Nat.Volatiles&Essent.Oils,2021;8(5):5393-5404

# Predictive Analytic Framework for Tools Creation and Driving Decisions in Livestock Production

**S.A. Shaik Mazhar[1] D.Akila[2]**

[1]Research Scholar, Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Email:shaik72ms@gmail.com
[2]Associate Professor, Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India.
Email : akiindia@yahoo.com

**Abstract:**
Big data can be provided through operational data collection or the use of remote animal tracking technology in animal production systems. The analytical method can be employed to produce data input systemically in order to improve the yield of animals, their health and the decision on welfare. the formulation of a theory of live animal architecture in the labs is similar to the concept of a goal variable. The use of a single dataset for the creation of a model is made possible by data partitioning., refinement, and evaluation of models. The predictive analytic method is completed with final model evaluations. The predictive approach concludes with the final model assessment, which includes an approximation of the forecast accuracy and an estimate of the chance to detect events and nonevents defined by the target variable. A thorough methodology is provided by a predictive analysis tool to the analysis of vast volumes of data to improve animal management and to improve animal prices.
**Keyword:** Big Data; Machine learning; Predictive analysis; Logistic regression; Decision tree; C4.5 decision tree; KNN ; K-means Algorithm

## INTRODUCTION

Many citizens in various countries around the world, especially in Asia, India, and South America, are becoming more financially able to purchase animal protein. This reality, coupled with shifting diets in those countries, would result in a 70% rise in global demand for animal products (meat, eggs, and milk) by 2050 [1]. The world's population of human being in the world is increasing at a rate of almost three people per second, with the global population expected to surpass 8 billion by 2025 and 9.6 billion by 2050 [2]. Some of the example of livestock farming all over the world: The hog breeding industry has always been a vital part of animal husbandry in China, and it has been the primary source of income for farmers aside from planting [3] and In Guangxi national economy animal husbandary plays an important rule especially Buffalo beef production [4].

Precision farming and decision support tools are becoming more common in the livestock industry. As a result, sensors, weather stations, individual animal tracking, feed control, and other sources generate massive amounts of data, the majority of which is used for a single purpose. Having on-farm data interoperable and open, as well as federating it with public data repositories, has unrealized potential benefits [5]. There is growing concern over poultry health at the moment, and poultry farming should be done humanely in terms of feeding and slaughtering. Around the same time, good welfare conditions are critical for healthy poultry and commodity efficiency, which can boost economic competitiveness. A balanced body, a strong emotional attitude and the ability to articulate normal behaviour helps to the well-being of a bird. There are a host of indices though, including behavioural metrics, clinical distress and performance proxies for the quantification of animal health. Many engage and evaluate each other, particularly at the commercial level[6], which makes assessment challenging and time consuming.

PLF gathers data (i.e. movement, body temperature, consumption and weight) from sensors for environmental control, human identification, and performance measurement, including RFID systems, accelerometers, cameras, microphones and temperature sensors. To detect animal behavior, a number of sensor techniques are available, welfare, and production by collecting data from various sources in an effective and continuous manner. Commercial livestock farming is a massive industrial industry that dwarfs other sectors. The carcass volume of annual meat production is calculated to be about one trillion USD. This massive size illustrates the gradual rise in meat consumption as developed countries become more reliant on animal products for protein [7].

Machine learning is a form of AI that allows machines to learn from their experiences. Its algorithms use statistical methods to learn directly from databases rather than relying on preset equations as models. If the number of training samples available grows, the algorithms gradually learn to improve their accuracy [8]. Machine learning is a subset that studies estimation and inference algerms, sometimes referred to as mathematical learning. Machine learning relies heavily on data learning. Data mining and machine learning have a similar spirit and are often discussed together. If we narrow our definition, data mining entails database structure analysis, which is important when dealing with massive databases [9].

Despite the fact that the ability to capture, process, and archive vast amounts of data has greatly improved, In livestock ecosystems, the use of predictive computational approaches is still uncommon. Statistical regression is sometimes used to determine if a sample size is small enough to quantify the likelihood of finding a significant variation where no discrepancy exists. The ability to track individual animaux and their surroundings has enabled vast volumes of spatiotemporal and animal behavioral data to be processed by advances in linked sensor systems.

This extensive database allows more detailed analyzes than just correlations, and the data can be used to assign findings into associated subgroups or foresee results. This paper describes how big data and predictive analytics can be used to decide on animal hygiene, wellbeing and safety. This paper considers how organizational data can be used to establish a forecast and classification system that can provide new insights to animal managers.

## LITERATURE SURVEY

Rodrigo García et al. [10] A detailed review of recent machine learning (ML) research in precision livestock farming (PLF) is submitted with an emphasis on two areas of concern: pasture and animal welfare. The above is a criticism: I recognize the potential for machine learning in the livestock industry; (ii) demonstrate new sensors, applications and processing techniques. In its early stages, the use of ML in PLF has been discovered to be difficult. Some examples of these challenges are given below: As a tool for disease prevention and control, I am focusing on hybrid models for disease and prescription in animals. (ii) combining the problem of grazing and animal welfare; (iii) providing autonomous data collection periods for PLF monitoring and meta-learning; and (iv) taking together variables in soil and pastures because both of animal health are a problem; and (v) bringing together variables in soil and pasture because animal health is a problem both.

Jorge A. Vázquez-Diosdado et al. [11] The authors propose a combination offline and online learning algorithm that addresses concept drift and it is considered a valuable way of managing long-term processes in fields. The proposed algorithm classifies the related behaviors with three axis gyroscopic and three axis gyroscope sensors, using embedded edge data. For the first time, the proposed approach has been published in precision livestock behavior monitoring, and it successfully classifies similar behavior in sheep in real-time, under rapidly evolving circumstances.

T. Norton et al. [12] discussed the primary approaches used in the development of precision livestock farming techniques. Precision livestock farming is a method that provides farmers with more reliable knowledge about the commodity, allowing them to make smarter decisions about the future of their production system. This paper highlights some of the core solutions and techniques used to create technology based on sound and image analysis.

Yongliang Qiao et al. [13] Any of the steps proposed for the technique include principal frame extraction (detection of large cattle motion frame), image enhancement (reduction of light and shade effects), Segmentation of animals and retrieval of body contours. On a complex cattle image dataset, we learned and checked the proposed solution. With the latest state-of-the-art segmentation approaches SharpMask and DeepMask, the proposed solution would provide relatively good accuracy in the cattle sector by providing 0,92 mean pixel accurately contouring (MPA) and 33,56 pixel ADE (average distance error).

Gota Morota et al. [14] mentioned that Initiatives to enhance animal agriculture have encouraged computer-led discoveries in animal science but the ever-increasing amount and variety of data provided by fully automated, high-performance data collection or phenotyping systems such as digital pictures, sensor and sound data, autonomous vehicles and real-time information from computer vision, presents a challenge to the effectiveness of animal precision. Machine learning and data extraction will play a significant role in tackling world agriculture's challenges. However, in the field of animal science, where knowledge is patchy, their relevance. This paper describes a system for computer and data mining and explains how they can be applied in order to overcome these gaps of expertise and solve pressing problems in animal science.

Yu Wang et al. [15] introduced Technically advanced and broadly applicable livestock plant management and regulation scheme. This system performs feed tracking and maintenance, RFID e-label identification, quality traceability, climate analysis of animal agriculture, growth monitoring and prediction and other activities. This article provides a brief description of the visualization, growth management and predictive algorithms of the animal agriculture ecosystem. This technique raises greatly the productivity and the survival rate of animal products and the off-take rate, resulting in shorter animal farming periods. This system provides an easy-to-use gui for standardizing livestock management and processing. Significant quantities of data on the farming of various types of livestock are produced by animal breeding in various areas and over several years. Constant research will contribute to optimizing animal farming activities to provide more scientific and accurate animal farming with technological assistance.

## PROPOSED METHOD

Precision livestock farming (PLF) suggests a solution to this issue by using animal space technologies to take automatic and real-time decisions on animal and population levels (Berckmans, 2017; Benjamin and Yik, 2019). Animal sensor or in their atmosphere data gathered in combination with sophisticated analytical technologies (for example, cameras, micro-phones, accelerometer systems, gas analyzers and spectrometers) provide effective instruments for animal tracks and optimization of resources such as feed, water, land and human work [16]. Animal caregivers may be alerted in real time by precision livestock farming devices, enabling them to offer individualized attention to an animal showing altered behavior due to disease, injury, or stress. PLF may also be used for a host of other livestock-related tasks, such as predicting estrus in beef and dairy cattle for better herd reproductive control, precision feeding by measuring daily feed consumption and weight gain, and so on.

Traditional inferential statistical techniques are often used in this approach to determine a result of interest, measure possible correlations with other factors, and, in some situations, make forecasts for new results. Regulated trials that are well constructed use procedures to prevent bias, and the obtained data is often interpreted using conventional inferential statistical methods. Many standard mathematical approaches cannot be used to analyse data until those assumptions are fulfilled. In well-designed experimental research trials, inferential statistics are a helpful method for evaluating possible correlations between variables. If no distinction occurs between these factors, the evaluation aim is also to ascertain the probability that observed variations are due to random chance. This is in contrast to the primary goal of predictive analytics, which is to use data that has already been compiled. This is in contrast to predictive analytics' primary aim, which is to use cumulative data to make a specific potential forecast.

Predictive analytical techniques are rarely mentioned as conventional test designs or measuring tools, but predictive analytical procedures adopt the decision-making process based on scientific information: Create a hypothesis/prediction, test it, review the observations, analyze them, update, and repeat the procedure.

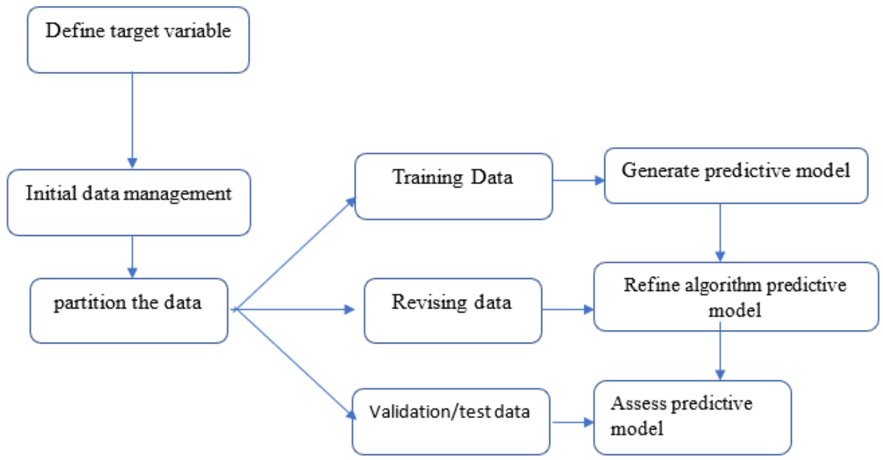



**Figure1. Block Diagram of the proposed method.**

The analytical predictive approach consists of choosing an objective variable, handling the data, partitioning data, constructing algorithms and optimizing algorithms (Figure 1). Each step is crucial to ensure the findings are valid internally and provide the evidence needed to improve future decision making. The preferred models, data types used and raw dataset management include many permutations. However, the previsionary analytic approach seen in Figure 1 promotes the systemic implementation and evaluation of big data analyses.

### *Define target variable*

Machine learning algorithms are those that need external assistance. The input dataset is split into two parts: train and analyses. Both algorithms learn patterns from the trial dataset and add them to the test dataset to predict or classify [17][24]. The train dataset contains an output vector that must be estimated or graded. The scheme provides a graphical overview of our whole sample and interpretation of the hypothesis of our dataset between dependent and independent variables (which has been discussed in the next section). Models of regression generate regression coefficients dependent on the complete dataset analysis[25].

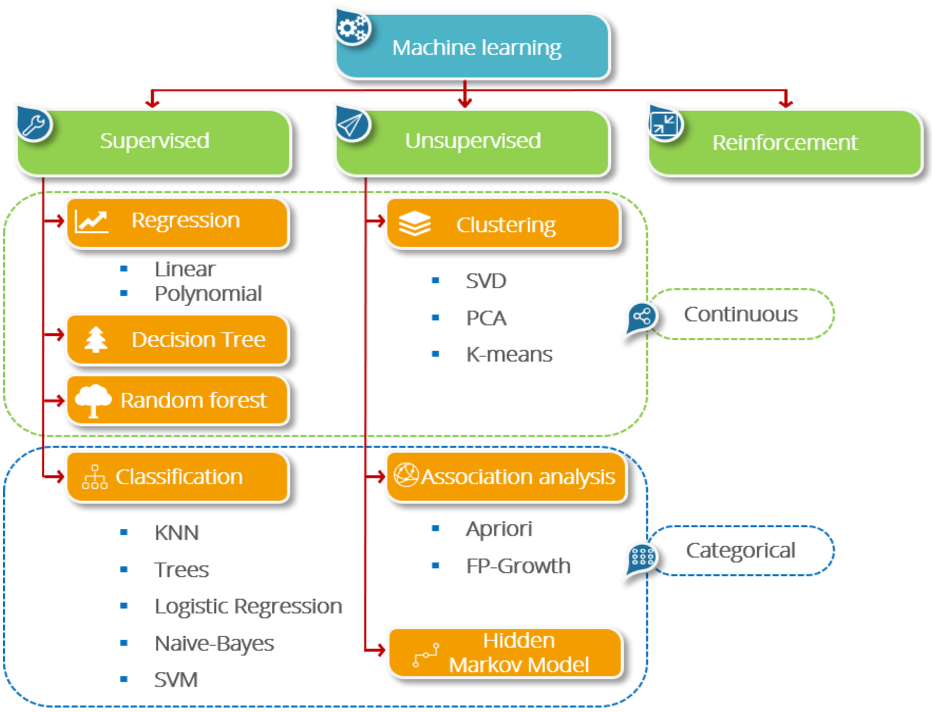



**Figure2. Machine Learning Types**

The next move should be to specifically specify the issue to be answered after defining the target attribute. Reducing the complexity of a problem to an abstract hypothesis or questionnaire reveals the exact information required, limits the models that can be used and decides the accuracy of which the model can be used for the making of decisions. Although no model is flawless, the rest of the project will be driven by realistic and consistent development objectives[18]. The model should rely on the prediction of a certain knowledge which is used for operational decisions with a well-defined decision point [19] [20].

### *Initial data management*

The scope of missing data and visual patterns can be evaluated using basic exploratory data analysis. The model array will be guided to a simple statistical overview, scattering plots, box plots, histograms, and 2-by-two tables. To get a better understanding, correlations between continuous variables may be checked of the relationships between variables in the dataset and to make choices on which variables to include or exclude.

After determining the target attribute, and validating the data structure with a simple data summary, the data must be preprocessed. Each analyst has a favorite preprocessing tool. The strategy generally involves detecting possible outliers, assessing the degree of missing data and employing methods to detect collinearity between variables if necessary.

### Multiple Linear Regressions (MLR):

MLR takes note of multiple independent variables, in contrast to a single linear regression formula. The dependent variables (expected growth rate) and separate parameters of the data set were used for our model to identify separate variables (current weight, climate, race, age, habit of food and origin). The most effective consequence of the variables with the highest correlation functions is for the effects [18].

### *Partitioning of DATA*

After deciding the variables will be used in the model, the next step is the preparation, review and validation / checking of data sets (Figure 1). In order to verify the durability of the predictive model and ensuring it is properly extended into new datasets, data must be divided in these data sub-sets prior to the model creation. The initial models are generated using the first data sub-set (training), the possibility of results differing from the original data array.

The second dataset (review) is used to preliminary inspections of classification algorithms and predictive models that are produced by the training data. Maximize efficiency by revising data between different classification models with modifications to model configuration or structure. Testing the model for guidance or reviewing subsets is not a smart idea, since it can lead to an increase in modal accuracy.

Data can be arbitrarily separated into these three data subsets before the template construction phase. An organized, randomized methodology can be considered in order to separate observations between data partitions and a hierarchical arrangement of observations that affect the process of partitioning of data. If the dataset contains several observations on true animals, for example, the individual animals can be randomly added to each data set, and all observations are associated with each animal record. The ultimate classification priorities and the underlying hierarchical data structure are inextricably tied to the data partitioning strategy.

### *Balancing of DATA*

When designing the predictive model, the final consideration is the frequency of results and when any changes in data are necessary to take care of the real objective result of the interest. Because deaths may have a major economic impact, considering their frequency, predictive modeling also focuses on unusual effects (e.g., death). Several analytical instruments can be helpful in developing models with unusual results.

The region under the receiving feature curve can be used to compare models to determine If there are differences in accuracy of the final classification or relative improvements over the normal target variable occurrence rate. In the same set of classification algorithms, mortality estimates were produced (Bayesian network, decision stump, filtered classifier, boosted logistic regression, logistic regression, multi-boosted regression, naive Bayesian, random forest and voted perceptron). When developing predictive models for unusual events, it is important to balance training data dependent on the probability of incident occurrence.

### *Selection of Model*

There are several kinds of statistical modeling that can be used to find data or forecast results based on a collection of data. Since big data includes a huge number of variables, each of which has complex relationships and interactions with the intended result, It is always difficult to pick a model dependent on the goal variable. The final evaluation, which model gives the most valuable prediction, focuses on model results in the accuracy assessment of model predictions in the validation/test data sub-sets, which also leads to various models attempting to predict the target results. The assessment can be achieved by contrasting different model results. It does not mean, when comparing the output of many models, that data must adhere to a certain type, and the best model is determined by final accuracy measurement rather than preconceived notions. The predictive analytic architecture provides the environment for testing and evaluating various classifying algorithms in order to assess the best match (in terms of accuracy) for a given scenario and target variable.

### C4.5

There have been several variants of decision tree algorithms. One of the most well-known decision tree induction algorithms is C4.5. C4.5 is a simple and efficient conventional classification prediction approach that only refers to a few possible attributes. and the objects data does not contain any contradictory information (Attribute values are the same, but they are in different categories) [21].

ID3, which is the predecessor to DY.5, is the source of C4.5. The significant difference is a shift in the evaluation classification function to replace the knowledge gain relationship with the gain of information. The main objective of this update is ID3, which picks greater values. The remedy of continuous care of attributes is another improve. C4.5 cannot easily build a more succinct and accurate decision tree if the attribute of an entity has a constant attribute [22].

In C4.5, the entropy of one of the objects in collection C that belongs to J separate classes is E(C) (3-1):

$$E(c) = -\sum Pj * log2Pj \qquad (1)$$

Log20=0 in this figure, where pj=(class J's number)/(class C's number).

Choosing Ai for the Decision Tree signifies the existence of m-Child nodes under this node (assuming that Ai has m-attributes), and each mother node entity is assigned a child node with the attribute of Ai. As a consequence, the entorpy of Ai is E. (Ai):

$$E(Ai) = \sum_k \left(\frac{nk}{n}\right) * E(Ck) \qquad (2)$$

Ck is an entity with its C array, with attributes of the same object The entropy of the target subset is Ai subset k; N is C number; Nk is Ck number, and N is C number.

Choose Properties for information acquisition nodes in the decision tree. The entropy of a list of items with Ai as a sub-tree entropy for the distance from the induced changes, i.e. entropy with Ai as an entropy sub-tree for the distance between the items of the tree node.

The systemic algorithm has provided the highest significance to attribute knowledge in this node during the construction of a Decision Tree for each tree node:

**Algorithm1.** C4.5 algorithm.

1. Work the new node from the root node. The root node is C, and all structures are set C at this stage.

2. To this category set node C and stop if all objects in C have the same form.
3. Select E for the C set item (C)
4. The results of Decision Tree Entropy E (Ai) and Income G (Ai) = E(C)-E(Ai) shall be defined for Ai or C for all nodes that have not yet been established from root to present nodes against Ai's attributes (called candidate attributes).
5. C's classification property will be the candidate feature with the largest infrmation benefit.
6. Child nodes 1C, 2C, ..., mC (if the selected category of attributes is M), were established in the C node and all object in C, based on their categorical attributes, were assigned to the corresponding children's node.
7. As the current node C, repeat step 2 for each child node Ci.

Data containing errors due to human error, expert misjudgment in classifying training examples, and so on. As noisy datasets are fed into the c4.5 algorithm, it still provides higher precision since it uses tree pruning methods to remove noise from the data [23]. When compared to other decision tree techniques, the precision is higher.

## Proposed methodology Algorithm

The aim of this research, as previously stated, was to look at a mechanism for behavioural classification in livestock that can respond to evolving conditions, also known as "concept drift." In the online version, the approach combines both offline and network algorithms. It was developed to combat paradigm shift and outperform standalone offline or online solutions.

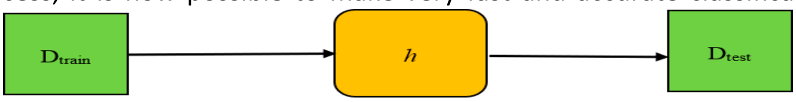
The computer classifies three-axis accelerometer data and gyroscope data from an IMU (PME). In the proposed system, two algorithms for a certain time window test the IMU sensing data using a third algorithm for each classification mark. The 3 algorithms are outlined in detail in the following pages, with high-level explanations:

1. The offline KNN algorithm: a model of KNN that produces a symbol like a move, stand, or lie (trained offline). It was carried out online with the SE C1000's PME.
2. On-line k- stands for the algorithm: an unregulated vector algorithm is used to describe walking, standing or lying by using the input functionality below:
3. Online Online Online MeanAMag measurement: a mean time window acceleration magnitude (MeanAMag) that can be used as a predictor for local operation.
4. Combined online algorithm: a joint algorithm based on preceding rules for the historical data assessment, with one and two outputs, to produce a full grade score. This is the hybrid algorithm suggested.

## Offline KNN Algorithm

An algorithm produces a model based on a Dtrain training kit and monitors an offline learning process using a data set The Ruin. In this analysis, the data obtained by Walton et al were used to construct a function vector model. [8] on an offline basis. The model is a vectorized representation of the 8455 sample dataset (7 s window, 16 Hz sampling frequency). This model was then loaded/flashed into the SC1000's Pattern Matching Engine (PME). The PME is a form of associative memory. In this associative memory, basically a bi-dimensional sequence, are stored the vectors acquired during the offline training stage (one byte per element).

At the same time, the array is contrasted with a new vector. The closest vector is located in the most fundamental form of the training set and the corresponding category is found. Up to 128 functional vectors can be stored, and many training sets may be simultaneously stored depending on the program (referred to as "context"). The number of recognitions per range has risen into the tens of thousands, and using this in-hardware process, it is now possible to make very fast and accurate classifications appropriate for in-field




scenarios.

**Figure3. Offline KNN learning scheme**

Figure 3 depicts a standard simplified offline algorithm scheme.

## Online K-means Algorithm

Compared with offline methods of learning, online learning algorithms process an endless number (xi) and outcome (yi) of potentially predictors before they meet each other. The YT is computed as y't = ht-1, as the goal of learning is the prediction of the present input yt by using the model HT-1 that is already known (xt-1). Another distinction is that the training and analysis data sets are not entirely isolated, but that a sample of cases is used for model testing until it is used for model training. Figure 4 depicts a typical online forecasting mark scheme as follows:

$$y't = h_{t-1}(x_{t-1}). \quad (3)$$

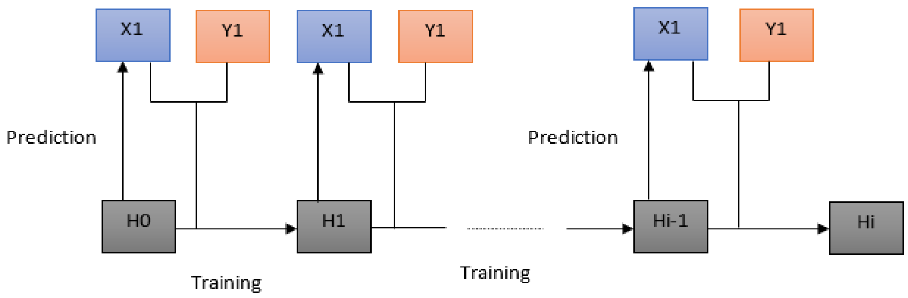



**Figure4. Online K-means learning scheme**

A lot of methods are available for this kind of on-line learning, such as increasing help vector machines, on-line random forestry, Progressive quantization of vectors and descent of stochastic gradients. In this analysis, we have used an unregulated K-means algorithm[29]. Using Matlab's integrated K Means function, the unattended online clustering algorithm was implemented [30]. The following protocols must be performed before uncontrolled k-means online clustering.

**Algorithm2.** K-means online clustering algorithm

1. Use previously acquired data values (n = 100 marks, where applicable). Use the current data set k C1, C2, C3.k centroid number.
2. Predict the new MeanAMag stage behavioral type by using the original centroid values. The class whose center is nearest to MeanAMag results in the expected behavioral class (as measured by Euclidean distance).
3. To upgrade the centroids, connect the most recently acquired MeanAMag datapoint to them.
4. Steps 2 and 3 shall be replicated until there is no additional detail. The classification predictor (YJ) with the centroids from the previous iteration shall be determined at every new iteration (C1j1, C2j1, C3j1).

An online estimate of the mean acceleration size called MeanAMag was used as an input in this algorithm. Since the SE C1000 supports local variable calculation in real-time, this variable has been selected. The formula below has been used.

$$MeanAmag = \frac{\sum_{i=1}^{n} Al}{n} \quad (4)$$

Where Al denotes the acceleration degree at each and every sampling point. The acceleration's amplitude was measured as follows:

$$Ai = \sqrt{A_{xi}^2 + A_{yi}^2 + A_{zi}^2} \quad (5)$$

Axi, Ayi and Azi describe the speed for each specimen along the aches x, y and z.

In order to obtain the discrete MeanAMag, the gravity value was deduced from the Al and then discretized in the range 1 to 20. The accelerometer and gyroscope sample frequencies are fixed at 16 Hz. The discrete MeanAMag is a discreet variant of the renowned "dynamic body acceleration," successfully used as a surrogate for animal power consumption and as one of only two attributes needed to distinguish the various cow behaviours. The discrete MeanAMag estimates often show a low computational complexity and detectable behavioral deviations from previously obtained data.

## Combined Offline and Online Algorithm

This research proposes how to combine classification labels with the KNN offline algorithm (manufactured by the PME) and the online KM algorithm using on-line MeanAMag measurement for input, all of which are an input into a roster of decisions, as previously discussed. Figure 4 shows the overall architectural scheme for the algorithmic solution, as is illustrated in the steps below.

**Algorithm3.** Combined Online K-means and Offline KNN Algorithm

1. Sensor data is obtained in its raw form, and variables are calculated.
2. MeanAMag and other function characteristics are calculated, and a vector is created to describe them.
3. The KNN supervised classification algorithm is fed the classifier vector.
4. The unsupervised learning algorithm K-means is fed MeanAMag.
5. The outputs of the K-means algorithms and KNN are used to characterize the three associated behaviors using a decision tree algorithm.

The results and the uncontrolled clustering k-means of the PME classification have been combined with the collection of previously investigated decision-making legislation. These rules were created by a decision tree classifier based on previously collected data. The set of judgment rules that was obtained is listed below. These decision rules are used to construct clusters in the combined algorithm.

1. The walking cluster contains KNN samples, the specimens which are supposed to obey the KNN algorithm with a high Means MeanAMag algorithm, Means KNN samples with a MeanAMag algorithm, and K-means with a high MeanAMag algorithm. K-means.
2. The standing cluster includes samples predicted to be lied to the low MeanAmag KNN algorithm by the k-means algorithm and samples predicted to support KNN algorithm.
3. The Lying Cluster provides examples that KNN with a medium or small MeanAMag prediction by k-means is predictable for.

Table1. online K-means and Offline KNN accuracy.

| | Offline KNN Accuracy | Online K-means Accuracy |
|---|---|---|
| Data 1 | 81.00 | 90.93 |
| Data 2 | 78.07 | 54.49 |
| Data 3 | 88.42 | 57.06 |
| Data 4 | 82.49 | 67.49 |

The table 1 shows the Offline KNN and online K-means Accuracy of various data's collected from the Big data.

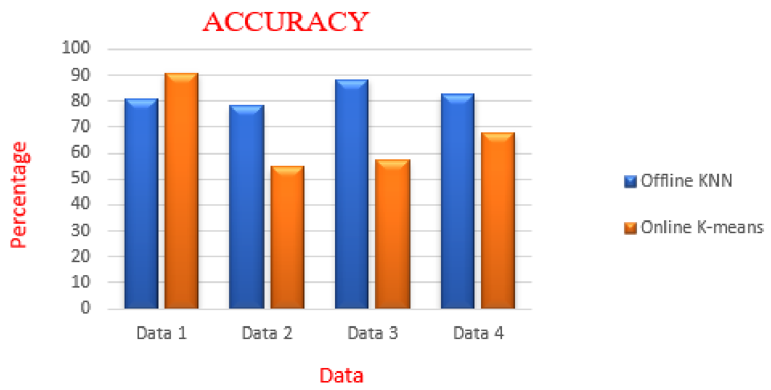



**Figure5. Offline KNN and online K-means accuracy.**

Figure 5 shows the precision of the various data sets collected using Offline KNN and Online K-means from Big Data. The results show the variation in the precision of the Offline KNN and the K-Medium algorithms.

**Table2.** C4.5 and Combined algorithm accuracy.

| | C4.5 accuracy | Combined algorithm Accuracy |
|---|---|---|
| Data 1 | 88.00 | 93.43 |
| Data 2 | 76.07 | 79.47 |
| Data 3 | 80.31 | 83.10 |
| Data 4 | 82.35 | 86.07 |

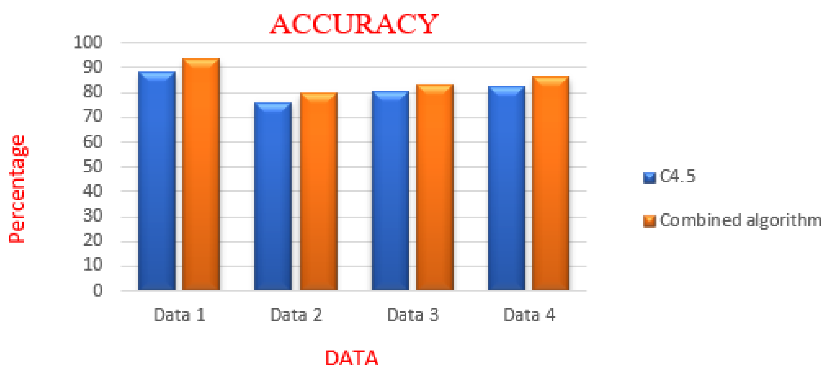



**Figure6. C4.5 and Combined algorithm accuracy.**

The Figure5 shows the C4.5 and Combined algorithm accuracy of various data's collected from the Big data. The resulted accuracy we got from the processing of C4.5 and combined algorithm accuracy shows the above represents that the combined algorithm is better in the precision analysis in livestock's than C4.5 algorithm.

## CONCLUSION

Big data can be provided through operational data collection or the use of remote animal tracking technology in animal production systems. The analysis approach for predicting cattle, health and welfare choices can be employed to extract systematic input from these results. The definition of an objective variable is like the construction of a hypothesis in a laboratory design for live animals. The predictive analytic method ends with final model evaluation, which includes an estimate of prediction precision, includes the reference variable's approximate chance of identifying events and nonevents. The Combined algorithm from combining offline KNN and Online K-means Algorithm show promising results that existing algorithm in the precision analysis in livestock farming. The predictive analytic architecture offers a comprehensive approach for analyzing big data in order to improve livestock decision making and facilitate precision animal management.

## REFERENCES


1. Berckmans, D. (2017). General introduction to precision livestock farming. *Animal Frontiers*, *7*(1), 6-11.
2. Jo, S. K., Park, D. H., Park, H., & Kim, S. H. (2018, October). Smart livestock farms using digital twin: Feasibility study. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)* (pp. 1461-1463). IEEE.
3. Cheng, L. (2011, November). An Analysis of Hog Production Prediction in Liaoning Province. In *2011 International Conference on Information Management, Innovation Management and Industrial Engineering* (Vol. 3, pp. 236-239). IEEE.
4. Li, D. (2012, October). The Analysis of Guangxi Beef Production Forecast Based on GM (1, 1) Model. In *2012 Second International Conference on Business Computing and Global Informatization* (pp. 569-571). IEEE.
5. Bahlo, C., Dahlhaus, P., Thompson, H., & Trotter, M. (2019). The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Computers and electronics in agriculture*, *156*, 459-466.
6. Li, N., Ren, Z., Li, D., & Zeng, L. (2020). Automated techniques for monitoring the behaviour and welfare of broilers and laying hens: towards the goal of precision livestock farming. *animal*, *14*(3), 617-625.
7. Vaughan, J., Green, P. M., Salter, M., Grieve, B., & Ozanyan, K. B. (2017). Floor sensors of animal weight and gait for precision livestock farming. In *2017 IEEE SENSORS* (pp. 1-3). IEEE.
8. Mekonnen, Y., Namuduri, S., Burton, L., Sarwat, A., & Bhansali, S. (2019). Machine learning techniques in wireless sensor network based precision agriculture. *Journal of the Electrochemical Society*, *167*(3), 037522.
9. Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., & Fernando, S. C. (2018). Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of animal science*, *96*(4), 1540-1550.
10. García, R., Aguilar, J., Toro, M., Pinto, A., & Rodríguez, P. (2020). A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, *179*, 105826.
11. Vázquez-Diosdado, J. A., Paul, V., Ellis, K. A., Coates, D., Loomba, R., & Kaler, J. (2019). A combined offline and online algorithm for real-time and long-term classification of sheep behaviour: Novel approach for precision livestock farming. *Sensors*, *19*(14), 3201.
12. Norton, T., Chen, C., Larsen, M. L. V., & Berckmans, D. (2019). Precision livestock farming: Building 'digital representations' to bring the animals closer to the farmer. *Animal*, *13*(12), 3009-3017.
13. Qiao, Y., Truman, M., & Sukkarieh, S. (2019). Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. *Computers and Electronics in Agriculture*, *165*, 104958.
14. Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., & Fernando, S. C. (2018). Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of animal science*, *96*(4), 1540-1550.
15. Wang, Y., Yong, X., Chen, Z., Zheng, H., Zhuang, J., & Liu, J. (2018, May). The design of an intelligent livestock production monitoring and management system. In *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)* (pp. 944-948). IEEE.
16. Rosa, G. J. (2021). Grand Challenge in Precision Livestock Farming. *Frontiers in Animal Science*, *2*, 3.
17. Dey, A. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, *7*(3), 1174-1179.
18. Tawheed, B. M., Masud, S. T., Islam, M. S., Arif, H., & Islam, S. (2019, October). Application of Machine Learning Techniques in the Context of Livestock. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)* (pp. 2029-2033). IEEE.
19. Athmaja, S., Hanumanthappa, M., & Kavitha, V. (2017, March). A survey of machine learning algorithms for big data analytics. In *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* (pp. 1-4). IEEE.

20. Kejela, G., Esteves, R. M., & Rong, C. (2014, December). Predictive analytics of sensor data using distributed machine learning techniques. In *2014 IEEE 6th international conference on cloud computing technology and science* (pp. 626-631). IEEE.
21. Cherfi, A., Nouira, K., & Ferchichi, A. (2018). Very fast C4. 5 decision tree algorithm. *Applied Artificial Intelligence*, *32*(2), 119-137.
22. Jia, W., & Huang, L. (2010, August). Improved C4. 5 decision tree. In *2010 International Conference on Internet Technology and Applications* (pp. 1-4). IEEE.
23. Chauhan, H., & Chauhan, A. (2013). Implementation of decision tree algorithm c4. 5. *International Journal of Scientific and Research Publications*, *3*(10), 1-3.
24. R. Kiruthiga, **D.Akila, "** Phishing Websites Detection Using Machine Learning**",** International Journal of Recent Technology and Engineering, Vol.8,(2S11), September 2019, pp. 111-114
25. E.Kanomozhi, **D.Akila** ,"An Empirical Study on Machine Learning Algorithm for Plant Disease Prediction", Journal of Critical Reviews Vol 7, Issue 5, 2020 491-493