A Data-Driven Approach for Milk Quality Prediction using Machine Learning Techniques

Muhammad Ayoub Kamal[1, 2*], Huma Jamshed[1], Urooj Waheed[1], Yusra Mansoor[1], Laiq Muhammad Khan[3]

***Abstract*— Machine learning-based approaches can be extremely helpful in monitoring the quality of products and making rapid decisions in the food sector, where maintaining standards is crucial. Even a single gram of milk with poor quality can degrade large quantities, leading to significant financial losses. Contaminated milk can harbor millions of bacteria within just a few hours, posing serious health risks to consumers. Therefore, to ensure milk quality, it must be thoroughly examined for the presence of essential components and any potential contaminants. In this study, machine learning algorithms were employed to assess milk quality. Seven factors were considered for evaluation, and the dataset was sourced from the publicly accessible Kaggle data portal. The milk samples were classified into low, medium, and high-quality categories based on these seven characteristics. The K-Nearest Neighbor, Naive Bayes, Multilayer Perceptron, and Support Vector Machine techniques were utilized for classification and estimation. The findings of each method were presented and compared, demonstrating the classification accuracy achieved.**

***Keywords*— Milk Quality, Machine Learning, Quality Prediction, KNN Prediction, Data Driven, Naive Bayes, Multilayer Perceptron, Support Vector Machine**

INTRODUCTION

Modern methods for analyzing and evaluating milk quality, such as machine learning-driven milk quality testing, utilize cutting-edge algorithms [1]. Ensuring that milk meets the highest standards before reaching consumers is essential due to the increasing demand for high-quality dairy products. Traditional milk quality testing techniques have drawbacks, including being time-consuming, subjective, and prone to human error [2]. Milk quality refers to the features and standards that define the safety, freshness, and nutritional content of milk produced and consumed by humans [3]. Factors such as livestock wellness, milk handling and transportation, and the entire milk production process all impact milk quality. The well-being of dairy cows is the primary determinant of milk quality [4].

[1] DHA Suffa University, Karachi
[2] Multimedia University, Cyberjaya
[3] Institute of Business Management (IoBM), Karachi
Country: Pakistan, Malaysia
Email: ayoub.kamal@mmu.edu.my

High-quality milk is produced by healthy cows that are properly fed, regularly seen by veterinarians, and not exposed to harmful toxins. Routine testing for infections like mastitis is necessary to preserve milk quality and avoid contamination. Additionally, maintaining sanitation and hygiene during milking, storage, and shipping is crucial [5]. Preserving milk quality and preventing bacterial contamination requires regular maintenance of milk storage tanks, good hygiene practices by milkers, and proper sanitation of milking machinery, including udder washing. After milking, raw milk must be quickly cooled to a safe temperature to inhibit bacterial growth and prevent the development of spoilage enzymes. Proper storage practices, such as maintaining consistently low temperatures and protecting milk from light and odors, are essential for preserving its freshness and flavor [6]. Rigorous testing and quality control measures are crucial to ensuring milk quality. Regular monitoring of milk composition, including fat, protein, and bacterial counts, helps detect any abnormalities or contamination. Appropriate regulatory standards and testing methods are vital to guarantee consistent milk quality. With advancements in technology, machine learning algorithms are effectively employed in education [7], food [8], transportation [9], healthcare [10], localization [11], security [12], manufacturing [13], and agriculture [14] for quality control. By conducting essential quality tests at various stages, both milk dealers and consumers can be assured of the nutritional value of raw milk [15]. To protect children, especially newborns, from illness, it is crucial that milk is clean and safe. Various processed milk products, including butter, yogurt, cheese, and even cereals, are produced. As the milk market grows, so does the demand for food manufacturers. Consequently, these manufacturers are focusing on improving the quality of soured milk and addressing consumer concerns [16]. Traditional methods for determining milk quality can be error-prone and time-consuming. Therefore, assessing milk quality should involve multiple factors, as relying on a single criterion is inadequate. Implementing quality control measures and utilizing an intelligent system to analyze milk data through various characteristics can be highly effective. Machine learning techniques, which use large datasets to train models, offer a precise and efficient way to evaluate different aspects of milk quality [17].

These parameters may include taste, pH, smell, temperature, color, fat content, and clarity. By analyzing a combination of these parameters, machine learning algorithms can provide an objective assessment of milk quality. The process typically involves collecting milk samples from various sources and measuring these quality parameters accurately. In this investigation, the dataset from the Kaggle repository was used. Machine learning models are then trained on this data to identify patterns and correlations, enabling them to predict outcomes and classify milk samples based on their quality characteristics. One of the main advantages of machine learning-based milk assessment is its ability to detect subtle changes and irregularities in milk composition that may be challenging for human experts to discern [18]. The algorithms' capacity to identify outliers and deviations from the norm allows for the early detection of potential quality issues. Additionally, machine learning models can be regularly updated and refined with new data [19], ensuring that the system remains aligned with evolving milk quality standards. This adaptability contributes to more precise and reliable quality assessments.

LITRATURE REVIEW

Machine learning-based milk quality checking is an emerging field that aims to use computational techniques to improve the assessment of milk quality. This approach leverages advances in data analysis and pattern recognition to create automated systems that can quickly and accurately evaluate milk samples. Machine learning techniques offer an opportunity to overcome these limitations by providing objective and efficient ways of analyzing large datasets. These approaches use algorithms to learn patterns from training data and apply the acquired knowledge to new samples, enabling automated milk quality assessment. Various algorithms, such as Naive Bayes, Random Forest, KNN, and Logistic Regression, are employed in this study. Among them, Random Forest proves to be the most accurate. By utilizing four input features (color, turbidity, temperature, and pH), the suggested model achieves an impressive accuracy of 98.27%. This enables the development of a fully automated and reliable gadget that can be conveniently used to assess milk quality[20]. The study [21] indicates that the observed dairy farm, which operated under hot weather conditions, can serve as a representative example of large-scale dairy farms. The findings suggest that the dataset has low data structure and the key variables remained relatively stable throughout the year. The success of prerequisite programs, such as solid agricultural and sanitary practices, that were put in place at the farm under study is responsible for this stability. To adapt to climate change, other dairy farms that may be susceptible to vulnerability should evaluate their pre-requisite programs and possibly adopt more rigorous food safety measures. A mathematical model was developed using multigene symbolic regression genetic programming to score milk based on seven key input characteristics: temperature, taste, flavor, fat content, turbidity, color, and pH level [22]. The model may assign a score to milk samples by combining the outcomes of these attributes. The model was trained and tested using an online dataset. The generated model's R 2 value of 0.95441 demonstrated how well it foresaw the quality of milk samples. This study [23] examines how artificial intelligence may be used in food processing. It also discusses technologies and procedures that might be used to develop automated technology-aided processing. It is expanded on an idea for an automatic food processing line made up of multiple operational levels and procedures that is intended to improve the microbiological safeguarding and quality assessment of drinkable foods like milk and drinks. The classification outcome of fresh milk grades using ANN in this study [24] included low, medium, and high grade. The classification achieved has an accuracy value of 98.74%. Temperature and color were employed as grouping characteristics. The third cluster is the best clustering using K-Means. According to data analysis, the smart grading system saved consumers time by making it easier to determine the quality of fresh milk. This research [25] describes an Internet of Things-based approach for detecting contaminants in milk by monitoring its pH and electric resistivity. To do this, a system based on fuzzy logic was created in MATLAB employing the Fuzzy Logic Toolbox and used on an Arduino mega controller to analyses contaminants in milk samples using hardware. Authors in [26] the milk samples from 622 specific cows with known full protein composition, scientific characteristic data, and mid-infrared emissions were made available in order to assess the prediction potential of different regression and classification techniques. The accuracy of some features' predictions using mid-infrared spectroscopy may be increased by using contemporary statistical machine learning algorithms. The findings of this study [27] prove the SVM classifier's 95% accuracy, which depends on the fusion characteristics. A model is constructed using random forest (RF), extreme gradient boost (XGBoost), and gradients boosted decision tree (GBDT) to estimate the amount of dairy fat and protein based on E-nose characteristics. $R2 = 0.9399$ in milk fat and $R2 = 0.9301$ for protein from milk show that the RF algorithms perform the best, thus the recommended method is effective. A fundamental basis for predicting milk quality may be provided by this work, which might improve the predicted accuracy for milk fat and protein. The suggested approach [8] is utilized to anticipate the existence of contaminants in a binary categorization issue as well as to determine which of five contaminants discovered using multiclass classification was. In deep learning, we present a Convolutional Neural Network design that does not need spectral data preprocessing. Classifiers analyzed show encouraging outcomes, with classification accuracies reaching 98.76% and exceeding frequently used classical learning approaches.

MATERIALS AND METHODS

This study utilized machine learning algorithms to assess milk quality using selected feature sets. The steps involved in the machine learning process and the findings are discussed below and shown in the figure 1.

A. Data Collection

The open-access Kaggle Milk Quality dataset served as the source of the data used in this study as shown in figure 1. Manual observations were used to compile the dataset's information. It includes the following seven attributes of milk samples: turbidity, color, fat, taste, odor, temperature, and pH. Usually, these characteristics are used to evaluate the quality of milk. The goal is to classify the milk into three categories: Poor, Moderate, and Good. Taste, odor, fat, and turbidity can have values of either one or zero, while temperature, pH, and color each have their specific values. The precise values for the milk measurements are shown in Table 1.

Table. 1 Shows the values of milk samples

pH	Temperature	Taste	Odor	Fat	Turbidity	Color	Grade
6.6	35	1	0	1	0	254	high
6.6	36	0	1	0	1	253	high
8.5	70	1	1	1	1	246	low
9.5	34	1	1	0	1	255	low

6.6	37	0	0	0	0	255	medium
6.6	37	1	1	1	1	255	high
5.5	45	1	0	1	1	250	low
4.5	60	0	1	1	1	250	low
8.1	66	1	0	1	1	255	low
6.7	45	1	1	0	0	247	medium

B. Data Splitting

An important aspect of preparing datasets for machine learning applications is data splitting. This process involves dividing an existing dataset into two or more subsets for training, validation, and testing of machine learning models as shown in figure 1. Data splitting is crucial for evaluating a model's performance on unseen data and helps prevent overfitting [28]. Whenever a machine learning system overfits, it performs poorly on fresh instances because it has learned the noise and certain features of the initial training data too well. Scientists can evaluate a model's ability to generalize to new data by dividing the dataset [29]. The dataset is usually split into two sections: 30% is used to assess the machine learning model's performance, and 70% is used to train the model.

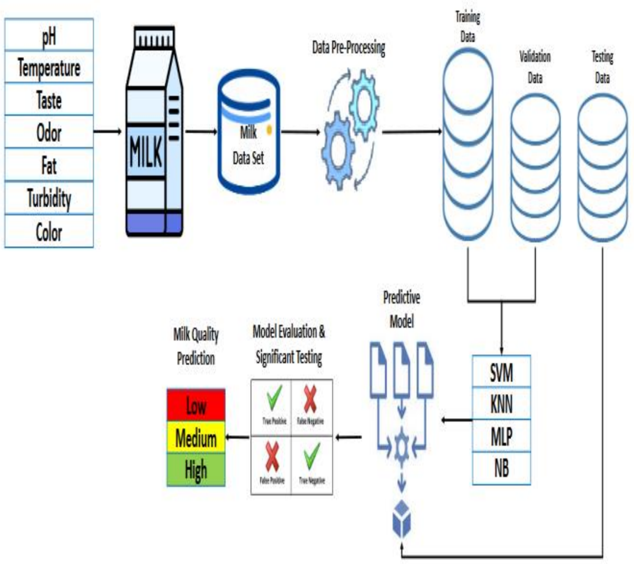

Fig 1. Proposed system for ML based Milk Quality Assessment

C. Model Selection

This study selects appropriate machine learning algorithms for assessing milk quality. Common algorithms for this task include support vector machines (SVM), multilayer perceptron's, k-nearest neighbors (KNN), and naive Bayes as shown in the figure 1.

- Support Vector Machine

One of the most well-known supervised machine learning techniques for regression and classification is the Support Vector Machine method. Both linearly and non-linearly distinct issues may be handled by SVMs by identifying the optimal hyperplane or decision boundary that efficiently separates the data points of different classes [30].

$$w^T x + b = 0$$

Where:

The weight vector is denoted by w.

The input vector is denoted by x.

The bias term is b.

The goal is to increase the margin $\frac{2}{\|w\|}$, subject to:

$$y_i(w^T x_i + b) \geq 1$$

for all training samples i, where y_i is the class label.

- K Nearest Neighbor

A simple yet effective classification and regression approach is the K-nearest neighbors (KNN) algorithm. Being non-parametric, it makes no assumptions on the distribution of the underlying data. The KNN method is based on distance metrics, and for each query example, its "k" nearest neighbors decide its class or projected value[30]. Although the Manhattan distance and Minkowski distance are additionally employed, the Euclidean distance is the one that is most frequently used. KNN predicts the classification label for the classification job using the majority of the category among its k nearest neighbor. It forecasts the average or the median among the k nearest neighbor for regression problems [31].

$$d(x_a, x_b) = \sqrt{\sum_{k=1}^{n} (x_{ak} - x_{bk})^2}$$

Where:

d(x_a,x_b) is the Euclidean distance between points x_a and x_b

n is the number of features.

- Multilayer Perceptron

A multilayer perceptron, or MLP, is a form of artificial neural network composed of many layers of linked artificial neurons (nodes). As a feed forward neural network, there are no loops or cycles because data moves straight through its input level to the output level[32]. An artificial neuron serves as the fundamental building block of an MLP. Each neuron receives inputs, weights them, and then processes the weighted total using a function of activation to generate an output. A network is formed when the result of one layer of neurons is used as the input for the subsequent layer. In general, an MLP is composed of three types of layers: an input layer, several hidden layers, and an output layer. The output layer receives the incoming data and produces the final output. The input data must be processed, pertinent characteristics must be extracted, and predictions must be made using the hidden layers. The amount of layers with the quantity of neurons within each layer determine the layout of an MLP [33].

$$z = \sum_{i=1}^{n} w_i x_i + b$$

Where:

w_i Weights are denoted.

x_i These are the inputs

b is the bias term.

z Weighted sum is denoted.

The output of the neuron is obtained by applying an activation function f(z):

$$a = f(z)$$

The sigmoid function is a common activation function

$$\sigma(z) = \frac{1}{1+e^{-z}},$$

The ReLU function ReLU(z)=max⁡(0,z), and the softmax function for the output layer in classification tasks.

- Naive Bayes Algorithm

Based on the Bayes theorem and the premise of feature independence, Naive Bayes is a popular and simple probabilistic classification method [34]. It is frequently used for sentiment analysis, spam filtering, text categorization, and recommendation systems[35].

$$P(C|\,X) = \frac{P(X|C) * P(C)}{P(X)} \quad (10)$$

Where:

Given feature set X, the subsequent likelihood of forming class C is denoted by P(C| X).

Given class C, the probability of the feature set X is P(X|C).

Class C's prior probability is denoted by P(C).

The feature set X's probability is denoted by P(X).

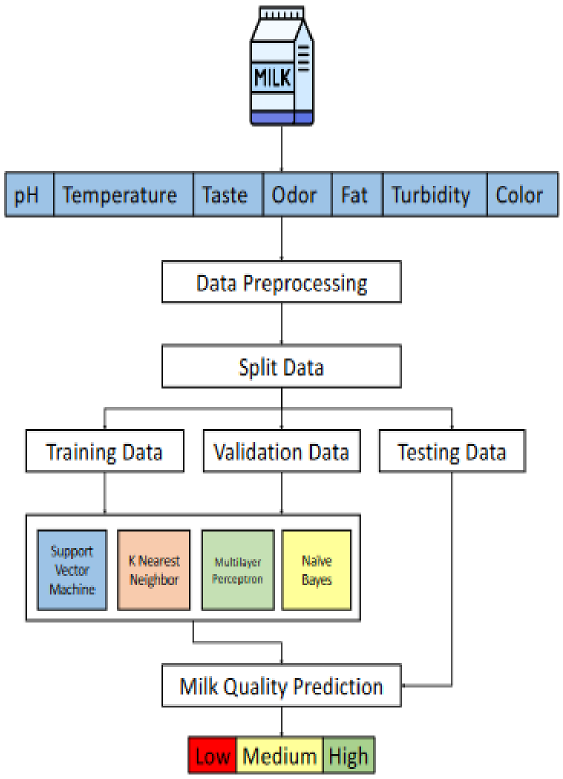

Fig 2. Overall architecture of Proposed model for ML based Milk Quality Assessment

RESULTS

The study revealed that crucial factors in assessing milk quality include the milk's pH, temperature, taste, odor, fat content, turbidity, and color as shown in figure 2.

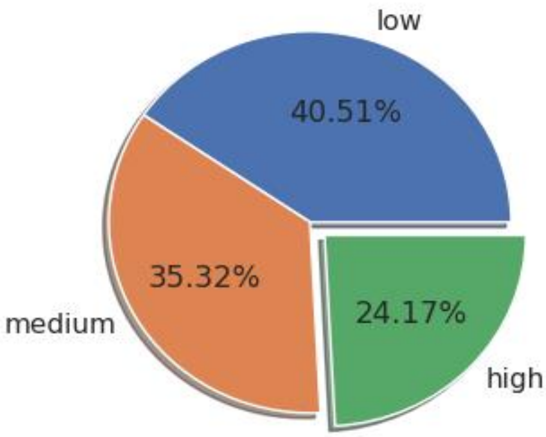

Figure 3. Milk Grade Distribution

According to observations, the pH of high and medium-quality milk ranges from 6-7, the temperature is at most 45 degrees, the taste and smell values must both be 1, the turbidity value must also be 0, and the fat value must be 1. In low quality milk, it was observed that pH value is more than 7, the temperature value is more than 45, taste value is equal to 0, the smell value is also equal to 1, turbidity value also equal to 1 and fat value is equal to 0. The distribution of the results obtained is shown on figure 1.

In this study results are shown through graphical representation and confusion matrix. Graphical representation refers to the use of charts, graphs, and diagrams to present data or information in a graphical or pictorial format. It provides a visual way to interpret and analyze data, making complex information easier to understand and communicate and confusion matrix is used to describe the performance of a classification model. Based on the predictions of a classifier, it shows the number of fake positives, incorrectly identified actual positives, and false negatives and positives for a given dataset. With the confusion matrix, we can calculate many metrics to assess a classifier's performance:

Accuracy: The overall accuracy rate of the classifier, calculated by

$$\boldsymbol{Accuracy.} = \frac{\boldsymbol{True\ Positive\ (TP) + True\ Negative(TN)}}{\boldsymbol{True\ Positive\ (TP) + True\ Negative(TN) + False\ Positive(FP) + False\ Negative(FN)}}$$

Precision: How many of the positive predictions are actually true positive, calculated by

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ Positive\ (TP)}}{\boldsymbol{True\ Positive\ (TP) + False\ Positive(FP)}}$$

Recall (Sensitivity or True Positive Rate): How many of the actual positive instances were identified correctly, calculated by

$$\boldsymbol{Recall.} = \frac{\boldsymbol{True\ Positive\ (TP)}}{\boldsymbol{True\ Positive\ (TP) + False\ Negative(FP)}}$$

F1-score: A combined metric that balances recall and precision, as determined by

$$\boldsymbol{F1 - Score.} = \frac{\boldsymbol{2 * Precision * Recall}}{\boldsymbol{Precision + Recall}}$$

As shown in figure 4, the KNN algorithm predicted milk quality values with a prediction accuracy of 96.85%. Through the use of a confusion matrix, the performance of the KNN algorithm with respect to milk quality assessment was examined

Figure 4. Confusion Matrix of KNN Algorithm

Figure 5 depicts the outcomes obtained using the naive Bayes method. Analyzing predicted with actual milk quality scores revealed that the Nave Bayes method has a prediction accuracy of 92.13%. The efficiency achieved by the naive Bayes method was examined using the confusion matrix. Figure 6 displays the outcomes of the multilayer perceptron method. The milk quality was predicted using a multilayer perceptron algorithm with a prediction accuracy of 56.91%. Figure 6 depicts the multilayer perceptron algorithm's confusion matrix, which allows us to evaluate the model's performance.

Figure 4. Confusion Matrix of KNN Algorithm

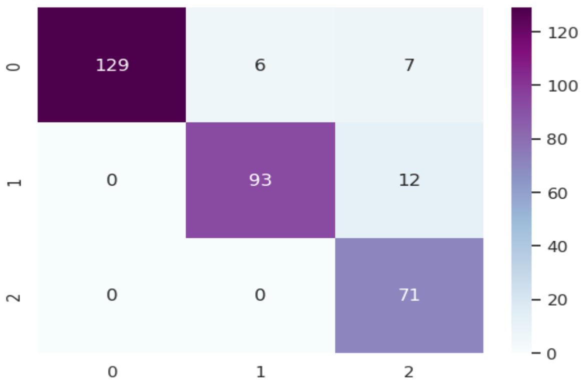

Figure 5. Confusion Matrix of naïve bayes Algorithm

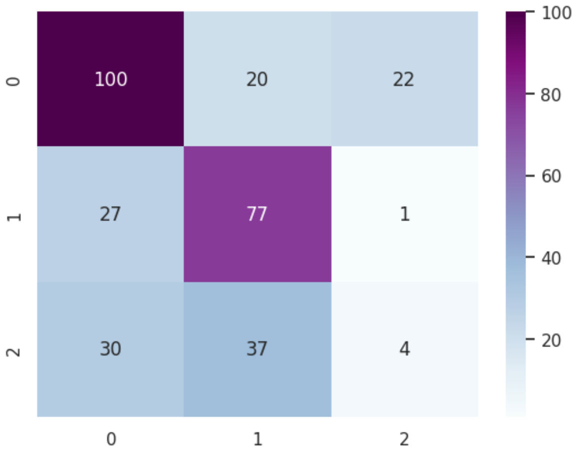

Figure 6. CM of multilayer perceptron Algorithm

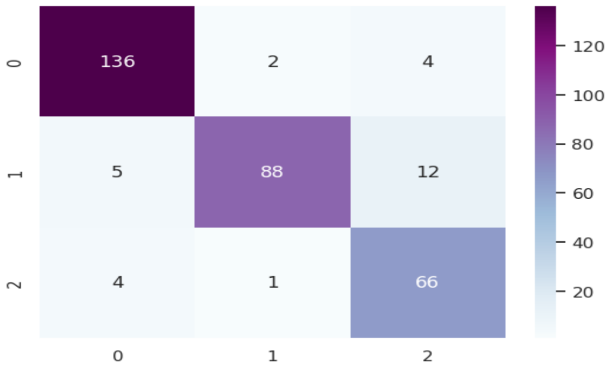

Figure 7. CM of support vector machine Algorithm

Figure 7 displays the outcomes of the support vector machine technique. The prediction accuracy of the milk quality assessment using the support vector machine technique was 91.19%. Figure 7 depicts the support vector machine algorithm's confusion matrix, which allows us to evaluate the model's effectiveness. Figure 8 compares the accuracy of milk quality prediction. As observed in the figure, KNN outperforms SVM, MLP, and NB in terms of accuracy when predicting milk quality.

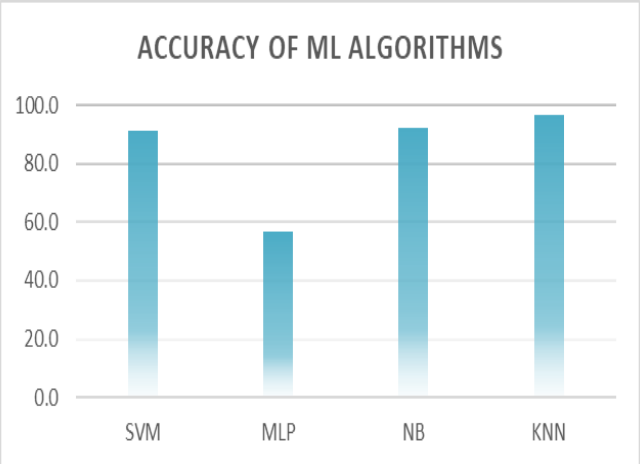

Figure 8. Accuracy Comparison of ML algorithms

In comparison to these, the accuracy of the support vector machine, multilayer perceptron, and naive bayes is 91.19%, 56.91%, and 92.13 respectively, while the accuracy of the k-nearest neighbor is 96.85%. As seen in figure 9, the regression metric known as root mean square error (RMSE) calculates the average size of the error that exists between predicted and true continuous values. The mean of the squared differences' square root is computed. The RMSE is dependent on outliers and gives greater mistakes more weight. Better model performance is indicated by a lower RMSE value.

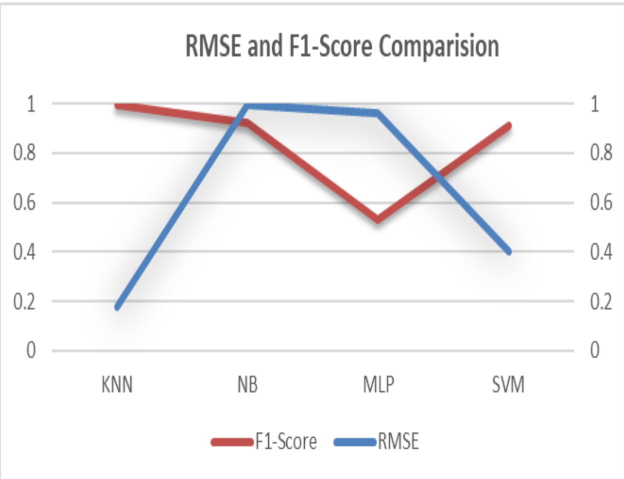

Figure 9. Comparison b/w RMSE and F1-Score

The F1 score, on the opposing hand, is a classifying statistic that assesses the precision (the percentage of properly predicted samples that are positive among all projected positive samples) and accuracy (the real successful rate or recall) of the model in detecting positive samples. To give a single indicator of overall model performance, the F1 score combines these two indicators into a harmonic mean. F1 scores vary between 0 to 1, with 1 denoting flawless recall and precision. Better model performance is denoted by a higher F1 score.

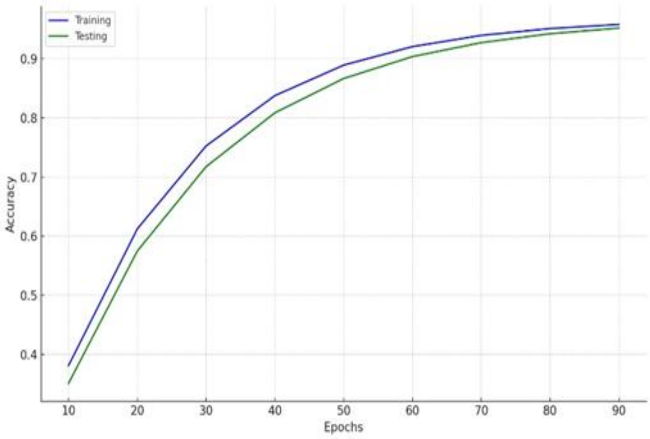

Figure 10. Curve of Accuracy

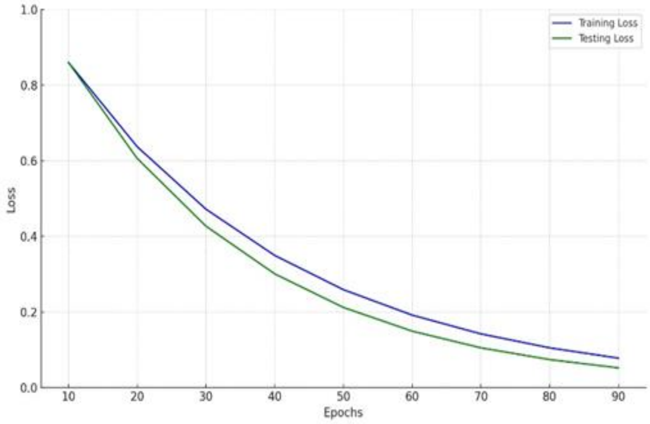

Figure 11. Curve of Loss

Figure 10 depicts the model's accuracy on the training data is shown by the blue line, while its accuracy on the data used for testing is shown by the green line. Both accuracy in training and testing rise rapidly in the early epochs (about 10–20), suggesting that the model is picking things up quickly. After around 50 epochs, the testing accuracy reaches a stage or even begins to decline, but the training accuracy keeps rising continuously. This indicates overfitting, a condition in which the model performs poorly on unknown data because it has grown too specific to the training set. According to the graph, the model appears to function best between 40–50 epochs, when testing accuracy is at its peak and the difference among accuracy in training and testing is minimal. As shown in the figure 11 the model's loss on its training data is shown by the blue line, while its loss on the data from testing is shown by the green line. Both the testing and training losses drop off fast in the initial epochs (about 10–20), suggesting that the model has been picking circumstances up quickly. After around 50 epochs, the testing loss reaches a level or even begins to rise, but the training loss keeps declining gradually. This indicates overfitting, a condition in which the model performs poorly on unknown data because it has grown too specific to the training set. The graph indicates that the model appears to function best between 40–50 epochs, when the testing loss is at its lowest and the difference across testing and training loss is minor.

CONCLUSIONS

Support vector machines, naïve bayes, multilayer perceptron, and k-nearest neighbors have all been employed as machine learning algorithms in this study. The downloaded dataset from the Kaggle repository was utilized. The dataset had seven features for each sample. There were several qualities utilized, including the pH level, temperature, taste, odor, fat, clarity, and color. On the Google colab platform, results are produced using Python. The pre-processed dataset is divided by the author into sets for training and testing. The dataset is divided into two parts, with 30% used to assess the machine learning model's performance and 70% used to train the model. Results show that KNN accuracy outperforms the MLP, SVM, and NB model built using a dataset on the quality of milk. In terms of classification accuracy, KNN scored 96.85%, MLP scored 56.91%, NB scored 92.13 percent, and SVM scored 91.1 percent. The matrix of confusion comparisons are done, and the results are shown graphically. This study has shown that grade prediction using machine learning techniques is accurate.

Conflict of Interest

There is no conflict of interest between all the authors

REFERENCES

[1] S. Neethirajan, "The role of sensors, big data and machine learning in modern animal farming," Sens. Bio-Sensing Res., vol. 29, p. 100367, 2020.

[2] J. Rodriguez Alvarez et al., "Estimating body condition score in dairy cows from depth images using convolutional neural networks, transfer learning and model ensembling techniques," Agronomy, vol. 9, no. 2, p. 90, 2019.

[3] M. Ahmedsham, N. Amza, and M. Tamiru, "Review on milk and milk product safety, quality assurance and control," Int. J. Livest. Prod., vol. 9, no. 4, pp. 67–78, 2018.

[4] A. S. Paraffin, T. J. Zindove, and M. Chimonyo, "Perceptions of factors affecting milk quality and safety among large-and small-scale dairy farmers in Zimbabwe," J. Food Qual., vol. 2018, pp. 1–7, 2018.

[5] A. Bekuma and U. Galmessa, "Review on hygienic milk products practice and occurrence of mastitis in cow's milk," Agric. Res. Technol. Open Access J., vol. 18, no. 2, pp. 1–11, 2018.

[6] M. Lu and N. S. Wang, "Spoilage of milk and dairy products," in The microbiological quality of food, Elsevier, 2017, pp. 151–178.

[7] M. A. Kamal and A. Ali, "Role and Effectiveness of IOT in E-Learning: A Digital Approach for Higher Education," Innov. Comput. Rev., vol. 3, no. 1, 2023.

[8] H. A. Neto, W. L. F. Tavares, D. C. S. Z. Ribeiro, R. C. O. Alves, L. M. Fonseca, and S. V. A. Campos, "On the utilization of deep and ensemble learning to detect milk adulteration," BioData Min., vol. 12, no. 1, p. 13, 2019, doi: 10.1186/s13040-019-0200-5.

[9] H. Khawar, T. R. Soomro, and M. A. Kamal, "Machine learning for internet of things-based smart transportation networks," in Machine Learning for Societal Improvement, Modernization, and Progress, IGI Global, 2022, pp. 112–134.

[10] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," Big Data Min. Anal., vol. 2, no. 4, pp. 288–305, 2019, doi: 10.26599/BDMA.2019.9020007.

[11] M. A. Kamal, M. Shahid, and H. Khawar, "The Mathematical Model for searching the Shortest Route for TB Patients with the help of Dijkstra's Algorithm," Sukkur IBA J. Comput. Math. Sci., vol. 5, no. 2, pp. 41–48, 2021, doi: 10.30537/sjcms.v5i2.772.

[12] M. Shafiq, Z. Tian, A. K. Bashir, X. Du, and M. Guizani, "CorrAUC: A Malicious Bot-IoT Traffic Detection Method in IoT Network Using Machine-Learning Techniques," IEEE Internet Things J., vol. 8, no. 5, pp. 3242–3254, 2021, doi: 10.1109/JIOT.2020.3002255.

[13] M. A. Kamal and M. M. Alam, "Impact of LoRA and 5G on Smart Manufacturing from Automation Perspective Impact of LoRA and 5G on Smart Manufacturing from Automation Perspective," no. March, 2022, doi: 10.13052/jmm1550-4646.1852.

[14] M. Swain, R. Singh, F. Hashmi, and others, "Spade to Spoon: An IoT-Based End to End Solution for Farmer Using Machine Learning in Precision Agriculture," in Applications of Artificial Intelligence in Engineering, 2021, pp. 387–396.

[15] E. Palupi, A. Jayanegara, A. Ploeger, and J. Kahl, "Comparison of nutritional quality between conventional and organic dairy products: a meta-analysis," J. Sci. Food Agric., vol. 92, no. 14, pp. 2774–2781, 2012.

[16] M. Henchion, M. Hayes, A. M. Mullen, M. Fenelon, and B. Tiwari, "Future protein supply and demand: strategies and factors influencing a sustainable equilibrium," Foods, vol. 6, no. 7, p. 53, 2017.

[17] M. Yoosefzadeh Najafabadi, M. Hesami, and M. Eskandari, "Machine learning-assisted approaches in modernized plant breeding programs," Genes (Basel)., vol. 14, no. 4, p. 777, 2023.

[18] J. Berryhill, K. K. Heang, R. Clogher, and K. McBride, "Hello, World: Artificial intelligence and its use in the public sector," 2019.

[19] C. E. Handford, K. Campbell, and C. T. Elliott, "Impacts of milk fraud on food safety and nutrition with special

emphasis on developing countries," Compr. Rev. Food Sci. Food Saf., vol. 15, no. 1, pp. 130–142, 2016.

[20] H. V T, S. S., S. Jha, and B. S., "MilkSafe: A Hardware-Enabled Milk Quality Prediction using Machine Learning," in 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), 2023, pp. 1–6. doi: 10.1109/ViTECoN58111.2023.10157863.

[21] R. J. Feliciano, G. Boué, F. Mohssin, M. M. Hussaini, and J.-M. Membré, "Raw milk quality in large-scale farms under hot weather conditions: Learnings from one-year quality control data," J. Food Compos. Anal., vol. 117, p. 105127, 2023, doi: https://doi.org/10.1016/j.jfca.2023.105127.

[22] J. M. Custodio, J. V Cortez, A. E. Chua, and R. Concepcion, "Development of a Quality Grading Model for Processed Milk through Sensor Data and Symbolic Genetic Programming," in 2023 8th International Conference on Business and Industrial Research (ICBIR), 2023, pp. 483–488. doi: 10.1109/ICBIR57571.2023.10147437.

[23] K. S. Kyaw, S. C. Adegoke, C. K. Ajani, O. F. Nwabor, and H. Onyeaka, "Toward in-process technology-aided automation for enhanced microbial food safety and quality assurance in milk and beverages processing," Crit. Rev. Food Sci. Nutr., pp. 1–21, 2022.

[24] W. Habsari, F. Udin, and Y. Arkeman, "An analysis and design of fresh milk smart grading system based on internet of things," in IOP Conference Series: Earth and Environmental Science, 2022, vol. 1063, no. 1, p. 12059.

[25] P. P. Lal et al., "IoT integrated fuzzy classification analysis for detecting adulterants in cow milk," Sens. Bio-Sensing Res., vol. 36, p. 100486, 2022, doi: https://doi.org/10.1016/j.sbsr.2022.100486.

[26] M. Frizzarin et al., "Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods," J. Dairy Sci., vol. 104, no. 7, pp. 7438–7447, 2021, doi: https://doi.org/10.3168/jds.2020-19576.

[27] F. Mu, Y. Gu, J. Zhang, and L. Zhang, "Milk source identification and milk quality estimation using an electronic nose and machine learning techniques," Sensors (Switzerland), vol. 20, no. 15, pp. 1–14, 2020, doi: 10.3390/s20154238.

[28] S. Theocharides, G. Makrides, G. E. Georghiou, and A. Kyprianou, "Machine learning algorithms for photovoltaic system power output prediction," in 2018 IEEE International Energy Conference (ENERGYCON), 2018, pp. 1–6.

[29] X. Ying, "An overview of overfitting and its solutions," in Journal of physics: Conference series, 2019, vol. 1168, p. 22022.

[30] I. H. Sarker, M. F. Faruque, H. Alqahtani, and A. Kalim, "K-nearest neighbor learning based diabetes mellitus prediction and analysis for eHealth services," EAI Endorsed Trans. Scalable Inf. Syst., vol. 7, no. 26, pp. e4–e4, 2020.

[31] O. O. Olatunji, S. Akinlabi, N. Madushele, and P. A. Adedeji, "Property-based biomass feedstock grading using k-Nearest Neighbour technique," Energy, vol. 190, p. 116346, 2020.

[32] A.-N. Sharkawy, "Principle of neural network and its main types," J. Adv. Appl. Comput. Math., vol. 7, pp. 8–19, 2020.

[33] G. Panchal, A. Ganatra, Y. P. Kosta, and D. Panchal, "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers," Int. J. Comput. Theory Eng., vol. 3, no. 2, pp. 332–337, 2011.

[34] L. M. Gladence, M. Karthi, and V. M. Anu, "A statistical comparison of logistic regression and different Bayes classification methods for machine learning," ARPN J. Eng. Appl. Sci., vol. 10, no. 14, pp. 5947–5953, 2015.

[35] P. Ajitha, A. Sivasangari, R. Immanuel Rajkumar, and S. Poonguzhali, "Design of text sentiment analysis tool using feature extraction based on fusing machine learning algorithms," J. Intell. Fuzzy Syst., vol. 40, no. 4, pp. 6375–6383, 2021.