



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 13, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Prediction of Milk Quality Using a Supervised Machine Learning Technique with Hardware Testing

Mr. V. Sornagopal<sup>1</sup>. Mr. S.Senthilkumar<sup>2</sup>

<sup>1&2</sup> Associate Professor, Department of Electronics and Communication Engineering, GRT Institute of Engineering and Technology, Tiruvallur District, Tamil Nadu, India.

**ABSTRACT:** 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.

**KEYWORDS:** Arduino, Milk Adulteration, Conductive Sensors.

## I. INTRODUCTION.

AI has contributed significantly to providing effective solutions to major human and societal problems under various fields including natural language processing (NLP), which employs computational and linguistics techniques to aid computers understand and sometimes generate human languages in the form of texts and speech/voice.1 Prominent contributions in the field of NLP under active research include translation systems, information retrieval (IR), questions and answering (Q & A) systems, text summarization systems, sentiment analysis (SA),1, 2 and so on. Branching from the field of SA whose core intent is to analyze human language by extracting opinions, ideas, and thoughts through the assignment of polarities either negative, positive, or neutral is the subfield of emotion detection (ED), which seeks to extract finer-grained emotions such as fear, sad, angry, and so on, from human languages rather than coarse-grained and general polarity assignments in SA.

## IMPLEMENTATION OF MACHINE LEARNING.

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated in to three categories. There are supervised learning, unsupervised learning and reinforcement learning. Supervised learning program is both given the input data and the corresponding labelling to learn data has to be labelled by a human being beforehand. Unsupervised learning is no labels. It provided to the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

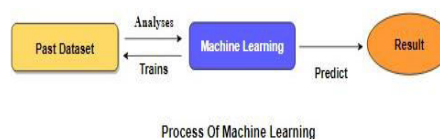


Fig [1]: Process of Machine Learning

## II. LITERATURE SURVEY

**General--** 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. Usually, it has an organizational pattern and combines both summary and synthesis.

[1] *Mid infrared spectroscopy and milk quality traits (2021)*: A chemometric data analysis challenge has been arranged during the first edition of the “International Workshop on Spectroscopy and Chemometrics”, organized by the Vistamilk SFI Research Centre and held online in April 2021. The aim of the competition was to build a calibration model in order to predict milk quality traits exploiting the information contained in mid-infrared spectra only. Three different traits have been provided, presenting heterogeneous degrees of prediction complexity thus possibly requiring trait-specific modelling choices. In this paper the different approaches adopted by the participants are outlined and the insights obtained from the analyses are critically discussed.

[2]. *Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods (2021)* : Numerous statistical machine learning methods suitable for application to highly correlated features, as those that exist for spectral data, could potentially improve prediction performance over the commonly used partial least squares approach. Milk samples from 622 individual cows with known detailed protein composition and technological trait data accompanied by mid-infrared spectra were available to assess the predictive ability of different regression and classification algorithms. The regression-based approaches were partial least squares regression (PLSR), ridge regression (RR), least absolute shrinkage and selection operator (LASSO), elastic net, principal component regression, projection pursuit regression, spike and slab regression, random forests, boosting decision trees, neural networks (NN), and a post-hoc approach of model averaging (MA).

[3]. *Prediction of Raw Milk Microbial Quality Using Data Mining Techniques (2014)*: Data mining techniques were applied to predict raw milk quality in terms of methylene blue reduction time (MBRT) from the independent parameters of raw milk inspection parameters such as travel time, temperature of milk, solid-not-fat, %fat, acidity and specific gravity. Predictive models were developed and the performance of 3 data mining algorithms namely, Multiple Linear Regression (MLR), Artificial Neural Network (ANN) and K-Nearest neighbour (KNN), was measured in terms of average error and Root Mean Square Error (RMSE).

## III. 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:*

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

## IV. 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, a better algorithm is used for the prediction of the outcome.

Advantages:

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

SYSTEM ARCHITECTURE:

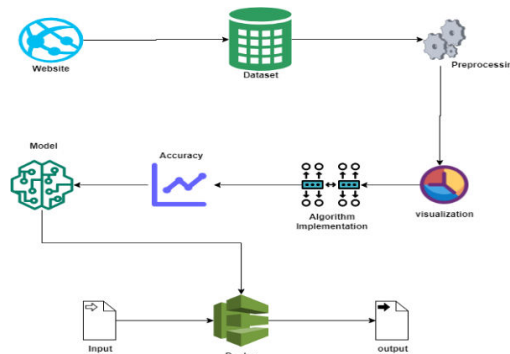


Fig [2]: System Architecture.

ALGORITHM IMPLEMENTATION:

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. A way to do this is to use different visualization methods to show the average accuracy, variance, and other properties of the distribution of model accuracies.

*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.).

*Random Forest Classifier:* 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.

*MLP Classifier:* Multilayer perceptron (MLP) is a supplement of feed forward 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. Like a feed forward 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 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.

*XG Boost Classifier:* XG Boost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted 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. 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.

DEPLOYMENT:

*Django :* 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. Django follows the "Batteries included" philosophy and provides almost everything developers might want to do "out of the box". Because everything you need is part of the one "product", it all works seamlessly together, follows consistent design principles, and has extensive and up-to-date documentation.

OUTPUT SCREENSHOT:



Fig [3]: Output Screensort

V. BLOCK DIAGRAM

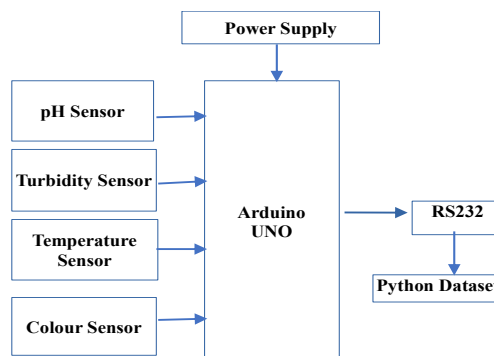


Fig [4]: Block Diagram for Milk Quality Prediction.

VI. HARDWARE REQUIREMENTS

The UNO is the best board to get started with electronics and coding. If this is your first experience tinkering with the platform, the UNO is the most robust board you can start playing with. The UNO is the most used and documented board of the whole Arduino family.

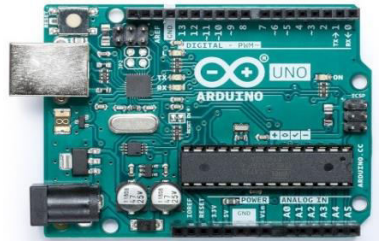


Fig [5]: Arduino Uno

Arduino Uno is a microcontroller board based on the ATmega328P (datasheet). It has 14 digital input/output pins (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and a reset button. It contains everything needed to support the microcontroller; simply connect it to a computer with a USB cable or power it with a AC-to-DC adapter or battery to get started. You can tinker with your UNO without worrying too much about doing something wrong, worst case scenario you can replace the chip for a few dollars and start over again. "UNO" means one in Italian and was chosen to mark the release of Arduino Software (IDE) 1.0. The Uno board and version 1.0 of Arduino Software (IDE) were the reference versions of Arduino, now evolved to newer releases. The Uno board is the first in a series of USB Arduino boards, and the reference model for the Arduino platform; for an extensive list of current, past or outdated boards see the Arduino index of boards.

#### pH Sensor

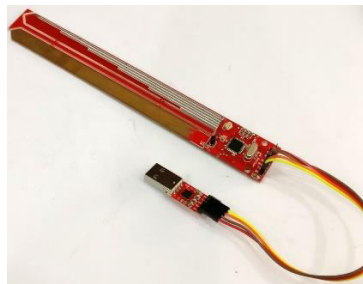


Fig [6]: pH Sensor.

pH sensors measure the level of pH in sample solutions by measuring the activity of the hydrogen ions in the solutions. This activity is compared to pure water (a neutral solution) using a pH scale of 0 to 14 to determine the acidity or alkalinity of the sample solutions. A pH Meter is a device used for potentiometrically measuring the pH, which is either the concentration or the activity of hydrogen ions, of an aqueous solution. It usually has a glass electrode plus a calomel reference electrode, or a combination electrode.[1] pH meters are usually used to measure the pH of liquids, though special probes are sometimes used to measure the pH of semi-solid substances.

*pH Measurement with Arduino* We use analog pH meter sensor, specially designed for Arduino controllers and has built-in simple, convenient and practical connection and features. It has an LED which works as the Power Indicator, a BNC connector and PH2.0 sensor interface. To use it, just connect the pH sensor with BNC connector, and plug the PH2.0 interface into the analog input port of any Arduino controller. Attention: In order to ensure the accuracy of the pH probe, you need to use the standard solution to calibrate it regularly. These standard calibration solutions come with this pH sensor. Generally, the period is about half a year. If you measure the dirty aqueous solution, you need to increase the frequency of calibration.

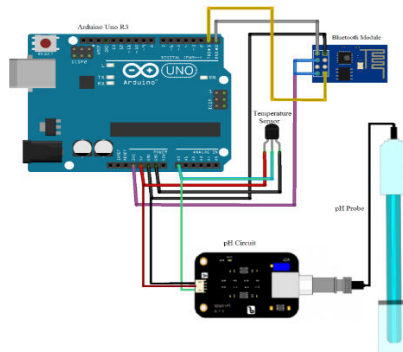


Fig [7]: pH Measurement With Sensor.

### Colour Sensor



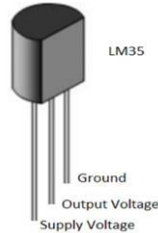
Fig [8]: Colour sensor TCS3200

White light is a mixture of three basic colours known as primary colours. They are red, blue, and green. These colours have different wavelengths. Combinations of these colours at different proportions create different types of colours. When the white light falls on any surface, some of the wavelengths of the light are absorbed by the surface while some are reflected based on the properties of the surface material. Colour of the material is detected when these reflected wavelengths fall on the human eye. A material reflecting wavelengths of red light appears as red. The component used to detect colours is the Colour sensor.



Fig[9]: Turbidity Sensor.

The turbidity sensor detects water quality by measuring the levels of turbidity. It uses light to detect suspended particles in water by measuring the light transmittance and scattering rate, which changes with the amount of total suspended solids (TSS) in water. As the TSS increases, the liquid turbidity level increases. Turbidity sensors are used to measure water quality in rivers and streams, wastewater and effluent measurements, control instrumentation for settling ponds, sediment transport research and laboratory measurements. This sensor provides analog and digital signal output modes. The threshold is adjustable when in digital signal mode. You can select the mode according to your MCU.

*Temperature Sensor**Fig [10]: LM35-Temperature Sensor*

The LM35 series are precision integrated-circuit temperature devices with an output voltage linearly proportional to the Centigrade temperature. The LM35 device has an advantage over linear temperature sensors calibrated in Kelvin, as the user is not required to subtract a large constant voltage from the output to obtain convenient Centigrade scaling. The LM35 device does not require any external calibration or trimming to provide typical accuracies of  $\pm 1/4^\circ\text{C}$  at room temperature and  $\pm 3/4^\circ\text{C}$  over a full  $-55^\circ\text{C}$  to  $150^\circ\text{C}$  temperature range. Lower cost is assured by trimming and calibration at the wafer level. The low-output impedance, linear output, and precise inherent calibration of the LM35 device makes interfacing to readout or control circuitry especially easy. The device is used with single power supplies, or with plus and minus supplies. As the LM35 device draws only  $60\ \mu\text{A}$  from the supply, it has very low self-heating of less than  $0.1^\circ\text{C}$  in still air. The LM35 device is rated to operate over a  $-55^\circ\text{C}$  to  $150^\circ\text{C}$  temperature range, while the LM35C device is rated for a  $-40^\circ\text{C}$  to  $110^\circ\text{C}$  range ( $-10^\circ$  with improved accuracy). The LM35-series devices are available packaged in hermetic TO transistor packages, while the LM35C, LM35CA, and LM35D devices are available in the plastic TO-92 transistor package. The LM35D device is available in an 8-lead surface-mount small-outline package and a plastic TO-220 package.

**VII. CONCLUSION**

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 of higher accuracy score algorithm will be found out. The founded one is used in the application which can help to find the quality of the milk.

**REFERENCES**

- [1] S. R. Karunathilaka, B. J. Yakes, K. He, L. Brückner, and M. M. Mossoba, "First use of handheld Raman spectroscopic devices and on-board chemometric analysis for the detection of milk powder adulteration," *Food Control*, vol. 92, pp. 137–146, Oct. 2018.
- [2] M. Di Domenico, M. Di Giuseppe, J. D. Wicochea Rodríguez, and C. Cammà, "Validation of a fast realtime PCR method to detect fraud and mislabeling in milk and dairy products," *J. Dairy Sci.*, vol. 100, no. 1, pp. 106–112, Jan. 2017.
- [3] A. Azmi et al., "Performance of coating materials on planar electromagnetic sensing array to detect water contamination," *IEEE Sensors J.*, vol. 17, no. 16, pp. 5244–5251, Aug. 2017.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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, Art. no. 100433.
- [8] 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.





# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details