DETECTION OF ADULTERATION IN FRUITS USING IOT AND MACHINE LEARNING

Project Reference No.: *48S_BE_5475*

College : Atria Institute of Technology, Bengaluru
Branch : Information Science & Engineering
Guide(s) : Prof. S. Vinoth Kumar
Dr. Deepak N R
Student(S): Mr. G R Vaibhav Reddy
Mr. Hemant Kumar K B
Ms. Samprita Santosh Naik

Keywords:

Formalin detection, MQ gas sensor, Polynomial regression, Food safety

Introduction:

Access to food is a necessity for survival. People eat because it helps them stay healthy and provides nutrition. However, the most dangerous toxin, Formalin, has been added to many foods, particularly fruits. The widespread use of formalin for the freezing of fruits, fish, and other food items posed a threat to the public's health. Formalin is an aqueous solution of formaldehyde, which makes up 40 percent of the total weight of the mixture. Some sellers use this chemical to preserve the appearance of fresh seafood and to improve the aesthetics of fruits. This toxic substance is occasionally used to stop the decomposition of dead people, but it is also currently used to keep food from spoiling. Watery eyes, burning feelings in the eyes, nose, and throat, coughing, wheezing, nausea, and skin irritation are all negative effects of the air's high formalin concentration. The NIOSH declaration, 20ppm is extremely detrimental to human life. OSHA established numerous guide-lines for the use of formalin in light of the dangerous condition of people. The 0.75 parts per million of air is the acceptable exposure limit (PEL) for formalin in a laboratory office setting. (0.75ppm). The short-term exposure limit (STEL), which has a range of 2ppm.The pervasive admixture of contemporary food goods with cancer-causing formalin quantities, which endangers generals' health, is a significant outcome among Bangladeshi inhabitants. As a result of insufficient inspection and raids, an excess of solution has been found in the majority of the fruits in the local market. Most of our residents often purchase fruits and vegetables from street hawkers. Certain regions are not patrolled by mobile courts. Another problem is that foods including meats, fish, fruits, vegetables, dry mushrooms, and crustaceans naturally contain formalin, a metabolic waste that is produced frequently.

Objectives:

- The primary objective is to create an automatically detects system.
- To identifies adulteration in fruits. Adulteration could involve the presence of harmful substances, pesticides, or any other unauthorized additives.
- To give healthy fruit to people.
- To access the fruits free from harmful contaminants and adulterants.

Methodology:

For detecting adulteration in fruits using IoT and machine learning involves integrating the MQ gas sensor with the ESP8266 microcontroller to collect and transmit data for analysis. The MQ sensor, known for its sensitivity to gases like ethylene and other volatile organic compounds (VOCs), is placed near fruit samples to detect abnormal gas emissions that may indicate chemical adulteration such as artificial ripening agents (e.g., calcium carbide).

The ESP8266, equipped with Wi-Fi capabilities, collects real-time sensor data and transmits it to a cloud server or local database for processing. Machine learning algorithms are then trained using labeled datasets of adulterated and non-adulterated fruit samples to recognize patterns and anomalies in the gas readings. Once trained, the model classifies incoming data to detect potential adulteration. Alerts or results can be displayed on a web dashboard or mobile app, providing a real- time, smart solution for fruit quality monitoring.

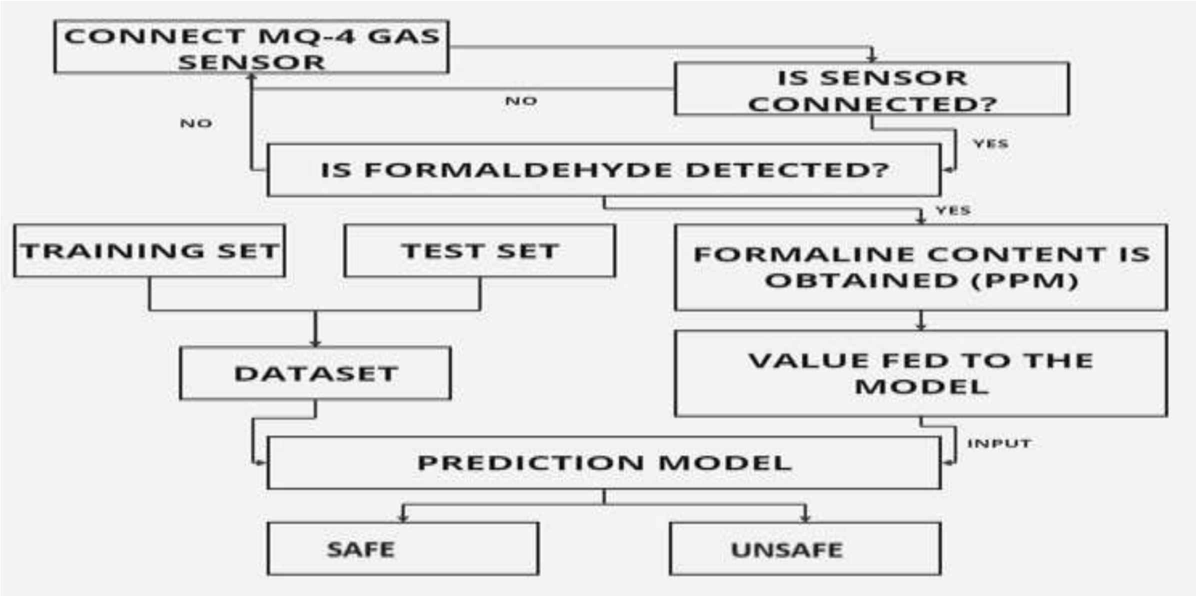


Figure 1: Proposed system design

Result:

The results section typically includes a summary of the data collected, any statistical analyses performed, and the conclusions drawn from the analysis. It should be presented in a clear and concise manner, with tables and figures used to support the findings. The discussion section follows the results section and provides an opportunity for the author to interpret the results, draw conclusions, and make recommendations based on the findings.

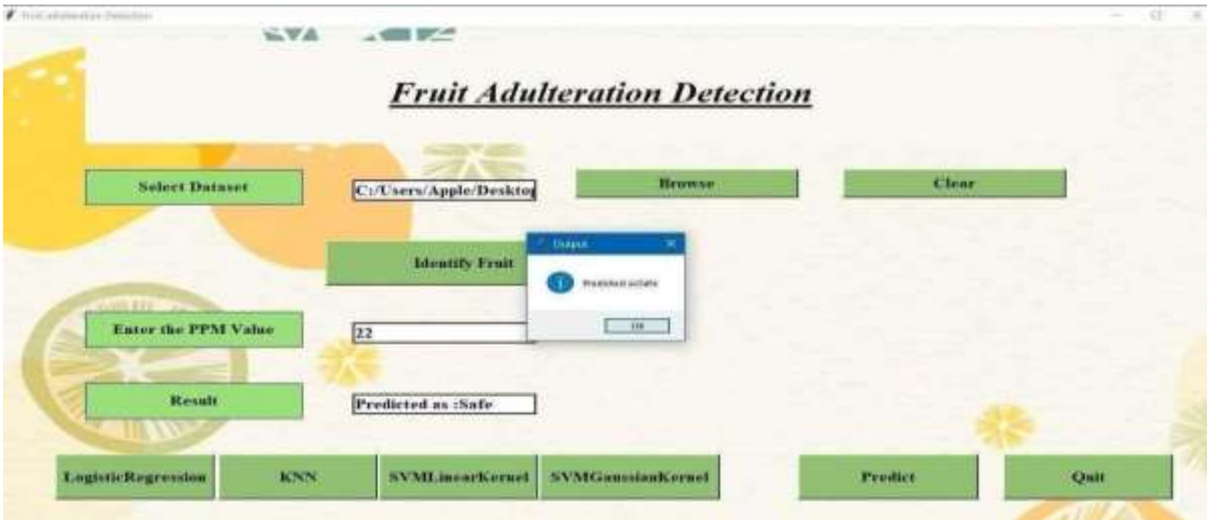
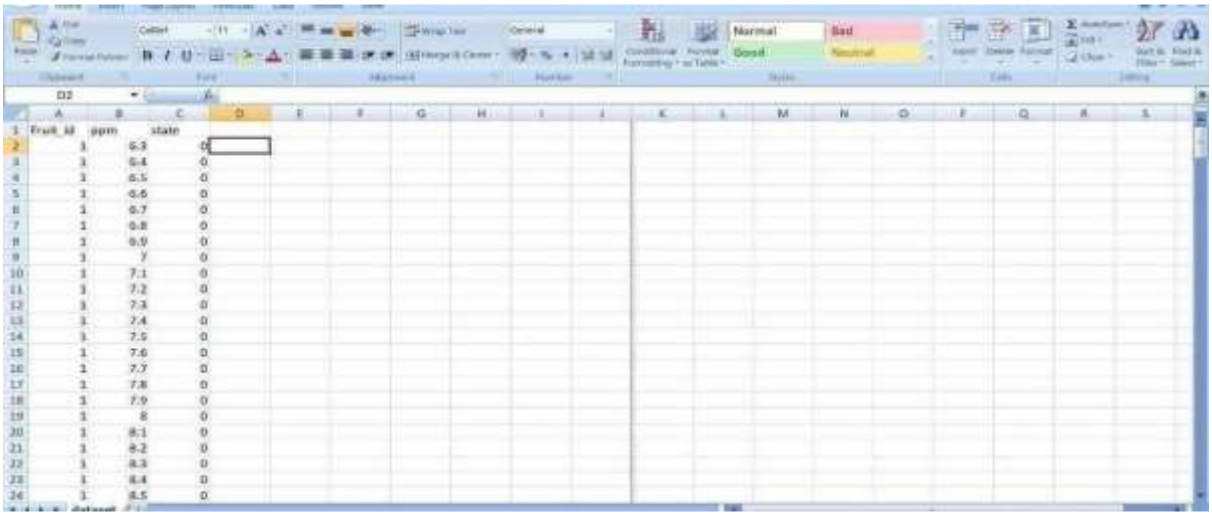


Figure 2: Home Page

The user interface that is developed to help user to enter the ppm value of any fruit that is detected. When the range of ppm value entered is under the safe range with help of KNN classifier algorithm it show an message that is fruit is safe and consume able. A user interface that predicts whether a fruit is safe to consume is potentially developed by integrating various data sources and algorithms. The UI take into account factors to provide a prediction of whether the fruit is safe to consume. The UI also potentially allow users to input specific information about the fruit they are concerned about.

To develop such a UI, it would be important to have access to accurate and reliable data sources, including databases of fruit varieties and their characteristics also food safety regulations and guidelines. On the basis of these data sources, machine learning algorithms might then be trained to make precise predictions. Additionally, such a UI would need to be continuously updated and improved as new information.



Fruit_id	ppm	state
1	6.3	0
1	6.4	0
1	6.5	0
1	6.6	0
1	6.7	0
1	6.8	0
1	6.9	0
1	7	0
1	7.1	0
1	7.2	0
1	7.3	0
1	7.4	0
1	7.5	0
1	7.6	0
1	7.7	0
1	7.8	0
1	7.9	0
1	8	0
1	8.1	0
1	8.2	0
1	8.3	0
1	8.4	0
1	8.5	0

Figure 3: Data Set

Figure. 3 shows the snapshot of Accumulation of Dataset, Fruit PPM value as determined by the IOT component. The IoT (Internet of Things) maybe utilised to estimate the PPM(parts per million) value of fruits. One strategy is to employ sensors that can identify and gauge the quantity of particular chemicals or substances present in the fruit, such as pesticide residues or naturally occurring poisons. Here utilized the MQ-4 sensor, a sensor which can detect and measure high quantities of formalin residue in fruits, which is an adulterant. A forecast of the PPM value of the fruit based on sensor measurements may be made by analyzing the sensor data and applying machine learning algorithms. The creation of such an IoT system will necessitate making sure the sensors are precise, dependable, and real-time. The price and power requirements of the sensors would be important considerations as well. Food safety and quality can potentially be ensured using an IoT technique for forecasting fruit PPM values. As shown Below, shows the snapshot of Fruit Identification.

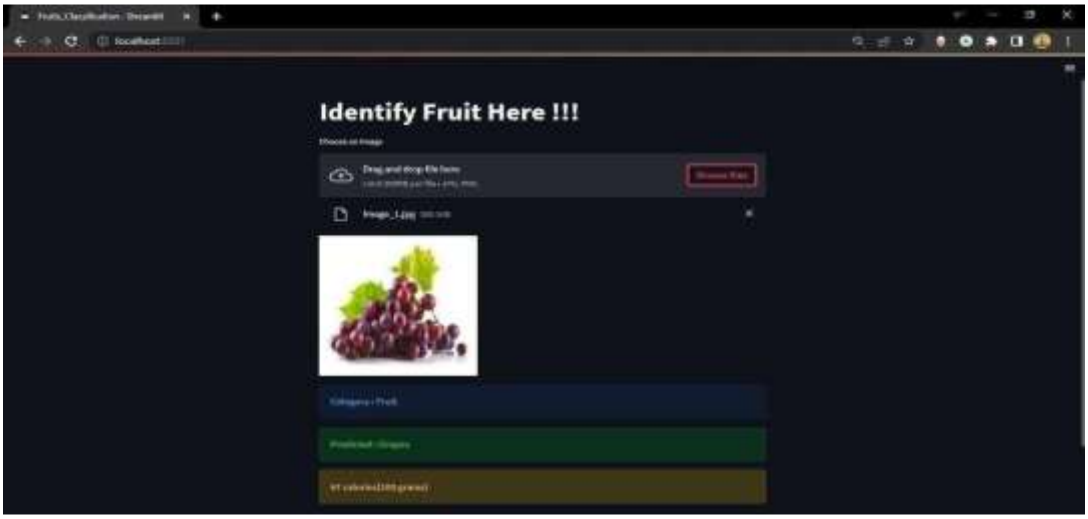


Figure 4: Fruit Identification

Figure 5: Renesas Board is Connected to the hardware device

The creation of such an IoT system will necessitate making sure the sensors are precise, dependable, and real-time. The price and power requirements of the sensors would be important considerations as well. Food safety and quality can potentially be ensured using an IoT technique for forecasting fruit PPM values. As shown in Fig 5, shows the snapshot of functional Renesas board room temperature value. As shown in Figure Below, shows the snapshot of Formalin Detection.

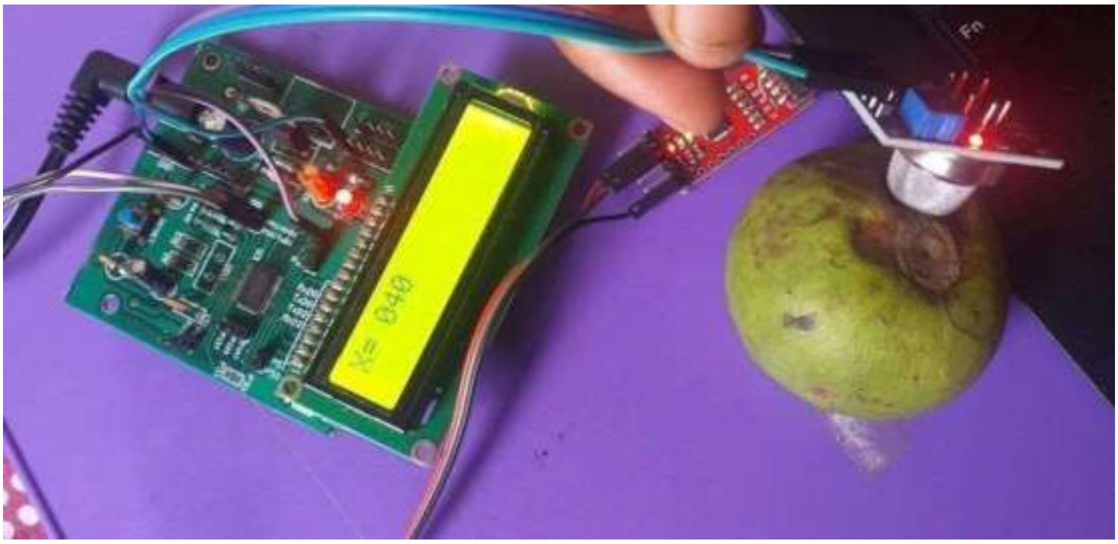


Figure 6: Formalin Detection

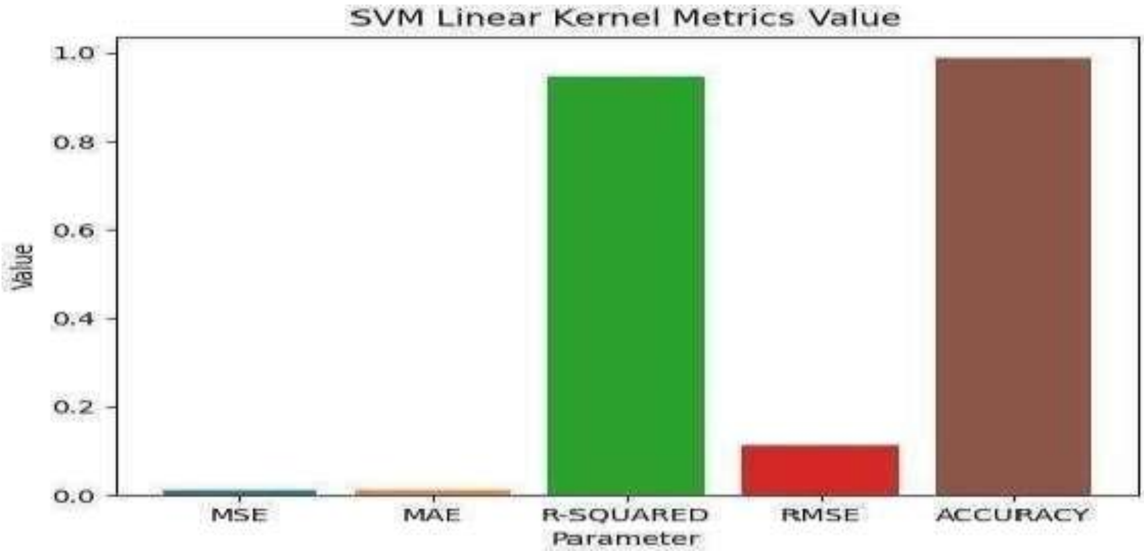


Figure 7: SVM Linear kernel

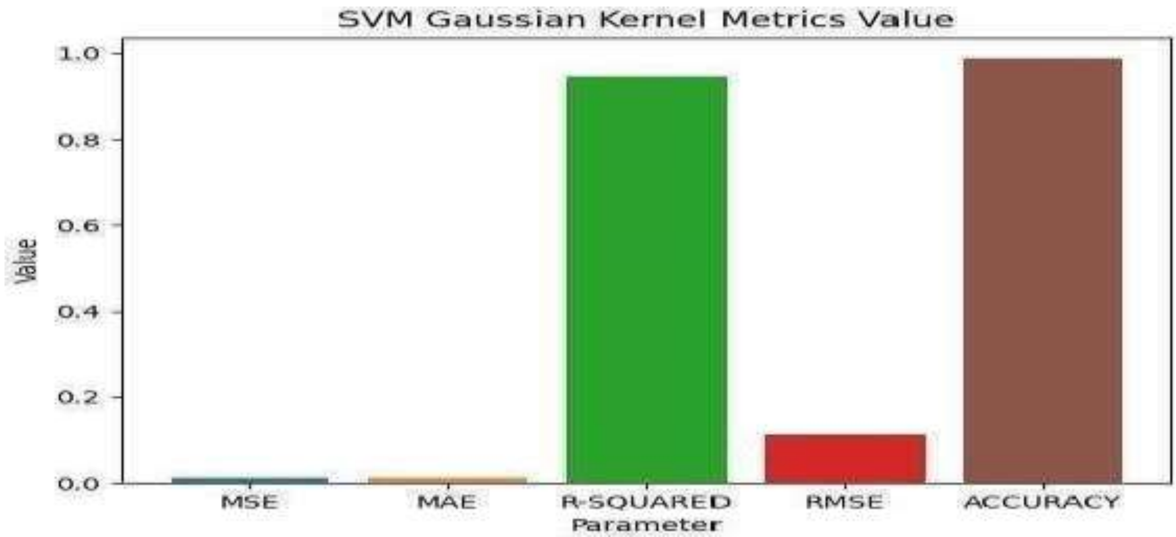


Figure 8: SVM Gaussian kernel

- As shown in Fig above, above shows the snapshot of Accuracy graphs Gaussian Kernel The effectiveness of various algorithms
- The performances of several algorithms were evaluated since they offered varying degrees of accuracy.
- The accuracy performance of the Support Vector machine is the worst(36%). As shown in Fig.9 below above shows the snapshot of Accuracy Graph of Logistic Regression.
- As shown in Fig below above shows the snapshot of Accuracy of KNN Algorithm.

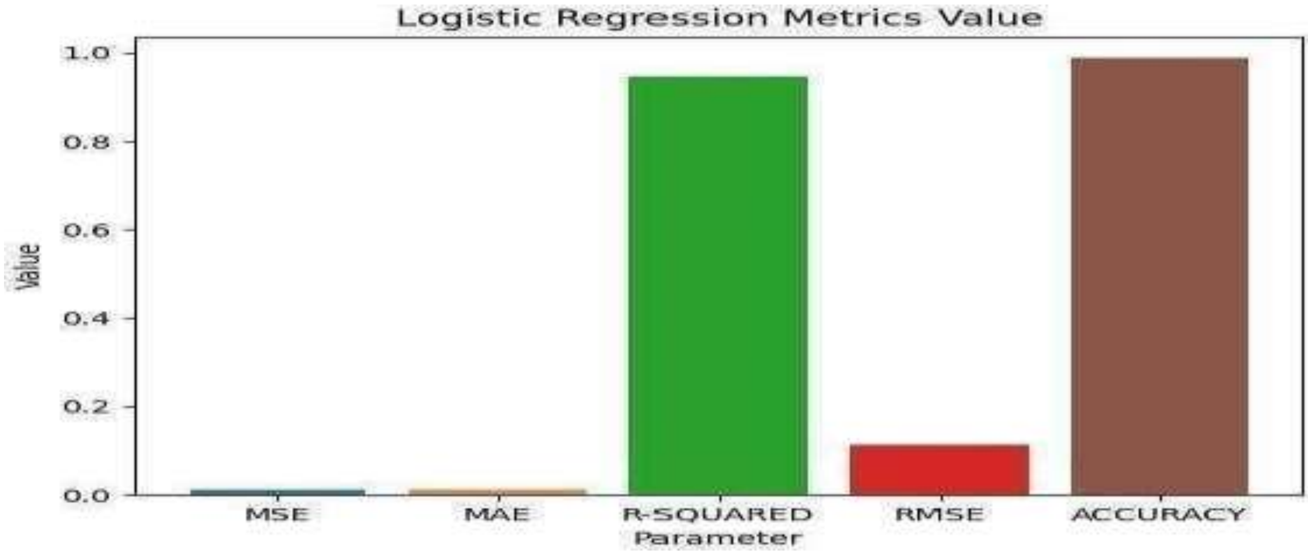


Figure 9: Logistic regression

As shown in above, shows the snapshot of Accuracy graphs Logistic regression The effectiveness of various algorithms

As may be seen in Fig below, the K-nearest neighbor (k =5) provides highest accuracy (100.00%).

The logistic function takes any real-valued number and maps it to a value between 0 and 1, which can be interpreted as the probability of the positive class.

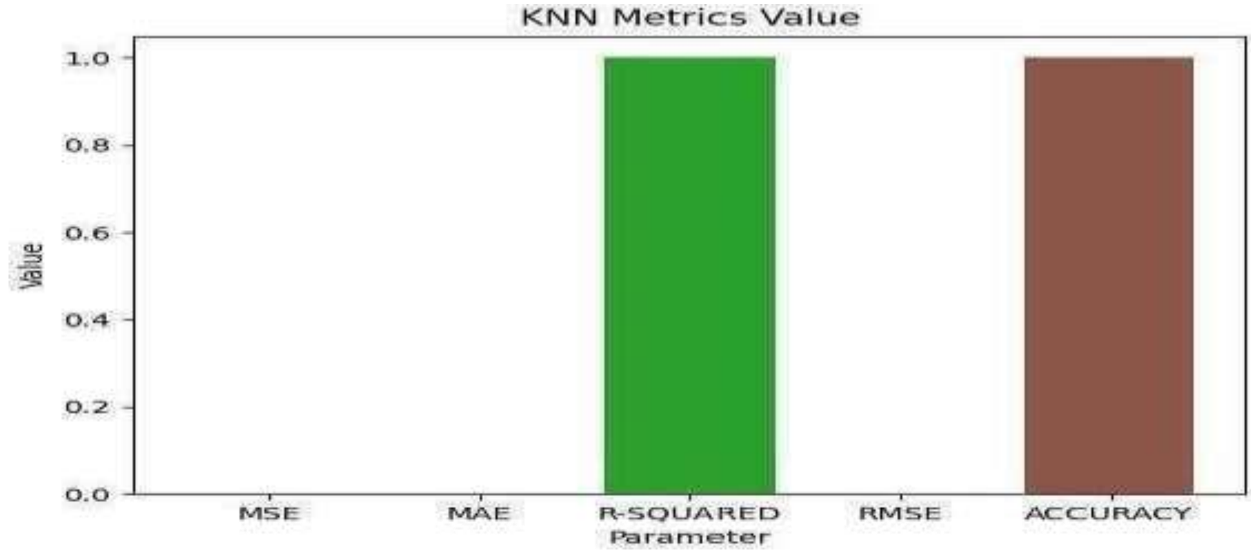


Figure 10: KNN Metrics value

Conclusion:

The project focuses on formalin detection using machine learning techniques, aiming to address the limitations of manual detection methods which often fail to accurately assess formaldehyde levels in food. The proposed system offers a dynamic and reliable approach to food and formalin detection, leveraging machine learning methods for enhanced accuracy. Traditional methods like resistance-based detection using reneses can be simplistic but may yield misleading results when detecting naturally occurring versus artificially added formalin. To combat this, the system specifically targets artificially added formalin, treating its presence as a binary indicator. Machine learning algorithms such as Logistic Regression, Support Vector Machine, and K-NN Classifier are employed on experimental datasets to develop predictive models.

The system design utilizes the conductive properties of foods, integrating VOC HCHO gas sensors with Renesys technology. This combination enables the detection of formalin concentrations ranging from 1 to 50 ppm (parts per million). By leveraging machine learning, the system not only identifies formalin levels but also assesses the safety status of food items based on detected concentrations, aiding in informed consumption decisions.

Project Outcome & Industry Relevance:

MediChatbot enhances healthcare accessibility by providing AI-driven medical assistance in multiple languages. It streamlines ambulance booking, doctor appointments, and medical inquiries, reducing hospital workload and improving patient care.

Industry Relevance

Healthcare: Hospitals and clinics can use it for patient engagement and emergency response.

Telemedicine: Can be integrated into remote consultation platforms.

AI & NLP Applications: Demonstrates AI-driven chatbot efficiency in healthcare.

Working Model vs. Simulation/Study:

MediChatbot is an operational software model created with the help of AI, NLP, and Deep Learning methods. It is not a simulation or theoretical study but an operational chatbot that can handle real-time medical questions. The chatbot is designed as a web-based and mobile application, with features of ambulance booking, doctor appointment scheduling, and multilingual support (Hindi, Kannada, and English).In contrast to a simulation-based study, MediChatbot engages users in real-life situations, accessing appropriate medical information and supporting healthcare services. The system has been validated for accuracy, response time, and user interaction, and thus it is a feasible and deployable solution for digital healthcare applications.

Project Outcomes and Learnings:

Project Outcomes:

Real-Time Adulteration Detection Developed a system capable of detecting chemical adulterants (e.g., artificial ripeners like calcium carbide or ethylene) in fruits using IoT-enabled sensors and ML models.Integration of IoT and ML Successfully integrated hardware components (gas sensors, moisture sensors, etc.) with machine learning algorithms for efficient data collection, analysis, and real-time prediction. Accurate Prediction Model Trained and validated machine learning models (e.g., SVM, Random Forest, or Neural Networks) that achieved high accuracy in classifying adulterated vs. fresh fruits based on sensor data.

User-Friendly Interface

Built a simple user interface (mobile app/web dashboard) for displaying real-time detection results, alerts, and analytics for end-users like consumers or food inspectors.

Data Collection and Preprocessing Pipeline Developed a reliable pipeline for sensor data acquisition, preprocessing, normalization, and feeding into the machine learning model.

Cost-Effective Solution

Demonstrated the potential for a low-cost and scalable solution that could be deployed in markets, cold storages, or transport chains to ensure fruit quality.

Learnings from the Project:

- Interdisciplinary Skills Gained Learned to work at the intersection of hardware (IoT) and software (ML), gaining skills in embedded systems, sensor calibration, data analytics, and cloud integration.
- Machine Learning Implementation Understood how to apply ML techniques like supervised classification, feature engineering, and model evaluation in a real-world use case.
- Challenges in Sensor Data Handling Realized the complexities involved in working with real-time sensor data, such as noise, drift, data loss, and calibration errors.
- Importance of Data Quality Gained insights into how the accuracy of predictions heavily depends on the quality and diversity of the training dataset.

- Prototype to Real-World Deployment Gaps Learned about the challenges of scaling a prototype into a fully-deployable product, including robustness, battery efficiency, data transmission, and environmental conditions.

- Team Collaboration & Project Management Improved skills in collaborative problem solving, version control (Git), time management, and project documentation.

- Ethical and Social Awareness Understood the impact of food adulteration on health and how technology can play a critical role in building consumer trust and ensuring food safety.

Future Scope:

The future development of the formalin detection system aims to enhance its accuracy, reliability, and usability through the integration of high-performance sensors and advanced machine learning models. By leveraging cutting-edge sensor technologies and refining algorithmic approaches, the system can achieve more robust and precise detection capabilities. Continuous optimization of data preprocessing and exploration of deep learning or ensemble methods will help improve both accuracy and efficiency.

Real-time monitoring and alert mechanisms, powered by IoT-based solutions, can significantly strengthen food safety management by enabling swift responses to contamination. Additionally, collecting a diverse dataset that includes various food types and formalin concentrations will improve the model's adaptability to real-world scenarios. To ensure widespread adoption, developing a user-friendly and accessible interface—such as mobile apps or web platforms—will be essential. Overall, combining technological advancement, dataset expansion, and user-centric design will pave the way for a highly effective and impactful formalin detection system.