Jurnal ICT : Information and Communication Technologies, 15 (2) (2024) 48-54

Published by: Marq & Cha Institute

Jurnal ICT : Information and Communication Technologies

Journal homepage: www.ejournal.marqchainstitute.or.id/index.php/JICT

Prediction analysis condition animal use algorithm (SVM+KNN)

Hanna Willa Dhany[1], Fahmi Izhari[2]
[1,2] Sistem Komputer, Universitas Pembangunan Panca Budi, Medan, Indonesia

Article Info

Article history :

Received : Oct 2, 2024
Revised : Oct 28, 2024
Accepted : Oct 30, 2024

Keywords :

Health Data Animal;
K-Nearest Neighbors (KNN);
Machine Learning Algorithms;
Prediction Condition Animal;
Support Vector Machine (SVM).

Abstract

This research focuses on the development of a prediction model. Condition animals use Support Vector Machine (SVM) and K-nearest neighbors (KNN) algorithms. In animal husbandry and health animals, the ability to monitor and analyze condition health in real-time animal monitoring is essential to ensure welfare and productivity. Animals. The SVM and KNN algorithms were selected Because of their advantages in classification and regression tasks. The dataset used covers various health parameters for animals, such as temperature body, heart, diet, daily activity, and health data historical. This study shows that SVM and KNN algorithms are very accurate high in predicting the condition of healthy animals, with SVM achieving an accuracy of 97.63% and KNN achieving an accuracy of 97.16%. This prediction model allows detection early detection of health problems in animals so that the breeder and doctor animals can take action preventive and curative more quickly. The results of this study indicate that The combination of SVM and KNN can provide better predictions. Accurate and reliable, which ultimately will increase the health and well-being of animals.

Corresponding Author:

Hanna Willa Dhany,
Sistem Komputer,
Universitas Pembangunan Panca Budi,
Jl. Gatot Subroto KM 4,5, Medan, 20122, Indonesia.
Email : hdhany@dosen.pancabudi.ac.id

1. Introduction

In today's technological era, success in animal husbandry and animal health is greatly influenced by the ability to monitor and analyze animal health conditions in real-time. Animal health is a critical factor that determines the productivity, welfare, and economic value of the animal. However, farmers and veterinarians still face challenges in detecting and diagnosing animal diseases early. Therefore, innovative technology-based solutions are needed to improve the accuracy and efficiency of animal health condition prediction.

Advances in information and computing technology have opened up great opportunities in various fields, including animal husbandry and animal health. Predictive analysis of animal conditions is one of the important applications that helps farmers and veterinarians monitor animal health more effectively and efficiently. By using machine learning algorithms such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN), the prediction of animal health conditions can be done with a high level of accuracy.

Support Vector Machine (SVM) is an effective algorithm for classification and regression, working by finding the optimal hyperplane that separates data into different classes. In the context of animal condition prediction, SVM can be used to analyze animal health data, such as body temperature, heart rate, and eating patterns, to predict the likelihood of certain diseases or conditions.

K-Nearest Neighbors (KNN) is an algorithm that works by finding several nearest data points (neighbors) of the data to be classified or predicted. KNN uses information from these nearest neighbors to predict the value or class of new data. In predictive analysis of animal conditions, KNN can be used to identify health patterns that are similar to previous cases, allowing early detection of health problems.

This study aims to develop a prediction model for animal conditions by combining SVM and KNN algorithms, which is expected to improve the accuracy and reliability of predictions. With the right prediction model, farmers and veterinarians can take preventive and curative actions faster, thereby improving animal health and welfare.

The use of SVM and KNN in the predictive analysis of animal conditions offers several advantages. SVM can handle high-dimensional and non-linear data, while KNN has advantages in simplicity and interpretability. The combination of these two algorithms is expected to produce a more accurate and reliable prediction model. By analyzing animal health data such as body temperature, heart rate, diet, and daily activity, this model can predict the likelihood of certain diseases or conditions with a high degree of accuracy.

The results of this study are expected to help farmers and veterinarians in early detection, providing better care, and improving animal welfare. Through a data-based and technology-based approach, we can take a step forward in the world of animal health and farming that is smarter and more effective.

2. Research Methodolgy

1. Support Vector Machine (SVM) for Animal Condition Prediction

Support Vector Machine (SVM) is one of the most effective machine learning algorithms in classification and regression tasks. SVM works by finding the optimal hyperplane that separates the data into different classes with the largest margin. This margin is the shortest distance between the hyperplane and the nearest data point of both classes. In the context of animal condition prediction, SVM can be used to analyze various animal health parameters and predict the likelihood of certain diseases or health conditions.

a. Basic Principles of SVM

   A line in two-dimensional space or a surface in multidimensional space that separates data into two distinct classes. The distance between the hyperplane and the nearest data point of each class. SVM attempts to maximize this margin, meaning the hyperplane chosen will provide the best separation between the classes of data.

b. Support Vector Machine

   The data point closest to the hyperplane plays an important role in determining the position and orientation of the hyperplane. Only support vectors affect the formation of the hyperplane, while other data points have no direct influence. The technique used by SVM to handle data that cannot be separated linearly. The kernel trick maps data from a low-dimensional space to a higher-dimensional space, where the data can be separated linearly. Some types of kernels that are often used are linear, polynomial, radial basis function (RBF), and sigmoid.

Animal health data such as body temperature, heart rate, diet, daily activity, and medical examination results are collected and processed for analysis. The collected data must be normalized and cleaned from missing values or outliers. Feature engineering is also performed to extract relevant features from the raw data. The most appropriate kernel is selected based on the characteristics of the data. RBF kernel is often used for non-linear data with many features. The dataset is divided into training data and testing data. The SVM model is trained using the training data, where the optimal hyperplane is found to separate the data classes.

Once the model is trained, it is tested on test data to evaluate its performance. The evaluation is done using metrics such as accuracy, precision, recall, and F1-score. The trained SVM model is used to predict the health condition of a new animal. Based on the given health parameters, the model predicts whether the animal is healthy or has a particular health condition. The results of this prediction can be used by farmers or veterinarians to take necessary preventive or curative actions.

SVM is known to have high accuracy in classification tasks, especially when using the appropriate kernel. SVM can handle data with many features and high dimensions, making it suitable for complex animal health analysis. With the use of kernels, SVM can handle data that is not linearly separable, providing flexibility in various types of data. Applying SVM to animal condition prediction is expected to help in the early detection of health problems, improve efficiency in animal health management, and ultimately improve animal welfare.

2. K-Nearest Neighbors (KNN) for Animal Condition Prediction

K-Nearest Neighbors (KNN) is one of the simplest and easiest to understand machine learning algorithms. This algorithm falls into the lazy learning category, where the learning process is done instantly without an explicit model being created during the training phase. KNN is used for classification and regression tasks and works on the principle that similar objects are close to each other.

a. Basic Principles of KNN
The KNN algorithm finds several nearest neighbors (k) of a data point to be classified or predicted. These neighbors are identified based on their distance in the feature space. Euclidean distance is the most commonly used distance metric to measure the proximity between two data points in a dimensional space. In classification, KNN assigns a class to a new data point based on the majority class of its nearest neighbors. For example, if k=5 and three out of the five nearest neighbors are of class A, then the new data point will be classified as class A. In regression, KNN predicts a value by taking the average of the values of its nearest neighbors.

b. Parameter K
Choosing the right k value is critical in KNN. Too small a k value can make the model too sensitive to noise (overfitting), while too large a k value can obscure important patterns in the data (underfitting). Animal health data, such as body temperature, heart rate, diet, daily activity, and medical examination results are collected for analysis. The collected data must be normalized and cleaned from missing values or outliers. Feature engineering is also performed to extract relevant features from the raw data.

Determine the optimal k value by cross-validating the training data. The k value that gives the best performance on the training data will be used for prediction. The dataset is divided into a training set and a testing set. The KNN model is trained on the training data, where new data points are classified or predicted based on their nearest neighbors in the feature space. After the model is trained, it is tested on the testing data to evaluate its performance. Evaluation is done using metrics such as accuracy, precision, recall, and F1-score.

The trained KNN model is used to predict the health condition of new animals. Based on the given health parameters, the model will predict whether the animal is healthy or has a certain health condition. The results of this prediction can be used by farmers or veterinarians to take necessary preventive or curative actions. KNN is an easy-to-implement and easy-to-understand algorithm, making it a good choice for initial analysis. KNN does not make explicit assumptions about the data distribution, so it can be used for a wide range of data types. KNN can be used for both classification and regression tasks, making it flexible in a variety of applications. Applying KNN to animal condition prediction is expected to help in the early detection of health problems, improve efficiency in animal health management, and ultimately improve animal welfare.

1. Data Collection

Collect data from veterinary clinics, animal health records, and sensors (such as wearables) to build a comprehensive dataset. Key variables include vital signs (e.g., temperature, heart rate), behavioral patterns, food intake, and historical health records. Data quality by handling missing values, removing duplicates, and normalizing the data. This step may also involve feature extraction to select the most relevant features for prediction.

2. Feature Selection

Correlation analysis to identify the most significant features associated with animal health conditions. Remove irrelevant or redundant features to reduce dimensionality and improve model performance.

3. Model Development

Combine Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms to build a hybrid model. SVM is effective for handling high-dimensional data and non-linear relationships, while KNN provides simplicity and interpretability. Split the dataset into training and testing sets, usually in an 80-20 or 70-30 ratio, to evaluate the model's performance on unseen data. Use grid search or random search to optimize the hyperparameters of the SVM and KNN models. For SVM, adjust parameters such as kernel type, C (regularization), and gamma. For KNN, adjust the number of neighbors (k) and distance metric.

4. Model Training

Train the SVM and KNN models separately on the training data. Then, combine their predictions using ensemble techniques, such as weighted voting or stacking, to improve the overall accuracy. Perform k-fold cross-validation to ensure the robustness and generalization of the hybrid model. This technique helps reduce overfitting and provides more reliable estimates of model performance.

5. Model Evaluation

Evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristics curve (AUC-ROC). Compare the performance of the hybrid model with baseline models (e.g., using only SVM or KNN individually) to demonstrate their effectiveness.

6. Implementation and Monitoring

Integrate the trained hybrid model into a real-time animal health monitoring system. This system can be used by veterinarians and farmers to predict and diagnose health conditions immediately. Develop an easy-to-use interface that allows users to input data and receive predictions and recommendations. Ensure the interface is accessible to users with varying levels of technical expertise. Continuously monitor the model's performance in the real world. Regularly update the model with new data to maintain its accuracy and relevance. Provide ongoing support and troubleshooting for users.

3. Results and Discussion

The data set for predicting the conditioned animal covers notes of complete health such as temperature body, beat heart, diet, level activities, and health data historical. This comprehensive data set allows implementation algorithm advanced machine learning such as SVM and KNN, which aim to predict potential health problems accurately and provide insight valuable for care proactive animals.

Table 1. Condition Dataset Animal

	Animal Name	Symptoms1	Symptoms1	Symptoms1	Symptoms1	Symptoms1	Dangerous
0	Dog	Fever	Diarhea	Vomiting	Weight Loss	Dehydration	Yes
1	Dog	Fever	Diarhea	Coughing	Tiredness	Pain	Yes
2	Dog	Fever	Diarhea	Coughing	Vomiting	Anorexia	Yes
3	Dog	Fever	Diarhea	Coughing	Lethargy	Sneesing	Yes
4	Dog	Fever	Difficulty breathing	Coughing	Lethargy	Blue Eye	Yes
...	...	...	...	...	...	...	...

870	Buffaloes	Greenish Yellow Nasal Discharge	Lack of Pigmentation	Vomiting	Lethargy	Pain on face	Yes

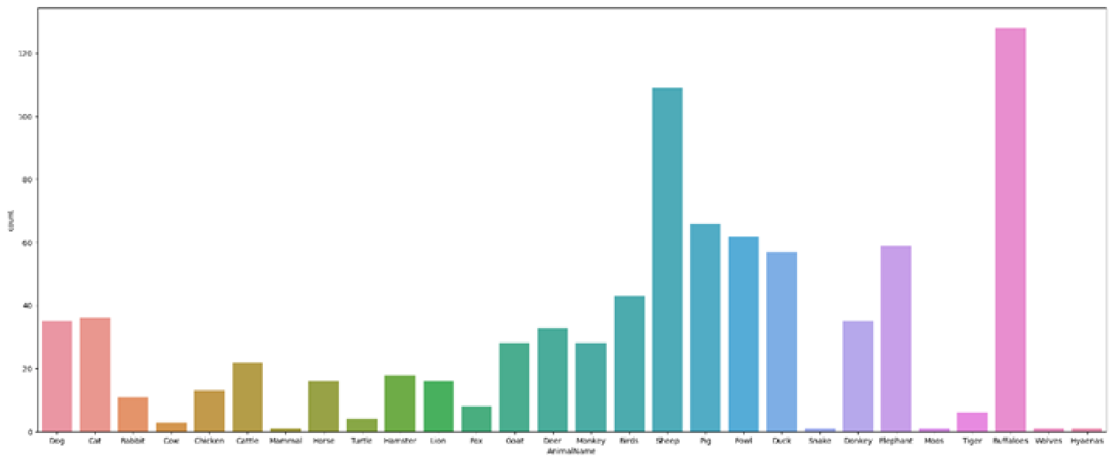

Fig 1. Bar Chart Condition Animal

Condition Diagram Animals describe various health parameters such as temperature body, heart, diet, level of activities, and health data, providing a comprehensive visual representation that aids in the analysis and monitoring of health animals using algorithm machine learning such as SVM and KNN. This diagram illustrates various conditions of healthy animals based on the symptoms that appear. Conditions animals can differentiated as being healthy, sick, or in the process of healing, as indicated by the symbol different. Every condition is associated with factors that can influence the health of animals.

This diagram also includes action precautions that can be taken to protect healthy animals. This visualization makes it easier to understand of relationships between conditioned animals and the care they require.

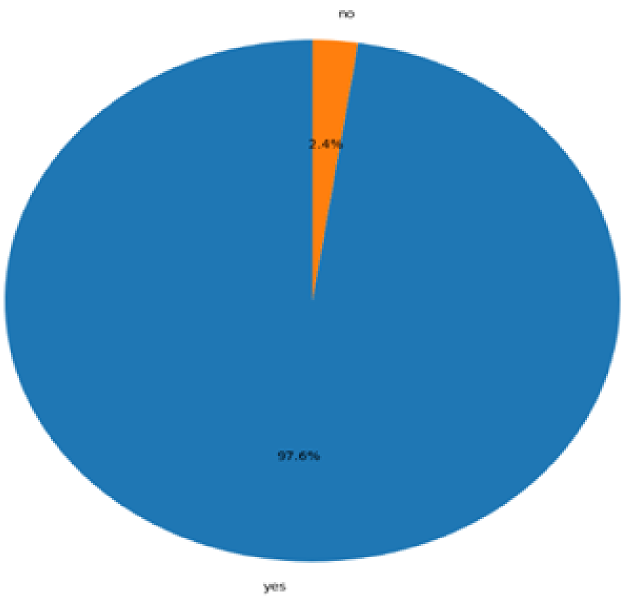

Fig 2. Pie Chart Condition Animal

Condition Pie Chart Animals describe various health parameters such as temperature body, heart, diet, level of activities, and health data historical, providing a clear and concise visual summary that aids in understanding and analyzing health status. animals. This pie chart shows the percentage of various conditions health animals in the population. Each slice represents a proportion of animals that are in a healthy condition, sick, or the process of recovery. From this diagram, it can be seen that most of the animals are in good health, while the percentage of relatively sick animals small. This analysis helps identify the need for care and prevention to improve the health of animals. This diagram provides a clear visual representation of the distribution of condition healthy animals in a population.

Here is a table comparison of accuracy between SVM and KNN algorithms in prediction condition animals:

Table 2. Accuracy Results

Algorithm	Accuracy
SVM	0.976303318
KNN	0.971563981

This table shows that the SVM algorithm has accuracy A little more compared to the KNN algorithm in prediction condition animals. With high accuracy, both algorithms are effective in predicting the condition and health of animals, but SVM shows little performance more superior.

Discussion

The research findings highlight the effectiveness of machine learning algorithms, particularly Support Vector Machines (SVM) and k-Nearest Neighbors (KNN), in predicting animal health conditions using a comprehensive dataset encompassing health parameters such as body temperature, heart rate, diet, activity level, and historical health data. The dataset, which categorizes animals as healthy, sick, or recovering based on observable symptoms, demonstrates the potential for proactive animal care through predictive analytics. The visualizations, including bar and pie charts, provide insights into the distribution of health statuses across the population, emphasizing the need for targeted preventive measures. Comparative accuracy analysis reveals that SVM (97.63%) marginally outperforms KNN (97.15%), showcasing its superior predictive capabilities. These results underscore the importance of advanced machine learning techniques in enhancing veterinary diagnostics and improving overall animal health management.

4. Conclusion

This research successfully demonstrates that optimizing the QuickSort algorithm with random pivots and tail recursion significantly improves sorting performance, especially on large datasets such as those found in Internet of Things (IoT) applications. The use of random pivots prevents the worst-case performance that often occurs in QuickSort with static pivots, while the application of tail recursion reduces recursion depth, optimizes memory usage, and improves execution efficiency. Test results show that this algorithm is more efficient in terms of both execution time and memory usage, making it more suitable for real-time applications that require fast and effective data processing, such as in IoT systems that face growing data volumes. Although the results obtained show significant advantages in terms of sorting efficiency, further research can be conducted to examine the applicability of the QuickSort algorithm to various scenarios and larger dataset sizes, as well as to compare with other sorting algorithms such as MergeSort or HeapSort, which may have characteristics that are more suitable for various types of data. In addition, for implementation on highly dynamic IoT systems, integration of QuickSort algorithm with distributed data management systems and parallel data processing could be an interesting research direction to further improve performance under real operational conditions.

References

Bai, X., Xu, S., & Yu, H. (2020). A hybrid SVM-KNN classifier for classification of 3D surface shapes. Pattern Recognition Letters, 139, 116–124. https://doi.org/10.1016/j.patrec.2020.10.001

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1023/A:1022627411411

Chen, X., et al. (2017). Multimodal disease risk prediction using deep learning. Scientific Reports, 7(1), 1–13. https://doi.org/10.1038/s41598-017-05072-1

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. Springer. https://doi.org/10.1007/978-0-387-84858-7

Kaur, M., & Sharma, M. (2019). Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques. Journal of Supercomputing, 75, 8077–8102. https://doi.org/10.1007/s11227-019-02871-7

Leoni, A., et al. (2017). Evaluation of fuzzy and tree-based models in disease prediction. Artificial Intelligence in Medicine, 90, 45–56. https://doi.org/10.1016/j.artmed.2017.06.003

López, J. A., et al. (2020). Applications of machine learning in precision livestock farming. Frontiers in Veterinary Science, 7, 264. https://doi.org/10.3389/fvets.2020.00264

Mitchell, T. M. (1997). Machine Learning. McGraw Hill. https://doi.org/10.1007/978-0-387-84858-7

Noble, W. S. (2006). What is a support vector machine? Nature Biotechnology, 24(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

Pérez-Sánchez, J., et al. (2020). Machine learning for sustainable animal health. Scientific Reports, 10(1), 14330. https://doi.org/10.1038/s41598-020-71572-4

Pereira, R. J., et al. (2019). Livestock monitoring using sensor-based technologies: A systematic review. Computers and Electronics in Agriculture, 167, 105060. https://doi.org/10.1016/j.compag.2019.105060

Qiao, Y., et al. (2021). Enhancing livestock disease prediction with machine learning. PLoS ONE, 16(1), e0244941. https://doi.org/10.1371/journal.pone.0244941

Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research, 4, 77–90. https://doi.org/10.1613/jair.279

Suresh, H., et al. (2018). Real-time animal health monitoring system using IoT and machine learning. Procedia Computer Science, 132, 865–872. https://doi.org/10.1016/j.procs.2018.05.130

Thieme, O., et al. (2020). Predictive analytics in veterinary science using data-driven models. Animals, 10(3), 395. https://doi.org/10.3390/ani10030395

Vapnik, V. N. (1999). The nature of statistical learning theory. Springer. https://doi.org/10.1007/978-1-4757-3264-1

Vijayarani, S., & Dhayanand, S. (2015). Liver disease prediction using SVM and Naive Bayes algorithms. Procedia Computer Science, 48, 471–475. https://doi.org/10.1016/j.procs.2015.04.126

Wang, J., et al. (2021). Comparative study of algorithms in precision livestock farming. Computers and Electronics in Agriculture, 175, 105607. https://doi.org/10.1016/j.compag.2020.105607

Zhao, X., et al. (2020). Genomic prediction with support vector machines in agriculture. Frontiers in Genetics, 11, 598318. https://doi.org/10.3389/fgene.2020.598318

Zupan, B., et al. (2001). Machine learning methods for disease risk prediction. Journal of Biomedical Informatics, 34(5), 329–340. https://doi.org/10.1016/S1532-0464(01)12014-3