International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 1
ISSN: 2088-5334

Evaluating Machine Learning Models for Optimal Livestock Environment Prediction in Smart Farming Applications to Enhance Food Security

Sharifah Nabila S Azli Sham [a], Emmerich Wong [a], Adriana Arul Yacub [a], Deventheren Kamala Nathan [a], Law Kah Hou [a], Liew Tien Yew [a], Nur Atiqah Malizan [a], Normaizeerah Mohd Nor [a], Norulzahrah Mohd Zainudin [a], Noor Afiza Mat Razali [a,*]

[a] *Faculty of Defense Science and Technology, National Defense University of Malaysia, Kuala Lumpur, Malaysia*

Corresponding author: [*]noorafiza@upnm.edu.my

***Abstract*— Livestock is a vital protein source for the global population, and any supply disruptions can significantly threaten national food security. Therefore, ensuring a stable and continuous livestock supply is essential. Previous studies have highlighted a strong link between livestock production and environmental factors and proposed several smart farming solutions to be adopted to monitor and optimize livestock production effectively. In this study, we propose adopting sensor technology and machine learning to establish optimal environmental conditions for livestock in Malaysia. Machine-learning techniques are evaluated to determine the most effective model for enhancing livestock production in smart farming systems. This research simulates the livestock living environment equipped with sensors and selected parameters for data collection to train the machine learning chosen models: Decision Tree, Naïve Bayes, and K-Nearest Neighbors. The trained machine learning models are then applied to predict the optimum environment for livestock using the dataset of the simulated environment. Then, performance evaluation on the machine learning models was carried out. The accuracy results for Decision Tree, Naïve Bayes, and K-Nearest Neighbors are 99%, 63%, and 89%, respectively. The research shows that the Decision Tree model is the best-performing model at predicting the optimum environment for livestock. These findings provide invaluable insight to advance research on optimum livestock environment prediction in smart farming for the Malaysia use case. They will enable precise adjustments and monitoring to achieve ideal conditions for livestock growth to provide consistent livestock production.**

***Keywords*— Food security; environment; livestock; machine learning; smart farming.**

Manuscript received 8 Jun. 2024; revised 10 Sep. 2024; accepted 11 Nov. 2024. Date of publication 28 Feb. 2025.

I. INTRODUCTION

Food security is intrinsically linked to national security because a nation's stability and well-being are directly connected to a stable and reliable food supply. Food security is also described as accessibility to physical and economic availability to safe, adequate, and nourishing food to sustain nutritional requirements for an active and healthy life for everyone at any time [1], [2]. Food security is one of the most fundamental human rights, considering humans require food to survive and function. Food security has raised significant concerns around the world as more countries have started to realize its importance at the turn of the 21st century during the COVID-19 pandemic, which forces many countries to reevaluate their food security strategy in times of emergency due to the widespread famines in areas of crisis and natural disasters [3], [4] .

Researchers and policymakers discussed that food security can be achieved when a country can consistently produce or source more food than it consumes [2]. Having stable food security is crucial to ensure a well-fed population. To combat food insecurity, the Malaysian government enacted a five-point security plan. The plan establishes tactics and programs to be carried out to ensure food security in Malaysia's agricultural sector, named "Malaysia's Agro-Food Policy (NAP 2011–2020) Performance and New Direction" [5]. Thus, food security, which forms the basis of national security, is our main motivation for this research. Meanwhile, livestock, which includes poultry, cattle, goats, and others, are the leading providers of proteins, comprising more than a third of human protein consumption [6]. However, livestock are exposed to numerous risks, including diseases. Thus, current research primarily focuses on disease detection, such as detecting zoonotic diseases, how to prevent a disease outbreak, and the effects of livestock's greenhouse gas emissions on the environment [6], [7]. Our study found a significant gap in research regarding the determination of optimal environments for livestock farming despite the environment's critical role in enhancing livestock production and stabilizing prices to ensure availability for the global population at affordable rates. While previous studies on smart farming have primarily focused on integrating sensor technology into livestock farming, there has been insufficient emphasis on fully leveraging data analytics derived from the sensor data. Thus, to address this gap, this research focuses on using the collected data to determine the best machine-learning technique to predict the optimum environment for livestock to increase livestock production. In this research, we choose Decision Tree, Naïve Bayes, and K-Nearest Neighbors for the machine learning model to understand the suitability of machine learning models for data for smart farming by measuring the performance, accuracy, and precision of each machine learning technique using collected data. For data collection, we selected the most essential parameters, temperature, humidity, carbon dioxide, and ammonia, based on findings in the literature review. Sensors were installed in our simulated smart farming to record temperature, humidity, carbon dioxide, and ammonia data.

This paper has four sections: Section I introduces the research. Section II highlights the materials and methods used in this research. Section III evaluates the performance of each machine learning technique used by providing an analytical comparison and a discussion of the results. Section IV highlights the summary and the potential of this research, as well as the conclusion attained.

II. Material and Method

A. Food Security

The four pillars of food security serve as a guideline to evaluate a nation's food security [2]. The four pillars of food security are as follows:

1) Food availability: The availability of food in sufficient amounts and of decent quality, either sourced through domestic production or imported from other countries. Availability does not correlate to fully domestic produce and countries can have good food availability from imported sources.

2) Food access: The accessibility of nutritious food in adequate amounts by individuals. The access should not be limited by the individual's legal, political, social, and economic status.

3) Utilization: Achieving a state of nutritional health involves meeting all physical needs through proper nutrition, including a balanced diet, access to clean water, sanitation, and healthcare. This shows that non-food inputs are crucial in food security.

4) Stability: To achieve a state of food stability, a nation must ensure that its population can always access sufficient food. They should not experience famine or lose access to food supplies even through crises like wars or cyclical events, such as seasonal monsoons or droughts. The stability concept can, therefore be attributed to both food security aspects of availability and accessibility [2].

Malaysia's Food security is above average compared to other countries, as Malaysia was ranked 41st out of 113 countries worldwide according to the Global Food Security Index. Malaysia achieved an overall score of 69.9%, availability score of 59.5%, and quality and safety score of 74.7% [2]. According to the figures above, Malaysia can sustain two-thirds of its population at any given point. However, this does not translate to the ability to self-sustain its population. According to the Supply and Utilization Accounts of Selected Agricultural Commodities, Malaysia's national self-sufficiency ratio (SSR) is only 57%. Moreover, Malaysia still heavily relies on imports for its most important food source, including crops, livestock, and fisheries. Additionally, livestock, such as mutton and beef, have the highest import dependency ratio (IDR) at 89.4% and 81.6%, respectively in 2021. This shows that Malaysia has only managed to produce 10.6% of the mutton and 18.4% of the mutton that the population consumes. Fortunately, the poultry industry in Malaysia is faring better than its ruminant counterparts, producing almost all the chicken meat consumed at 99.9% and 114.4% of chicken eggs. The IDR for poultry is virtually non-existent, barring just 6.1%, while none for chicken eggs. The pork industry in Malaysia is pretty much self-sustaining, with 93.4% SSR and an IDR of 7.1%.

B. Livestock

Livestock in Malaysia is widely consumed as the primary source of meat-based protein. Per capita, Malaysians consume 92.3 kg of livestock products per year in 2021. From the overall 92.3 kg, 46.6 kg is the production of chicken meat, 20.8 kg is the production of chicken and duck eggs, 5.5kg is the production of beef, while duck meat and mutton are 1.6 kg and 1 kg, respectively. Livestock growth is affected by many factors [8]. The parameters used in this research are temperature, humidity, carbon dioxide, and ammonia levels, as they are the most essential parameters [9], [10], [11], [12], [13]. Carbon dioxide and ammonia are considered greenhouse gases. Much of the research into the relationship between greenhouse gases and livestock is done to study the effect of the greenhouse gases produced by livestock to the environment and not to the livestock themselves [6], [14], [15], [16], [17], [18]. Most research on livestock using Machine Learning techniques is that of disease detection [7], [13], [19], [20] and livestock monitoring [13], [21].

From the literature review, we can conclude that livestock requires specific ranges of carbon dioxide and ammonia levels, as well as particular temperature and humidity levels, to be healthy. The optimal environment for livestock is a temperature of 30°C and lower, humidity level of between 40-60%, carbon dioxide levels between 400 and 1000ppm, ammonia levels of below 25ppm [9], [10], [11], [12], [13]. Higher temperatures will cause livestock to be stressed, and their growth will be impeded. A high humidity level and a high level of carbon dioxide and ammonia level will cause the environment to be a suitable hotbed for bacteria to thrive, potentially spreading zoonotic disease. The disease will quickly spread in a humid climate, leading to premature death of the livestock.

C. Smart Farming

Smart farming is rooted in the concept that fresh food can be grown year-round in urban areas while utilizing fewer resources, such as land, water, and chemicals. Smart farming also aims to minimize the distance food travels to reach consumers, thus reducing food miles. The Smart Farming implementation discussed can increase efficiency compared to traditional farming. The United Nations (UN) stated that the growing global population catalyzes the surge in both food consumption and production requirements, with a potential increase of up to 70% by the year 2050. A shift from conventional agriculture to smart farming is necessary to meet these demands.

Therefore, utilizing IoT technologies that would enable accurate data collection for data-driven decision-making processes is essential. The transition to smart farming will allow remote monitoring that can decrease production costs. It can also promote adopting efficient and sustainable agricultural practices that align with the demand and resource efficiency objectives of smart farming. Despite the challenges associated with adopting smart farming, including the increased use of sensors that may pose security risks, smart farming remains the most promising approach to enhancing livestock production. Implementing smart farming techniques offers unparalleled opportunities for optimizing agricultural practices and ensuring the sustainability of food resources. Although the smart farming solution poses several potential vulnerabilities, including opening the attached vectors to the IoT devices, sensors, and actuators through the hardware layer, network layer, and application layer, a comprehensive framework will be able to address this issue. Therefore, the benefits of smart farming in improving livestock yields and contributing to food security outweigh the potential drawbacks, making it a critical strategy for the future of agriculture [22], [23], [24].

However, smart farming introduces additional risks and threats due to the data storage and communication nature inherited. The limitations of IoT devices necessitate storing data in cloud computing facilities, making cloud security crucial [25]. Ensuring cloud infrastructure security is also crucial for smart farming environment. Furthermore, understanding cloud security implications is essential to protect data and address access control challenges [26], [27]. Human factors, like cloud computing acceptance, must also be considered to mitigate security risks in training and analysis [28].

Researchers in [9], [11], [21] discuss the implementations of sensory technology in monitoring farming environments while research in [10] discusses fuzzy logic in calculating the ideal temperature and humidity for chicken coop. However, our research discusses a gap from previous research, whereby the gathered data was not fully utilized. The collected data were only implemented so the farmers could remotely monitor their farming environment. We observe that previous research only uses collected information to adjust conditions to create an optimal setting and has not adequately addressed the utilization of data for predicting the environmental changes in the livestock environment. Hence, to address this gap, this research utilizes machine learning to predict the optimum livestock environment.

D. Machine Learning

Machine learning (ML) is a system that makes decisions and predictions through adaptability rather than being explicitly programmed. It automatically obtains and absorbs data through the understanding of patterns and relationships in data via the creation of algorithms [29]. The basis of this method is learning from training, observation, analysis, and experience obtained from the data [30]. The outcomes from the analysis in ML demonstrate its ability to self-improve and its effectiveness. Through this technique, the future behavior of data can be predicted without depending on a preset equation as a model. ML can be divided into two categories, namely supervised and unsupervised learning. Supervised learning generates predictions or classifications for new input data by utilizing labeled data instead of unsupervised learning, which uses unlabeled data. Supervised learning is used for tasks that require classification and regression [31]. Supervised learning is more accurate for prediction tasks as they are trained directly with labeled datasets, while unsupervised learning finds patterns and structures from the given data. Since this research focuses on training ML models to predict the optimum condition of the livestock environment using the labeled datasets, supervised learning ML models are used. The ML models used in this research are the most commonly used ML models for supervised learning, which are Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbors (KNN).

Decision Tree: The Decision Tree model is a flowchart-like tree-structured model of every possible decision and their consequences [32]The decision tree model contains decision nodes to denote its attributes and edges to represent the attribute values. It is an effective technique for constructing classifiers from data. Depicting the Decision Tree model as a tree structure helps classify new data input by allowing it to construct decision rules.

Naïve Bayes: The Naïve Bayes model considers each feature as unique from one another. The model utilizes the concept of a Bayesian approach to its probability calculations. The Bayesian approach uses the Bayes' Theorem to form predictions as more training is given. It combines the prior probability and conditional probability in a formula. The mathematical equation of the Bayes' Theorem is as follows:

$$P(x|Y) = \frac{P(Y|x)P(x)}{P(Y)} \tag{1}$$

The equation displays P(x) as the prior probability of the feature set x regardless of the label. P(x|Y) is the prior probability of feature set x given that the label is Y. P(Y) is the prior probability that the label Y has occurred. The equation shows the naive prediction process of assuming that all features are unrelated, hence the "Naïve" in its name.

K-Nearest Neighbors: K-Nearest Neighbors model is an altered version of an instance-based learning algorithm that relies on feature distinctions within a labeled dataset. The K-Nearest Neighbors model employs distance metrics to identify a group of K-samples nearest to the new and unknown samples. This means that the K-Nearest Neighbors model identifies the K most similar instances nearest to the current point of data using a labeled dataset. The model retains the complete training set in the training phase to be used in the next steps. As a result, the model forms a prediction by comparing the similarity of the newly input data to the training data. This is done by comparing each instance of the training data with labels of unknown samples or new input data, and the prediction is obtained by calculating the mean of the response variables [32].

Based on the related work we referred to in this study, the conceptual structure for Optimal Livestock Environment Prediction in Smart Farming to Enhance Food Security is proposed in Figure 1.

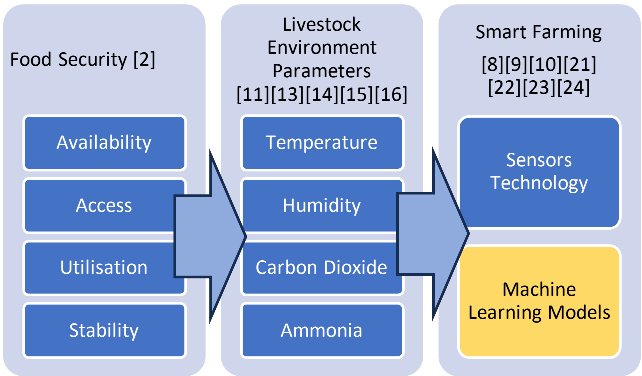

Fig. 1 Conceptual Structure for Optimal Livestock Environment Prediction in Smart Farming to Enhance Food Security

Based on the related works discussed in the previous section, we developed a conceptual structure substantially aligned with improving key production stability and supply elements to enhance food security. This alignment depends on acquiring sufficient data on the most suitable machine learning models to ensure effectiveness in predicting optimal living environments for livestock. Based on this significant gap, we conducted an experimental analysis to evaluate the effectiveness of various machine-learning models in predicting optimal living environments for livestock. We seek to generate important reference data by contributing to the existing knowledge body through this approach. The experimental analysis is discussed in the following section.

E. Method

Fig. 2 shows the proposed Smart Livestock Farming System algorithm for farmers to optimize their livestock production. In this proposed Smart Livestock Farming System, the farm uses DHT22 sensors to measure temperature and humidity levels and MQ135 Air Quality sensors to measure carbon dioxide and ammonia levels. The sensors above are used as they are compatible with the ESP32 microcontroller, an Arduino module. The sensors are then connected to the ESP32 microcontroller. The data collected by these sensors in the farm is stored in the database from the ESP32 microcontroller via Hypertext Transfer Protocol (HTTP) for viewing and generating the training dataset. The dataset is then fed to the Machine Learning models. The trained Machine Learning models will then analyze the data and provide predictions that the farmer can utilize for decision-making. The optimum environment of the farm is determined by the combined measurement of temperature, humidity, carbon dioxide, and ammonia levels in the farm from the values obtained in *B. Livestock*, with class 0 being not optimum and class 1 being optimum. The classes 0 and 1 are used to make it easier to train the Machine Learning models. Then, the farmer can carry out corrective measures on the farm based on the predictions in real time. This means that the farmer could rectify the conditions of the environment on the farm immediately without having the livestock experience a stressful environment.

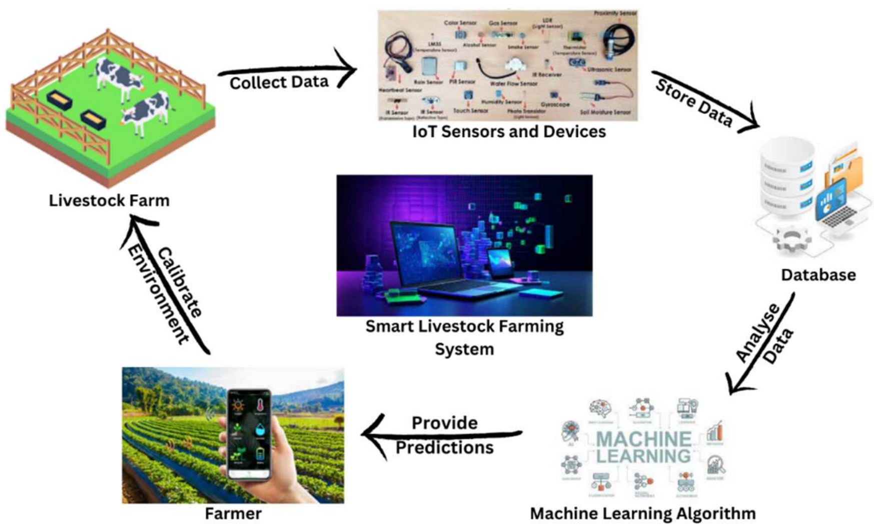

Fig. 2 Proposed Smart Livestock Farming System

This research collected 5000 rows of temperature, humidity, carbon dioxide, and ammonia levels data from sensors throughout the farm over one week in different weather conditions to simulate various conditions. The dataset is then filtered to remove duplicate data and contain only 5000 rows. The dataset is then identified as class 0 or 1 using the algorithm in Table I. The dataset is then used as the training dataset. The training dataset is then fed into the Machine

Learning models for training. Once the Machine Learning models are trained, they are tested with another dataset of 5000 rows of temperature, humidity, carbon dioxide, and ammonia levels data collected throughout another week to evaluate their performance.

TABLE I
ALGORITHM OF FARM ENVIRONMENT CONDITION GENERATION

```
Algorithm: Generate the output of the farm environment for the
training dataset
Output: Condition of Environment for Livestock Farm
Start:
Input Data: Training set (Dataset collected from sensors in farm)
For each row in Training set do
If temperature = optimum AND humidity = optimum AND
carbon dioxide = optimum AND ammonia = optimum then Farm
environment = 1
Else Farm environment = 0
End if
End for
End
```

The Machine Learning models used in the experiment are having their performance evaluated in the final phase of this research as validation. After training the Machine Learning models, the models are tested to determine which are the best at predicting the optimum livestock condition when given the test dataset. Next, a confusion matrix is generated to evaluate the performance of each Machine Learning model. This is to determine the best model that produces the best performance based on the parameters in the training data. The confusion matrix is used to visualize the machine learning models' performance by showing the machine learning models' effectiveness in distinguishing between different classes, in this case, optimum and not optimum classes. The confusion matrix evaluates the True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) values of the predictions made by Machine Learning models. The aforementioned values are defined as:

a. TN: The rate of correct predictions where the instances are of the negative class.
b. TP: The rate of correct predictions where the instances are of the positive class.
c. FN: The rate of incorrect predictions where the instances are falsely classified as the negative class when they should be positive.
d. FP: The rate of incorrect predictions where the instances are falsely classified as the positive class when they should be negative.

The confusion matrix uses these values to calculate the Machine Learning models' accuracy, precision, recall, and F1 score. These paint a clearer picture for understanding the performance and behavior of the Machine Learning models. The formulas used for calculating the accuracy, precision, recall, and F1-score values of the Machine Learning models are as follows:

Accuracy is the ratio of the total amount of correct predictions (TN and TP) to the total number of predictions (TN, TP, FN, and FP) the model made. This is to evaluate the model's overall accuracy at correctly classifying all instances into negative and positive classes.

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \tag{2}$$

Precision is the ratio of True Positive (TP) predictions to the total number of total positive predictions (TP and FP) made by the model. This is used to evaluate the model's ability to correctly predict positive predictions.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall is the ratio of True Positive (TP) predictions to the dataset's actual number of positive instances (TP and FN). This evaluates the model's effectiveness at predicting all positive instances.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

The F1 score is a metric for assessing a Machine Learning model's overall performance and is used to balance the trade-off between precision and recall scores.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{5}$$

This research analyzes the confusion matrix and concludes that the best model for determining the optimum livestock environment is in the next section.

III. RESULT AND DISCUSSION

The test results were accessed by comparing the true negative, true positive, false negative, and false positive values, as well as the score of accuracy, precision, recall, and F1 score of the three Machine Learning models (Decision Tree, Naïve Bayes, and K-Nearest Neighbors). The accuracy, precision, recall, and F1 scores are crucial for measuring and comparing the performance of one model with another. The evaluation results of the Machine Learning models are shown in the following tables. Tables II, III, and IV show the performance evaluation of the Naïve Bayes, Decision Tree, and K-Nearest Neighbors model.

TABLE II
NAÏVE BAYES PERFORMANCE

DT	False Negative	False Positive
True Negative	47.02%	0.08%
True Positive	0.18%	52.72%
Accuracy	99.74%	
Precision	99.84%	
Recall	99.66%	
F1-score	99.75%	

TABLE III
DECISION TREE PERFORMANCE

NB	False Negative	False Positive
True Negative	47.10 %	0 %
True Positive	36.36 %	16.54 %
Accuracy	63.64 %	
Precision	100 %	
Recall	31.27 %	
F1-score	47.64 %	

TABLE IV
K-NEAREST NEIGHBORS PERFORMANCE

KNN	False Negative	False Positive
True Negative	45.04 %	2.06 %
True Positive	8.68 %	44.22 %
Accuracy	89.26 %	
Precision	95.54 %	
Recall	83.59 %	
F1-score	89.16%	

Table II shows the performance of the Naïve Bayes model. This model has a 0% false positive rate, which means it correctly classified all negative instances. The model has a high false negative rate of 36.36%, indicating that the model is falsely classified 36.36% of positive instances as negative. Overall, the model achieved an accuracy score of 64%, correctly classifying 64% of instances. The model scored 100% in precision, accurately identifying the positive class 100% of the time. The model has a low recall score of 31%, only managing to classify 31% of all positive instances. Subsequently, the F1 score of the Naïve Bayes model is at a relatively low score of 47%. This is due to the high disparity between the recall and precision scores.

By just looking at the perfect precision score, we can conclude that the model predicts positive instances well. However, that is not the case. By analyzing the recall score, the Naïve Bayes model is, in fact, incapable of identifying positive instances, having only done so 31% of the time. This means that the model will wrongly classify an optimum environment as non-optimum, leading the farmer to implement corrective measures incorrectly in an otherwise optimum environment. The overall performance of the Naïve Bayes model is not up to par since it has a high rate of false negatives, and its inability to identify all positive classes results in a low recall and F1 score.

Table III shows the Decision Tree model achieving a near-perfect 0% score for both false positive and false negative rates. The negligible false positive and false negative rate means the model has an overall accuracy of 100%, hence correctly classifying all instances. Subsequently, the model scored 99.84% in precision, accurately identifying the positive instances 99.84% of the time. The model achieved a near-perfect recall score of 99.66%, accurately identifying 99.66% of all positive instances. The low disparity between the recall and precision scores means that the model achieved an F1 score of 99.75%. The Decision Tree model can correctly classify negative and positive instances 99.74% of the time, with minimal error. The exceptional performance of the Decision Tree model is due to its simple yet effective algorithm.

Table IV shows the K-Nearest Neighbors model with a False Negative rate of 8.68% and a false positive rate of 2.06%. The model would wrongly classify a positive class 8.68% of the time, while a negative class is wrongly classified 2.06% of the time. As a result, the model achieved a modest overall accuracy of 89.26%, accurately classifying 89.26% of instances. The model scored 95.54% in precision, accurately identifying the positive instances 95.54% of the time. The model achieved a recall score of 83.59%, correctly classifying 83.59% of all positive instances. The K-Nearest Neighbors model has a recall score of 89.16%, indicating that the trade-off between precision and recall is not high. Overall, the K-Nearest Neighbors model has a moderate false negative rate while maintaining a low false positive rate.

In summary, the models performed relatively well in the test, except for the Naïve Bayes model. The models are exceptional at predicting negative instances, with all models having low false positive scores, with the Naïve Bayes model achieving a 0% false positive score. However, accurately classifying positive instances is somewhat ineffective, with the Naïve Bayes model recording the lowest recall score of 31.27%. The Decision Tree and K-Nearest Neighbors models are on par with the requirement for the best machine learning model in predicting the optimum livestock environment. They both have relatively low false positive and false negative rates, and both models possess high accuracy, precision, recall, and F1-score scores. However, a clear gap between both models indicates that the Decision Tree model is the better model. The false positive and false negative rate of the Decision Tree is negligible. However, the same cannot be said for the K-Nearest Neighbors model; while low, it is still somewhat of a noticeable error. The Decision Tree model's accuracy, precision, recall, and F1 scores are near perfect 100%. In comparison, the K-Nearest Neighbors model's score is only around the 90% mark, still considerably better than the Naïve Bayes model.

The most important factor when considering the best machine learning model for predicting the optimum livestock environment is the ability of the machine learning model to classify instances into different classes accurately. The inability and the ineffectiveness will lead to high false positive and false negative rates, which will mislead the farmer when monitoring and managing the livestock environment. A false positive prediction will lead the farmer into falsely believing that the farm's environment is optimum and not carrying out any actions to calibrate the environment. A false negative prediction will lead the farmer to mistakenly carry out corrective measures to optimize the farm's environment. By considering all the factors, this research concludes that the Decision Tree model is the best machine-learning technique for predicting the optimum livestock environment.

IV. Conclusion

This research aims to improve food security by increasing livestock production, thus lowering costs and making livestock more affordable. By adopting a smart farming solution that incorporates machine learning, this research focused on the gap in selecting the most accurate machine learning model by evaluating 3 models.

This research compares and evaluates the performance of machine learning models at predicting optimal conditions for livestock in smart farming environments for the selected parameters of temperature, humidity, carbon dioxide, and ammonia levels. This research successfully compared and determined the best machine learning model to predict the optimum livestock environment, the Decision Tree model. This research concludes that the Decision Tree model best predicts both instances without positive bias. The simplicity of the Decision Tree algorithm allows it to correctly and effectively classify all instances into positive and negative classes with minor false positives and negatives.

In future work, more parameters will be used in our smart farming simulation for more comprehensive data collection for Malaysia's use case in predicting the optimum livestock environment. The current parameters of temperature, humidity, ammonia, and carbon dioxide levels form the rudimentary basis of environmental factors affecting livestock, which is also the limitation of this research. More comprehensive parameters can be added to adequately represent the factors affecting the livestock environment, thus improving the accuracy of predicting the optimum environment for livestock. Additionally, the adoption of technology can also be another perspective that can be considered. Understanding constructs for technology adoption, acceptance, and readiness levels in smart farming is another area that has the potential for future areas to be studied [33]

Examining the cybersecurity aspects of smart farming solutions is crucial to comprehensively understanding the associated threats. More studies to address this gap will provide insights into potential vulnerabilities and inform strategies to mitigate risks, ensuring the secure implementation of smart farming technologies that will also improve the food security strategy of a nation. Also, security issues, including protecting infrastructure, understanding security implications, and addressing human factors towards digital hygiene, are essential to mitigate risks in this advanced agricultural environment.

Overall, this research possesses the potential to drastically enhance the capabilities of predicting the optimum livestock environment. This can, in turn, optimize livestock production, which is crucial for maintaining the food security of the livestock sector for anyone who relies on it as their diet, source of income, and livelihood. Proposed future work can provide a more comprehensive structure to the body of knowledge.

ACKNOWLEDGMENT

The authors of this paper would like to whole-heartedly acknowledge the approved funds from the National Defense University of Malaysia (UPNM) and the Ministry of Higher Education Malaysia (MOHE) that have made this research achievable and practical. This research is endorsed by the Research Grant UPNM/2023/GPPP/ICT/2.

REFERENCES

[1] T. A. Naji, M. A. Teka, and E. A. Alemu, "The impact of watershed on household food security: A comparative analysis," *J. Agric. Food Res.*, vol. 15, no. January, p. 100954, 2024, doi:10.1016/j.jafr.2023.100954.

[2] N. R. Abdul Halim, H. Hashim, S. Mutalib, and M. Abd Ghani, "Food safety regulations implementation and their impact on food security level in Malaysia: A review," *Int. Food Res. J.*, vol. 31, no. 1, pp. 20–31, Mar. 2024, doi: 10.47836/ifrj.31.1.02.

[3] Y. Feng, C. Dong, and Q. Bao, "Impact of urbanization on forest ecological security in China," *Shengtai Xuebao*, vol. 42, no. 7, pp. 2984–2994, 2022, doi: 10.5846/STXB202105241352.

[4] J. Fanzo, B. de Steenhuijsen Piters, A. Soto-Caro, A. Saint Ville, M. Mainuddin, and J. Battersby, "Global and local perspectives on food security and food systems," *Commun. Earth Environ.*, vol. 5, no. 1, pp. 1–4, 2024, doi: 10.1038/s43247-024-01398-4.

[5] R. Alfred *et al.*, "Modelling and Forecasting Fresh Agro-Food Commodity Consumption Per Capita in Malaysia Using Machine Learning," *Mob. Inf. Syst.*, vol. 2022, 2022, doi:10.1155/2022/6106557.

[6] X. Zhang, S. Sun, and S. Yao, "Influencing factors and spatiotemporal heterogeneity of livestock greenhouse gas emission: Evidence from the Yellow River Basin of China," *J. Environ. Manage.*, vol. 358, no. March, p. 120788, 2024, doi: 10.1016/j.jenvman.2024.120788.

[7] V. Lavanya, A. D. Kannan, S. Kamaleshwaran, R. Nishanth, and S. M. Kamatchi, "Detecting Disease In Livestock and Recommending Medicines using Machine Learning," *Proc. 2nd Int. Conf. Edge Comput. Appl. ICECAA 2023*, no. Icecaa, pp. 517–522, 2023, doi:10.1109/icecaa58104.2023.10212385.

[8] R. Roy, M. M. Baral, S. K. Pal, S. Kumar, S. Mukherjee, and B. Jana, "Discussing the present, past, and future of Machine learning techniques in livestock farming: A systematic literature review," *2022 Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. COM-IT-CON 2022*, vol. 1, no. May, pp. 179–183, 2022, doi: 10.1109/COM-IT-CON54601.2022.9850749.

[9] K. S. Dhanalakshmi, J. S. Jeyanathan, K. C. Anjineyulu, K. M. Babu, M. Mahendra, and N. P. Reddy, "Cloud IoT based Poultry Environment Analysis System," in *Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 76–81. doi: 10.1109/icais56108.2023.10073851.

[10] Y. A. Liani *et al.*, "The Broiler Chicken Coop Temperature Monitoring Use Fuzzy Logic and LoRAWAN," in *Proceeding - ICERA 2021: 2021 3rd International Conference on Electronics Representation and Algorithm*, Institute of Electrical and Electronics Engineers Inc., Jul. 2021, pp. 161–166. doi: 10.1109/icera53111.2021.9538771.

[11] F. J. Adha, M. Gapar Md Johar, M. H. Alkawaz, A. Iqbal Hajamydeen, and L. Raya, "IoT based Conceptual Framework for Monitoring Poultry Farms," in *2022 12th IEEE Symposium on Computer Applications and Industrial Electronics, ISCAIE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 277–282. doi:10.1109/iscaie54458.2022.9794471.

[12] A. Ur Rehman *et al.*, "Implementation of an Intelligent Animal Monitoring System Using Wireless Sensor Network and IoT Platform," *Int. Conf. Cyber Resilience, ICCR 2022*, pp. 1–11, 2022, doi: 10.1109/iccr56254.2022.9996080.

[13] A. A. Chaudhry, R. Mumtaz, S. M. Hassan Zaidi, M. A. Tahir, and S. H. Muzammil School, "Internet of Things (IoT) and Machine Learning (ML) enabled Livestock Monitoring," *HONET 2020 - IEEE 17th Int. Conf. Smart Communities Improv. Qual. Life using ICT, IoT AI*, no. May 2004, pp. 151–155, 2020, doi:10.1109/honet50430.2020.9322666.

[14] X. Díaz de Otálora *et al.*, "Influence of farm diversity on nitrogen and greenhouse gas emission sources from key European dairy cattle systems: A step towards emission mitigation and nutrient circularity," *Agric. Syst.*, vol. 216, no. March 2023, 2024, doi:10.1016/j.agsy.2024.103902.

[15] M. F. Díaz Baca, L. Moreno Lerma, N. Triana Ángel, and S. Burkart, "The relationships between land tenure, cattle production, and climate change – A systematic literature review," *Land use policy*, vol. 141, no. April, 2024, doi: 10.1016/j.landusepol.2024.107169.

[16] D. F. Sandoval, J. J. Junca Paredes, K. J. Enciso Valencia, M. F. Díaz Baca, A. M. Bravo Parra, and S. Burkart, "Long-term relationships of beef and dairy cattle and greenhouse gas emissions: Application of co-integrated panel models for Latin America," *Heliyon*, vol. 10, no. 1, p. e23364, 2024, doi: 10.1016/j.heliyon.2023.e23364.

[17] S. M. Leitner *et al.*, "Greenhouse gas emissions from cattle enclosures in semi-arid sub-Saharan Africa: The case of a rangeland in South-Central Kenya," *Agric. Ecosyst. Environ.*, vol. 367, no. November 2023, p. 108980, 2024, doi: 10.1016/j.agee.2024.108980.

[18] H. Zhao, X. Jia, J. Yang, Y. Wu, X. Wu, and L. Du, "Spatiotemporal variations and influencing factors of methane emissions from livestock in China: A spatial econometric analysis," *Sci. Total Environ.*, vol. 931, no. May, p. 173010, 2024, doi: 10.1016/j.scitotenv.2024.173010.

[19] K. Kumar Ghosh, M. F. Ul Islam, A. A. Efaz, A. Chakrabarty, and S. Hossain, "Real-Time Mastitis Detection in Livestock using Deep Learning and Machine Learning Leveraging Edge Devices," *Int. Symp. Med. Inf. Commun. Technol. ISMICT*, vol. 2023-May, pp. 1–6, 2023, doi: 10.1109/ISMICT58261.2023.10152110.

[20] D. Kaur and A. Kaur, "IoT and Machine Learning-Based Systems for Predicting Cattle Health Status for Precision Livestock Farming," *2022 Int. Conf. Smart Gener. Comput. Commun. Networking, SMART GENCON 2022*, pp. 1–5, 2022, doi:10.1109/smartgencon56628.2022.10083995.

[21] R. R. Samantaray, A. Azeez, and N. Hegde, "Efficient Smart Farm System Using Machine Learning," *IEEE Int. Conf. Adv. Electron. Commun. Comput. Intell. Inf. Syst. ICAECIS 2023 - Proc.*, pp. 576–581, 2023, doi: 10.1109/icaecis58353.2023.10169974.

[22] N. M. Noor, N. A. M. Razali, S. N. S. A. Sham, K. K. Ishak, M. Wook, and N. A. Hasbullah, "Decentralised Access Control Framework using Blockchain: Smart Farming Case," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, 2023, doi: 10.14569/ijacsa.2023.0140560.

[23] S. N. B. S. A. Sham, K. K. Ishak, N. A. M. Razali, N. M. Noor, and N. A. Hasbullah, "IoT Attack Detection Using Machine Learning and Deep Learning in Smart Home," *Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 510–519, Mar. 2024, doi: 10.62527/joiv.8.1.2174.

[24] N. M. Noor, N. A. M. Razali, N. A. Malizan, K. K. Ishak, M. Wook, and N. A. Hasbullah, "Decentralized Access Control using Blockchain Technology for Application in Smart Farming," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, Jan. 2022, doi:10.14569/ijacsa.2022.0130993.

[25] Z. Ishak, N. Rajendran, O. I. Al-Sanjary, and N. A. Mat Razali, "Secure Biometric Lock System for Files and Applications: A

Review," *Proc. - 2020 16th IEEE Int. Colloq. Signal Process. its Appl. CSPA 2020*, no. Cspa, pp. 23–28, 2020, doi:10.1109/CSPA48992.2020.9068689.

[26] M. Noorafiza, H. Maeda, R. Uda, T. Kinoshita, and M. Shiratori, "Vulnerability analysis using network timestamps in full virtualization virtual machine," *ICISSP 2015 - 1st Int. Conf. Inf. Syst. Secur. Privacy, Proc.*, pp. 83–89, 2015, doi: 10.5220/0005242000830089.

[27] M. Noorafiza, H. Maeda, T. Kinoshita, and R. Uda, "Virtual machine remote detection method using network timestamp in cloud computing," *2013 8th Int. Conf. Internet Technol. Secur. Trans. ICITST 2013*, pp. 375–380, 2013, doi:10.1109/icitst.2013.6750225.

[28] W. N. Wan Muhamad et al, "Enhance Multi-factor Authentication Model for Intelligence Community Access to Critical Surveillance Data," *Springer Int. Publ.*, vol. 11870, pp. 560–569, 2019, doi:10.1007/978-3-030-34032-2_49.

[29] S. Ommi and M. Hashemi, "Machine learning technique in the north zagros earthquake prediction," *Appl. Comput. Geosci.*, vol. 22, no. April, p. 100163, 2024, doi: 10.1016/j.acags.2024.100163.

[30] N. A. M. Razali *et al.*, "Political Security Threat Prediction Framework Using Hybrid Lexicon-Based Approach and Machine Learning Technique," *IEEE Access*, vol. 11, no. January, pp. 17151–17164, 2023, doi: 10.1109/access.2023.3246162.

[31] M. Çakir, M. Yilmaz, M. A. Oral, H. Ö. Kazanci, and O. Oral, "Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture," *J. King Saud Univ. - Sci.*, vol. 35, no. 6, 2023, doi: 10.1016/j.jksus.2023.102754.

[32] A. Abubakar Mas'ud, "Comparison of three machine learning models for the prediction of hourly PV output power in Saudi Arabia," *Ain Shams Eng. J.*, vol. 13, no. 4, p. 101648, 2022, doi:10.1016/j.asej.2021.11.017.

[33] M. R. A. Bakar, N. A. M. Razali, M. Wook, M. N. Ismail, and T. M. T. Sembok, "Exploring and Developing an Industrial Automation Acceptance Model in the Manufacturing Sector Towards Adoption of Industry4.0," *Manuf. Technol.*, vol. 21, no. 4, pp. 434–446, 2021, doi:10.21062/mft.2021.055.