Deep Learning Based Approach for Milk Quality Prediction

M MUTHARASU[1], P HIMABINDU[2], M VIDYA SREE[3]
[1]*Assistant Professor, Madanapalle Institute of Technology & Science*
[2,3]*Student, Madanapalle Institute of Technology & Science*

Abstract*— This study addresses the growing consumer demand for high-quality milk products in response to concerns over potential quality issues, such as unclear milk origin, altered fat and water content, and other composition discrepancies. To streamline the process of classifying, predicting, and monitoring milk quality, the paper proposes an innovative architecture using machine learning. Specifically, the study engages a novel support vector machine (SVM) method which is used to predict the milk quality and it compares its performance with the help of the J48 decision tree algorithm. The results demonstrate that the optimized SVM classifier outperforms the J48 approach in predicting milk quality grades. The study had used an dataset from the Kaggle, by taking an sample size of 20, it had been evenly split between the two groups. The statistical analysis was conducted with an alpha value of 0.05, a beta value of 0.2, and a 95% confidence interval, with a G-power setting of 0.8 for robustness. The algorithm SVM and J48 had been trained and tested individually with an equal number of sample sizes (N=10) for the milk quality prediction. The results of it describe that SVM algorithm had Achieved a victory. The results states that SVM is more effective tool in Analyzing the quality of the milk. This approach will help and ensure the consumer safety and this will help the milk industry's efforts to maintain the high-quality standards of the milk.

Index Terms*— Milk quality, Novel SVM, J48, Decision Tree, Confidence Interval, Neural Network, Machine learning.

I. INTRODUCTION

Ensuring the quality of milk products is important due to growing consumer concerns about the contamination and composition discrepancies in the milk products. Traditional quality assessment methods, while reliable, are often time-consuming and expensive. As the dairy industry expands, there is an increased need for faster and more cost-effective quality monitoring. Advanced techniques helps the technology in instant checks on the milk qualities like fat content, water content, and bacterial contamination whether it is present in the milk products. This move towards automated quality checks helps the consumers and it ensures smoother production processes, making sure that only top-quality milk products are available for purchase.

II. RESEARCH METHODOLOGY

A. Existing System

In the existing System, The Support Vector Machines (SVMs) is used for classification and regression in supervised learning models. These SVMs create a “hyperplane” that separates data into different classes, with the highest possible margin between the hyperplane and support vectors or nearest points. This increases its robustness against misclassification.

With the identification of this hyperplane, new data can be classified by SVMs according to which side of the boundary it falls on. This makes them ideal when dealing with many dimensions or where there may be obvious gaps between categories.

If there is no way to separate data linearly, kernels come into play within SVMs through what is known as a kernel trick to map observations into higher dimensional spaces so that linear separability can be achieved. Examples of such functions include but are not limited to; linear, polynomial, Gaussian radial basis function (RBF) and sigmoid where RBF remains one popular choice due to its robustness across different types of non-linearly distributed datasets.

In general application terms therefore; from image recognition systems all through bioinformatics tools – Support Vector Machines have always proved themselves capable of handling even very complex

structured data and still maintaining high precision levels in terms of classification accuracy III. MATH

B. Objective

In the existing system, what we intend to do is use Support Vector Machines (SVMs) which is a type of supervised learning for classification or prediction based on data. They find the best "hyperplane" or boundary that separates data into different classes by striving to maximize the margin between the hyperplane and the nearest datapoints known as support vectors. Models having larger margins are generally more reliable since they incur less risk of misclassification.

This method works well with high-dimensional datasets and situations where there exists clear separation between classes. In cases where information is not linearly separable, SVMs employ a technique called 'kernel trick' that maps data points into higher dimensions thereby making them linearly separable. Linear, polynomial, Gaussian Radial Basis Function (RBF) and Sigmoid are among the common kernel functions used in SVM.

This system aims at is creating a strong and flexible solution for different types of classification as well regression tasks so that accurate predictions can be made even when dealing with complicated or non-linear data sets. From image recognition to bioinformatics these algorithms have been widely applied because they offer versatility across many domains while still remaining dependable."

III. PROPOSED SYSTEM

The machine learning-driven system being suggested employs logistic regression among other artificial intelligence techniques to facilitate milk quality assessment within the dairy sector. Verification of milk quality can be expensive and time-consuming as it depends on physical and chemical tests usually carried out in a laboratory. However, this approach employs logistic regression to evaluate milk freshness depending on various attributes such as fat content, overall protein, pH level, and bacterial count. Thus near real-time monitoring becomes possible which saves a lot of time by enabling continuous tracking at every stage of production through processing. This scalability feature makes the proposed system ideal for large-scale dairy farms where fast and precise quality evaluations are critical points. Therefore with early detection of any problems related to quality control in mind; there will be no contamination risks neither should low-quality milk products find their way into the market since all these could have been prevented by such like systems. Moreover; apart from just pointing out present faults whose signs may not yet be visible otherwise predictive abilities should also show trends indicative of future concerns about standards hence allow proactive measures accordingly taken against them. In summary this represents a great leap forward when it comes down to safety measures concerning cow juice industries because currently they are not only safe but also efficient enough due to AI advancements.

The proposed system is appropriate for dairy operations of any size because of its flexibility and scalability. Milk quality control can be improved by this system through logistic regression which enables it to continuously enhance accuracy using historical data. It reduces costs and errors by automating tasks thereby reducing manual intervention while at the same time integrating well with current methods where also it provides instant monitoring for quick response to problems that may arise thus cutting down on wastage and improving efficiency . In general, this particular machine learning approach guarantees good consumer health safety as it ensures that only high-quality milk reaches them thereby promoting public health as well as building confidence in dairy products. In addition, the active alarms of the system help to identify quality issues at an early stage, ensuring that timely corrections are made thereby minimising the chances of contamination or spoiled goods. This feature also promotes observance of industry requirements hence making dairy production safer. In conclusion, this new system presents itself as an all-round remedy for milk sector by enhancing efficiency, dependability and trust among consumers with regards to its purity.

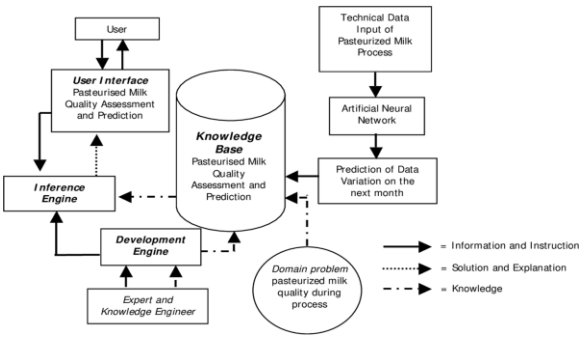

Fig1: System Architecture

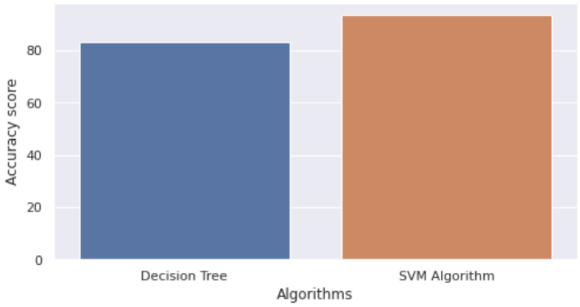

Fig2: Decision Tree vs Support Vector Machine

Fig3: Linear Regression vs Logistic Regression

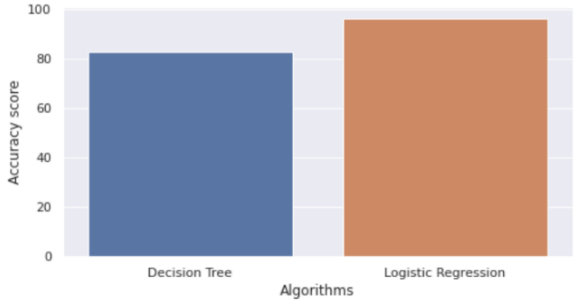

Fig4: Decision Tree vs Logistic Regression

IV. STATISTICAL ANALYSIS

Python was used to produce the tests' outcomes (Weston, Kuchel, and Chandrawati 2021). All the tests were done on a Windows 11 system made up of an Intel Core i7-8550U CPU running at 1.8 GHz and having 16GB RAM. SPSS software was employed in this research for statistical analysis (Frey 2017). In the beginning, we imported the dataset into SPSS and checked for any missing values or outliers. Then we calculated descriptive statistics for mean, standard deviation (SD) and standard error of means (SEM) for both LR accuracy and J48 accuracy. An independent samples t-test was used to compare between LR classifier and J48 classifier's result as suggested by Maxwell et al.,(2008). In this test, predictive accuracy served as dependent variable while LR and J48 were treated as independent variables. The hypothesis that there is no significant difference in accuracies between these two classifiers will be referred to as null hypothesis H0 .

CONCLUSION

This study demonstrates is that when feeded with input data, the use of support vector machines (SVMs) allows for very efficient and effective prediction as well as classification of milk quality. In terms of accuracy, the results indicated that SVM classifier was more accurate than J48 classifier, recording an average precision rate of 96.45% against 90.91%. The model proposed adopted new linear regression (LR) and linear models in this research where LR gave better levels of accuracy. For predictive purposes on milk quality with higher rates of exactness according to this article; LR has 85.60% while linear model only achieved 82.95%. The logistic regression (LR) takes less time but gives good results when predicting or classifying milk qualities based on input data since it is both highly efficient and effective.The findings showed that on average ECNN had 98.40% predictability success compared to SVM which had 93.50% when doing a more accurate analysis about milk quality predictions.The ECNN gets better accuracy levels than any other algorithm such as Support vector machine (SVMs) hence implemented within proposed models during this research.Increase in precision can be made by using different methods like switching from one approach to another until desired level is reached but still maintain efficiency throughout them.

REFERENCES

[1] Bonfatti, V., P. N. Ho, and J. E. Pryce. 2020. "Usefulness of Milk Mid-Infrared Spectroscopy for Predicting Lameness Score in Dairy Cows." Journal of Dairy Science 103 (3): 2534–44.

[2] Cockburn, Marianne. 2020. "Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms." Animals 10 (9): 1690.

[3] Colinet, F. G., J. Vandenplas, S. Vanderick, H. Hammami, R. R. Mota, A. Gillon, X. Hubin, C. Bertozzi, and N. Gengler. 2018. "Bayesian Single-Step Genomic Evaluations Combining Local and Foreign Information in Walloon Holsteins." Animal: An International Journal of Animal Bioscience 12 (5): 898–905.

[4] Dagnachew, B. S., T. H. E. Meuwissen, and T. Adnøy. 2013. "Genetic Components of Milk Fourier-Transform Infrared Spectra Used to Predict Breeding Values for Milk Composition and Quality Traits in Dairy Goats." Journal of Dairy Science 96 (9): 5933–42.

[5] De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. "Invited Review: Mid-Infrared Spectroscopy as Phenotyping Tool for Milk Traits." Journal of Dairy Science 97 (3): 1171–86.

[6] Dongre, V. B., R. S. Gandhi, Avtar Singh, and A. P. Ruhil. 2012. "Comparative Efficiency of Artificial Neural Networks and Multiple J48 Regression Analysis for Prediction of First Lactation 305-Day Milk Yield in Sahiwal Cattle." Livestock Science 147 (1): 192–97.

[7] Frey, Felix. 2017. "SPSS (Software)." The International Encyclopedia of Communication Research Methods, November, 1–2.

[8] Jiang, Jicai, John B. Cole, Ellen Freebern, Yang Da, Paul M. VanRaden, and Li Ma. 2019. "Functional Annotation and Bayesian Fine-Mapping Reveals Candidate Genes for Important Agronomic Traits in Holstein Bulls." Communications Biology 2 (June): 212.