLE DATASET

### B1.SAMPLE CODING

* 1. **MODULE -1**

### DATA PREPROCESSING

import pandas as p import numpy as n import warnings

warnings.filterwarnings('ignore') df=p.read\_csv('A.csv')

df.head()

df.tail() df.shape df.size df.isnull()

df['Grade'].unique() df = df.dropna()

df = df.drop\_duplicates() sum(df.duplicated()) p.Categorical(df["Colour"]).describe() df["Grade"].value\_counts() df.groupby(["Turbidity","Colour"]).groups p.crosstab(df.Odour, df.pH) p.crosstab(df["Temperature"], df["Odour"]) df.info()

### MODULE-2

**DATA VISUALIZATION AND DATA ANALYSIS**

import warnings warnings.filterwarnings('ignore') import pandas as pd

import numpy as np

import matplotlib.pyplot as plt import seaborn as sns milk=pd.read\_csv('milk.csv') milk.head() fig,ax=plt.subplots(figsize=(15,8))

sns.heatmap(milk.corr(),ax=ax,annot=True) p=sns.factorplot('Odour',data=milk,kind='count',hue='Turbidity') p.set\_xlabels('Odour')

p.set\_ylabels('No. of Values') p.fig.set\_figheight(3)

k=sns.factorplot('Odour',data=milk,kind='count',hue='Fat') k.set\_xlabels('Odour')

k.set\_ylabels('No. Of Values') k.fig.set\_figheight(3)

p=sns.factorplot('Taste',data=milk,kind='count',hue='Fat')

plt.title('Grades of milk')

### MODULE-3

**RANDOM FOREST ALGORITHM**

import warnings warnings.filterwarnings('ignore') import pandas as pd

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import confusion\_matrix, accuracy\_score, plot\_confusion\_matrix

import matplotlib.pyplot as plt milk=pd.read\_csv('milk.csv') milk.head() milk['Grade'].unique()

from sklearn.preprocessing import LabelEncoder le=LabelEncoder() milk['Grade']=le.fit\_transform(milk['Grade']).astype(int) milk.head()

milk['Grade'].unique()

features=['pH', 'Temperature', 'Taste', 'Odour', 'Fat', 'Turbidity', 'Colour'] print(features)

target=['Grade'] print(target) x=milk[features]

y=milk[target]

from sklearn.model\_selection import train\_test\_split

x\_train,x\_test,y\_train, y\_test=train\_test\_split(x,y,test\_size=0.30,random\_state=4 2, stratify=y)

RFC=RandomForestClassifier() RFC.fit(x\_train,y\_train) y\_prediction=RFC.predict(x\_test) CM=confusion\_matrix(y\_test,y\_prediction) print(CM) AC=accuracy\_score(y\_test,y\_prediction) print(AC\*100) fig,ax=plt.subplots(figsize=(10,10))

PCM=plot\_confusion\_matrix(RFC,x\_test,y\_test,ax=ax) plt.title('Confusion matrix of RFC')

plt.show()

### MODULE-4

**LOGISTIC REGRESSION ALGORITHM**

import warnings warnings.filterwarnings('ignore') import pandas as pd

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import confusion\_matrix, accuracy\_score, plot\_confusion\_matrix

import matplotlib.pyplot as plt milk=pd.read\_csv('milk.csv') milk.head()

from sklearn.preprocessing import LabelEncoder le=LabelEncoder() milk['Grade']=le.fit\_transform(milk['Grade']).astype(int)

features=['pH', 'Temperature', 'Taste', 'Odour', 'Fat', 'Turbidity', 'Colour'] print(features)

target=['Grade'] print(target) x=milk[features] y=milk[target]

from sklearn.model\_selection import train\_test\_split AC=accuracy\_score(y\_test,y\_prediction

print (AC\*100)

### MODULE-5

**XG BOOST CLASSIFIER ALGORITHM**

import warnings warnings.filterwarnings('ignore') import pandas as pd

from xgboost import XGBClassifier

from sklearn.metrics import confusion\_matrix, accuracy\_score, plot\_confusion\_matrix

import matplotlib.pyplot as plt milk=pd.read\_csv('milk.csv') milk.head()

from sklearn.preprocessing import LabelEncoder le=LabelEncoder() milk['Grade']=le.fit\_transform(milk['Grade']).astype(int)

features=['pH', 'Temperature', 'Taste', 'Odour', 'Fat', 'Turbidity', 'Colour'] print(features)

target=['Grade'] print(target) x=milk[features] y=milk[target]

from sklearn.model\_selection import train\_test\_split

from sklearn.model\_selection import train\_test\_split XGB=XGBClassifier()

XGB.fit(x\_train,y\_train) y\_prediction=XGB.predict(x\_test) CM=confusion\_matrix(y\_test,y\_prediction) print(CM) AC=accuracy\_score(y\_test,y\_prediction) print(AC\*100) fig,ax=plt.subplots(figsize=(10,10))

PCM=plot\_confusion\_matrix(XGB,x\_test,y\_test,ax=ax) plt.title('Confusion matrix of XGB')

plt.show()

### MODULE-6

**MLP CLASSIFIER ALGORITHM**

import warnings warnings.filterwarnings('ignore') import pandas as pd

from sklearn.neural\_network import MLPClassifier

from sklearn.metrics import confusion\_matrix, accuracy\_score, plot\_confusion\_matrix

import matplotlib.pyplot as plt milk=pd.read\_csv('milk.csv')

milk.head()

from sklearn.preprocessing import LabelEncoder le=LabelEncoder() milk['Grade']=le.fit\_transform(milk['Grade']).astype(int)

features=['pH', 'Temperature', 'Taste', 'Odour', 'Fat', 'Turbidity', 'Colour'] print(features)

target=['Grade'] print(target) x=milk[features] y=milk[target]

from sklearn.model\_selection import train\_test\_split

x\_train, x\_test, y\_train, y\_test=train\_test\_split(x,y,test\_size=0.36,random\_state=4 2)

MLP=MLPClassifier() MLP.fit(x\_train,y\_train) y\_prediction=MLP.predict(x\_test) CM=confusion\_matrix(y\_test,y\_prediction) print(CM) AC=accuracy\_score(y\_test,y\_prediction) print(AC\*100)

plt.title('Confusion matrix of MLP') plt.show()

### MODULE-7 DEPLOYMENT

from django.shortcuts import render, redirect

from django.contrib.auth.forms import UserCreationForm, AuthenticationForm from django.contrib.auth import login as auth\_login, authenticate, logout

from. import forms from. import models import numpy as np import joblib

model = joblib.load('E:/Diwakar/Project/Machine learning/Milk/Deploy/latest/new/rf.pkl')

*# Create your views here.*

def home\_view(request):

if request.method == "POST":

username = request.POST['username'] print(username)

password = request.POST['password'] print(password)

name = request.POST['user'] if name == "user":

user = authenticate (request, username=username, password=password)

*#print(user)*

if user is not None: auth\_login (request, user)

return render (request, 'new/index.html')else

msg = 'Invalid Credentials'

form = AuthenticationForm(request.POST)

return render (request, 'new/user\_login.html', {'form': form, 'msg': msg})else:

user = authenticate(request, username=username, password=password)

*#print(user)*

if user is not None: auth\_login(request, user)

model = models.UserPredictDataModel.objects.latest('id') form = forms.FeedForm(request.POST)

*#print(model)*

return render(request, 'new/last.html', {'model':model,'form':form}) else:

msg = 'Invalid Credentials'

form = AuthenticationForm(request.POST)

return render(request, 'new/user\_login.html', {'form': form, 'msg':

msg})

form = forms.FeedForm()

model = models.UserPredictDataModel.objects.latest('id')

return render(request, 'new/last.html', {'model':model,'form':form}) def feedback(request):

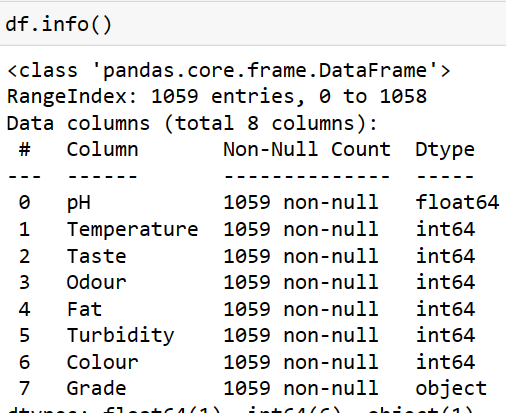
model = models.FeedModel.objects.latest('id') return render (request, 'new/feedback.html',

{'model':model})def apredict(request): return render(request, 'new/index.html')

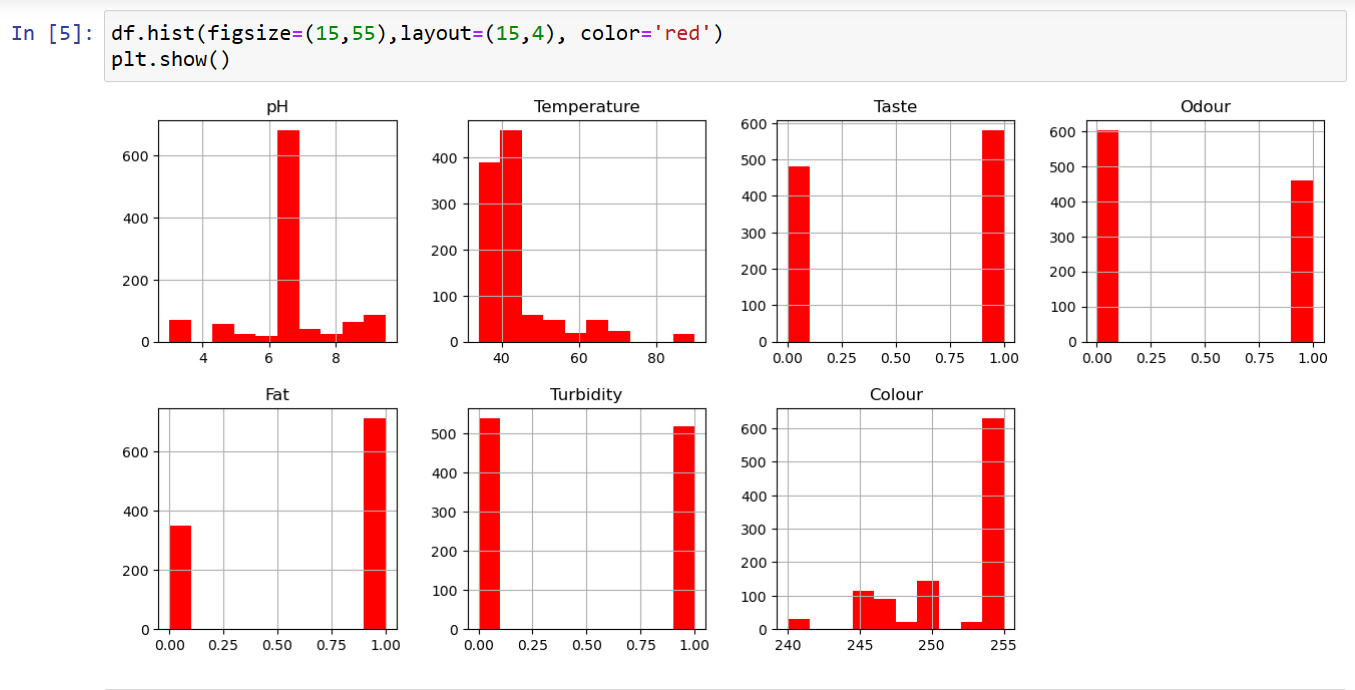
def logout\_view(request): logout(request)

return redirect('home\_view')

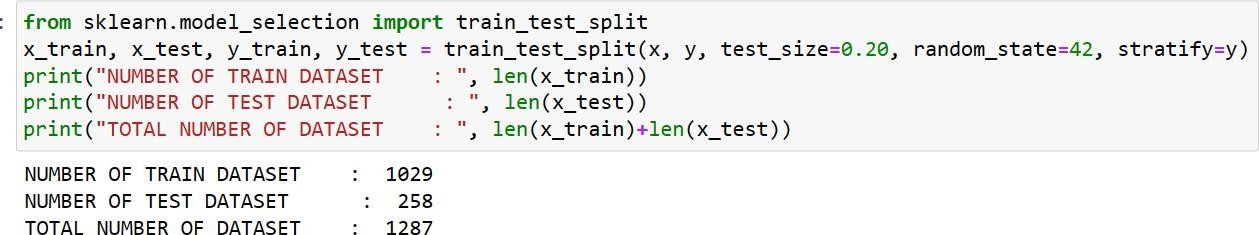
### B2 SAMPLE SCREENS



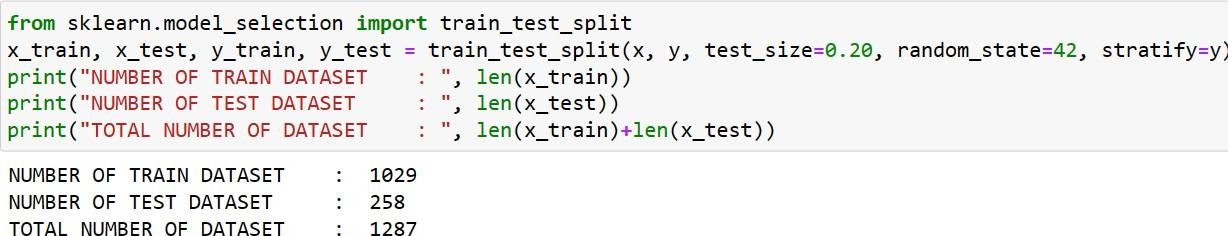
**Figure B2.1** Data Preprocessing Coding



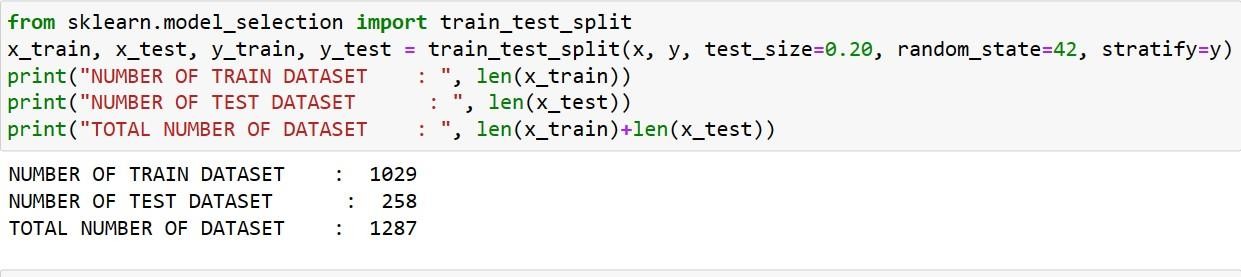
**Figure B2.2** Data Visualization Coding



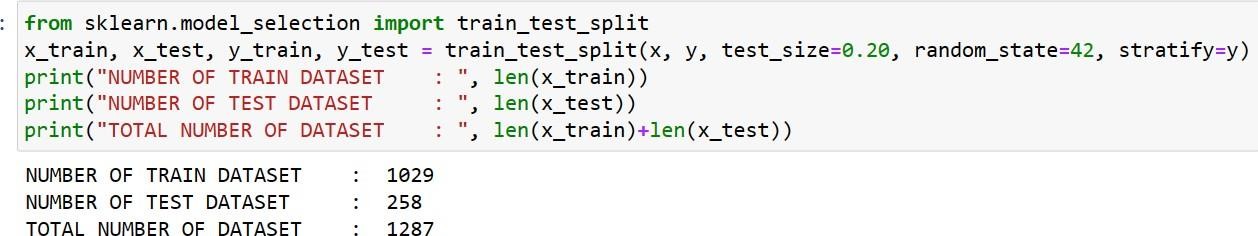
**Figure B3.3** Random Forest Algorithm coding



**Figure B4.4** Logistic Regression Algorithm Coding



**Figure B5.5** XG Boost Classifier Algorithm Coding



**Figure B6.6** MLP Classifier Algorithm Coding

### REFERENCES

1. R. Nagraik, A. Sharma, D. Kumar, P. Chawla, and A. P. Kumar,“Milk adulterant detection: Conventional and biosensor based approaches: A review,” Sens. Bio-Sens. Res., vol. 33, Aug. 2021.
2. M. Chakraborty and K. Biswas, “Milk tester: Simultaneous detection of fat content and adulteration,” IEEE Trans. Instrum. Meas., vol. 69, no. 5, pp. 2468– 2476, May 2020.
3. X. Hu et al., “Use of a smartphone for visual detection of melamine in milk based on Au@Carbon quantum dots nanocomposites,” Food Chem.,vol. 272, pp. 58–65, Jan. 2019.
4. F. Trimboli et al., “Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis,” J. Dairy Sci., vol. 102, no. 7,pp. 5962–5970, Jul. 2019.
5. Z. H. Zargar and T. Islam, “A novel cross-capacitive sensor for non con-tact microdroplet detection,” IEEE Trans. Ind. Electron., vol. 66, no. 6,pp. 4759– 4766, Jun. 2019.
6. T. Islam, A. Salamat, S. K. Singh, and M. Rehman, “A cross-conductive sensor to measure bottled water quality,” in ProcIEEE Int. Instrum. Meas. Technol. Conf., May 2021.
7. U. Salmaz, A. Salamat, and T. Islam, “Fractional order sensor for measuring the quality of milk,” Mater. Today Proc., vol. 18, pp. 1077–1089,Jan. 2019.
8. R. Kunes et al., “In-line technologies for the analysis of important milk parameters during the milking process: A review,” Agriculture, vol. 11,no. 3, pp. 239–246, 2021.
9. A. Costa, G. Visentin, M. De Marchi, M. Cassandro, and M. Penasa “Genetic relationships of lactose and freezing point with minerals and coagulation traits predicted from milk mid-infrared spectra in Holstein cows,” J. Dairy Sci., vol. 102, no. 8, pp. 7217–7225, Aug. 2019.
10. Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K.L. Ingvartse M. Crowe, M. T. Sorensen, E. Froidmont, C. P.Ferris, C. Marchitelli, 2019. Potential of milk mid-IR spectra to predict meta-bollic status of cows through blood components and an innovative clustering approach.

[11]P.S.Venkateswaran,A.Sharma,S.Dubey,“Rapid and automated measurement of milk adulteration using a 3D printed optofluidic microviscometer (OMV),” IEEE Sensors J., vol. 16, no. 9, pp. 3000–3007, May 2016.