International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 55-61, January-2024

https://doi.org/10.47001/IRJIET/2024.801007

Survey of Dairy Products Using Machine Learning Techniques

[1]Nour Abd AL Khaliq Fadel, [2]Baydaa Sulaiman Bahnam

[1]Student, Department of Software, College of Computer Science and Mathematics, University of Mosul, Iraq

[2]Assist. Professor, Department of Software, College of Computer Science and Mathematics, University of Mosul, Iraq

***Abstract* - The quality of dairy is of great importance in the food industry, as dairy is one of the main sources of proteins, calcium, vitamins and minerals that the human body needs. Dairy is found in many products, including milk, cream, and cheese, the most important of which is milk. It represents the basic and important element in people's lives, especially children, because it is an indispensable source for their growth, building their bones, and strengthening their bodies. Milk is a perishable product. Every gram of milk of poor quality or structure can cause tons of milk to spoil, causing significant financial losses and can lead to poisoning. Therefore, the quality and safety standards of this product are essential to ensure the provision of healthy and safe products to consumers. Determining the quality of the milk product is crucial for the purpose of monitoring to reduce potential losses and damages. The aim of this research is to provide a brief survey of the methods used in classifying milk quality using artificial intelligence methods, which include the use of machine learning algorithms through Identifying the most important and accurate methods and displaying the results reached using the known metrics: Accuracy, Precision, Recall, and F1_score.**

Keywords: Dairy Quality, Milk Quality, Artificial Intelligence, Machine learning.

I. INTRODUCTION

The quality of dairy is of great importance in the food and nutrition industry, as dairy is one of the main sources of proteins, calcium, vitamins and minerals that the human body needs. Dairy is found in a variety of products such as milk, yogurt, cheese, and their derivatives [1]. Milk plays an important role in diets around the world, as it is an animal product that is sold at reasonable prices in most low- and middle-income countries [2]. Milk production is an important source of livelihood for small farm owners [3]. The demand for milk production is increasing to meet the needs of population growth and in accordance with food safety standards [4]. Individuals engaged in informal sector enterprises seek avenues to enhance their businesses by improving milk quality, ensuring safety, and minimizing spoilage. The imperative is to facilitate the delivery of safer and superior quality milk to markets [5]. Given its perishable nature, milk poses a unique challenge. Even a small quantity of subpar or compromised milk can lead to significant volumes spoiling, resulting in substantial financial setbacks. The rapid proliferation of millions of bacteria in spoiled milk exacerbates this issue. Consequently, consumption of such compromised milk or dairy products may give rise to situations jeopardizing human health. Dairy quality affects the nutritional value of dairy products and public health, so quality and safety standards are essential to ensure healthy and safe products for consumers [6][7]. With the widespread developments in technology, there has been a need to use Artificial Intelligent (AI) in the processes of detecting and classifying the quality of dairy products, as they have an important role in their usefulness. Therefore, it is very necessary to monitor the quality of these products produced and determine them in a short time using these techniques.

The inception of artificial intelligence (AI) dates back to the mid-twentieth century when Alan Turing envisioned the possibility of machines exhibiting cognitive capabilities. Since that time, the field of AI, within the realm of computer science, has experienced rapid growth. The aspiration to develop machines with human-level intelligence gained momentum, guided by the implementation of programs simulating intelligent behavior [8]. The primary objective was to design computer systems capable of learning, interacting, and making decisions within intricate and dynamic environments. Presently, artificial intelligence has become an indispensable component of our contemporary daily lives, contributing significantly across various sectors such as the economy, industry, banking, homeland security, government, and more [9]. The field of artificial intelligence includes many sub-fields including natural language processing, machine learning, data mining, expert systems, vision systems, planning and scheduling, robotics, etc. AI continues to evolve over time and in this field a lot of innovations are being made. It includes some of the latest inventions and discoveries in the dairy industry to improve product quality and increase production efficiency using smart robots, drones, and self-driving cars [10].

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 55-61, January-2024

https://doi.org/10.47001/IRJIET/2024.801007

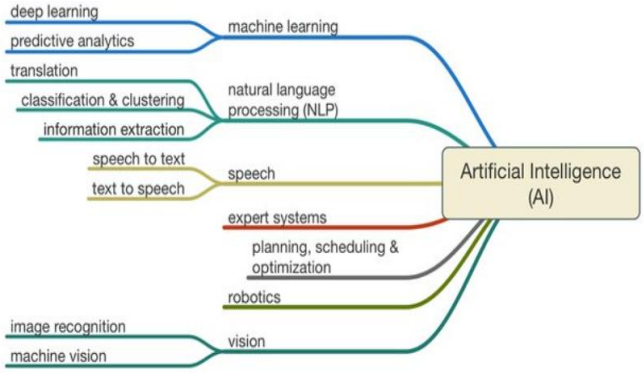

Figure 1: Fields of AI [10]

Machine learning (ML) is a subset of artificial intelligence [11]. In machine learning, the required issue is determined, then a data set is given to investigate and solve that issue, followed by extracting information that can be used in making decisions and discovering patterns [12]. The primary goal of ML is to support intelligent systems to learn a function by using a set of learning data [9]. ML algorithms are successfully used in many fields for the purpose of performing estimation, classification, clustering, etc. operations [13]. The use of machine learning algorithms in detecting and classifying food products and knowing the quality of these products, and this is what attracted researchers' interest in this field.

II. RELATED WORKS

Various studies have been presented recently in the field of dairy product quality based on machine learning methods. The researchers presented. The researchers (Xiao et al. 2019) used three models: the Random Forest (RF) model, the Logistic Regression (LR) model, and the Adaptive Boosting (AB) model and conducted tests to find the most appropriate model to classify the milk data. They used three characteristics to classify the quality of milk: milk color and its smell and taste. They showed that these characteristics have the ability to identify the quality of milk. According to the results they obtained, the highest success rate was 96.8% using the random forest algorithm [14].

The researchers (Sambasivam et al., 2020) estimated the rate of vitamin D deficiency (VDD) by using several machine learning algorithms: k-nearest neighbor algorithm (KNN), decision tree (DT), random forest (RF), AdaBoost adaptive boosting (AB), classifier. Bagging (BC), Extra Trees (ET), Stochastic Gradient Descent (SGD), Gradient Boosting (GB), Support Vector Machine (SVM), and Multilayer Perceptron (MLP) algorithm. These algorithms were applied to data collected for 3,044 university students between the ages of 18 and 21. According to the results they obtained, the highest percentage was 96% using the random forest algorithm [15].

The researchers (Kavitha and Deepa, 2021) presented a comparative analysis for detecting pure milk from adulterated milk using several ML algorithms: LR, Naive Bayes, RF, SVM, and gradient boosting (GB) [16].

The researchers (Frizzarinet al., 2021) used several statistical machine learning methods to predict the characteristics of raw cow's milk based on chemical analyses. It was found that the Model Average (MA) approach was the best in its ability to predict 6 of the 14 traits examined [17].

The researcher (ÇELIK, 2022) used a neural network (NN) and adaptive boosting (AB) to classify milk quality into three levels (high, medium, low) and used the orange platform, which is open source and written in Python, as the platform for the application. According to the results obtained; The highest classification accuracy was 99.9% using the AdaBoost algorithm and 95.4% using the neural network algorithm [18].

Researchers (Zhang et al., 2022) developed an electronic nose model to distinguish the source of milk, estimate the fat and protein content of milk, recognize the authenticity and accuracy of milk and evaluate milk quality. And using three machine learning algorithms such as LR, SVM, and RF to build the milk source (cow farm) and evaluate and compare the classification effects. The results show that the classification effect of the SVM-LDA fusion model based on LDA is better than other individual models, as the accuracy of the test set reached 91.5% [19].

Through studying previous works, it was found that the quality of milk was not determined as either high-quality or low-quality milk. We also find that most previous works use only one, two, or three features in the classification process. Table 1 presents a summary of previous.

III. MACHINE LEARNING ALGORITHMS

The aim of this study is to conduct a survey to determine and classify dairy products using ML algorithms for training and testing. This research was created to help speed up and facilitate the process of automatically detecting the quality of dairy products and knowing whether this quality is good or bad by building an intelligent system that facilitates the process of detecting milk quality by creating models based on ML algorithms. Using ML algorithms to make important decisions will save effort, time and money, as well as business efficiency. Below is an explanation of the most important types of machine learning algorithms that are used to classify milk quality and estimate the rate of vitamin D deficiency.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 1, pp 55-61, January-2024

https://doi.org/10.47001/IRJIET/2024.801007

Table 1: Results of previous studies in the field of milk quality assessment

No.	Researcher	Data set	Algorithms	Metrics	Results	Notes
1	Xiao et al., 2019[14]	Milk data set	1.Random Forest (RF) 2.Linear Regression (LR) 3.AdaBoost (AB)	Accuracy	The highest success rate obtained using random forest was 96.8%.	Using three characteristics: color, odor and taste of milk to classify milk quality
2	Sambasivam et al., 2020[15]	Primary data containing blood vitamin D levels were collected from a total of 3,044 university students aged 18–21 years.	1.K-nearest Neighbor (KNN) 2.Decision tree (DT) 3. RF 4.AB 5.Bagging Classifier (BC) 6. Extra Tree (ET) 7.Stochastic Gradient Descent (SGD) 8.Gradient Boosting (GB) 9. Support Vector Machine (SVM) 10.Multi-Layer Perceptron (MLP)	Accuracy Precision Recall F1-score	According to the results they obtained, the highest percentage was 96% using the RF Random Forest algorithm	11parameters were used to predict the risk of vitamin D deficiency with specific age groups
3	Kavitha and Deepa, 2021[16]	The researcher collected samples of pure and adulterated milk	1.LR 2. Naive Bayes 3.RF 4.SVM 5.Gradient Boosting (GB)	Accuracy Precision	It turned out that the random forest was the most efficient	Comparative analysis to detect pure milk from adulterated milk
4	Frizzarin et al., 2021[17]	Milk samples taken from 622 cows	1.Partial least squares Regression (PLSR) 2. ridge regression (RR) 3.least absolute shrinkage and selection operator (LASSO) 4. Model Averaging (MA) 5. Neural Network (NN) 6.Partial least squares discriminant analysis (PLSDA) 7. RF 8.Boosting Decision trees 9. SVM	Accuracy Root mean square error (RMSE)	It was found that the customized approach Model Averaging (MA) was the best in its ability to predict the quality of raw cow's milk for 6 of the 14 attributes.	In chemical analyses, statistical methods are considered better than others because their results are more interpretable and their parameters are less adjustable than other methods.
5	ÇELIK, 2022[18]	Milk data set	1. NN 2.AB	Accuracy Precision Recall F1-score	The highest classification accuracy was 99.9% using the AB algorithm, while it was 95.4% using the NN	Using the Orange platform, which is open source and written in Python, to classify the quality of milk into three categories using two attributes: the pH and temperature of the milk.
6	Zhang et al., 2022[19]	1000sets of milk data were collected from 10 farms in different places	1.LR 2. RF 3. SVM	Accuracy	The SVM-LDA fusion model based on LDA is better than other individual models, as the accuracy of the test set reached 91.5%.	Develop an electronic nose model to distinguish the source of milk, estimate the fat and protein content of milk, recognize the authenticity and accuracy of milk and evaluate milk quality.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 8, Issue 1, pp 55-61, January-2024
https://doi.org/10.47001/IRJIET/2024.801007

1) Support Vector Machine (SVM)

The SVM algorithm was designed by Vladimir Vatnik and his colleagues in the early 1990s [20]. This algorithm is a directed ML method used in classification, which is based on statistical learning theory and is used to solve the problem of data classification and regression control [21]. This algorithm finds lines or surfaces that separate different classes in a data set. It seeks an optimal separator level that matches the data's dimensions for binary classification, categorizing training into two classes. In the case of more than two classes, the algorithm employs multiclass SVM [22]. The fundamental concept of SVM involves constructing an optimal level in a space to address classification challenges and distinguish between models. A larger optimal level size enhances algorithm efficiency and classification accuracy. To establish the maximum margin, the SVM algorithm creates two parallel levels on either side of the margin, representing the distance between them. Consequently, SVM's machine learning approach is grounded in decision levels, defining decision boundaries. The decision level acts as a boundary separating entities with distinct classification affiliations. The SVM algorithm identifies the optimal level with the largest margin to separate categories, preventing local disruptions and ensuring optimal performance [23] [24].

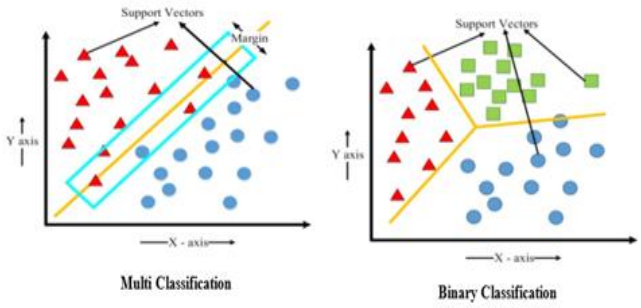

Figure 2: The ideal level for binary and multiple classification operations based on the SVM [25]

2) Decision Tree (DT) algorithm

Decision tree is a machine learning technique used to solve classification and prediction problems. A decision tree is based on dividing data into different categories or classifications using a series of binary decisions. The decision tree operates on a series of if-else statements, organized based on specific conditions. Illustrated in Figure (3), it comprises multiple nodes, referred to as leaves. Each leaf undergoes a test, directing a query through the node's branches. This process iterates until reaching the terminal leaf, associating a value with each leaf node in the tree. Emphasis is placed on constructing the smallest tree by prioritizing key features, such as organizing samples into groups. Following the initial attribute's sample splits, the remaining samples yield similar decision tree (DT) problems but with fewer samples and one fewer attribute. These subtrees, with less critical attributes, aid in managing complexity. A node with a higher sample count indicates greater complexity, while a homogeneous node containing samples of one class reduces complexity. Node objectives include growing trees by consistently striving to achieve clearer leaf nodes, thereby reducing sample class complexity [26][27].

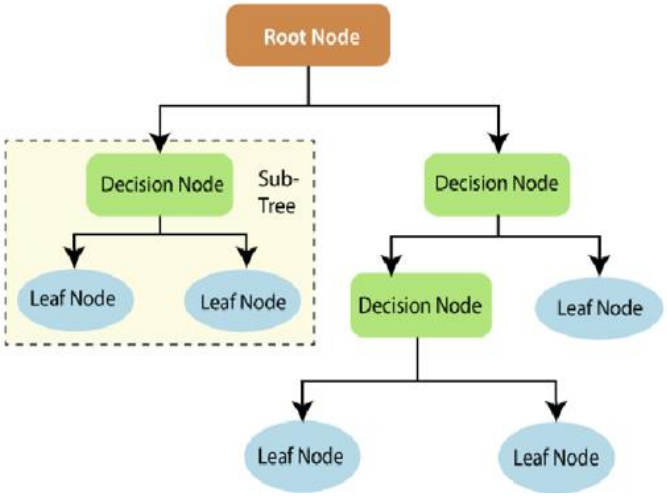

Figure 3: Decision Tree Flowchart [28]

3) K-Nearest Neighbors (KNN) algorithm

The KNN classifier stands out as one of the most straightforward and extensively utilized classifiers within classification algorithms. Originally introduced in 1951 by Fix and Hodges and subsequently modified by Cover and Hart, KNN functions effectively for both classification and regression purposes. The fundamental principle of KNN revolves around computing distances between tested samples and training data samples to identify their nearest neighbors. Subsequently, the tested sample is straightforwardly assigned to the class of its closest neighbor [29].

In the context of KNN, the variable K signifies the number of nearest neighbors, playing a pivotal role in this classifier. The chosen value of K dictates the number of neighbors influencing the classification outcome. For instance, when K equals 1, the new data object is assigned to the class of its nearest neighbor. These neighboring objects are derived from a set of training data objects with known classifications. The KNN algorithm seamlessly handles numerical data, employing distance concepts like Euclidean and Manhattan distances to classify data objects. Among these, Euclidean distance is the most commonly used with KNN [30]. The algorithm's flow chart is depicted in Figure (4) [31].

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 55-61, January-2024
https://doi.org/10.47001/IRJIET/2024.801007

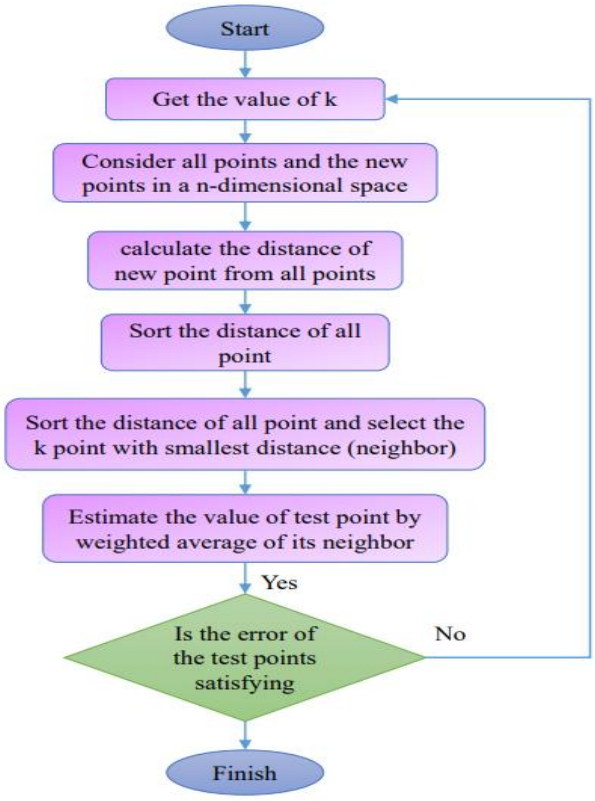

Figure 4: Flowchart of the KNN algorithm [31]

4) Adaptive Boosting

Adaptive Boosting, commonly known as AdaBoost, stands as an ensemble learning algorithm crafted to enhance the performance of weak learners and construct a potent, precise predictive model. Yoav Freund and Robert Schapire pioneered its development in 1996. The operational mechanism of AdaBoost involves the iterative training of a sequence of weak classifiers on data subsets. In each iteration, it assigns elevated weights to misclassified instances, guiding subsequent weak classifiers to prioritize previously misclassified samples. This iterative procedure persists, culminating in a final model that amalgamates the individual weak classifiers into a resilient and accurate ensemble [32].

Salient attributes of AdaBoost encompass its capacity to adapt to intricate datasets through the assignment of varied weights to instances based on their difficulty in classification. Notably effective in handling noisy data and outliers, AdaBoost finds widespread application in diverse machine learning domains such as face detection, object recognition, and bioinformatics. Although AdaBoost may exhibit sensitivity to noisy data, its overall efficacy lies in its ability to construct a robust classifier by systematically accentuating the shortcomings of earlier classifiers, ultimately leading to an enhanced and more precise model [33].

IV. CONCLUSION

The importance of milk quality in the food and nutrition industry makes it important to monitor this quality of milk to ensure its safety and reduce spoilage in order to contribute to obtaining safer and higher quality milk in the markets. The researchers have conducted several studies on multiple models of milk data and using several intelligent algorithms, including machine learning algorithms to classify milk quality It was found that the adaptive boosting algorithm obtained the highest classification accuracy of 99.9% using the open-source orange platform written in Python, to classify milk quality into three categories using two attributes (pH and temperature) using milk data from Kaggle warehouse. Therefore, we propose to use other machine learning algorithms on the milk dataset in order to obtain the highest accuracy in detecting milk product quality.

ACKNOWLEDGEMENT

I would like to admit that the authors thank the University of Mosul for supporting them and providing all means to advance knowledge and advance science and knowledge.

REFERENCES

[1] Han, J. and Wang, J. Dairy Cow Nutrition and Milk Quality. Agriculture 2023, 13,70. https://doi.org/10.3390/agriculture13030702

[2] Munda, E., Mtimet, N., Schneider, F., Wanyoike, F., Dominguez-Salas, P., & Alonso, S.(2021). Could the new dairy policy affect milk allocation to infants in Kenya? A bestworst scaling approach. Food Policy, January, 102043. https://doi.org/10.1016/j.foodpol.2021.102043

[3] Msalya, G. (2017). Contamination levels and identification of bacteria in milk sampled from three regions of Tanzania: Evidence from literature and laboratory analyses. Veterinary Medicine International, 1–10. https://doi.org/10.1155/2017/9096149

[4] Ramsing, R., Santo, R., Kim, B.F. et al. Dairy and Plant-Based Milks: Implications for Nutrition and Planetary Health. Curr Envir Health Rpt 10, 291–302 (2023). https://doi.org/10.1007/s40572-023-00400-z

[5] Alonso S., Muunda E., Ahlberg S., Blackmore E., Grace D., Beyond food safety: Socio-economic effects of training informal dairy vendors in Kenya, Global Food Security, Vol.18, 2018,pp. 86-92.

[6] Lemma, H. D., Mengistu, A., Kuma, T., Kuma, B., Lemma, D. H., Mengistu, A., Kuma, T., & Kuma, B. (2018). Improving milk safety at farm-level in an intensive dairy production system: Relevance to

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 55-61, January-2024
https://doi.org/10.47001/IRJIET/2024.801007

smallholder dairy producers. Food Quality and Safety, 2(3), 135–143. https://doi.org/10.1093/fqsafe/fyy009

[7] Tong, L., Yi, H., Wang, J., Pan, M., Chi, X., Hao, H., & Ai, N. (2019). Effect of Preheating Treatment before Defatting on the Flavor Quality of Skim Milk. Molecules (Basel, Switzerland), 24(15), 2824. https://doi.org/10.3390/molecules24152824

[8] Papa R, Jackson KM , 2021. "Artificial Intelligence, Human Agency and the Educational Leader": Springer Nature.

[9] Zohuri, B., and Zadeh, S. (2020). "Artificial Intelligence Driven by Machine Learning and Deep Learning" Publisher: Nova Science Publishers.

[10] Villanueva, M. B., and Salenga, M. L. M. (2018). "Bitter melon crop yield prediction using machine learning algorithm." International Journal of Advanced Computer Science and Applications, 9(3).

[11] Eggers, S. L., and Sample, C. (2020). Vulnerabilities in Artificial Intelligence and Machine Learning Applications and Data. Idaho National Lab.(INL), Idaho Falls, ID (United States).

[12] Serrano, L. (2021). "Grokking Machine Learning" Publisher: Simon and Schuster.

[13] Sen J. et al., 2021, "Machine Learning - Algorithms, Models and Applications", Artificial Intelligence, http://dx.doi.org/10.5772/intechopen.94615

[14] L. Xiao, K. Xia, and H. Tian, "Research on Classification Model of Fermented Milk Quality Control Based on Data Mining," in 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2019, pp. 324-327: IEEE.

[15] G. Sambasivam, J. Amudhavel and G. Sathya, "A Predictive Performance Analysis of Vitamin D Deficiency Severity Using Machine Learning Methods," in IEEE Access, vol. 8, pp. 109492-109507, 2020, https://doi.org/10.1109/ACCESS.2020.3002191

[16] P. V. KavithaP. V. Deepa, "A comparative analysis of the machine learning methods for milk adulteration detection", AIP Conference Proceedings 2408, 030008 (2021), vol 2408, Issue1.

[17] M. Frizzarin et al., "Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods,"Journal of Dairy Science, vol. 104, no. 7, pp. 7438-7447, 2021.

[18] A.ÇELIK, "Using Machine Learning Algorithms to Detect Milk Quality," Eurasian Journal of Food Science and Technology, vol. 6, no. 2, pp. 76-87, 2022.

[19] Y. Zhang, L. Zhang, Y. Ma, J. Guan, Z. Liu, and J. Liu, "Research on dairy products detection based on machine learning algorithm," MATEC Web Conf., vol. 355, p. 03008, 2022. https://doi.org/10.1051/matecconf/202235503008

[20] Menyhart J. and Szabolcsi R. (2016): "Support Vector Machine and Fuzzy Logic", Acta Polytechnica Hungarica Vol. 13, No. 5, page: 205-220.

[21] Kulkarni A. A. ,Hundekar V. A., Sannakki S. S. and Rajpurohit V. S. (2017): "Survey on Opinion Mining Algorithms and Applications", International Journal of Computer Techniques – Volume 4 Issue 3 ، ISSN :2394-2231، page: 9.

[22] Bahnam B. S. and Dawwod S. Abd, "A proposed model for diabetes mellitus classification using coyote optimization algorithm and least squares support vector machine," IAES International Journal of Artificial Intelligence (IJ-AI), vol. 11, no. 3, p. 1164, 2022, https://doi.org/10.11591/ijai.v11.i3.pp1164-1174

[23] Siemers F.M. and Bajorath J.," Differences in learning characteristics between support vector machine and random forest models for compound classification revealed by Shapley value analysis.," Sci Rep 13, 5983, 2023. https://doi.org/10.1038/s41598-023-33215-x

[24] Khairnar J. and Kinikar M. "Machine Learning Algorithms for Opinion Mining and Sentiment Classification", International Journal of Scientific and Research Publications, vol. 3, Issue 6, pp.1-6, 2013.

[25] Muzzammel R.and RazaA., "A Support Vector Machine Learning-Based Protection Technique for MT-HVDC Systems," Energies, vol. 13, no. 24, p. 6668, Dec. 2020, https://doi.org/10.3390/en13246668

[26] Tanyildizi H., "Prediction of the Strength Properties of Carbon Fiber-Reinforced Lightweight Concrete Exposed to the High Temperature Using Artificial Neural Network and Support Vector Machine," Adv. Civ. Eng. 2018, 5140610.https://doi.org/10.1155/2018/5140610

[27] Chithra S., Kumar S.S., Chinnaraju K., Ashmita F.A.A.," comparative study on the compressive strength prediction models for High Performance Concrete containing nano silica and copper slag using regression analysis and Artificial Neural Networks," Constr. Build. Mater. 2016, 114, 528–535. https://doi.org/10.1016/j.conbuildmat.2016.03.214

[28] Nafees A., et al.," Modeling of Mechanical Properties of Silica Fume-Based Green Concrete Using Machine Learning Techniques," Polymers, vol.14, no.30, 2022, https://doi.org/10.3390/polym14010030

[29] M. Mohammed, M.B. Khan and E.B.M. Bashier," Machine learning: algorithms and applications,". CRC Press, Boca Rato, 2016.

[30] N. Ali, D. Neagu, &Trundle," Evaluation of k-nearest neighbor classifier performance for heterogeneous data sets,". SN Appl. Sci. 1, 1559, 2019, https://doi.org/10.1007/s42452-019-1356-9

ISSN (online): 2581-3048
Volume 8, Issue 1, pp 55-61, January-2024
https://doi.org/10.47001/IRJIET/2024.801007

[31] Muradian, E. Khamehchi, S. Hjirezaei and A. Hemmati-Sarapardeh, "Modeling viscosity of crude oil using k-nearest neighbor algorithm," ADVANCES IN GEO-ENERGY RESEARCH, 2020, https://doi.org/10.46690/ager.2020.04.08

[32] Colakovic I. and Karakatič S., "Adaptive Boosting Method for Mitigating Ethnicity and Age Group Unfairness," SN Computer Science, vol. 5, no. 1, p. 10, 2023/11/15 2023, doi: 10.1007/s42979-023-02342-7

[33] Ding, Y.; Zhu, H.; Chen, R.; Li, R. An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification. Appl. Sci. 2022, 12, 5872. https://doi.org/10.3390/app12125872

Citation of this Article:

Nour Abd AL Khaliq Fadel, Baydaa Sulaiman Bahnam, "Survey of Dairy Products Using Machine Learning Techniques" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 8, Issue 1, pp 55-61, January 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.801007

****

