Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 2, April 2024, pp. 509-518

DOI: https://doi.org/10.52436/1.jutif.2024.5.2.1521
p-ISSN: 2723-3863
e-ISSN: 2723-3871

CATTLE BODY WEIGHT PREDICTION USING REGRESSION MACHINE LEARNING

Anjar Setiawan[*1], Ema Utami[2], Dhani Ariatmanto[3]

[1,2,3]Magister of Informatics Engineering, Universitas AMIKOM Yogyakarta, Indonesia
Email: [1]anjar_setiawan@students.amikom.ac.id, [2]ema.u@amikom.ac.id, [3]dhaniari@amikom.ac.id

(Article received: November 11, 2023; Revision: December 28, 2023; published: April 15, 2024)

Abstract

Increasing efficiency and productivity in the cattle farming industry can have a significant economic impact. Cow health and productivity problems directly impact the quality of the meat and milk produced. In the cattle farming industry, it can help predict cow weight oriented to beef and milk quality. The importance of predicting cow weight for farmers is to monitor animal development. Meanwhile, for traders, knowing the animal's weight makes it easier to calculate the price of the animal meat they buy. This research aims to predict cow weight by increasing the results of smaller MAE values. The methods used are linear Regressor (LR), Random Forest Regressor (RFR), Support Vector Regressor (SVR), K-Neighbors Regressor (KNR), Multi-layer Perceptron Regressor (MLPR), Gradient Boosting Regressor (GBR), Light Gradient boosting (LGB), and extreme gradient boosting regressor (XGBR). Producing cattle weight predictions using the SVR method produces the best values, namely mean absolute error (MAE) of 0.09 kg, mean absolute perception error (MAPE) of 0.02%, root mean square error (RMSE) of 0.08 kg, and R-square of 0.97 compared to with other algorithm methods and the results of statistical correlation analysis showed several significant relationships between morphometric variables and live weight.

Keywords: *Cow Weight, Prediction, SVR.*

1. INTRODUCTION

Cattle are the primary source of meat and milk production in many regions worldwide. Increasing efficiency and productivity in the cattle farming industry can have a significant economic impact [1]. The health and productivity of cows have a direct impact on the quality of the meat and milk produced [2]. The importance of the cattle farming industry is because it can help predict the weight of cows, which is oriented to the quality of beef and milk [3]. Weighing cows is very important for producers [4]. Activities such as nutrition, management, genetics, health, and the environment can benefit from cow weight control. Cows have high economic value. Therefore, increasing accuracy and efficiency in weighing cattle can have a significant impact on productivity and animal welfare.

Cow body weight is an important indicator that has an accurate and efficient method for estimating cattle weight [5]. Increased cattle weights can help identify the best time to market animals, as animals that have reached slaughter point can represent a burden for feedlots. Predicting cow weight based on deep learning measurements using a convolutional neural network (CNN) algorithm provides the best performance with a Mean Absolute Error (MAE) of 23.19 kg. However, the MAE value results with the application of the CNN algorithm and the data used can still be improved for training segmentation models [6].

It can affect the economic value of cows because the meat and milk cows produce are based on weight. Used to plan marketing and sales strategies, farmers can predict cow weight. This can help develop more effective sales strategies and management. The goal of cattle weight prediction is to provide farmers and the livestock industry with an accurate and efficient tool for monitoring cattle health and productivity. By understanding cow body weight, farmers can make better decisions about feeding, health management, sales, and breeding [7]. Predicting animal weight is very important for breeders to monitor animal development. Meanwhile, for traders, knowing the animal's weight makes it easier to calculate the price of the animal meat they buy. Several studies have applied machine learning (ML) and deep learning (DL) to predict animal weights as a way of technological innovation.

Predicting the weight of cattle based on the use of 3D scanning technology and machine learning analytics can be used to predict live weight (LW). From the experimental results, it was found that the artificial neural networks (ANN) method produced a prediction model value with an R2 accuracy of 0.7 and an RMSE of 42. However, the results of the R2 and RMSE values by applying 3D images of live animals and the ANN algorithm can still be improved [8].

Other research on cow predictions in determining the economic index (EI) and calving interval (CI) approach [9]. The results showed that the best EI prediction was obtained with a model built using NN MLA (MAE: 20.72; RMSE: 29.35). The best CI prediction was obtained with a model built using GB MLA (MAE: 0.79; RMSE: 1.27). However, the data set used does not reflect the number of cows. By adding a more extensive and more diverse training data set, accuracy can be improved.

Research conducted [10] aims to predict the weight of sheep using images, a non-invasive method that can potentially increase the efficiency of weight management on farms. Images are processed and analyzed using various techniques, including Analysis of Variance (ANOVA) and Tukey's Test. The results show a Mean Absolute Error (MAE) of 3.099 kg, indicating promising potential for the random forest regressor (RFR) method. However, the set of images for training can still be added to improve the accuracy of other models with neural networks.

A machine learning algorithm for predicting the body weight of Balochi sheep. This research found that the random forests method gave the best results in predicting the body weight of Balochi sheep, with a coefficient of determination (R^2) of 0.988 for the training dataset and 0.916 for the testing dataset. The accuracy results from this research show that the random forests method has the lowest Mean Absolute Error (MAE) value, namely 3,275 for the test dataset. However, to test the reliability of the random forests method in predicting the body weight of livestock at various stages of growth, it can still be improved [11].

In this research, the StackingRegressor algorithm gave the best results with a Mean Absolute Error (MAE) of 4,331 and a Mean Absolute Percentage Error (MAPE) of 4,296 on the test dataset. Shows that machine learning methods can provide better results than traditional linear regression algorithms for predicting the live weight of pigs. However, further research can be done to improve prediction accuracy by applying more data preprocessing techniques, such as outlier detection and normalization, which can help improve prediction quality [12].

This study used a deep learning algorithm to estimate pig weight using images of the pig's back from above. The algorithm is based on the faster R-CNN object detection algorithm, improved by introducing a regression neural network. The MAE estimate for pig weight is 0.644 kg, and the relative error is 0.374%. This algorithm can recognize and locate the pig, as well as accurately predict the weight of the pig when the overlap area in the image is less than 30%. However, different pig postures will affect the accuracy value of the pig's weight. By adding training data, it is hoped that accuracy can be increased, and we can focus on implementing a non-contact pig weighing system [13].

This research aims to analyze correlation and regression models and determine the best and most accurate regression model to predict the body weight of Sakub ewes using body size. The best BW prediction using two predictors (BL and GC) is BW = - 56.522 + 0.509BL + 0.843CG, followed by using three predictors (BL et al.) is BW = - 57.897 + 0.505BL + 0.839CG + 0.034 WH, and using the only predictor (CG) is BW = - 28.443 + 0.905CG. However, the value of other combinations can still be improved [14].

Based on the background and previous research, this research will be carried out to predict cow weight by increasing the results of smaller MAE values. It can contribute to research in Computer Vision and Machine Learning.

2. RESEARCH METHODS

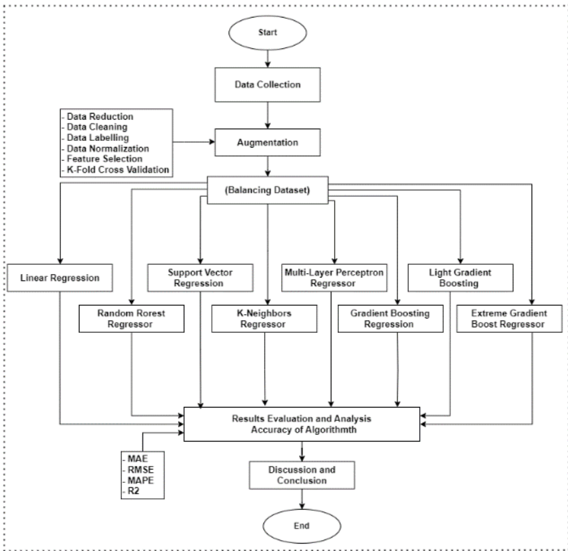

Figure 1. Research Flow Diagram

Figure 1 shows the research flow diagram in the process stages, namely data collection, augmentation, algorithm model scenarios, and results evaluation. In the first stage, the data used is the full-cow-promer dataset obtained from GitHub. The second stage is pre-processing, where the processes carried out are data reduction, data cleaning, data labeling, data normalization, feature selection, and k-fold cross-validation. The third stage is a machine learning scenario, where a design is created to determine the best accuracy using data balancing with linear regression (LR), random forest regressor (RFR), support vector regression (SVR), k-neighbors regressor (KNR), multi-layer perceptron regressor (MLPR), gradient boosting regression (GBR), light gradient boosting (LGB), and extreme gradient boost regressor (XGBR). Finally, the fourth stage is evaluation and analysis, which includes carrying out research, assessing results, and drawing conclusions based on the experiments carried out.

2.1. Dataset Collection

The used dataset comprises measurements of saplings taken manually with measuring sticks and recorded in a single centimeter [22]. The dataset has 150 data points, each corresponding to ten variables:

live weight, withers height, sacrum height, chest depth, chest width, maclock width, hip joint width, oblique body length, oblique rear length, and chest girth. This study will use the Full Cow Promer (FCP) dataset, which was obtained from GitHub and is a sample of commercial dairy farming in the Nizhny Novgorod region of Russia [23].

2.2. Data Preprocessing

Data reduction is a process that aims to decrease the complexity and amount of the obtained data. The objective of the reduction is to eliminate extraneous data about cows. Researchers might enhance their attention on the most significant and pertinent data by reducing the quantity of livestock data sent for analysis [10].

The process of data cleaning is conducted in order to ensure the quality of the data. The objective is to eliminate data that is invalid, incomplete, or irrelevant. Additionally, it offers precise and dependable findings in research [24].

The process of data labeling is conducted in order to assign categorizations to the data about each cow. The objective is to ascertain and distinguish data by specific qualities. The significance of data labeling in this study lies in its ability to provide more targeted and pertinent categorization, modeling, and statistical examination [25].

Data normalization is implemented to transform data into a standardized format, facilitating the processing and analysis of cattle data. The primary objective of data normalization is to mitigate scale disparities, ensuring that each attribute makes a proportionate contribution towards getting more precise research findings [12].

The objective of feature selection is to ascertain the subset of attributes in the cattle dataset that is most pertinent and meaningful. The objective of feature selection is to decrease the dimensionality of data, enhance computing efficiency, eliminate redundant features, and enhance the performance of prediction models [24].

K-fold cross-validation is conducted to enhance the accuracy and reliability of model performance evaluation. This is achieved by partitioning the data into k subsets of equal size. The primary objective of k-fold cross-validation is to assess the robustness and generalizability of the model when applied to previously unseen cow data [10].

2.3. Machine Learning Scenario

This research predicts cow weight using eight machine-learning models. The eight models examined in this research are linear regression (LR), random forest regressor (RFR), support vector regression (SVR), k-neighbors regressor (KNR), multi-layer perceptron regressor (MLPR), gradient boosting regression (GBR), light gradient boosting (LGB), and extreme gradient boost regressor (XGBR).

LR can work well on datasets with a linear relationship between independent and dependent variables [10]. The multiple linear regression approach has the capability to yield optimal performance in determining the most accurate prediction line [27]. There are several components, including B, which is the dependent variable or predicted value; b, which is a constant; K, which is the independent variable; and c, which is the regression coefficient.From this equation, a line can be drawn that is able to predict the dependent variable based on the independent variable, which is defined as equation (1).

$$B = b + b_1K_1 + c_2K_2 + .. + c_nK_n \tag{1}$$

RFR can handle large datasets with high dimensions and unrelated features [15]. The Random Forest algorithm is derived by aggregating the outcomes of multiple separate decision trees [26]. For an RF consisting of J trees, where X is the indicator function and an is the jth tree of the RF defined as equation (2).

$$l(y) = ar\ gmax_c(\textstyle\sum_{j=1}^{J} X_{an(y)=c}) \tag{2}$$

SVR also has flexibility in selecting kernel functions, which can adapt to various data types and relationships [16]. In the context of regression, Support Vector Machines (SVM) endeavor to construct a hyperplane that exhibits the least possible distance to the given data points. $K(x_z, x_i)$ is the dot-product kernel defined as $K(x_z, x_j) = \phi^T(x_z)\phi^T(x_j)$. Equation (3) provides an explicit definition of the regression function using Lagrange multipliers and optimality requirements.

$$f(x) = \textstyle\sum_{z=1}^{\ell}(a_z - a_z^*)K(x, x_z) + b \tag{3}$$

There are 2 SVR kernel functions; the first is the linear kernel function equation (4), and the second is the polynomial kernel function equation (5).

$$k(x, y) = x^Ty + C \tag{4}$$

$$k(x, y) = (ax^Ty + C)^d \tag{5}$$

KNR can adapt to changes in input data [17]. Predicting output results from a set of independent factors using provided variables is the primary goal of regression issues. Based on the outcomes of the k neighbors closest to the location, kNR generates forecasts. Equation (6) is the equation of Euclidean distance. $Z(j, k)$ = distance between points j and k, n = number of data, and $i = i - th$ data index

$$Z(j, k) = \sqrt{\textstyle\sum_{i=1}^{n}(j_i - k_i)^2} \tag{6}$$

MLP is an artificial neural network that can learn from data and make accurate predictions [18]. MLP uses a backpropagation algorithm to optimize weights to produce output that matches the desired target. The function equation is as follows (7).

$$Largey = f(sum_{i=1}^{n} w_i x_i + b) \quad (7)$$

The equation in the formula is x: input from the MLP, w: weight that connects each neuron, b: bias added to each neuron, f: activation function used on each neuron, y: output from MLP.

GBR has the advantage of optimizing complex loss functions and handling high-dimensional data [19]. Carry out a residual calculation, which is the value of the prediction, using equation (8).

$$\tilde{y}_{im=-}\left[\frac{\partial\Psi(y_i,F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (8)$$

LGBM is designed to be faster and more efficient in memory usage [20]. The generalization of the algorithm equation for enhancing regression trees is denoted by equation (9).

$$f(x) = \sum_{b=1}^{B} f^b(x) \quad (9)$$

XGBR benefits from its efficiency and ability to handle sparse data [21]. The essential components of the objective function comprise two distinct elements, specifically the training loss and the regularization term, as depicted in equation (10).

$$obj(\theta) = L(\theta) + \Omega(\theta) \quad (10)$$

The formulation of the training loss function is often represented by equation (11).

$$L(\theta) = \sum_{i=1}^{n} l(y_{i,}\hat{y}_i) \quad (11)$$

As represented by equation (12), the cross-entropy loss is commonly employed as a general formula for quantifying training loss.

$$L(\theta) = -[y_i\, log(\hat{y}_i) + (1 - y_i)\, log(1 - \hat{y}_i)] \quad (12)$$

2.4. Model Evaluation

The present study involved evaluating model performance to ascertain the optimal model among the eight models that had been constructed. The evaluation of model performance encompasses several metrics, including mean absolute error (MAE), root mean square error (RMSE), mean fundamental percentage error (MAPE), and the R2 value. The R2 value is employed to assess the accuracy of predictions generated by the two models. The modeling that has been previously developed will subsequently undergo evaluation, as denoted by equations (13), (14), (15), and (16).

$$MAE = \frac{\sum_{i=1}^{n}|xi-yi|}{n} \quad (13)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(xi-yi)^2}{n}} \quad (14)$$

$$MAPE = \frac{100\%}{n}\sum\left|\frac{y-\hat{y}}{y}\right| \quad (15)$$

$$R^2 = 1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2} \quad (16)$$

3. RESULTS AND DISCUSSIONS

The data about all the chosen cows was gathered within the enclosure to conduct manual body measurements [22]. This dataset comprises information about cattle owned by private farms in Russia's Nizhny Novgorod area.

As depicted in Figure 2, The nine anthropometric measurements were manually obtained by a skilled practitioner with a measuring tape and then documented in cm.

White paint was applied to the cow's body to create hand measurement markers. Furthermore, the automatic method uses anatomical markers based on the cow's body parameters. Anatomical markers in the form of bony protrusions and depressions on the surface of the cow's body can be measured [28]. Figure 2 displays the dimensions of each cow, including: (1) live weight, (2) withers height, (3) sacrum height, (4) chest depth, (5) chest width, (6) maclock width, (7) hip joint width, (8) oblique body length, (9) oblique rear length, and (10) chest girth.

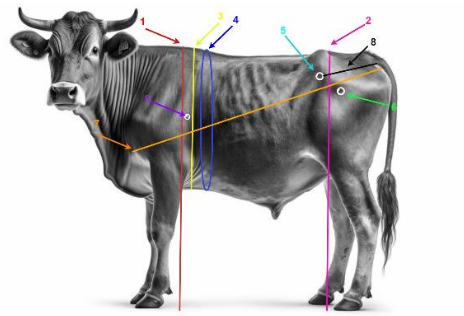

Figure 2. Dimensions of the cow's body

The primary aim of the researcher is to determine the optimal model for predicting cow weight and identify the model that yields the lowest mean absolute error (MAE) value. The application of regression machine learning techniques.

The cow dataset has benefits in every dimension of cow body measurements, providing helpful information in evaluating cows' body condition and health, which can be seen as a whole in Table 1.

Table 1. Cow Dataset

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
415	117	122	62	40	43	42	145	43	172
407	116	121	60	39	42	44	127	37	171
448	114	121	60	43	41	40	128	41	176
443	118	123	63	46	44	44	150	46	176
410	124	127	66	41	42	44	140	45	178
441	120	124	60	43	43	44	138	45	178

427	117	121	62	41	43	45	149	45	172
380	115	117	60	48	41	41	137	43	172
416	118	120	62	44	42	40	140	44	175
424	119	122	65	46	44	40	151	46	182
450	120	124	64	43	43	45	138	42	180
429	118	120	60	46	40	41	140	41	178
410	117	119	61	41	41	40	143	44	179
464	119	122	62	45	44	41	133	39	186
407	116	129	64	40	39	40	139	41	175
312	105	110	55	31	38	37	137	40	153
412	118	120	61	43	43	45	143	45	183
414	120	124	63	41	42	44	144	45	181
378	118	120	60	35	43	42	143	43	170
441	121	124	64	40	43	49	152	46	181
373	117	120	58	38	36	43	138	41	168
394	122	125	63	44	41	49	148	45	177
439	122	125	63	44	41	49	148	45	177
441	119	123	61	43	41	49	148	45	177
434	119	121	58	42	44	43	154	46	174
414	118	121	65	46	41	44	142	45	174
400	118	118	62	42	42	42	142	43	174
464	125	128	65	45	42	45	159	48	183
442	120	124	61	44	44	48	144	47	174
416	121	124	62	45	44	43	149	46	180
444	122	126	66	43	44	46	452	47	184
395	119	122	60	40	42	43	148	45	166
293	115	118	55	37	37	40	133	42	157
444	120	123	61	47	44	48	145	47	181
440	120	122	64	42	43	41	142	45	175
313	113	116	56	36	36	46	143	38	160
445	122	127	62	45	46	48	144	44	186
444	122	124	67	44	44	47	151	46	189
437	120	125	62	43	42	48	153	44	179
375	117	119	62	42	43	43	150	45	180
436	120	122	61	43	43	45	152	44	183
428	121	125	61	42	44	45	151	45	175
362	112	115	61	42	40	43	138	44	173
470	118	125	62	43	45	52	156	44	180
381	110	116	62	43	42	44	142	44	176
428	115	119	61	42	43	42	148	43	182
382	114	117	62	38	40	40	149	42	172
538	123	129	67	43	47	46	170	48	188
462	117	120	64	42	45	51	151	46	183
410	116	120	62	45	47	47	144	43	184
431	120	125	65	43	43	43	147	45	176
467	121	128	63	45	45	47	151	44	185
468	116	122	62	46	44	43	161	45	181
469	117	123	65	44	44	53	155	42	185
454	118	121	63	43	43	44	148	43	182
474	125	131	62	45	45	45	160	46	179
485	120	122	66	45	46	46	164	47	184
489	120	126	61	43	46	46	459	45	186
513	121	124	63	49	48	48	162	45	188
458	118	120	63	45	44	44	150	45	183
452	117	123	59	42	42	48	152	44	175
462	117	122	65	46	44	50	148	45	184
454	125	130	64	41	43	48	161	42	179
468	118	120	63	47	46	52	155	43	185
482	120	121	60	53	50	53	157	46	190
459	119	121	64	44	43	49	156	44	188
560	123	129	64	47	48	53	163	49	182
477	122	126	65	45	46	49	152	42	185
432	121	127	60	40	45	43	158	44	180
455	121	123	61	43	42	49	149	43	182
480	120	125	61	47	47	50	157	44	188
456	118	120	62	46	43	45	154	41	184
428	121	123	64	42	43	48	152	42	182
478	121	127	64	43	46	48	162	45	185
480	123	125	65	46	45	49	154	45	186
463	118	119	62	43	46	47	152	45	185
495	120	124	65	47	48	50	162	45	184
489	129	124	67	40	45	48	152	47	161
486	120	125	62	50	45	48	155	41	191
497	120	125	65	50	43	4.	157	45	196
440	120	121	62	46	45	47	154	45	182
445	121	122	62	45	44	48	456	46	189
453	118	121	64	47	47	49	152	48	187
445	120	124	66	44	44	45	154	42	183
401	117	120	60	40	41	41	150	43	174
486	120	124	65	47	45	49	152	45	190
451	123	125	67	45	44	50	162	49	190
466	120	120	65	44	45	50	147	43	190
440	119	120	66	40	46	48	153	47	181
417	118	120	60	42	42	46	149	43	183
471	118	120	65	44	43	44	161	43	186
506	122	127	64	43	47	53	155	45	189
450	120	120	59	44	44	51	157	46	180
451	120	120	63	46	44	50	149	45	183
462	120	121	63	43	44	53	147	44	185
487	121	122	64	47	47	52	159	45	182
462	122	124	63	44	45	53	158	46	188
346	100	100	60	41	41	46	146	41	167
407	118	119	62	39	43	46	150	42	175
366	118	121	60	42	40	44	141	46	170
312	116	119	62	44	45	48	148	44	176
378	119	125	62	36	40	44	149	43	171
382	115	122	59	38	42	43	148	42	172
408	120	126	63	38	40	48	151	42	175
323	100	100	56	40	36	40	135	37	164
275	100	100	57	40	35	40	124	40	165
409	122	125	63	39	43	49	151	45	177
331	110	117	55	32	40	43	138	40	157
416	118	121	60	40	44	47	141	40	177
417	122	127	60	43	44	47	149	44	176
243	100	107	50	29	35	35	141	35	145
399	121	127	61	39	41	46	149	40	176
492	121	125	68	44	44	50	158	46	194
605	127	132	70	50	51	55	172	49	202
495	122	128	65	46	47	52	155	48	190
565	125	137	68	47	49	52	161	49	193
399	115	122	61	40	43	46	142	42	175
486	120	127	62	42	44	50	155	44	186
495	121	130	67	50	45	52	157	50	193
493	123	132	65	46	43	48	159	47	188
540	122	127	69	50	47	56	153	46	203
486	126	133	66	45	46	47	158	46	189
545	120	131	65	45	47	55	158	48	192
484	124	130	65	44	46	48	161	45	184
411	120	129	65	46	46	51	145	47	194
512	122	130	66	42	44	52	155	49	190
551	125	130	66	47	48	44	156	49	197
479	121	126	62	50	44	52	145	44	190
514	126	132	62	46	45	51	153	50	191
548	122	129	65	50	45	56	143	50	202
548	126	128	65	47	44	52	147	58	196
418	125	130	61	40	41	45	154	45	179
531	130	135	67	45	45	50	162	45	198
512	118	123	65	46	47	54	155	44	188
413	124	127	61	41	44	50	155	43	177
563	122	130	64	43	47	53	152	47	207
511	123	126	64	48	42	44	156	45	197
495	124	128	66	43	45	42	162	45	182
519	121	130	62	50	46	53	165	48	199
549	127	130	70	44	45	52	170	46	196
488	122	127	64	43	45	48	155	44	187
526	115	125	65	46	45	52	147	46	195
573	130	136	68	48	48	54	156	50	197
543	123	128	66	48	49	55	152	48	195
547	125	135	68	48	48	53	148	48	202
466	118	122	65	46	45	49	145	44	190
530	127	121	62	50	48	55	161	45	193
550	118	125	65	48	49	53	154	45	194
520	122	127	65	51	45	54	152	46	195
511	125	128	67	48	46	51	157	51	192

3.1. Algorithm Graph Results

Figure 3 shows performance metrics results with nine variables and 100-fold cross-validation. The support vector regression algorithm produces the best accuracy values, namely MAE of 0.09 kg, MAPE of 0.02%, RMSE of 0.88 kg, and R-square of 0.97, different from research [8], which used the artificial neural network (ANN) method to produce a prediction model value with an R2 accuracy of 0.7 and an RMSE of 42. However, the results of the R2 and RMSE values by applying live 3D animal images and the ANN algorithm could still be improved.

The research conducted in [12] using the StackingRegressor algorithm gave the best results with a Mean Absolute Error (MAE) of 4.331 and a Mean Absolute Percentage Error (MAPE) of 4.296 on the test dataset. Demonstrates that machine learning methods can provide better results than traditional linear regression algorithms in predicting the live weight of pigs.

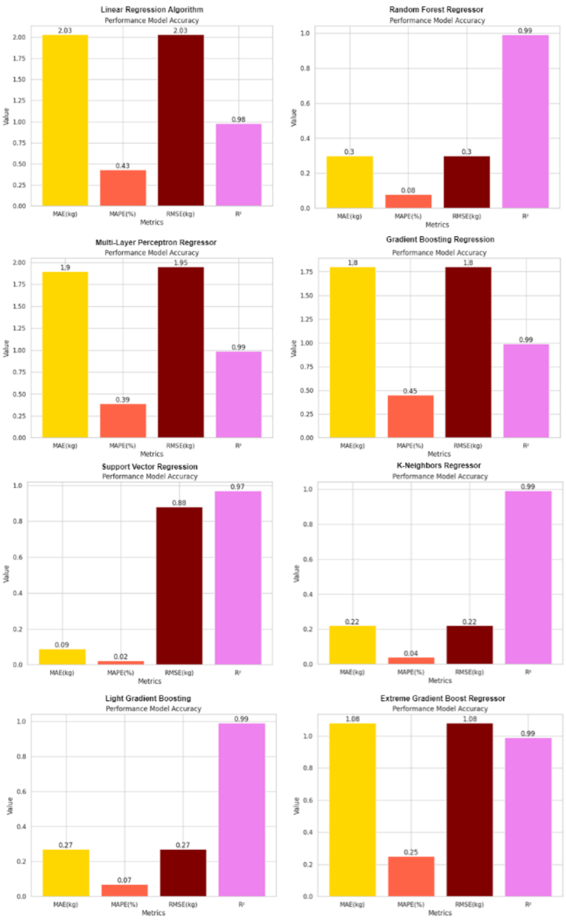

Figure 3. Algorithm model accuracy results

3.2. Relationship Features and Live Weithg

The pattern of positive relationship between the features variable and the overall live weight variable is in Figure 4. Live body weight was used as the primary target variable for analysis as an essential parameter in evaluating cattle's physical condition and welfare.

Data collected included a wide range of body sizes and dimensions carefully measured to understand morphological characteristics and potential correlations with live body weight. The variables measured include withers height, height in the sacrum, chest depth, chest width, width in maclocks, hip joint width, oblique length of the body, oblique rear length, and chest girth in Figure 4.

The statistical analysis results showed several significant relationships between morphometric variables and live body weight. In particular, several body dimensions such as withers height, chest depth, and hip joint width had a relatively strong positive correlation with live body weight. These results provide further insight into how specific morphological characteristics may contribute to live weight variability in cattle livestock. In contrast to research [14], it aims to analyze correlation and regression models and determine the best and most accurate regression model to predict the body weight of Sakub sheep using body size. The best BW prediction using two predictors (BL and GC) is BW = - 56.522 + 0.509BL + 0.843CG.

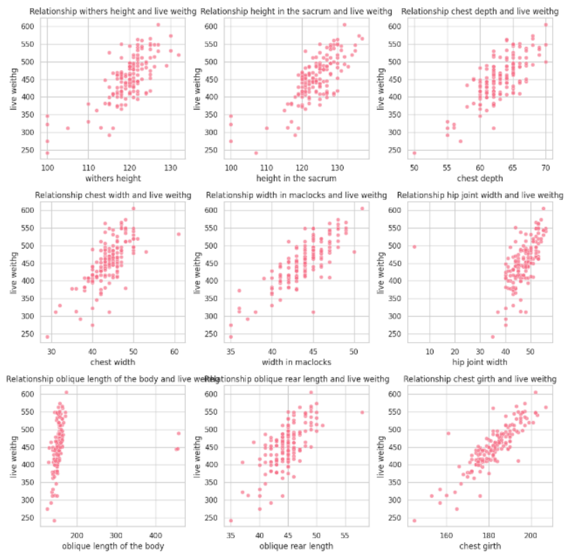

Figure 4. Relationship Features and Live Weithg

3.3. Algorithm model accuracy results

Table 2. Algorithm model accuracy results

No.	Algorithm	MAE	MAPE	RMSE	R^2
1	linear regression	2.03 kg	0.43%	2.03 kg	0.98
2	random forest regressor	0.3 kg	0.8%	0.3 kg	0.99
3	support vector regression	0.09 kg	0.02%	0.88 kg	0.97
4	k-neighbors regressor	0.22 kg	0.04%	0.22 kg	0.99
5	multi-layer perceptron regressor	1.9 kg	0.39%	1.95 kg	0.99
6	gradient boosting regression	1.8 kg	0.45%	1.8 kg	0.99
7	light gradient boosting	0.27 kg	0.07%	0.27 kg	0.99
8	extreme gradient boost regressor	1.08 kg	0.25%	1.08 kg	0.99

The research results in Table 2 show that the SVR model succeeded in providing very accurate predictions of body weight, with an MAE value of

0.09 kg, indicating a low level of prediction error. A MAPE of 0.02% confirms that this model has a high level of accuracy in predicting body weight, providing information that all predictions are very close to the actual values. In addition, the RMSE of 0.88 kg indicates a low degree of deviation between predicted and actual values, confirming the model's overall accuracy. Further analysis using R-Square indicates that the SVR model can explain 97% of the variation, providing strong evidence that this algorithm can accurately describe the pattern of relationships between input and output variables.

This is in contrast to cow weight prediction based on deep learning measurements using a convolutional neural network (CNN) algorithm, which provides the best performance with a Mean Absolute Error (MAE) of 23.19 kg [6]. The positive results of this research indicate that using the Support Vector Regression algorithm can be an effective and accurate choice in predicting individual body weight. The practical implications of these findings can be applied in various contexts, including weight management, health monitoring, and the development of more effective intervention programs. This research significantly contributes to developing cattle weight prediction techniques that can be used in the livestock industry.

4. CONCLUSION

The results of the cattle weight prediction experiment using the SVR method produced the best values, namely mean absolute error (MAE) of 0.09 kg, mean absolute perception error (MAPE) of 0.02%, root mean square error (RMSE) of 0.08 kg, and R-square of 0.97 compared with other algorithm methods and the results of statistical correlation analysis show several significant relationships between morphometric variables and live body weight. This research focuses on producing a minor mean absolute error (MAE) error value so that the optimization of the model and training data can still be improved, thus opening up opportunities for future researchers.

REFERENCES

[1] E. M. M. Van Der Heide, R. F. Veerkamp, M. L. Van Pelt, C. Kamphuis, I. N. Athanasiadis, and B. J. Ducro, "Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle," *Journal of Dairy Science*, vol. 102, no. 10, pp. 9409–9421, Oct. 2019, doi: 10.3168/jds.2019-16295.

[2] D. Giannuzzi et al., "Prediction of detailed blood metabolic profile using milk infrared spectra and machine learning methods in dairy cattle," *Journal of Dairy Science*, vol. 106, no. 5, pp. 3321–3344, May 2023, doi: 10.3168/jds.2022-22454.

[3] D.-H. Lee et al., "Estimation of carcass weight of Hanwoo (Korean native cattle) as a function of body measurements using statistical models and a neural network," Asian-australasian *Journal of Animal Sciences*, vol. 33, no. 10, pp. 1633–1641, Oct. 2020, doi: 10.5713/ajas.19.0748.

[4] G. Bretschneider, A. Cuatrín, D. Arias, and D. Vottero, "Estimation of body weight by an indirect measurement method in developing replacement Holstein heifers raised on pasture," *Archivos De Medicina Veterinaria*, vol. 46, no. 3, pp. 439–443, Jan. 2014, doi: 10.4067/s0301-732x2014000300014.

[5] R. A. Gomes, G. R. Monteiro, G. J. De Freitas Assis, K. C. Busato, M. M. Ladeira, and A. Chizzotti, "Technical note: Estimating body weight and body composition of beef cattle trough digital image analysis1," *Journal of Animal Science*, vol. 94, no. 12, pp. 5414–5422, Dec. 2016, doi: 10.2527/jas.2016-0797.

[6] M. Gjergji et al., "Deep Learning Techniques for Beef Cattle Body Weight Prediction," In *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, doi: 10.1109/ijcnn48605.2020.9207624.

[7] C. G. Dang et al., "Machine Learning-Based live weight Estimation for Hanwoo Cow," *Sustainability*, vol. 14, no. 19, p. 12661, Oct. 2022, doi: 10.3390/su141912661.

[8] G. A. Miller, J. J. Hyslop, D. Barclay, A. J. Edwards, W. Thomson, and C.-A. Duthie, "Using 3D imaging and machine learning to predict liveweight and carcass characteristics of live finishing beef cattle," *Frontiers in Sustainable Food Systems*, vol. 3, May 2019, doi: 10.3389/fsufs.2019.00030.

[9] J. Wełeszczuk, B. Kosińska-Selbi, and P. Cholewińska, "Prediction of Polish Holstein's economical index and calving interval using machine learning," *Livestock Science*, vol. 264, p. 105039, Oct. 2022, doi: 10.1016/j.livsci.2022.105039.

[10] D. A. Sant'Ana et al., "Weighing live sheep using computer vision techniques and regression machine learning," *Machine Learning With Applications*, vol. 5, p. 100076, Sep. 2021, doi: 10.1016/j.mlwa.2021.100076.

[11] Z. E. Huma and F. Iqbal, "Predicting the body weight of Balochi sheep using a machine learning approach," *Turkish Journal of Veterinary & Animal Sciences*, vol. 43, no. 4, pp. 500–506, Aug. 2019, doi: 10.3906/vet-1812-23.

[12] A. Ruchay et al., "A comparative study of

machine learning methods for predicting live weight of Duroc, Landrace, and Yorkshire pigs," *Animals*, vol. 12, no. 9, p. 1152, Apr. 2022, doi: 10.3390/ani12091152.

[13] Y. Cang, H. He, and Y. Qiao, "An intelligent pig weights estimate method based on deep learning in sow stall environments," *IEEE Access*, vol. 7, pp. 164867–164875, Jan. 2019, doi: 10.1109/access.2019.2953099.

[14] A. Ibrahim, W. T. Artama, I. G. S. Budisatria, R. Yuniawan, B. A. Atmoko, and R. Widayanti, "Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia," *Biodiversitas*, vol. 22, no. 7, Jun. 2021, doi: 10.13057/biodiv/d220721.

[15] T. K. Ho, "Random decision forests," *N Proceedings of 3rd International Conference on Document Analysis and Recognition*, Nov. 2002, doi: 10.1109/icdar.1995.598994.

[16] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Neural Information Processing Systems*, vol. 9, pp. 155–161, Dec. 1996, [Online]. Available: https://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf

[17] A. Navot, L. Shpigelman, N. Tishby, and E. Vaadia, "Nearest Neighbor Based Feature Selection for Regression and its Application to Neural Activity," *Neural Information Processing Systems*, vol. 18, pp. 996–1002, Dec. 2005, [Online]. Available: https://papers.nips.cc/paper/2848-nearest-neighbor-based-feature-selection-for-regression-and-its-application-to-neural-activity.pdf

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv (Cornell University)*, Dec. 2014, doi: 10.48550/arxiv.1412.6980.

[19] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Annals of Statistics*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.

[20] G. Ke et al., "LightGBM: a highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3149–3157, Dec. 2017.

[21] T. Chen and C. Guestrin, "XGBoost," *In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, doi: 10.1145/2939672.2939785.

[22] A. Ruchay, V. Kober, K. Dorofeev, В. И. Колпаков, A. Gladkov, and H. Guo, "Live weight prediction of cattle based on deep regression of RGB-D images," *Agriculture*, vol. 12, no. 11, p. 1794, Oct. 2022, doi: 10.3390/agriculture12111794.

[23] A. Ruchay, V. Kober, K. Dorofeev, В. И. Колпаков, and С. А. Мирошников, "Accurate body measurement of live cattle using three depth cameras and non-rigid 3-D shape recovery," *Computers and Electronics in Agriculture*, vol. 179, p. 105821, Dec. 2020, doi: 10.1016/j.compag.2020.105821.

[24] V. A. M. Weber et al., "Cattle weight estimation using active contour models and regression trees Bagging," *Computers and Electronics in Agriculture*, vol. 179, p. 105804, Dec. 2020, doi: 10.1016/j.compag.2020.105804.

[25] T. Hastie, R. Tibshirani, and J. H. Friedman, The elements of statistical learning. 2009. doi: 10.1007/978-0-387-84858-7.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Jan. 2001, doi: 10.1023/a:1010933404324.

[27] E. M. M. Van Der Heide, R. F. Veerkamp, M. L. Van Pelt, C. Kamphuis, I. N. Athanasiadis, and B. J. Ducro, "Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle," *Journal of Dairy Science*, vol. 102, no. 10, pp. 9409–9421, Oct. 2019, doi: 10.3168/jds.2019-16295.

[28] X. Song, E. A. M. Bokkers, S. Van Mourik, P. W. G. G. Koerkamp, and P. P. J. Van Der Tol, "Automated body condition scoring of dairy cows using 3-dimensional feature extraction from multiple body regions," *Journal of Dairy Science*, vol. 102, no. 5, pp. 4294–4308, May 2019, doi: 10.3168/jds.2018-15238.

[29] G. Sherwin, R. Hyde, M. A. Green, J. Remnant, E. Payne, and P. Down, "Accuracy of heart girth tapes in the estimation of weights of pre-weaned calves," *Vet Record Open*, vol. 8, no. 1, Aug. 2021, doi: 10.1002/vro2.16.

[30] A. Dakhlan, A. Saputra, M. D. I. Hamdani, and S. Sulastri, "Regression Models and Correlation Analysis for Predicting Body Weight of Female Ettawa Grade Goat using its Body Measurements," *Advances in Animal and Veterinary Sciences*, vol. 8, no. 11, Jan. 2020, doi: 10.17582/journal.aavs/2020/8.11.1142.1146.

[31] H. Ghotbaldini, M. Mohammadabadi, H. Nezamabadi-pour, O. Babenko, M. Bushtruk, and S. Tkachenko, "Predicting breeding value of body weight at 6-month age using Artificial Neural Networks in Kermani sheep

breed," *Acta Scientiarum. Animal Sciences*, vol. 41, no. 1, p. 45282, Jun. 2019, doi: 10.4025/actascianimsci.v41i1.45282.

[32] A. Ibrahim, W. T. Artama, I. G. S. Budisatria, R. Yuniawan, B. A. Atmoko, and R. Widayanti, "Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia," *Biodiversitas*, vol. 22, no. 7, Jun. 2021, doi: 10.13057/biodiv/d220721.

[33] A. Ibrahim, W. T. Artama, I. G. S. Budisatria, R. Yuniawan, B. A. Atmoko, and R. Widayanti, "Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia," *Biodiversitas*, vol. 22, no. 7, Jun. 2021, doi: 10.13057/biodiv/d220721.

[34] M. Sun, M. Hossain, T. Islam, M. Rahman, M. Hossain, and M. Hashem, "Different body measurement and body weight prediction of jamuna basin sheep in Bangladesh," *SAARC Journal of Agriculture*, vol. 18, no. 1, pp. 183–196, Jul. 2020, doi: 10.3329/sja.v18i1.48392.

[35] Z. N. Jannah, B. A. Atmoko, A. Ibrahim, M. A. Harahap, and P. Panjono, "Body weight prediction model analysis based on the body size of female Sakub sheep in Brebes District, Indonesia," *Biodiversitas*, vol. 24, no. 7, Jul. 2023, doi: 10.13057/biodiv/d240702.

[36] Y. Cang, H. He, and Y. Qiao, "An intelligent pig weights estimate method based on deep learning in sow stall environments," *IEEE Access*, vol. 7, pp. 164867–164875, Jan. 2019, doi: 10.1109/access.2019.2953099.

[37] M. Gjergji et al., "Deep Learning Techniques for Beef Cattle Body Weight Prediction," *In 2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, doi: 10.1109/ijcnn48605.2020.9207624.

[38] R. E. Uhrig, ''Introduction to artificial neural networks,'' *Proceedings of IECON '95 - 21st Annual Conference on IEEE Industrial Electronics, Orlando, FL, USA*, 1995, pp. 33-37 vol.1, doi: 10.1109/IECON.1995.483329.

[39] C. S. Vui, G. K. Soon, C. K. On, R. Alfred, and P. Anthony, "A review of stock market prediction with Artificial neural network (ANN)," In *2013 IEEE International Conference on Control System, Computing and Engineering*, Nov. 2013, doi: 10.1109/iccsce.2013.6720012.

[40] S. Bhardwaj, A. Tarafdar, M. Baghel, T. Dutt, and G. K. Gaur, "Determining Point of Economic Cattle Milk Production through Machine Learning and Evolutionary Algorithm for Enhancing Food Security," *Journal of Food Quality*, vol. 2023, pp

[41] M. S. Haris, M. I. Anshori, and A. N. Khudori, "PREDICTION OF STUNTING PREVALENCE IN EAST JAVA PROVINCE WITH RANDOM FOREST ALGORITHM," *Jurnal Teknik Informatika*, vol. 4, no. 1, pp. 11–13, Feb. 2023, doi: 10.52436/1.jutif.2023.4.1.614.

[42] D. A. Rachmawati, N. A. Ibadurrahman, J. Zeniarja, and N. Hendriyanto, "IMPLEMENTATION OF THE RANDOM FOREST ALGORITHM IN CLASSIFYING THE ACCURACY OF GRADUATION TIME FOR COMPUTER ENGINEERING STUDENTS AT DIAN NUSWANTORO UNIVERSITY," *Jurnal Teknik Informatika*, vol. 4, no. 3, pp. 565–572, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.920.

[43] C. Cahyaningtyas, D. Manongga, and I. Sembiring, "ALGORITHM COMPARISON AND FEATURE SELECTION FOR CLASSIFICATION OF BROILER CHICKEN HARVEST," *Jurnal Teknik Informatika*, vol. 3, no. 6, pp. 1717–1727, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.493.

[44] D. S. Hamdani, M. Jazman, M. L. Hamzah, and A. Anofrizen, "THE THE EFFECT OF ABDURRAB UNIVERSITY LIBRARY WEBSITE QUALITY ON USER SATISFACTION USING MULTIPLE LINEAR REGRESSION AND IMPORTANCE PERFORMANCE ANALYSIS," *Jurnal Teknik Informatika*, vol. 3, no. 6, pp. 1675–1687, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.454.

[45] R. Kurniawan and R. E. Indrajit, "THE INFLUENCE OF DIGITAL MARKETING AND LIFESTYLE ON THE DECISIONS OF E-COMMERCE (SHOPEE) USERS IN INDUSTRY 4.0," *Jurnal Teknik Informatika*, vol. 3, no. 6, pp. 1699–1707, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.468.

[46] I. Yunanto and S. Yulianto, "TWITTER SENTIMENT ANALYSIS PEDULILINDUNGI APPLICATION USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE," *Jurnal Teknik Informatika*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.292.

[47] S. Sumayah, F. Sembiring, and W. Jatmiko, "ANALYSIS OF SENTIMENT OF INDONESIAN COMMUNITY ON METAVERSE USING SUPPORT VECTOR MACHINE ALGORITHM," *Jurnal Teknik Informatika*, vol. 4, no. 1, pp. 143–150, Feb. 2023, doi: 10.52436/1.jutif.2023.4.1.417.

[48] L. I. Mustamu and Y. Sibaroni, “FUEL INCREASE SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE WITH PARTICLE SWARM OPTIMIZATION AND GENETIC ALGORITHM AS FEATURE SELECTION,” *Jurnal Teknik Informatika*, vol. 4, no. 3, pp. 521–528, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.881.

[49] A. Manik, E. B. Nababan, and T. Tulus, “IMPROVED SUPPORT VECTOR MACHINE PERFORMANCE USING PARTICLE SWARM OPTIMIZATION IN CREDIT RISK CLASSIFICATION,” *Jurnal Teknik Informatika*, vol. 3, no. 6, pp. 1739–1746, Dec. 2022, doi: 10.20884/1.jutif.2022.3.6.615.

[50] W. Sofiya and E. B. Setiawan, “FINE-GRAINED SENTIMENT ANALYSIS IN SOCIAL MEDIA USING GATED RECURRENT UNIT WITH SUPPORT VECTOR MACHINE,” *Jurnal Teknik Informatika*, vol. 4, no. 3, pp. 511–519, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.855..