**Integrated Milk Purity Tester: An Arduino Uno-Based Project with AI-Driven Milk Grading**

***Abstract:***

The *“Shuddh”* Milk Grader” presents a comprehensive solution for assessing the quality of milk using Arduino Uno as the microcontroller. The system integrates multiple sensors, including a pH sensor, turbidity sensor, gas sensor,colour sensor and temperature sensor, to measure various parameters for assessment. Additionally, the data is processed by an AI model to provide accurate grading of milk .

The Arduino Uno serves as the central processing unit, controlling the interaction between the various sensors and facilitating the collection of data parameters. The pH sensor measures the acidic/basic nature of the milk, while the turbidity sensor detects any suspended particles, ensuring transparency and cleanliness. The gas sensor is used to check for odour,which is important for determination of milk quality. The temperature sensor provides insights into the thermal conditions of the milk.

Furthermore, artificial intelligence (AI) is incorporated into the system for automated milk grading. The AI algorithm, trained on a diverse dataset of milk quality parameters, evaluates the sensor data and assigns a grade to the milk. This helps automate this process, ensuring fast and seamless grading without any human error or effort.

**Index**

[**Introduction 4**](#_2amcf2b6rft)

[1.1.1 A Chemical analysis of milk 5](#_a33ps1pkknie)

[1.1.2 An Outline of Chemical Reactions in Milk Affecting Composition and Quality 8](#_m14ed7iermhm)

[1.1.3 Evaluation and Selection of Factors for milk quality grading 9](#_8fwrup4nzdav)

[**Literature Review 11**](#_lrqktwo55dkc)

[2.1.1 Overview of similar existing projects and their implementation 11](#_p5bzmmv8d24m)

[2.2.1 Introduction to K-Nearest Neighbors (KNN) 13](#_4sa89ru0if3t)

[**Materials and methodology 17**](#_8j9b4m4sruzj)

[3.1.1 Materials for the proposed model 17](#_rycqb8ktlv7q)

[3.2.1 Dataset of parameters for milk grading 23](#_tkllg1uko41v)

[3.2.2 Implementation of KNN Classification for Milk Grading 26](#_gv1o39ki7eyy)

[**Results and Discussion 29**](#_j2i5su4hdbds)

[4.1. Key Findings pertaining to Milk Quality Parameters 29](#_d4jd9yfq4i1y)

[4.2 Performance of the KNN Classifier 31](#_smr2a4c066sm)

[**Objectives of Next Model 32**](#_e3ea6yd7wgtg)

[5.1 Project Goals 32](#_sjiosl3pd55k)

[5.2 Project Motivations 32](#_xzk2rvfb0f3e)

[**Future Scope of the Project 33**](#_ag4e534xxnhr)

[**Conclusion 34**](#_i3bdoniqwzp2)

[**References 35**](#_dghogztxhpvq)

# Introduction

Milk, a fundamental and nutritionally rich fluid secretion in mammals, notably in dairy animals such as cows, goats, and sheep. It serves as the primary source of nourishment for the offspring of these mammals during the early stages of life. Milk has long been a dietary staple for humans , renowned for its balanced composition of essential nutrients,”a complete meal”.

From a chemical perspective, milk is a complex colloidal suspension of water, fats, proteins, lactose , vitamins, and minerals. Moreover, milk is a noteworthy source of calcium, phosphorus, potassium, and vitamins, such as vitamin D and B-complex vitamins. These elements play vital roles in bone health, nerve function, and metabolic processes within the human body.

Milk has been closely linked to agriculture and industry, with dairy farming practices evolving over centuries to meet the growing demands of the world. Technological advancements have helped in the large-scale production of milk products, ranging from butter and cheese to yoghurt and powdered milk.

Milk is an extremely versatile food, finding its way into countless recipes, beverages, and cultural traditions worldwide.It is a cornerstone of human diet and nutrition. Milk holds immense nutritional significance and is a dietary staple in many cultures, including India. In the Indian context, where dairy is central in daily nutrition,consumption and livelihoods, the need for rigorous milk purity testing becomes even more critical.

*To determine the parameters to determine the highest standards and quality of milk, let us look into the chemical composition, reactions and conditions that affect the condition of milk.*

## 

## 

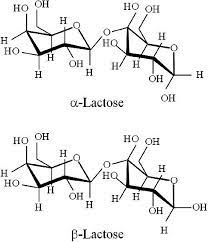
## 1.1.1 A Chemical analysis of milk

Milk is a complex colloidal suspension of more than 100+ components most important of which are water, fats, proteins, lactose , vitamins, and minerals. Fats in the form of globules are a good energy source, while proteins such as casein and whey provide amino acids which serve as building materials for the body. Lactose, a disaccharide, imparts a natural sweetness to milk and serves as a readily digestible carbohydrate.

1. Milk Fat: It is usually considered as the most important component of milk. Various types of milk are priced in the market depending on the amount of fat they each contain. It serves as a rich source of energy (around 9 kcal per gram), and carrier of fat-soluble vitamins (A, D, E, and K). Level of fat can vary from below 3 % to more than 6%. *Triglycerides* are the primary components of milk fat,around 98% of the fat content. Triglycerides are made up of three fatty acid chains attached to a glycerol molecule. Milk fat contains a wide variety of fatty acids, including saturated, monounsaturated, and polyunsaturated fatty acids. *Phospholipids* are a minor component of milk fat and contribute to the emulsification of fat in milk, allowing fat to be suspended in the watery portion of milk without separating. Milk fat also contains a small amount of cholesterol.

As suggested in the work put out by *B.M. Mehta (2015),* milk fats play an important role in creation of favourable or unfavourable flavours in dairy products and in endowing milk with a velvety, creamy texture. In absence of fat, milk shows some traits of being watery, flat and grainy.

1. Milk Proteins: The proteins in milk are considered to be a major driving force in human nutrition. Milk contains more than 200 types of proteins and around 95% of Nitrogen of milk is in the form of proteins. The primary proteins in milk are casein and whey. *Casein* makes up about 80% of the total protein content in cow's milk. It exists in the form of micelles dispersed in the liquid phase of milk. Casein has excellent nutritional value and is a good source of essential amino acids. It also plays a role in cheese production due to its ability to coagulate. *Whey* proteins constitute around 20% of the total protein in cow's milk. Some of the major whey proteins include beta-lactoglobulin, alpha-lactalbumin, serum albumin, and immunoglobulins. These proteins are water-soluble and also provide essential amino acids, particularly cysteine and leucine.
2. Milk sugar (Lactose): Lactose is the primary sugar found in milk. It belongs to the disaccharide group of sugars, composed of two simple sugar molecules: glucose and galactose joined by a β-glycosidic bond. Lactose is only a fourth as sugary as sucrose.



*Reference No. 9*

*Diagram 1.1.1 A: Isomeric forms of Lactose*

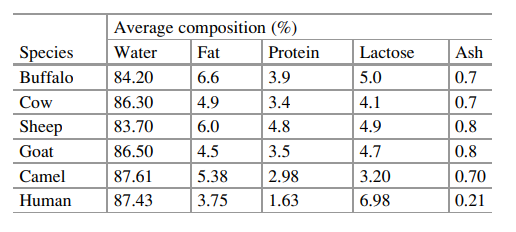
*Lactose is fermented by lactic acid bacteria to lactic acid that reduces the pH of milk.* Uncontrolled or unwanted fermentation clearly results in spoilage of the milk. However controlled fermentation is the basis of production of cheese and yoghurt. Lactose undergoes Maillard reactions at high temperature, leading to colour changes in milk heated to very high temperatures. Thermal degradation of lactose also leads to production of various organic acids,affecting stability of milk to further heat.

1. Enzymes: Milk contains small amounts of lipases, which can cause alteration in milk composition when milk is exposed to elevated temperatures or in contact with air. Lipases catalyse the breakdown of milk fat (triglycerides) into free fatty acids and glycerol causing the development of off-flavours and rancidity in milk, altering its taste and aroma. Proteases are enzymes that break down proteins and can affect milk quality by causing protein degradation. However, these enzymes typically have a minimal impact on milk rancidity compared to lipases. Peroxidases can contribute to off-flavours in milk by catalysing the oxidation of lipids. They work in conjunction with free fatty acids and oxygen to produce compounds that contribute to undesirable flavours and odours in milk.

*Factors such as temperature, improper storage conditions, or contamination with microorganisms can enhance enzyme activity leading to accelerated breakdown of fats and subsequent rancidity.*

1. Vitamins: Milk contains various fat-soluble vitamins (A, D, E, and K). Vitamin A is a precursor of β-carotene which is responsible for the yellow color of cow milk.

*Table 1.1.1 A: Average composition characteristics of milk obtained from various sources*

*Reference No. 2*

## 1.1.2 An Outline of Chemical Reactions in Milk Affecting Composition and Quality

1. *Fermentation:* Fermentation in milk is the conversion of lactose into lactic acid by lactic acid bacteria (LAB). The presence of LAB initiates the fermentation process, leading to changes in milk composition and taste. *The conversion of lactose to lactic acid decreases the pH of the milk, resulting in increased acidity resulting in a tangy taste.*

*Conditions for likely fermentation of milk:*

* Lactic acid bacteria are best active between 30°C to 45°C
* The optimal pH for most lactic acid bacteria is between 4.0 and 5.0
* Fermentation often occurs in anaerobic conditions with limited oxygen exposure.

1. Maillard Reactions: Maillard reactions occur between reducing sugars like lactose and free amino acids from milk proteins when exposed to *higher temperatures*. These reactions lead to the formation of various flavour compounds and browning of milk. In milk, Maillard reactions occur during heat processing. Maillard reactions can result in caramelization of milk, resulting in a taste which may or may not be desirable.
2. Lipolysis: Lipolysis is the breakdown of milk fat into free fatty acids and glycerol, catalysed by lipases. Lipolysis can occur due to various factors, including temperature, enzymatic activity, or microbial contamination. Lipolysis contributes to rancidity and off-flavours in milk.
3. Oxidation: Milk is susceptible to oxidation, especially because of its fat component. Oxidation reactions involve the interaction of oxygen with milk components, primarily unsaturated fatty acids in milk fat. This leads to the formation of off-flavours and a decrease in nutritional quality and shelf life.
4. Proteolysis: Proteolysis involves the breakdown of milk proteins by proteolytic enzymes.
5. Microbial activity: The growth of bacteria like Pseudomonas, Enterobacter, and others, in milk is a common cause of spoilage. These bacteria can grow at refrigeration temperatures and break down milk leading to off-flavours, gas production and curdling. Yeasts and moulds can also spoil the milk, causing changes such as souring and spoilage of milk.

## 1.1.3 Evaluation and Selection of Factors for milk quality grading

1. **pH**: The pH of normal milk varies from 6.5 to 6.7. A pH lower than the usual range, i.e. acidic milk is an indicator of fermentation and microbial activity. A noticeable deviation from the normal range could also indicate adulteration of milk with other substances.
2. **Temperature**: Milk quality is best preserved at lower temperatures. Temperatures between 30 to 45 celsius could encourage bacterial growth. Temperature abuse at high temperatures during processing could result in initiation of Maillard reactions if the milk contains lots of free amino acids. Free amino acids are created by breakdown of proteins due to considerable changes in temperature,pH and by enzymatic activity. These free amino acids, along with milk sugars can potentially take part in Maillard reactions causing caramelization, change in milk taste and brown alteration to colour of milk.
3. **Odour**: Microbial activity and spoilage of milk can be noticeable by a change of odour and aroma caused by the release of volatile compounds after the breakdown of milk components.
4. **Turbidity**: Turbidity can be defined in terms of deflection of light which passes through a medium by the undissolved particles. This parameter of milk can show deviation by possible impurities and adulteration of milk.
5. **Colour**: Color, although slightly unconventional, can also be used as a parameter for the grading of milk. Colour can help differentiate samples with adulteration, impurities ,higher fat percentages and possible temperature abuse (browning of milk).

**The main goal of our effort is to make a proof of concept of a *‘Shuddh’* milk purity grader using Arduino Uno infrastructure to measure the various above mentioned parameters using sensors and then using these parameters to instantly grade the quality of milk using an pre-trained AI model to eliminate human error.**

# 

# 

# 

# 

# 

# 

# Literature Review

## 2.1.1 Overview of similar existing projects and their implementation

Similar projects leveraging electronic parameter measurement have been initiated in the past. In a similar endeavour by *Goswami and Dangi (2021)*, parameters like pH, temperature, odour and fat(%) were used. Various sensors were used to read values which were then connected to an Arduino Uno Microcontroller. These readings are then evaluated to grade the milk.

*Goswami and Dangi* and team leveraged the use of Arduino Uno microcontroller which is relatively inexpensive as well as light weight. Using the Arduino ecosystem, a variety of sensors which have compatibility can be connected to collect and send data for processing. In this proposed framework, customers can use their smartphones to access milk parameter data.

A temperature sensor is used to check the temperature and ensure it is in the normal range suitable for milk storage. Prolonged storage at moderately high temperatures could encourage bacterial activity in the milk.

A pH sensor is used to check whether the parameter falls within the accepted rate. A gas sensor is used to check for odours and other gases present in the milk due to bacterial activity or possible adulteration. An interesting feature of this project is the setting up of a LED-LDR combination to make a ‘Fat detection module’. The light from the LED dispersed by the milk added in the test tube is measured by the LDR. From the degree of scattering of light, the fat content is estimated.

This project proposal has several merits. It aims to be a low cost solution to ensuring good quality of milk. It is suitable for both small scale uses such as in cooperatives and can be implemented on a moderately large scale as well. It is a portable and modular unit that can be easily moved around.

However there are a few limitations as well. Using an Arduino compatible pH sensor is not entirely suggestible due to its high price, compared to other components. Besides this, this type of sensor needs to be calibrated with another solution before use, which if needed to be repeated regularly could be a hindrance. Moreover this sensor can also be prone to damage with regular use.

The Fat detection module can help give an idea of the relative fat present in milk. However it has to be implemented well with the test tube LED-LDR fixed in a closed setup for best results. Specifically in the case of testing adulteration of milk with water, a Lactometer could be an easier and more cost effective solution.

In a similar project by *S. Verulkar (2019)* and team put forward the idea of an ‘Embedded Milk Analysis System’ meant to be a cheap and suitable solution for farmers and industry alike. The proposed system works on IoT application compatible Arduino ESP32 Microcontroller.The temperature sensor helps monitor the temperature of the milk. They have proposed using a viscosity sensor to measure the viscosity and a gas sensor to check the presence of odour.A salinity sensor is used to detect the salt content of the milk.Along with this, an ultrasonic sensor is used to find the level of milk with respect to the size of the container, in order to be able to measure the quantity of milk as well.

*Verulkar* and team acknowledge the fact that the various electronic parts require regular maintenance, especially if used regularly by a large group of people, such as in milk cooperatives.If any of the sensors stop working, replacement would be easier than analysing and repairing the damage. The main intention of their project is to reduce the adulteration practice in milk. They have put forward the opinion that if implemented at a low cost, it can be used by a large number of farmers and consumers alike. It can help increase awareness about the importance of drinking pure,safe,unadulterated milk especially since milk plays a huge role in nutrition in India, which has a large vegetarian population.

*Manisarathi and Raveesh (2022)* take it a step further by adding IoT capabilities as well as an automated billing system to price the milk according to the grade of the milk.

*To enhance the capabilities of our proposed project, we shall explore the topic of the K-Nearest Neighbours Classification model.*

## 2.2.1 Introduction to K-Nearest Neighbors (KNN)

K-Nearest Neighbours, commonly abbreviated as KNN, is a versatile and intuitive machine learning algorithm used for both classification and regression tasks. It comes under supervised learning, where the algorithm learns patterns and relationships from labelled training data to make predictions on unseen or new instances.

At its core, KNN operates based on the idea of proximity, indicating that instances that are similar in a given space share corresponding qualities or results. What sets this algorithm apart is its non-parametric nature, meaning it doesn't make any assumptions about the data distribution, making it versatile and suitable for a wide range of problems.

The primary concept behind KNN is to determine a class label or predict a target variable for a given data point by examining the classes or values of its nearest neighbours in the feature space. The "K" represents the number of neighbours taken into consideration for these predictions. Essentially, the algorithm looks at the K closest data points and determines the majority class for classification tasks or calculates the average for regression tasks.

KNN stands out for its simplicity and ease of interpretation, making it a popular option for beginners in the field of machine learning. However, the effectiveness of this algorithm relies heavily on the choice of distance metric, value of K, and the specific characteristics of the data being analysed. Understanding these variables and their impact on KNN is crucial for achieving success in real-world applications.

Despite its simplicity, KNN has proven to be a strong and dependable algorithm in various scenarios such as image recognition, recommendation systems, and anomaly detection. Its adaptable nature to different data distributions and straightforward implementation process make it a valuable tool in the arsenal of machine learning techniques. As we deepen our understanding of KNN, we will delve into its strengths, weaknesses, and practical implications.

As discussed in an article written by Hari Antoni Musril and other Indonesian scientists, the K-Nearest Neighbours (KNN) algorithm is a highly adaptable and effective approach to machine learning, commonly employed for both classification and regression tasks. It works by identifying the closest training instances to a new data point in the feature space, using the assumption that similar instances tend to cluster together.

Unlike other machine learning models, KNN does not require explicit training, making it particularly useful for dealing with complex or ambiguous data distributions. The algorithm then selects the K nearest neighbours to the new data point for classification purposes. While KNN has proven to be a beneficial tool in various domains, such as image recognition and bioinformatics, it does have its limitations. One such limitation is its tendency to be computationally expensive when handling large datasets.

Instance-based learning, also known as lazy learning, involves the use of the k Nearest Neighbour (k-NN) algorithm. This method focuses on approximating the function locally and postpones computation until classification. It utilises the Euclidean and Manhattan distances to measure the distance between the input sample and each training instance. Rather than just considering a small subset of the data, the algorithm compares the input sample to all training instances. Then, it selects the k most similar examples to the input sample. By taking into account the opinions of these nearest neighbours through a voting process, the resulting prognosis is the average of their results.

In an article about efficient kNN classification, Shichao Zhang, Xuelong Li and Ming Zhong present that the landscape of k-Nearest Neighbors (kNN) methods has witnessed various strategies aimed at enhancing efficiency and performance. Traditionally, two predominant approaches have been explored: (1) the use of a fixed, expert-predefined k value for all test samples and (2) the assignment of different optimal k values for individual test samples. Notably, the fixed kNN methods have proven impractical in real-world applications, prompting a surge in interest in varied kNN methods.

In an effort to overcome the drawbacks of the fixed k-nearest neighbours (kNN) approach, early researchers proposed alternative methods that tailored k values to individual samples. For example, Lall and Sharma advocated for a single, predetermined k value, whereas Zhu et al. suggested determining the most suitable k value through tenfold cross-validation. Despite these innovations, the issue of the lengthy and tedious process of either learning optimal k values for each test sample or searching through the entire training set for nearest neighbours persisted.

To tackle these challenges, this paper introduces the innovative kTree method, a novel approach that incorporates a training stage into the traditional kNN method. The kTree method leverages a sparse-based reconstruction model during the training stage to efficiently predict optimal-k values for each training sample. Subsequently, a decision tree (kTree) is constructed using the training samples and their corresponding optimal-k values, facilitating a fast and offline learning process. During the test stage, the kTree model is navigated to assign optimal-k values to test samples, and traditional kNN classification is applied for label assignment.

Locally adaptive KNN algorithms, as highlighted by Zinnia Sultana et al. in their article about Improved kNN Algorithm for Pattern Classification, dynamically determine the value of k for categorising an interrogation by leveraging cross-validation calculations within the resident locality of the query. This local approach mirrors the conventional KNN methodology and has been experimentally validated across twelve commonly used datasets.

Deepti et al. introduced the Quad Division prototype, employing a Selection-based K-nearest Neighbour classifier to address uneven class distribution. The performance of this Quad Division Prototype based on KNN (QDPSKNN) is specifically assessed in fraud detection within mobile advertising. Comparative evaluations involve base model KNN and other selection methods, such as NearMiss-1, NearMiss-2, NearMiss-3, and Condensed Nearest Neighbour (CNN).

Suyanto et al. proposed the Multi-Voter Multi-Commission Nearest Neighbour (MVMCNN) as a variant of KNN to enhance the Local Mean based Pseudo Nearest Neighbour. Comparative analysis with single voter models (KNN and BMFKNN) suggests that the multi-voter model, MVMCNN, yields superior decision-making.

Armand et al. introduced a metaheuristic search algorithm, Simulated Annealing, to optimise the selection of k, presenting a more computationally efficient alternative to exhaustive searches. The results are compared across four classification methods to showcase substantial advancements in computational efficiency compared to traditional KNN methods.

D. Maruthi et al. proposed an adaptive KNN classifier for an effective classification system in MRI brain tumour images, emphasising accuracy, sensitivity, and specificity in the classification and segmentation process.

Jieying et al. focused on precise image interpolation using an adaptive KNN approach, particularly in searching for image patches on the input image and establishing non-linear mappings between low-resolution and high-resolution image patches.

Jianping et al. offered a local mean representation-based KNN classifier to enhance classification performance and overcome primary issues associated with conventional KNN classification. Comparative evaluations with UCI and KEEL databases highlight the superior performance of LMRKNN over KNN-based methods.

The literature dived into a discussion that offers valuable perspectives on the various applications and adaptations of KNN. However, the main focus of this research is to introduce a new algorithm that can effectively solve pattern recognition problems in the field of machine learning. Specifically, the algorithm aims to tackle problems associated with variance-based Euclidean distance metrics and FRW in order to enhance the performance of KNN in pattern classification tasks. Additionally, the study provides experimental comparisons with other popular algorithms such as KNN, DANN, and C4.5, covering a range of domains. This study serves to bridge a gap in research on AI and NLP, making a significant contribution to the field.

*For this project we shall be using a fixed k value of 3 representing the three grades of quality of milk of High, Medium and Low. These grades has already been assigned to the various datapoint values in the dataset used for this project.*

# Materials and methodology

## 3.1.1 Materials for the proposed model

Arduino boards are the physical computing platforms. They form the foundation of the Arduino open-source electronics platform. These boards provide the necessary hardware to run programs written using the Arduino programming language or can also be programmed through python, JAVA, C++ etc. Arduino boards are electronic platforms that are designed for hobbyists, students, artists, and anyone interested in creating interactive projects. They have both hardware and software components,and are widely used for building some prototypes. Arduino provides a simple, accessible way to experiment and help develop projects involving sensors and various other electronic components.

The most important component of an Arduino board is a microcontroller. The microcontroller is a small computer on a single integrated circuit that will interpret and execute commands. Arduino boards use various microcontroller models, such as the ATmega328P for the case of Arduino Uno.Arduino boards have a great variety of connectors and pins that allow us to connect sensors, LEDs and other parts. It has digital ,analog input/output pins, power pins, and communication points. These boards can be powered using a USB connection, an external power adapter, or a battery.

Most boards have a USB connection that allows them to connect to a computer for programming. The USB connection is used to upload code from the Arduino coding IDE to the board.

Arduino has an Integrated Development Environment (IDE) , a software tool used to write and upload code to the board. It is very user-friendly coding, compiling, and uploading sketches (Arduino programs).We can use user-friendly coding languages like python. Even languages like JAVA,C,C++ can be used to write the code.

A sensor is a device that detects the change in the substance and responds to some output on the other system. A sensor converts a physical phenomenon into a measurable analog voltage or sometimes digital signal for a display or transmitted for reading,further processing.

A good understanding of working principles, advancements,and challenges associated with such sensors is needed for designing a cost-effective and efficient project to measure the purity of milk.

1. **Temperature Sensors:** It measures the temperature of milk. It is also used in applications such as weather monitoring, HVAC systems, and industrial processes.

Estimated cost: It would cost around 600-1200 rupees.

Working principle: i) *Thermocouples* are composed of two different metals joined at one end.When a temperature difference occurs between the joined hot junction and the free cold junctions voltage is generated.The voltage is directly proportional to the temperature difference and is then being used to determine the temperature.

ii) *Thermistors* are semiconductor devices with a resistance that is highly sensitive to temperature changes.Negative Temperature Coefficient (NTC) thermistors decrease resistance with an increase in temperature, while Positive Temperature Coefficient (PTC) thermistors increases resistance.The resistance change is measured to determine the temperature.

iii) Digital temperature sensors often use integrated circuits (IC’s), that convert the analog signal from a temperature-dependent component into a digital output.

Make the connection of the temperature sensor to the Arduino Uno board through breadboard and jumper wires. Connect power,ground, and signal pins of the sensors to the given corresponding pins of the board. Write a test code in Arduino IDE to read the temperature parameters from the attached sensors and to display it on the monitor.Upload the code to the Arduino uno board and open the serial monitor to see the temperature readings from the sensor.

1. **Gas sensor** : The concentration of odour will vary from fresh milk to toxic milk. When the toxicity in milk is high it tends to release toxic gases which come out as bad odour from the milk when milk is preserved for a very long time or due to external contamination. So, we detect the gases releasing out from samples which are nothing but bad odour in general. It is mainly used to detect the presence of ammonia or other gases that indicate spoilage or contamination of the milk.

Estimated cost:The price ranges from 200 to 300

Working principle:

i)Metal Oxide Gas Sensors (MOx) generally use a semiconductor material, such as tin dioxide (SnO2). When exposed to a specific gas, the conductivity of the metal oxide changes.When the gas is present, it interacts with the surface of the metal oxide, causing changes in material conductivity/resistance.The magnitude of the resistance change is comparable with the concentration of the gas.

ii)Electrochemical gas sensors consist of a working electrode, counter electrode, a reference electrode, and an electrolyte. The sensor uses chemical reactions to generate an electrical current which is proportional to the concentration of a specific gas.Electrochemical sensors are commonly used for some gases like carbon monoxide, hydrogen,and various toxic gases.

Connect the sensor to the Arduino uno board in the same way as the temperature sensor. Write a test code in Arduino IDE to read the gas concentration parameters from the sensor and to have it displayed it on the monitor.Upload the code to the Arduino Uno board and test the working of the sensor. Use of this sensor might require a few trial cases to determine the ideal level of this parameter for good quality milk. If the gas concentration is very high, it may indicate milk spoilage or possible adulteration with other substances.

1. **Turbidity sensor**: Turbidity is the phenomenon whereby a specific portion of a light beam passing through a liquid medium is deflected from undissolved or large size particles.

Estimated cost: The cost of a turbidity sensor is around 500 rupees.

Working principle:

i)Turbidity sensors generally consist of a light source, often an LED (Light-Emitting Diode), that emits light into the liquid being measured.

ii) Opposite the light source, there is a photodetector or a set of photodetectors. These detectors are for capturing light that passes through the liquid after having an interaction with suspended particles.

iii) As light passes through the liquid, it interacts with the suspended particles. The presence of particles causes the light to scatter in different directions.The photodetector measures the intensity of the scattering light. The amount of scattering is directly proportional to the concentration of particles in the liquid. More suspended particles result in greater light scattering

iv) The final output of the turbidity sensor is usually expressed in nephelometric turbidity units (NTU ). The turbidity of milk commonly at room temperature is 0.295±0.015.

Connect the sensor to the Arduino uno board. Write a test code in Arduino IDE to read the turbidity data from the sensor and display it on the monitor.Then upload the code to the board using the USB connection. If the turbidity is a bit too high, it give indications of spoilage/adulteration.

1. **Colour sensor:** Colour sensors are used to analyse the colour quality of the milk which can help point out possible adulteration and relative fat content as well.

Estimated cost: 250-600 rupees

Working principle:

i) Colour sensors often contain one or more light-emitting elements (LEDs) to illuminate the object for analysis.Some of the types of colour sensing technologies include RGB (Red,Green, Blue) and CMY (Cyan, Magenta,Yellow).

ii) RGB sensors use separate channels for red, green, and blue light. By measuring the intensity of each colour, the sensor can determine the overall colour of the object.

iii) CMY sensors measure the intensity of cyan, magenta, and yellow light.

iv) Colour sensors have photodetectors that capture the reflected or transmitted light. These detectors convert the light into electrical signals.

Connect the sensor to Arduino uno board.Write a test code in Arduino IDE to read the color of milk parameter from the sensor and then display it on the monitor.Upload the written code to board using USB and open the monitor to see the colour readings.If the colour shows too much of a deviation from the whiteness value of 255, it may prove that the milk has been spoiled or adulterated.

1. **pH Sensors**: A pH sensor is a type of sensor designed to measure the acidity or alkalinity of a solution, which is expressed as its pH value. The pH scale ranges from 0 to 14, where a pH of 7 is considered neutral. Milk is slightly acidic having pH ranging from 6.7-6.9.pH.

Estimated cost: Sensors are around the price range of 2000-3000.

*Thus the use of universal pH paper has been considered for this project which is relatively a lot cheaper than pH sensor. Using pH paper has the added benefit of requiring no prior calibration before use unlike pH sensors which have to be calibrated every 7-10 measurements for accuracy.*

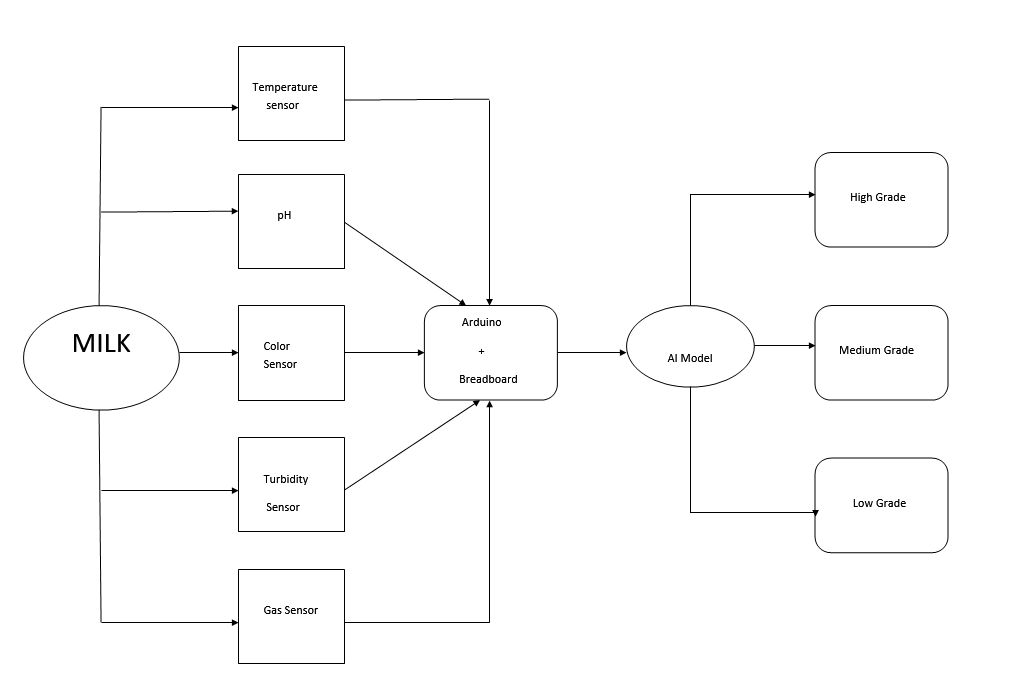
Working principle:

i) pH sensors use an ion-selective electrode, which is sensitive to the concentration of hydrogen ions (H⁺) in a solution. The potential difference across the ion-selective electrode is directly comparable to the pH.

ii) The glass membrane allows hydrogen ions to pass through, creating a potential difference.

iii) The pH electrode is paired with a reference electrode, which is oftentimes a Ag/AgCl electrode, which provides a good comparison potential.

iv) The potential difference across the glass membrane can be measured by the Nernst equation. This equation relates the concentration of hydrogen ions in the solution to the voltage measured by the pH electrode.



*Diagram 3.1.1 A: Block diagram of proposed system*

The results of the sensors are entered into the AI model interface for the milk grading result.

## 3.2.1 Dataset of parameters for milk grading

Kaggle is an online community platform of data scientists, researchers, and machine learning enthusiasts. The platform offers a vast repository of datasets covering various topics and fields. Users can explore, upload, and share datasets , collaborating to help others find valuable information for research and analysis.

For the proposed project, we shall be using the dataset contributed by author Shrijayan Rajendran on Kaggle, ‘Milk Quality Prediction’. (<https://www.kaggle.com/datasets/cpluzshrijayan/milkquality>)

It has a usability score of 7.65. Although this dataset has a low update frequency, it has a sufficient number of datapoints, large enough to train a statistical model.

This dataset has been manually collected from observations. This dataset takes into account seven independent variables of pH, Taste, Temperature, Turbidity, Fat, Odour, and Colour. All these parameters can help provide a proper depth to enable a higher efficiency of grading.

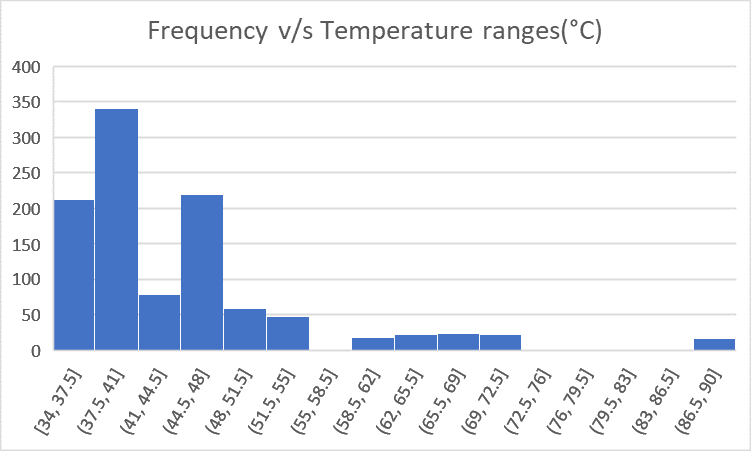
The dataset can be downloaded from Kaggle in the form of a .csv file which can be easily used in combination with AI packages. The target variable, i.e. the result variable of the dataset and statistical model is the Grade of Milk. Here we are classifying the milk into three grades: Good (3), Moderate (2) and Poor (1).

*Table 3.2.1 A: Snapshot of the project dataset*

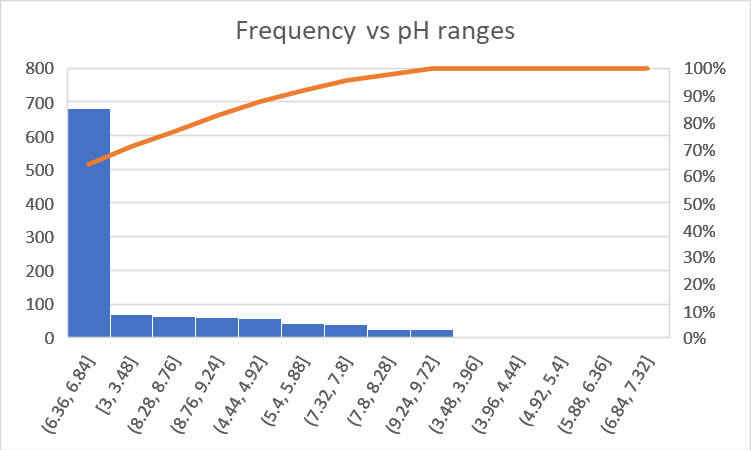
|  | **pH** | **temperature** | **taste** | **odour** | **fat** | **turbidity** | **colour** | **grade** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 6.6 | 35 | 1 | 0 | 1 | 0 | 254 | 3 |
| 1 | 6.6 | 36 | 0 | 1 | 0 | 1 | 253 | 3 |
| 2 | 8.5 | 70 | 1 | 1 | 1 | 1 | 246 | 1 |
| 3 | 9.5 | 34 | 1 | 1 | 0 | 1 | 255 | 1 |
| 4 | 6.6 | 37 | 0 | 0 | 0 | 0 | 255 | 2 |

There are in total 1059 data points with no null values or Not Available values, thus no cleaning is required prior to usage. pH is given in terms of float values, temperature is indicated in celsius. Taste, odour, fat, turbidity parameters are binary.

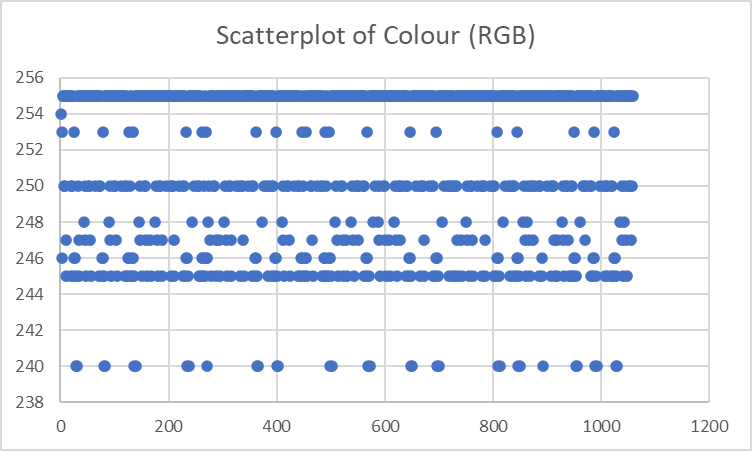
*Diagram 3.2.1 A: Plot of Frequency of datapoints v/s Temperature ranges*



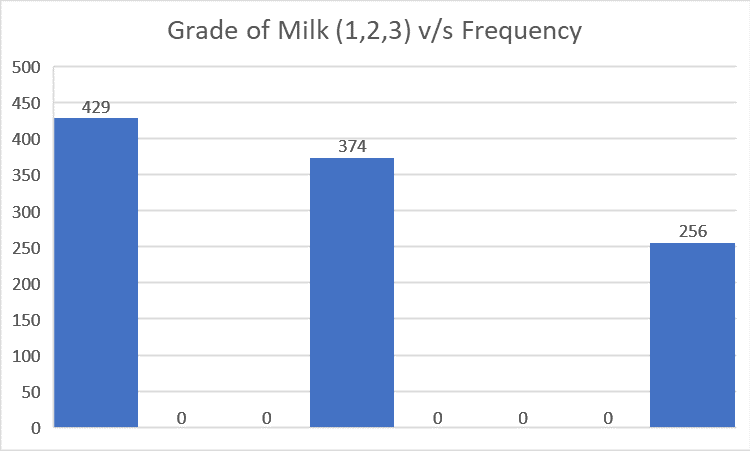
*Diagram 3.2.1 B: Plot of Frequency of datapoints v/s pH ranges*



*Diagram 3.2.1 C: Scatterplot of Colour values*



*Diagram 3.2.1 D: Plot of Frequency of data points v/s Grade of Milk*



## 3.2.2 Implementation of KNN Classification for Milk Grading

Scikit-learn is of the many python libraries that gives you a many tools and algorithms for doing data analysis. It can handle tasks relating to artificial intelligence and machine learning, including classification, regression, clustering, dimensionality reduction, etc. Another major plus point is that Scikit-learn integrates with other python libraries such as NumPy, SciPy, pandas, and matplotlib are used for graphing and data handling. It can help enable both supervised and unsupervised learning for many real world use cases.It uses packages NumPy and SciPy as its base.

The following model prototype has been developed on Google colab. The various steps in building this model have been listed below.

1. Packages such as numpy,pandas,matplotlib for data handling and visualisation and modules from sci-kit learn for model training is imported.

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import StandardScaler

from imblearn.over\_sampling import RandomOverSampler

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import classification\_report

1. The dataset is extracted from a csv file into a Pandas dataframe with labels for each parameter.

cols = ["pH","temperature","taste","odour","fat","turbidity","color","grade"]

df = pd.read\_csv("milknew.csv", names=cols)

1. The dataset is then split into training,validation and testing sets.

train, valid, test = np.split(df.sample(frac=1), [int(0.6\*len(df)), int(0.8\*len(df))])

1. The dataframes are then *scaled* to ensure each parameter receives equal importance despite the small or large magnitude of the parameter numerically when compared to other parameters. The dataframes also undergo *random-oversampling* to ensure that minority classes or values get good representation for training the models.

def scale\_dataset(dataframe, oversample=False):

X = dataframe[dataframe.columns[:-1]].values

y = dataframe[dataframe.columns[-1]].values

scaler = StandardScaler()

X = scaler.fit\_transform(X)

if oversample:

ros = RandomOverSampler()

X, y = ros.fit\_resample(X, y)

data = np.hstack((X, np.reshape(y, (-1, 1))))

return data, X, y

train, X\_train, y\_train = scale\_dataset(train, oversample=True)

valid, X\_valid, y\_valid = scale\_dataset(valid, oversample=False)

1. The scaled training datasets are then fitted on to the KNN classifier model for training the model, using the number of neighbours as 3 for high,medium and low grade of milk.

knn\_model = KNeighborsClassifier(n\_neighbors=3)

knn\_model.fit(X\_train, y\_train)

1. The input parameters for prediction are entered as a row for prediction by the model. The result is then given by the model as one of the three possible grades.

new\_row = [6.9, 20,1, 0, 1, 1, 255, 4]

test = test.append(dict(zip(cols, new\_row)), ignore\_index=True)

print(test.tail(1))

test, X\_test, y\_test = scale\_dataset(test, oversample=False)

y\_pred = knn\_model.predict(X\_test)

if y\_pred[-1] ==3:

print("HIGHEST QUALITY")

if y\_pred[-1] ==2:

print("MEDIUM QUALITY")

if y\_pred[-1] ==1:

print("LOW QUALITY")

# 

# 

# 

# 

# 

# 

# 

# 

# 

# 

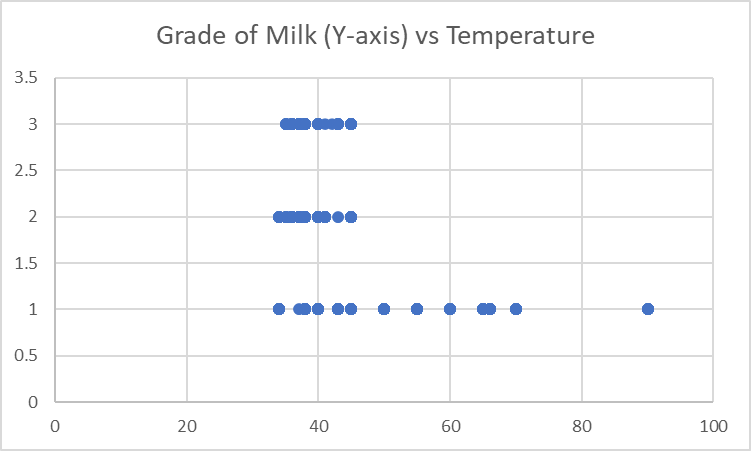
# 

# 

# Results and Discussion

## 4.1. Key Findings pertaining to Milk Quality Parameters

* Milk quality is usually best at lower temperatures; temperatures between 30 to 45 celsius could encourage bacterial growth as well as fermentation. High temperatures during industrial processing can cause browning of milk and even breakdown of certain milk components causing undesirable taste and could make it unsuitable for human digestion. Thus temperature deviation from the usual range of milk storage could indicate a lower quality milk. *Lower the temperature of milk, better is its assumed quality and grade.* Thus milk samples with lower temperatures are given higher preference by the algorithm.

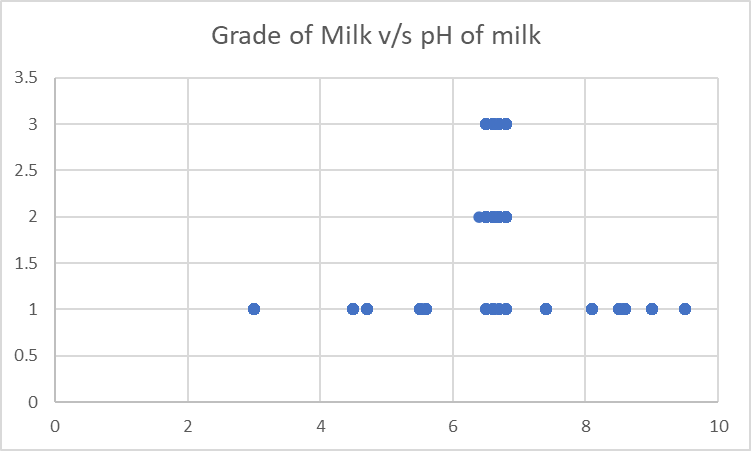


*Diagram 4.1.1 A: Plot of Grade v/s Temperature*

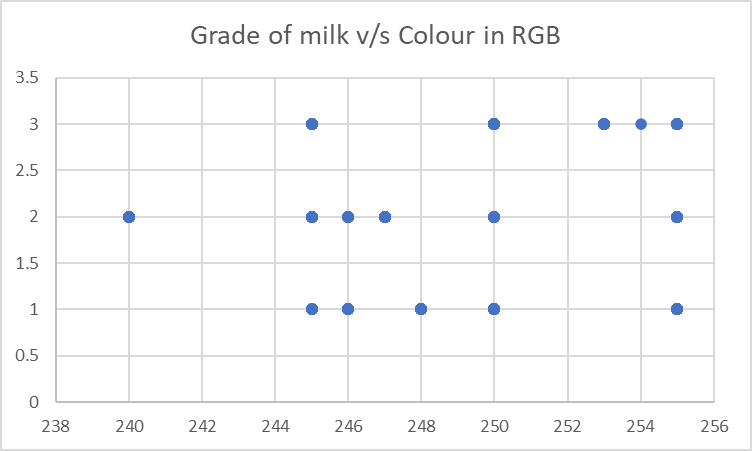
It is easy to notice from the graph that higher temperatures are generally graded lower on the algorithm.

* The pH of normal milk varies from 6.5 to 6.7. Acidic milk is an indicator of fermentation and microbial activity. A deviation from the normal range could also indicate adulteration of milk with other substances.

It is clear from the below diagram that the highest grade of milk is usually from the pH range of 6.7-6.9

*Diagram 4.1.1 B: Plot of Grade v/s pH*

* Microbial activity and spoilage of milk can be noticeable by a change of odour and aroma caused by the release of volatile compounds after the breakdown of milk components. Higher the gaseous content lower the assumed quality of milk.
* Turbidity parameter of milk can show deviation by possible impurities and adulteration of milk.
* Colour can help differentiate samples with adulteration, impurities ,higher fat percentages and possible temperature abuse (browning of milk).



*Diagram 4.1.1 C: Plot of Grade v/s Colour*

From the above plot we notice the relationship between colour and grade of milk is not completely concrete but milk with colour closest to 255 is usually rated higher by the algorithm.

## 4.2 Performance of the KNN Classifier

Using an inbuilt method in SK-Learn called classification report, we notice that the accuracy of this model is nearly 99%. This classification report method compares the validation datasets and and the predicted values and generates this report. From this we conclude that the KNN model is the appropriate model for this project for grading the milk.

y\_pred = knn\_model.predict(X\_test)

print(classification\_report(y\_test, y\_pred))

*Result:*

precision recall f1-score support

1.0 1.00 0.97 0.99 79

2.0 0.97 1.00 0.99 77

3.0 0.98 1.00 0.99 56

9.0 0.00 0.00 0.00 1

accuracy 0.99 213

macro avg 0.74 0.74 0.74 213

weighted avg 0.98 0.99 0.98 213

# 

# Objectives of Next Model

## 5.1 Project Goals

* To create a low-cost device capable of measuring various parameters of milk quality using Arduino-Uno microcontroller and various sensors and devices to measure pH, color,odour,temperature and turbidity.
* To make use of an AI model to eliminate any human error and to ensure near impartial grading of milk with a very high accuracy(~99%).
* To keep costs as low as possible while building this proof of concept using materials which are relatively easy to procure as well as assemble.
* To make the device and entire process as user friendly and easy to use as possible.
* To ensure the learning curve to use this device is not large and to enable anyone with no prior knowledge to be able to carry out this test with minimal instructions.
* To possibly document the assembly and building process of this device to help other enthusiasts and professionals make similar projects.
* The code should be written in such a way that it works for any other similar sensors with no or minimal changes to the code.
* To help make this project open-source, any designs, code,datasets,models used should be made available to public access.

## 5.2 Project Motivations

* Agricultural Impact: It will have a huge impact on dairy farmers and the agriculture industry as it will contribute to ensuring that farmers get a fair price for their milk based on quality and also enhancing agricultural practices.
* Innovation in Food-Technology: It is an exciting exploration into the intersection of technology and food safety. It can have a positive impact on public health and consumer confidence.
* Community Impact: A reliable milk purity tester can be used locally and potentially boost economic growth by promoting quality dairy production.
* Sustainability: It will lead to safer food production, reduced health risks, long term impact on agriculture and economic viability for farmers, all keeping in mind the effects on the environment.
* Real-World Relevance: It solves a common issue in not only India but also all over the world. It creates a consumer trust on the food industry as public health and safety is not ignored.
* It also enhances our knowledge as students, enlightening us about various theories, factors affecting food, sensors and a lot more.

# 

# Future Scope of the Project

* Integration and Implementation of IoT Sensors: Additional IoT sensors for real-time monitoring of environmental conditions, cow health and other factors affecting milk quality.
* ‘Smart Milk Cards’: RFID tagged cards could be provided to each milk farmer containing the details of the quality of the milk provided by him and his sales.
* Customer Connect: Customers can know the quality of the milk they are receiving and can assess the service being provided to them.
* Mobile Application Integration: A user-friendly mobile application can be developed to provide real-time data, trends and alerts to both the farmers and the consumers and establish a trusted link between both of them.
* Blockchain Implementation: Incorporation of blockchain technology for secure and transparent record-keeping and enhancing traceability and preventing tampering of records of milk quality, thus contributing to food safety and authenticity.
* Data Analytics for Quality Improvement: Implementation of advanced data analytics to derive insights from the collected data. Analysing trends in milk quality could help in identifying potential issues and lead to proactive measures for improvement and optimization of dairy farming practices.
* Global Adoption: Customise the product for different countries and geographical regions of the world to incorporate the different practices, climate and milk composition.

# Conclusion

In conclusion, the "Integrated Milk Purity Tester: An Arduino Uno-Based Project with AI-Driven Milk Grading" is a significant fusion of technology and agriculture, presenting an innovative response to the pressing problem of milk purity. Through its incorporation of Arduino Uno and artificial intelligence, specifically the k-Nearest Neighbors (kNN) algorithm, this project not only tackles the immediate challenge of maintaining the quality of dairy products, but also sets the stage for future breakthroughs in food safety technology.

The project's potential impact on public health, consumer trust, and community well-being truly emphasises its real-world significance. By utilising the powerful kNN algorithm, this system excels at accurately classifying and grading milk samples in relation to established quality standards, bolstering the precision and dependability of the purity testing procedure.

As we peek into the future, contemplating the possibilities of incorporating IoT and adopting blockchain technology, the potential for our "Integrated Milk Purity Tester" to revolutionise the dairy sector becomes clear. Its foundation in the versatile kNN algorithm makes it an indispensable tool for guaranteeing precise and situation-dependent milk classification.

With its potential for global implementation and efforts towards consumer education and satisfaction, the "Integrated Milk Purity Tester" equipped with kNN proves to be a formidable asset in not only transforming dairy farming methods but also in promoting the course of sustainability and food safety. This project, enhanced by kNN technology, serves as a compelling example of how innovative solutions can safeguard the production of reliable and sustainable food for communities around the world.

# 

# References

1. Goswami, S. and Dangi, A. 2021. “Arduino-Based Milk Quality Monitoring System”. IJAEB, 14(2): 245-249.
2. “Chemical Composition of Milk and Milk Products” ,BM Mehta , © Springer-Verlag Berlin Heidelberg 2015 P.C.K. Cheung, B.M. Mehta (eds.), Handbook of Food Chemistry, DOI 10.1007/978-3-642-36605-5\_31
3. Manisarathi A, Raveesh SJ ,“Automated Milk Quality Analyzer with Billing System”,International Journal of Environment, Agriculture and Biotechnology Vol-7, Issue-4; Jul-Aug, 2022
4. Shubhangi Verulkar, Gaurav Chavan, Kiran Patil, Harshal Chaudhary

JETIR1904G73-”Milk Quality and Quantity Checker”© 2019 JETIR April 2019, Volume 6, Issue 4

1. Shahadat Uddin, Ibtisham Haque, Haohui Lu, Mohammad Ali Moni, Ergun Gide Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci Rep* 12, 6256 (2022).
2. Hari Antoni Musril , S Saludin , Winci Firdaus , Usanto S , K Kundori , Robbi Rahim. Using k-NN Artificial Intelligence for Predictive Maintenance. Volume 10 Issue 6, 1-8, (2023).
3. Zinnia Sultana , Ashifatul Ferdousi , Farzana Tasnim and Lutfun Nahar. An Improved K-Nearest Neighbour Algorithm for Pattern Classification. International Journal of Advanced Computer Science and Applications, Vol. 13, No. 8, 2022.
4. Shichao Zhang, Xuelong Li, Ming Zong, Xiaofeng Zhu, Ruili Wang. Efficient kNN Classification With Different Numbers of Nearest Neighbors.IEEE Transactions on Neural Networks and Learning Systems ( Volume: 29, Issue: 5, May 2018).
5. Terahertz Time-Domain Spectroscopy to Identify and Evaluate Anomer in Lactose - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Molecular-structures-of-a-D-lactose-a-lactose-and-b-D-lactose-b-lactose\_fig2\_273751062 [accessed 12 Dec, 2023]