JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN APPROACH TO PREDICT THE QUALITY OF MILK BY UTILIZING MACHINE-LEARNING TECHNIQUES.

Dr.N.Rahul Pal[1]
Dep of Computer Science & Engineering
University College of Engineering
Nannaya University

Ms.Supriya Thontepu[2]
Dep of Computer Science & Engineering
University College of Engineering. Adikavi
Adikavi Nannaya University

Abstract- Milk is a product that benefits people whether raw or pasteurized. Ensuring the production of quality milk to meet consumer demands and expectations is an ongoing problem for the milk dairy sector. All living being's fundamental source of nutrition is milk. Adulterants should not be present in milk of superior grade. The food sector is highly significant today, and it would be helpful to monitor and quickly ascertain the quality of the products produced. Since using milk of poor quality can be detrimental to human health. The goal of this project is to create a system that can accurately forecast the quality of milk. The accuracy and efficiency of the milk quality prediction system will be improved by utilizing machine learning techniques.

Keywords: *Machine Learning, Milk Quality Prediction, Logistic Regression, Random Forest,Label Encoding.*

I.INTRODUCTION

One of the most significant sectors of the global agricultural industry is milk production. Testing dairy products, such as milk, is challenging because these products are usually produced in large quantities. There are numerous health risks to consumers when gallons of milk are spoiled by a single drop of contamination. Quality control should never be compromised, especially when it comes to food products.

Even a small amount of poor-quality milk can cause tons of milk to worsen, leading to significant financial losses. Millions of bacteria can form in spoiled milk in a very short amount of time. Manual methods of quality assessment can lead to shortcomings or puts off in the process. As technology advances, machine learning algorithms are being effectively applied to quality control in the food business.

Milk quality should be evaluated by considering numerous factors, as relying solely on one to establish its quality would not yield reliable results. Milk quality is predicted using characteristics of the milk such as pH, temperature, taste, odor, fat content, and color. The milk quality is categorized as Low, Medium, or High based on these characteristics. In this study, models that predict milk quality are created using logistic regression and random forest techniques.

II.EXISTING SYSTEM

A portion of machine learning techniques are employed in the current milk quality prediction system to forecast the quality of the milk. In this system, the milk quality is predicted using the Naïve Bayes method, which yields 73.25% accuracy, and the Support Vector Machine approach which yields 57.01% accuracy.

Drawbacks:

1. Issues with data quality
2. Interpretation and complexity
3. Overfitting of the model

III.PROPOSED SYSTEM

The dairy sector is responsible for giving consumers high-quality milk.so it's critical to more precisely check the milk's quality. Therefore, in order to obtain more accurate results, the proposed approach utilizes machine learning methods such as Random Forest and Logistic Regression to predict and classify the milk quality as low, moderate, and higher.

A. RANDOM FOREST:

Among the techniques used for supervised learning Random Forest is the widely recognized machine learning algorithm. It can be applied to ML issues involving both classification and regression. Its foundation is the idea of ensemble learning, which is the process of merging several classifiers to solve a challenging issue and enhance the model's functionality. According to its name, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the mean to enhance the predictive value of that dataset."

WORKING: Three primary hyper parameters of random forest algorithms must be established prior to training. These consist of the size of the nodes, the count of trees, and the quantity of characteristics sampled. Regression and classification issues can then be resolved using the random forest classifier. Each decision tree in the ensemble of decision trees used in the random forest technique is made up of a bootstrap sample, which is a sample of data taken from a training set with replacement. A third of that training sample—referred to as the out-of-bag (OOB) sample—is set aside for testing purposes. The majority vote—that is, the most frequently used categorical variable—will determine the projected class in a classification task, while the individual decision trees in a regression job will be averaged. Lastly, that prediction is confirmed using cross-validation using the OOB sample.

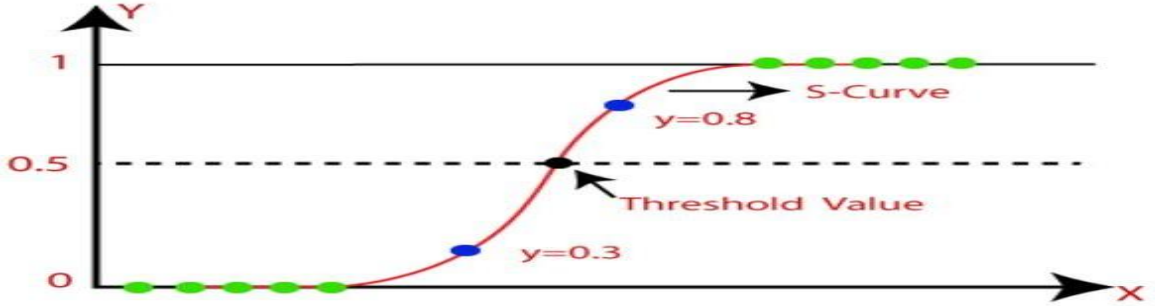
DESIGN: The Random Forest method is shown by the architecture below, which consists of several decision trees.

B.LOGISTIC REGRESSION:

One of the most widely used machine learning algorithms, under the category of supervised learning, is logistic regression. With a given collection of independent factors, it is used to predict the categorical dependent variable. With logistic regression, the result of a categorical dependent variable is predicted.

As a result, a discrete or category value must be the result. Instead of providing the exact values, which are 0 and 1, it provides the probabilistic values, which fall between 0 and 1. It can be either Yes or No, 0 or 1, true or False, etc.

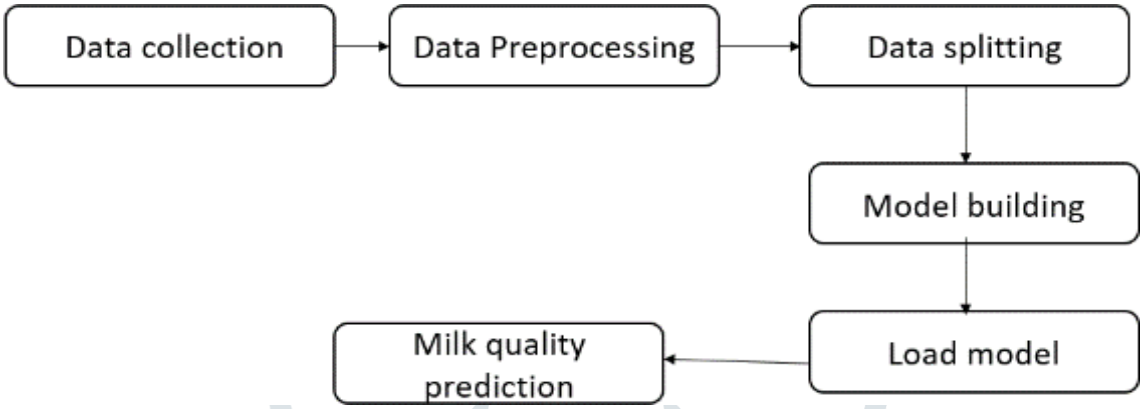
With the exception of how they are applied, logistic regression and linear regression are very similar. While logistic regression is used to solve classification difficulties, linear regression is used to solve regression problems. When using logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function.



IV.SYSTEM ARCHITECTURE:

For our milk quality prediction project, the system architecture consists of the interface between user, user interface, input, data structure, machine learning algorithm, output and output image Used interacts with the system using an intuitive user interface, providing data values for analysis.

Upon user input, the system leverages a pre-existingdataset to machine learning model, enhancing its proficiency in milk quality.



. Before training the model, input data is converted from string data intonumerical values for the model training, after converting the model trained by the input data. As the model analyses the input, it produces results highlighting the quality of the milk grade. These resultsare then displayed on the user interface.
This architectural design ensures a cohesive and efficient flow of information, enabling users to upload, analyze, and interpret milk quality seamlessly. It combines user interaction, advanced model processing, and result visualization to create acomprehensive milk quality prediction system

V.DATA-SET ANALYSIS:

The dataset for milk quality determination contains 1059 records with 6 different components of the milk for training and evaluation of the machine learning model To ensure robustness and generalizability, the dataset is carefully prepared, annotated and partitioned into 70-30 well-balanced training and testing purposes
1) Summary:
Breastfeeding data was collected from the Kaggle product. Based on these variables, data sets can be used to train models that can predict milk quality.

2) Pre-processing data:
The data set is in CSV file format. Clean the data in this step, removing the null values. The dataset contains string values that are not suitable for the system to make predictions, we use a label encoder to convert the string values to numeric values.

3) Classification of data:
In machine learning, it is standard to split data into training and testing sets. Splitting data into training and testing sets is key to testing data mining models. Generally, while separating the data set into the training set and test set makes up the bulk of the data set

They train, and a small portion of the data is used for testing. In this application, the training and test data sets are split in the ratio of 70:30. This means a 70% data set for training and a 30% data set for testing.

4) Model Training:
The Processed data set is used to train the model using machine learning algorithms that can predict milk quality. In this application, the model is trained using Random Forest and Logistic Regression algorithms.

5) Sample analysis:
The performance of the model can be evaluated in separate experiments, ensuring that it can make accurate predictions for new cases. The evaluation criteria for this classification task are precision, recall, F1-score, accuracy.

VI.OBSERVATIONS:

The figures shows the accuracy of the algorithmsused in this milk quality prediction project.

logistic regression

```
In [32]: from sklearn.linear_model import LogisticRegression
         model=LogisticRegression()
         model.fit(X_train,y_train)
Out[32]: LogisticRegression
         LogisticRegression()

In [18]: y_predicted = model.predict(X_test)
         model.score(X_test,y_test)
Out[18]: 0.6698113207547169
```

Fig .6.1. Accuracy of Logistic Regression

Random forest

```
In [23]: from sklearn.ensemble import RandomForestClassifier
         from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

         # Create and train the Random Forest model
         rfc = RandomForestClassifier(n_estimators=100, random_state=42)
         rfc.fit(X_train, y_train)
Out[23]: RandomForestClassifier
         RandomForestClassifier(random_state=42)

In [24]: # Make predictions on the test set
         predictions = rfc.predict(X_test)

         rfc.score(X_test,y_test)
Out[24]: 0.9874213836477987
```

Fig .6.2. Accuracy of Random Forest Algorithm

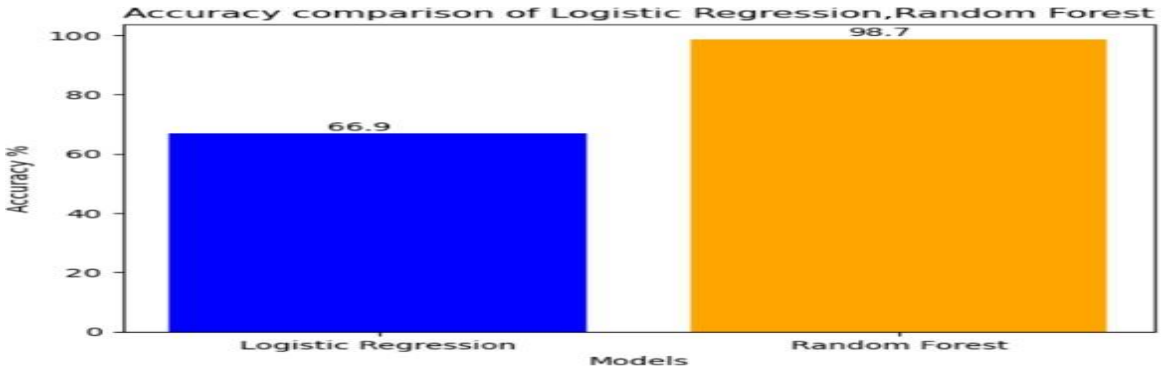


Fig.6.3. Accuracy comparison of algorithms by observing the accuracy of the logistic regression and random forest algorithms.

VII. PERFORMANCE ANALYSIS

In this proposed work Random Forest and Logistic Regression algorithms used for milk quality prediction. To calculate the effectiveness of the algorithm's evaluation metrics are used.

Algorithm	Accuracy	precision	recall
RF	0.98	0.97	0.99
Log. Reg	0.66	0.60	0.72

Table.7.1. Performance Score

1. Accuracy of a classifier predicts the Number of correct predictions from the total number of Predictions.
2. Precision is classifier predicts true positive fromtotal positive predicted.
3. Recall is classifier predicts true positives from actual predicted.

Accuracy=TP+TN / TP+TN+FP+FN
Precision=TP / TP+FP
Recall=TP / TP+FN.

VIII. WEB PAGE

The Milk Quality Prediction project is written as a web application using Streamlit, a Python web framework. The application allows users to record values from their devices, which are processed to determine milk quality. Implemented as a web application, the project provides a user-friendly interface for seamless communication.

The project is written in Python, with basic functionality in a Python file. In addition, SAV file named "milkquality.sav" has been created for the user interface. Python files handle the background processing, including the Streamlit framework for simple web application development. Streamlit is a powerful Python library

It is designed to quickly build interactive web applications with simple Python scripts. It simplifies the process of building data-driven applications by providing an intuitive interface that allows developers to easily turn data analysis and visualization scripts into shared web applications Streamlit allows developers to deliver their Python knowledge and existing libraries such as pandas, matplotlib, scikit-learn Create interactive use dashboards, machine learning prototypes, etc.

Its simple syntax and automated widget generation make it easy for users at all levels of expertise, enabling seamless deployment of data-centric applications without the need for extensive web development knowledge -Processes features efficiently and presents results in a user-friendly manner . The combination of Python and Streamlit adds a lot of functionality and user accessibility to the project.

IX. RESULTS:

The milk quality prediction system output webpagescreens are displayed below:

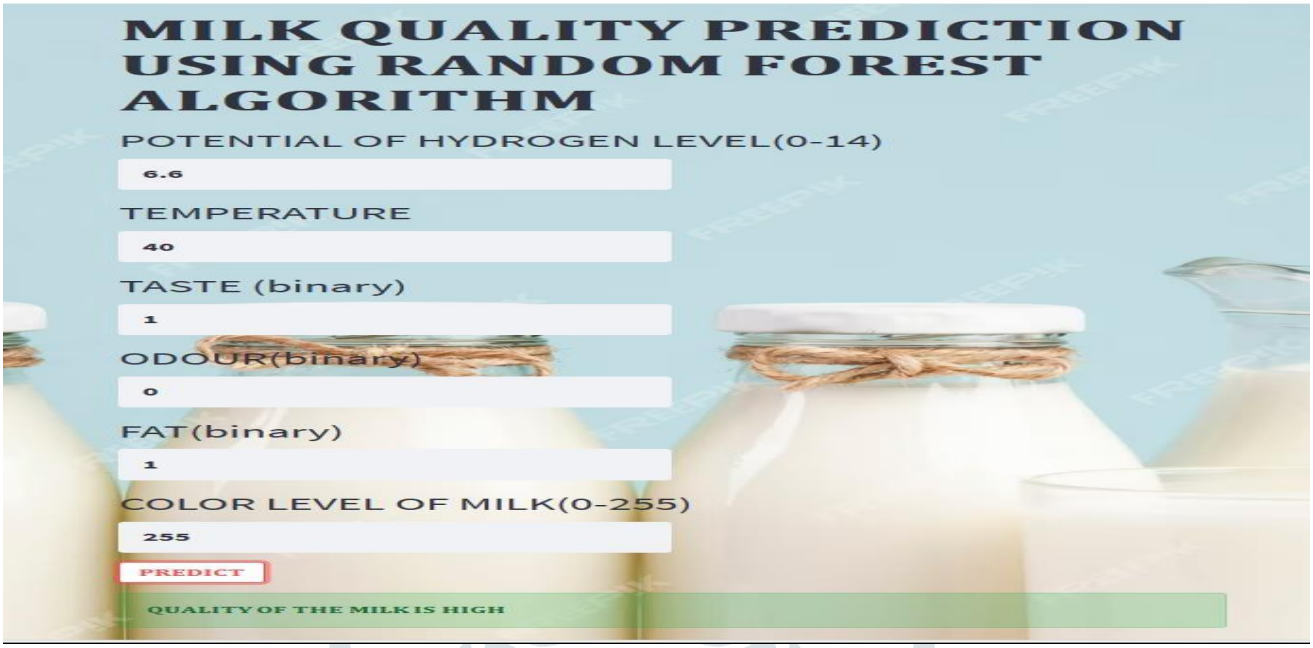


Fig.9.1. Prediction of milk quality as High

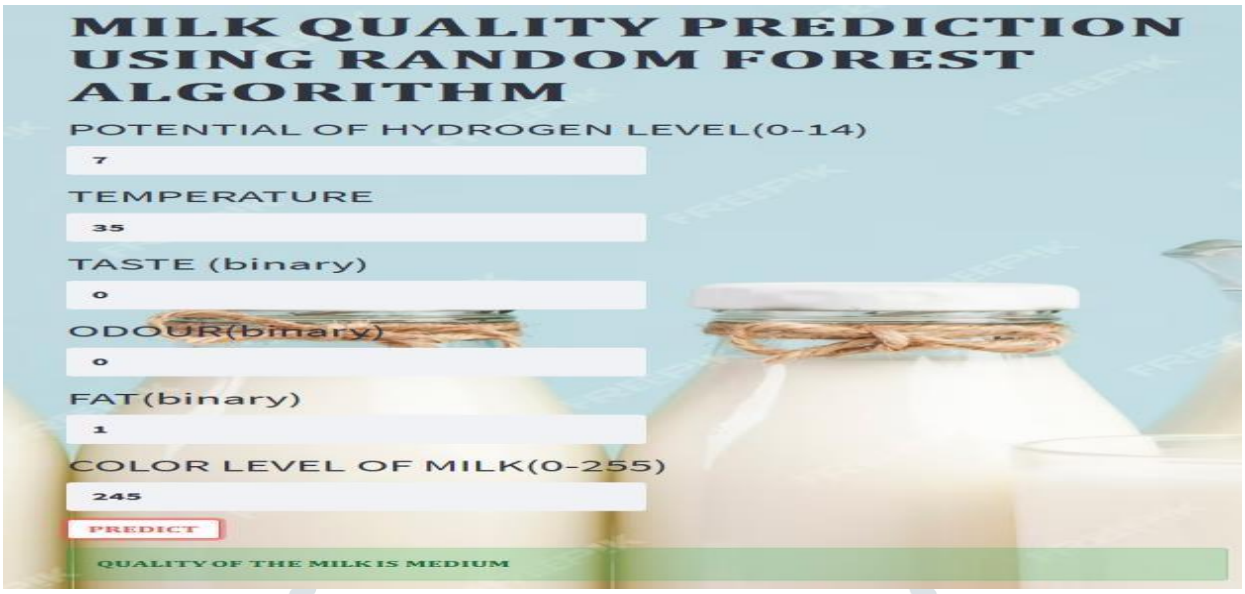


Fig.9.2. *Prediction of milk quality as Low*

Fig .9.3. Prediction of milk quality as Medium

By observing the above output screens, changes in any feature may affect the milk quality. It is important to maintain certain ranges of milk features.By this, we can easily predict milk quality.

X.CONCLUSION:

In conclusion, the proposed work demonstrates a large development in predicting milk first-rate. By leveraging algorithms like Logistic Regression and Random Forest, this work evolved a version that accurately assesses milk best. The accuracy of Random Forest is 98.7% and the accuracy of Logistic Regression is 66.9%. In this mission resulted, the Random Forest algorithm was carried out with greater accuracy than the Logistic Regression algorithm.

XI.REFERENCES:

[1]Anderson, Melisa, et al. "The microbial content of unexpired pasteurized milk from selected supermarkets in a developing country." Asian Pacific journal of tropical biomedicine 1.3 (2011): Volume 1, Issue 3,2011,

Pages 205-211, ISSN 2221-1691, doi:10.1016/S2221-1691(11)60028-2.

[2]Dhanashekar R, Akkinepalli S, Nellutla A. “Milk- borne infections. An analysis of their potential effect on the milk industry”. Germs. 2012 Sep 1;2(3):101-9. doi: 10.11599/germs.2012.1020. PMID: 24432270; PMCID: PMC3882853.

[3]Wenchuan Guo, Xinhua Zhu, Hui Liu, Rong Yue, Shaojin Wang,"Effects of milk concentration and freshness on microwave dielectric properties", Journal of Food Engineering, Volume 99, Issue 3,2010, Pages 344-350,ISSN 0260-8774, doi:10.1016/j.jfoodeng. 2010.03.015.

[4] L. W. Moharkar and S. Patnaik, "Detection and Quantification of Milk Adulteration by Laser Induced Instrumentation," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT),Bombay, India, 2019, pp. 1-5, doi:10.1109/I2CT45611.2019.9033883

ABOUT THE AUTHOR:

Dr.N.Rahul Pal, working as an Assistant Professor in the Department of CSE, University College of Engineering, Adikavi Nannaya University, has experience of 8 years in teaching, interested subjects include Machine Learning, Deep Learning, and Artificial Intelligence.

Ms.Supriya Thontepu, a student of final year B.Tech in Department of CSE, University College of Engineering, Adikavi Nannaya University, Interested in Machine Learning, Deep Learning.