KSÜ Tarım ve Doğa Derg 28 (1), 247-255, 2025
KSU J. Agric Nat 28 (1), 247-255, 2025
DOI:10.18016/ksutarimdoga.vi.1548951

Prediction of Beef Production Using Linear Regression, Random Forest and k-Nearest Neighbors Algorithms

Berkant İsmail YILDIZ[1], Kemal KARABAĞ[2]✍
[1,2]Akdeniz University, Faculty of Agriculture, Department of Agricultural Biotechnology, 07058, Antalya, Türkiye.
[1]https://orcid.org/0000-0001-8965-6361, [2]https://orcid.org/0000-0002-4516-6480
✉: karabag@akdeniz.edu.tr

ABSTRACT

The rapid increase in the global population and evolving dietary habits have significantly heightened the demand for high-quality protein sources. Beef, as a vital protein source, plays a crucial role in meeting this growing demand. This study aims to develop and evaluate a machine-learning model to predict beef production using meteorological, agricultural, and economic data. To achieve this, three different machine learning algorithms—Linear Regression, Random Forest, and k-Nearest Neighbors—were employed. The results indicate that the Random Forest algorithm outperformed the other methods in terms of R^2 and error metrics, demonstrating superior predictive accuracy. The study highlights the potential of machine learning techniques in predicting beef production, offering valuable insights for stakeholders involved in strategic decision-making to meet nutritional needs. As the global demand for protein continues to rise, the importance of such predictive models becomes increasingly significant, emphasizing the distinct advantages that machine learning approaches provide in this context.

Biostatistics

Research Article

Article History
Received : 12.09.2024
Accepted : 20.12.2024

Keywords
Beef production
Beef
Production prediction
Machine learning
Artificial intelligence

Doğrusal Regresyon, Rastgele Orman ve k-En Yakın Komşu Algoritmaları Kullanılarak Sığır Eti Üretiminin Tahmin Edilmesi

ÖZET

Küresel nüfusun hızla artması ve değişen beslenme alışkanlıkları, yüksek kaliteli protein kaynaklarına olan talebi önemli ölçüde artırmıştır. Önemli bir protein kaynağı olan sığır eti, bu artan talebin karşılanmasında kritik bir rol oynamaktadır. Bu çalışma, meteorolojik, tarımsal ve ekonomik veriler kullanarak sığır eti üretimini tahmin etmek için bir makine öğrenimi modeli geliştirmeyi ve değerlendirmeyi amaçlamaktadır. Bu amacı gerçekleştirmek için, üç farklı makine öğrenmesi algoritması—Doğrusal Regresyon, Rastgele Orman ve k-En Yakın Komşu—kullanılmıştır. Sonuçlar, Rastgele Orman algoritmasının R^2 ve hata metrikleri açısından diğer yöntemlerden daha iyi performans gösterdiğini ve üstün tahmin doğruluğu sağladığını göstermektedir. Çalışma, sığır eti üretiminin tahmin edilmesinde makine öğrenimi tekniklerinin potansiyelini vurgulamakta ve beslenme ihtiyaçlarını karşılamak için stratejik karar alma süreçlerine dahil olan paydaşlar için değerli bilgiler sunmaktadır. Küresel protein talebinin artmaya devam etmesiyle, bu tür tahmin modellerinin önemi giderek daha belirgin hale gelmekte ve makine öğrenmesi yaklaşımlarının bu bağlamda sunduğu belirgin avantajları öne çıkarmaktadır.

Biyoistatistik

Araştırma Makalesi

Makale Tarihçesi
Geliş Tarihi : 12.09.2024
Kabul Tarihi : 20.12.2024

Anahtar Kelimeler
Sığır eti üretimi
Sığır eti
Üretim tahmini
Makine öğrenmesi
Yapay zekâ

Atıf İçin: Yıldız, B. İ. & Karabağ, K. (2025). Prediction of Beef Production Using Linear Regression, Random Forest and k-Nearest Neighbors Algorithms. *KSÜ Tarım ve Doğa Derg 28* (1), 247-255. DOI: 10.18016/ ksutarimdoga.vi. 1548951.

To Cite: Yıldız, B. İ. & Karabağ, K. (2025). Prediction of Beef Production Using Linear Regression, Random Forest and k-Nearest Neighbors Algorithms. *KSU J. Agric Nat 28* (1), 247-255. DOI: 10.18016/ksutarimdoga.vi.1548951.

INTRODUCTION

The accelerating growth of the global population, coupled with evolving dietary preferences, has substantially increased the demand for high-quality protein sources (Henchion et al., 2017). Beef, with its rich protein content, plays a pivotal role in satisfying this demand and serves as a crucial source of essential micronutrients such as iron, zinc, and vitamin A (Li, 2017; Ruxton & Gordon, 2024). However, recent years have witnessed a deceleration in the growth rates of agricultural production and crop yields, raising concerns about the world's capacity to adequately feed its future population (FAO, 2024a). The increasing demand for meat has already posed significant challenges to the meat industry, further exacerbated by the fact that meat production is considerably more resource- and energy-intensive compared to other food sources, prompting the exploration of lab-grown artificial meats as potential alternatives (Marshall et al., 2011; Rout Srutee et al., 2021; Ching et al., 2022).

Accurate prediction of agricultural production data is crucial for both enhancing understanding of production processes and supporting the achievement of sustainable development goals (Bharadiya et al., 2023). Traditionally, production predictions have relied on robust statistical methods, including multivariate statistical techniques. However, these conventional approaches often fall short when dealing with complex, high-dimensional data characterized by intricate, nonlinear relationships, thereby limiting their predictive accuracy and flexibility (Yıldız et al., 2024). In response to these limitations, recent advancements in production prediction have increasingly turned towards artificial intelligence (AI) applications, particularly those based on machine learning (ML) techniques. Machine learning algorithms excel at analyzing large datasets and modeling complex interdependencies, often outperforming traditional statistical methods. In particular, supervised learning algorithms exhibit superior performance in utilizing historical data to capture the dynamic relationships within production processes (Kononenko, 2001; Ahmed & Hussain, 2022).

Numerous studies have highlighted the advantages of machine learning algorithms in agricultural production analysis. For instance, Nosratabadi et al. (2021) demonstrated that high accuracy in predicting animal food production could be achieved using machine learning algorithms such as the Adaptive Network-Based Fuzzy Inference System (ANFIS) and Multilayer Perceptron (MLP). Similarly, Alonso et al. (2013) employed Support Vector Regression (SVR) to predict the carcass weight of Asturiana de los Valles cattle, showing that carcass weight could be estimated 150 days before slaughter. Coşkun et al. (2023) successfully utilized eXtreme Gradient Boosting (XGB), Random Forest (RF), and Bayesian Regularized Neural Network (BRNN) data mining algorithms to predict the live weight of Anatolian Merinos lambs. Furthermore, Rahman et al. (2021) developed a machine learning-based prediction model for marine fish and aquaculture production by integrating Linear Regression (LR), Gradient Boosting Regression (GB), and Random Forest Regression (RFR) into an ensemble approach known as Voting Regression (VR), achieving high-performance outcomes. In another study, Yıldız et al. (2024) developed a model for predicting honey production in Turkey using various machine learning algorithms, including k-Nearest Neighbor (k-NN), Random Forest (RF), Linear Regression (LR), and Gaussian Naive Bayes (GNB). However, to date, there has been a noticeable gap in the literature specifically focusing on the application of machine learning algorithms to predict beef production.

The primary aim of this study is to develop and evaluate a machine-learning model for predicting beef production. This research seeks to compare the performance of LR, RF, and k-NN algorithms to identify the most effective predictive method. By leveraging the accuracy and reliability of machine learning, the study aims to facilitate more precise predictions of beef production, ultimately contributing to the optimization of production strategies and planning.

MATERIAL and METHOD

Material

The attributes were selected based on factors influencing beef production, as identified through a comprehensive review of the existing literature (FAO, 2024a; Van Kernebeek et al., 2016; Humer & Zebeli, 2017; Godfray et al., 2018; Nosratabadi et al., 2021; Çakan & Tipi, 2023). The study utilized a dataset comprising 62 annual average data points for each of the 18 attributes, collected between 1961 and 2022, as this period was chosen due to the availability of the most comprehensive and reliable data. These attributes include beef production, cattle population, beef price, total population, rural population, urban population, agricultural land area, pasture and meadow area, food price inflation, temperature, precipitation, Gross Domestic Product (GDP), Gross National Product (GNP), per capita GNP, barley production, corn production, barley price, and corn price. All attributes were treated as continuous variables, and no subgroup analyses were performed.

Agricultural data, such as beef production, cattle numbers, agricultural land area, pasture and meadow area, barley production, and corn production, were sourced from the Turkish Ministry of Agriculture and Forestry's official website (TMAF, 2024). Meteorological data, including temperature and precipitation, were obtained from the General Directorate of Meteorology under the Turkish Ministry of Environment, Urbanization, and Climate Change (GDM, 2024). Population and economic data—including total population, rural and urban population, food price inflation, GDP, GNP, per capita GNP, barley price, and corn price—were acquired from the Food and Agriculture Organization (FAO) of the United Nations website (FAO, 2024b). All available data from the mentioned sources between 1961 and 2022 were included in the analysis, without the use of any specific sampling technique. To address missing data within the dataset, mean imputation was employed, which involved filling the missing values with the average values calculated from the available data. The statistical properties of the attributes, including the mean, standard deviation, maximum, and minimum values, are summarized respectively in Table 1.

Table 1. Statistical properties for attributes
Çizelge 1. Özniteliklere ilişkin istatistiksel özellikler

Attributes	Mean	Standard deviation	Minimum	Maximum
Beef production (tonne)	432920.6950	378035.02030	90634.00	1572747.15
Cattle population (units)	13028781.8100	1971869.20700	9788102.00	17965482.00
Beef price (tonne/$)	6700.2241	1501.80565	2715.20	12251.40
Population (units)	56014.3943	17363.82823	28255.00	85279.55
Rural population (units)	21823.5306	1349.22945	19121.82	24723.16
Urban population (units)	33622.5860	17580.66390	9025.07	65453.23
Agricultural area (thousand ha)	38753.7032	1176.67056	36517.00	41223.00
Meadow and pasture area (thousand ha)	12415.6229	1797.07352	10000.00	14617.00
Food price inflation (%)	19.3578	12.19365	3.90	77.87
Temperature (°C)	12.0503	1.53099	9.58	15.10
Precipitation (mm)	606.1389	70.49718	460.69	793.80
GNP (million $)	349669.3695	283621.07450	904.19	941689.70
GDP (million $)	369093.8850	294729.94470	23609.87	957799.00
GDP per capita ($)	5286.1638	3465.49856	649.34	12507.80
Barley production (tonne)	6258980.6450	1868969.44400	2900000.00	9551000.00
Corn production (tonne)	2727806.4520	1921057.94700	800000.00	8500000.00
Barley price (tonne/$)	202.8625	49.39973	100.10	370.30
Corn price (tonne/$)	228.0594	44.37527	138.00	375.30

Method

In this study, commonly used machine learning algorithms—Linear Regression (LR), Random Forest (RF), and k-Nearest Neighbors (k-NN)—were selected to predict beef production. The analyses were conducted using the Python programming language, version 3.12.2 (Python Software Foundation, 2024), leveraging libraries such as Pandas (1.3.0) for data manipulation, Numpy (version 1.21.0) for numerical operations, Matplotlib (version 3.4.2) for data visualization, and Scipy (version 1.10.0) for scientific computations (Yıldız et al. 2024). The dataset, after addressing missing values through mean imputation, was split into training and test sets, with 70% of the data allocated for training the algorithms and the remaining 30% reserved for testing the predictive accuracy of the models. To ensure consistency and enhance the performance of the machine learning models, the data were standardized to balance the value differences among all attributes. This standardization process involved scaling the data to have a mean of zero and a standard deviation of one, thereby aligning the varying scales of different features and improving the models' convergence during training.

In addition, a hyperparameter optimization process was implemented using the Grid Search technique to further enhance the performance of the machine learning models. This approach entailed a systematic exploration of a range of hyperparameters for each algorithm, facilitating the discovery of the most effective combinations aimed at enhancing predictive accuracy. The selected hyperparameters for LR, RF and k-NN algorithms are detailed in Table 2.

The predictive performance of the algorithms was evaluated separately on both the training and test sets. Predictions generated by the models were compared against the actual values in the test set, and the performance was assessed using the following metrics: Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). R^2 (Coefficient of Determination) indicates the proportion of variance in the dependent variable explained by the independent variables, with values ranging from 0 to 1. The closer R^2 is to 1, the better the model explains the data. MAE (Mean Absolute Error) represents the average of the absolute differences between predicted and actual values, measuring how close the predictions are to the true values. A lower MAE indicates better model performance. MSE (Mean Squared Error) calculates the average of the squared differences between predicted and actual values, penalizing larger errors more heavily. It provides a measure of the model's overall error. Finally, RMSE (Root Mean Squared Error) is the square root of MSE and expresses the magnitude of the errors directly in the original data units. RMSE gives a precise measure of the model's overall accuracy. Each of these metrics provides a different perspective on model performance, offering a more comprehensive evaluation.

Table 2. Hyperparameter settings for the selected algorithms
Çizelge 2. Kullanılan algoritmalar için hiperparametre ayarları

Algorithm	**Hiperparameter**	**Value**
Linear Regression	C	1.0
	Solver	'lbfgs'
Random Forest	n_estimators	100
	max_depth	None
	min_samples_split	2
k-Nearest Neighbor	n_neighbors	5
	weights	'uniform'
	algorithm	'auto'

These error metrics were computed using the formulas outlined below:

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (1)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (2)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (3)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{y_i} - \hat{y}_{\mathrm{i}})^2} \qquad (4)$$

$(n) = \mathit{Total\ number\ of\ observations}$; $(yi) = (i) - \mathit{th\ observation\ of\ true\ values}$; $(\hat{y}_i) = (i) - \mathit{th\ observation\ of\ predicted\ values}$

To visually assess the alignment between the model predictions and the actual observed values, appropriate plots and graphs were generated. These visualizations facilitated a clearer understanding of how well the models performed in predicting beef production, highlighting areas of strong correlation as well as potential discrepancies.

Furthermore, a feature importance analysis was conducted using the RF algorithm. This analysis aimed to enhance the interpretability of the model by identifying and quantifying the relative impact of each attribute on beef production predictions. By assessing the contribution of each feature, the feature importance score analysis provided valuable insights into which factors most significantly influence the predictive outcomes, thus aiding in the understanding of underlying patterns and relationships within the data. This approach not only improves the model's transparency but also helps prioritize key variables that could be targeted in strategic interventions or policy formulations.

RESULTS and DISCUSSION

In this study, Linear Regression (LR), Random Forest (RF), and k-Nearest Neighbors (k-NN) algorithms were employed to predict beef production. The performance evaluations were conducted using key metrics, including the Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE). An R^2 value close to 1 indicates a strong alignment between the model's predictions and the actual data, reflecting a high level of accuracy in capturing the relationships among the variables. Conversely, error metrics such as MAE, MSE, and RMSE being close to zero suggest that the model's predictions have minimal error margins, demonstrating a high degree of concordance with the observed values. Based on these criteria, the performance of the algorithms was assessed, and the most successful models were identified by their high R^2 values and low error metrics, which indicate superior predictive accuracy and reliability (Gültepe, 2019; Rahman et al., 2021; Yıldız et al., 2024).

To comprehensively evaluate the performance of the algorithms, results from both the training and test datasets were meticulously analyzed. The training set results reflect the model's performance on data it was trained on, while the test set results provide an evaluation of the model's predictive capability on unseen, general data. High-performance values on the training set indicate that the model fits the training data well, suggesting that the model has effectively learned the underlying patterns within the data (Table 3).

This analysis ensures that the models not only excel in terms of training data but also maintain robustness and generalizability when applied to new datasets. Such thorough evaluation is crucial in establishing the models' utility in real-world applications, where the ability to generalize from past data to predict future outcomes is of paramount importance.

Table 3. Performance metrics of algorithms on the train set
Çizelge 3. Eğitim setinde kullanılan algoritmaların performans metrikleri

Algorithms	R^2	MAE	MSE	RMSE
Linear Regression	0.996	13103.153218	1249431000.0	35341.421005
Random Forest	0.997	8201.004167	663951600.0	25767.542022
k-Nearest Neighbor	0.96	47185.176091	19511660000.0	139666.572517

R^2, coefficient of determination; MAE, Mean Absolute Error; MSE, Mean Squared Error; RMSE, Root Mean Square Error

The results from the test set reflect the model's generalization ability and its performance on new data (Table 4). The minimal differences between the training and test results indicate that the model performs well on both the training data and general data, demonstrating its capability to make accurate predictions. Upon examining the test performances, the R^2 values for LR, RF, and k-NN were calculated as 0.98, 0.98, and 0.93, respectively. The highest R^2 value and the lowest error metrics (MAE, MSE, RMSE) were obtained with the RF algorithm, indicating higher accuracy. However, it is also observed that the LR algorithm exhibits a performance very close to this level of accuracy.

Table 4. Performance metrics of algorithms on the test set
Çizelge 4. Test setinde kullanılan algoritmaların performans metrikleri

Algorithms	R^2	MAE	MSE	RMSE
Linear Regression	0.98	46152.704444	4103773000.0	64060.696764
Random Forest	0.98	40239.739185	3867224000.0	62187.010216
k-Nearest Neighbor	0.93	84995.350154	13417670000.0	115834.650608

R^2, coefficient of determination; MAE, Mean Absolute Error; MSE, Mean Squared Error; RMSE, Root Mean Square Error

The alignment analysis between the actual and predicted values, as visualized in the fit plot, indicates that the predictions from the RF algorithm are closely aligned with the y = x line at a 45-degree angle (Figure 1). This strong alignment demonstrates a high level of accuracy in the RF predictions. Notably, while the RF algorithm consistently outperformed the other models in terms of overall accuracy and flexibility, it is important to highlight that the LR algorithm also exhibited commendable performance, with R^2 values being very close to that of RF. This proximity in results underscores the effectiveness of LR as a reliable alternative in predicting beef production.

The effectiveness of the machine learning algorithms used in the current research is consistent with findings from previous studies conducted in agricultural contexts. Li et al. (2018) evaluated the performance of three different machine learning algorithms—Random Forest (RF), Gradient Boosting Machine (GBM), and eXtreme Gradient Boosting (XGB)—in predicting genomic breeding values using SNP markers and body weight phenotypes in Brahman cattle. Among the three methods, RF and GBM were reported to consistently outperform XGB in terms of genomic prediction accuracy. Similarly, Maya Gopal and Bhargavi (2019) employed machine learning techniques to accurately predict crop yields. In their study, Artificial Neural Network (ANN), Support Vector Regression (SVR), k-Nearest Neighbors (k-NN), and RF algorithms were selected. Their results indicated that the RF algorithm achieved the highest accuracy, as determined by error analysis values.

Mishra et al. (2021) aimed to predict the most suitable agricultural crop to be grown in a specific region by using k-NN and RF machine learning algorithms along with data on soil quality, NPK values, moisture, and expected rainfall. They reported that the RF algorithm demonstrated superior accuracy. In another study, Bhardwaj et al. (2024) used machine learning algorithms such as Logistic Regression (LogR), XGB, CatBoost (CB), Gradient Boosting (GB), RF, and Support Vector Machine (SVM) for the prediction and classification of livestock diseases. Their findings indicated that RF and CB outperformed the other algorithms, with the RF algorithm achieving 83.56% accuracy, a precision and recall score of 0.84, and an F1 score of 0.82.

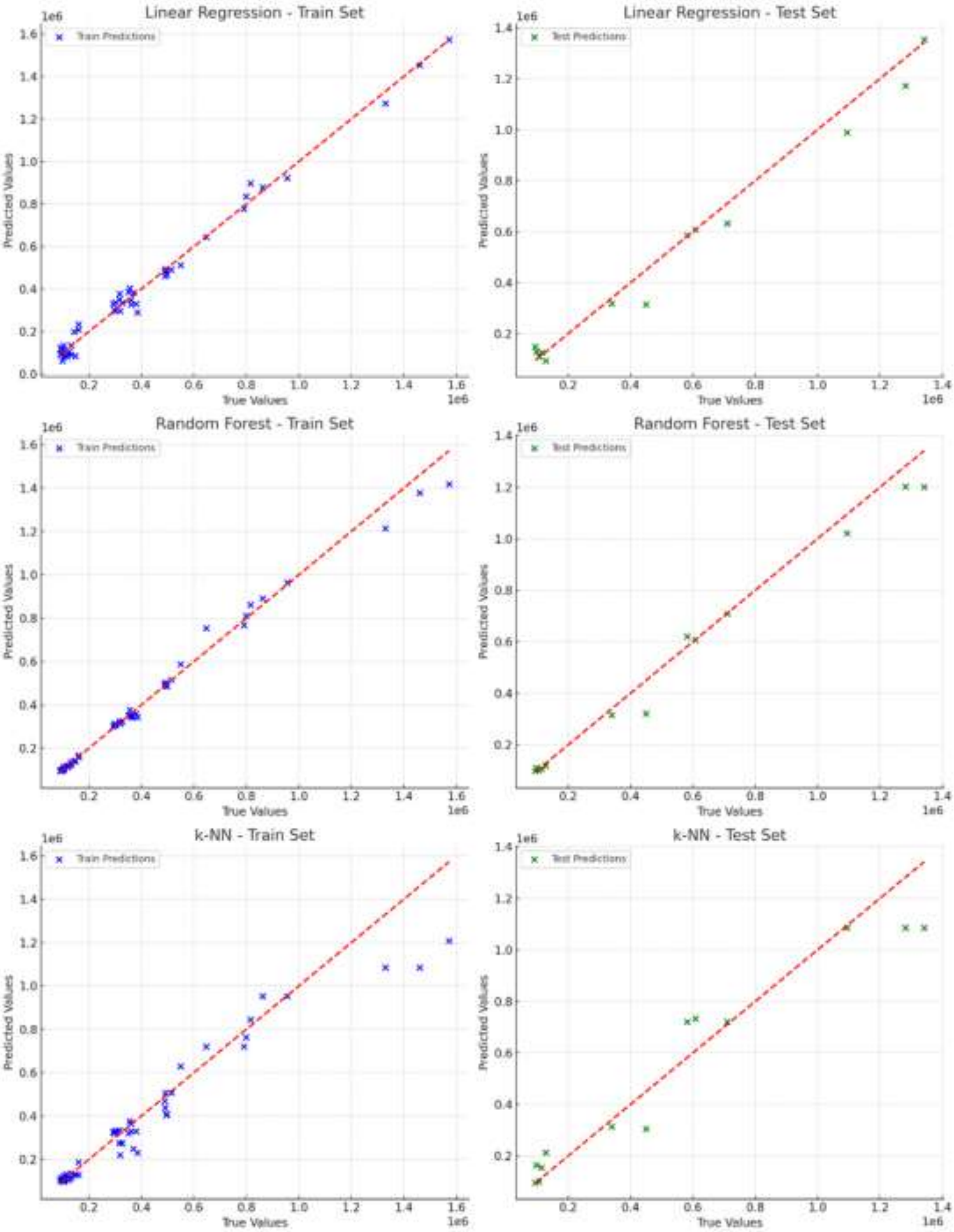

Figure 1. Alignment analysis between actual and predicted values
Şekil 1. Gerçek ve tahmin edilen değerler arasındaki uyum analizi

However, the performance of algorithms can vary across different studies. Yıldız et al. (2024) developed a model to predict honey production in Turkey using machine learning algorithms such as k-NN, RF, LR, and Gaussian Naive Bayes (GNB), based on data including honey production volume, number of farms, colony count, amount of pesticides used, agricultural and forest area, temperature, precipitation, and wind speed. Their results indicated that, among the four algorithms tested, LR was the most effective method for predicting honey production levels, with an R^2 value of 0.97. Similarly, Patel and Patel (2021) aimed to assist farmers in making informed crop selection decisions by developing a model to predict suitable crops for specific lands based on seasonal and soil parameters using popular supervised machine learning algorithms like SVM, k-NN, RF, and ANN. They found that the k-NN algorithm had superior performance metrics compared to the other approaches.

One possible reason for the poorer performance of the RF algorithm in these two studies could be its inherent random operational mechanism. As a tree-based method, RF trains each tree on a specific random subset, causing the model's performance to vary depending on certain characteristics of the data (Breiman, 2001). This randomness can sometimes lead to suboptimal results, particularly when the relationships among the dataset features are linear.

In studies that typically employ boosting algorithms, the RF algorithm has often lagged in terms of performance. For instance, Alsahaf et al. (2018) used four different machine learning algorithms—RF, Extremely Randomized Trees (ET), GBM, and XGB—to predict the age at which pigs reach a slaughter weight of 120 kg, and they reported that GBM and XGB, which are sequential ensemble methods, achieved lower error metrics than RF and ET. Similarly, Luo et al. (2021) applied three machine learning algorithms—Random Forest Regression (RFR), XGB, and CatBoost (CB)—to predict forest above-ground biomass (AGB). Their results showed that the CB algorithm outperformed both XGB and RFR in predicting AGB across all forest types. Additionally, it was noted that CB, unlike XGB, includes an algorithm to calculate leaf nodes when selecting a tree structure, which can help prevent overfitting.

In another study aimed at developing a machine learning-based prediction model for marine fish and aquaculture production, Rahman et al. (2021) found that the RFR algorithm performed worse than the Gradient Boosting Regression (GBR) algorithm. Ultimately, an ensemble approach called Voting Regression (VR) was used to combine these three machine learning algorithms, and the best performance metrics were achieved by VR ($R^2 = 0.81$). In a separate study, Srivastava et al. (2021) evaluated the performance of RF, EGB, and SVM algorithms in predicting carcass weight (CWT), marbling score (MS), backfat thickness (BFT), and eye muscle area (EMA) in Hanwoo cattle. They reported that EGB provided the lowest MSE for CWT and MS, while SVM yielded the lowest MSE for BFT and EMA.

Additionally, Coşkun et al. (2023) compared the performance of EGB, RF, and Bayesian Regularized Neural Network (BRNN) data mining algorithms in predicting the live weights of Anatolian Merinos lambs using body trait data collected at the onset of the fattening period. Their findings indicated that the XGB algorithm produced better results than the RF and BRNN algorithms across several performance metrics, including RMSE, standard deviation ratio (SDR), mean absolute percentage error (MAPE), and adjusted coefficient of determination (R^2).

To enhance the interpretability of the model and aid in the optimization of beef production strategies, a feature importance analysis was conducted. The analysis results are presented in Figure 2, where each feature is represented by columns that indicate a specific level of importance based on its contribution to the model. The height of the columns reflects the impact of these features on the model's performance; thus, taller columns indicate that the corresponding feature contributes more significantly to predictive power. The analysis revealed that the attributes with the highest impact were population, corn production, and urban population, each with importance scores exceeding 0.15. These high-importance scores indicate that these are the most critical factors influencing beef production. Notably, population had the highest impact on the model with a score above 0.25, highlighting its central role. Increasing population necessitates a direct increase in beef production due to higher demand (FAO, 2024a; Godfray et al., 2018). This underscores the significance of population as a key driver in beef production, suggesting that production strategies should be aligned with population growth considerations.

Corn production was also identified as a significant factor due to its role as a primary feed source in cattle nutrition (Klopfenstein et al., 2013). The high impact of corn on the model indicates a direct influence on the amount of beef production. Therefore, increasing corn production could potentially enhance cattle nutrition, thereby boosting beef output. Urban population was another attribute with a substantial impact, reflecting the changing dietary habits and increased demand for high-quality protein sources associated with urbanization (FAO, 2024a). This trend correlates with the rising demand for beef, which is a primary source of high-quality protein.

These findings illustrate the complex interrelationships among various factors affecting beef production, emphasizing the need for strategic planning that considers these influential attributes.

CONCLUSION

In this study, the effectiveness of LR, k-NN, and RF algorithms in predicting beef production in Türkiye was examined. The findings indicate that the RF algorithm outperformed the other algorithms, demonstrating higher performance with superior R^2 values and lower error rates. The RF algorithm was also effective in evaluating the relative importance of features, enhancing model interpretability, and providing more accurate and reliable predictions for future beef production.

Ensuring food security and planning agricultural policies sustainably will become increasingly important in the coming years. In this context, the high accuracy provided by the RF algorithm could play a critical role in predicting beef production outcomes and addressing potential food security challenges. Given factors such as population growth, climate change, and the reduction of agricultural lands, such predictions can serve as strategic tools in shaping agricultural policies and preventing future food deficits. The results of this study demonstrate the applicability of the RF algorithm in agricultural data analysis and production prediction, offering forward-looking solutions for agricultural sustainability.

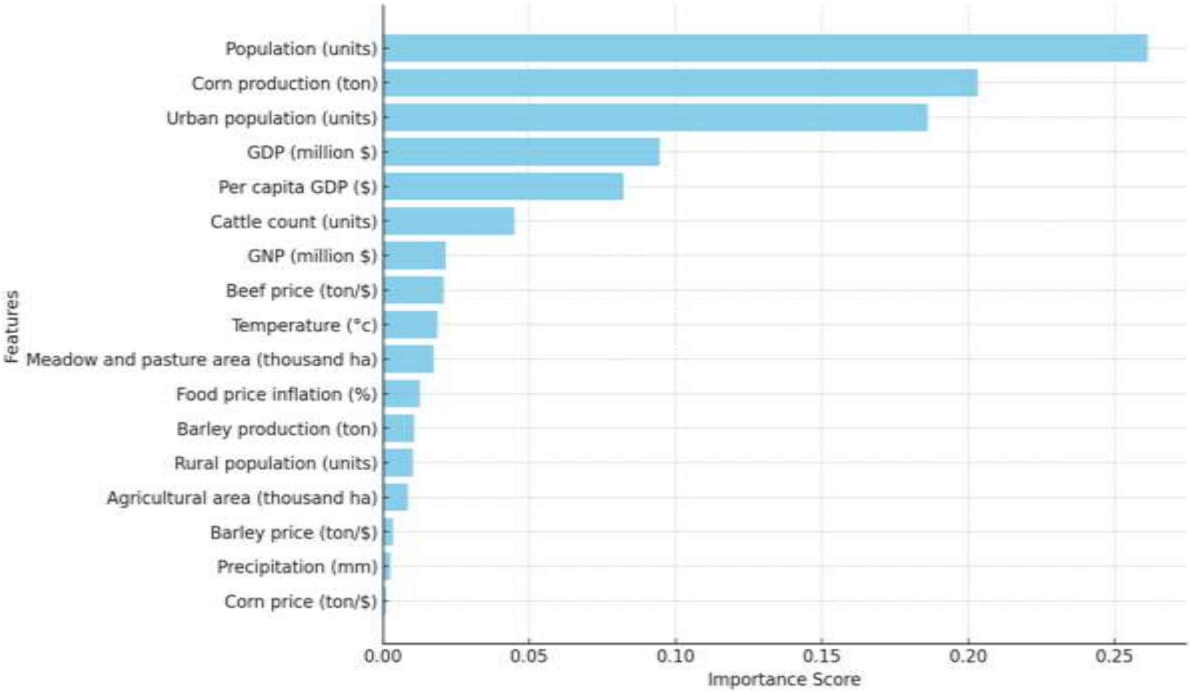

Figure 2. Feature importance analysis for predicting beef production

Şekil 2. Sığır eti üretimini tahmin etmek için özellik önem analizi

Contribution Rate Statement Summary of Researchers

The authors declare that they have contributed equally to the article.

Conflict of Interest

The authors have declared no conflict of interest.

REFERENCES

Ahmed, M. U. & Hussain, I. (2022). Prediction of wheat production using machine learning algorithms in northern areas of Pakistan. *Telecommunications Policy, 46,* 102370. https://doi.org/10.1016/j.telpol.2022.102370.

Alonso, J., Castañón, Á. R. & Bahamonde, A. (2013). Support Vector Regression to predict carcass weight in beef cattle in advance of the slaughter. *Computers and Electronics in Agriculture, 91,* 116-120. https://doi.org/10.1016/j.compag.2012.08.009.

Alsahaf, A., Azzopardi, G., Ducro, B., Veerkamp, R. F. & Petkov, N. (2018). Predicting slaughter weight in pigs with regression tree ensembles. *Frontiers in Artificial Intelligence and Applications, 310,* 1-9. https://doi.org/10.3233/978-1-61499-929-4-1.

Bharadiya, J. P., Tzenios, N. T. & Reddy, M. (2023). Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches. *Journal of Engineering Research and Reports, 24*(12), 29-44.

Bhardwaj, P., Kumar, S. J. K. J., Kanna, G. P. & Mithila, A. (2024). Machine learning-based approaches for livestock symptoms and diseases prediction and classification. *International Conference on Communication, Computer Sciences and Engineering (IC3SE),* Gautam Buddha Nagar, India, 2024, pp. 1-6.

Breiman, L. (2001). Random forests. *Machine Learning, 45,* 5-32. https://doi.org/10.1023/A:1010933404324.

Ching, X. L., Zainal, N. A. A. B., Luang-In, V. & Ma, N. L. (2022). Lab-based meat: The future food. *Environmental Advances, 10,* 100315. https://doi.org/10.1016/j.envadv.2022.100315.

Coşkun, G., Şahin, Ö., Altay, Y. & Aytekin, İ. (2023). Final fattening live weight prediction in Anatolian merinos lambs from some body characteristics at the initial of fattening by using some data mining algorithms. *Black Sea Journal of Agriculture, 6*(1), 47-53. https://doi.org/10.47115/bsagriculture.1181444.

Çakan, V. A. & Tipi, T. (2023). How does the change in feed prices affect meat prices? A case study of Turkey.

Atatürk Üniversitesi Ziraat Fakültesi Dergisi, 54(2), 68-74. https://doi.org/10.5152/AUAF.2023.22054.

FAO. (2024a). Global and regional food consumption patterns and trends. *Food and Agriculture Organization of the United Nations*. https://www.fao.org/4/ac911e/ac911e05.htm. (Accessed Date: 12 November 2024).

FAO. (2024b). FAOSTAT, demographic and economic and political stability data of Türkiye. https://www.fao.org/faostat/en/#country/223. (Accessed Date: 4 November 2024).

GDM. (2024). General Directorate of Meteorology, official climate statistics of Türkiye. https://www.mgm.gov.tr/veridegerlendirme. (Accessed Date: 4 November 2024).

Godfray, H. C. J., Aveyard, P., Garnett, T., Hall, J. W., Key, T. J., Lorimer, J., Pierrehumbert, R. T., Scarborough, P., Springmann, M. & Jebb, S. A. (2018). Meat consumption, health, and the environment. *Science, 361*(6399), eaam5324. https://doi.org/10.1126/science.aam5324.

Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M. & Tiwari, B. (2017). Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods, 6*(7), 53. https://doi.org/10.3390/foods6070053.

Humer, E. & Zebeli, Q. (2017). Grains in ruminant feeding and potentials to enhance their nutritive and health value by chemical processing. *Animal Feed Science and Technology, 226,* 133-151. https://doi.org/10.1016/j.anifeedsci.2017.02.005.

Klopfenstein, T. J., Erickson, G. E. & Berger, L. L. (2013). Maize is a critically important source of food, feed, energy and forage in the USA. *Field Crops Research, 153,* 5-11. https://doi.org/10.1016/j.fcr.2012.11.006.

Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine, 23,* 89-109. https://doi.org/10.1016/s0933-3657(01)00077-x.

Li, C. (2017). The role of beef in human nutrition and health. In M. Dikeman (Ed.), *Ensuring safety and quality in the production of beef* (pp. 1-10). Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2016.0009.16.

Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S. & Sun, Y. (2021). Combination of feature selection and catboost for prediction: The first application to the estimation of aboveground biomass. *Forests, 12*(2), 216. https://doi.org/10.3390/f12020216.

Marshall, B. M. & Levy, S. B. (2011). Food animals and antimicrobials: Impacts on human health. *Clinical Microbiology Reviews, 24*(4), 718-733.

Mishra, T. K., Mishra, S. K., Sai, K. J., Alekhya, B. S. & Nishith, A. R. (2021). Crop recommendation system using KNN and random forest considering Indian data set. *19th OITS International Conference on Information Technology (OCIT),* Bhubaneswar, India, December 2021, pp. 308-312.

Nosratabadi, S., Ardabili, S., Lakner, Z., Mako, C. & Mosavi, A. (2021). Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. *Agriculture, 11*(5), 408. https://doi.org/10.3390/agriculture11050408.

Patel, K. & Patel, H. B. (2021). A comparative analysis of supervised machine learning algorithm for agriculture crop prediction. *Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT),* Tamil Nadu, India, September 2021, pp. 1-5.

Python Software Foundation. (2024). The Python language version 3.12.2.

Rahman, L. F., Marufuzzaman, M., Alam, L., Bari, M. A., Sumaila, U. R. & Sidek, L. M. (2021). Developing an ensembled machine learning prediction model for marine fish and aquaculture production. *Sustainability, 13,* 9124. https://doi.org/10.3390/su13169124.

Ruxton, C. H. S. & Gordon, S. (2024). Animal board invited review: The contribution of red meat to adult nutrition and health beyond protein. *Animal, 18*(3). https://doi.org/10.1016/j.animal.2024.101103.

Srivastava, S., Lopez, B. I., Kumar, H., Jang, M., Chai, H. H., Park, W., Park, J. E. & Lim, D. (2021). Prediction of Hanwoo cattle phenotypes from genotypes using machine learning methods. *Animals, 11*(7), 2066. https://doi.org/10.3390/ani11072066.

Srutee, R. Sowmya, R. S. & Annapure, U. S. (2022). Clean meat: Techniques for meat production and its upcoming challenges. *Animal Biotechnology, 33*(7), 1721-1729. https://doi.org/10.1080/10495398.2021.1911810.

TMAF. (2024). Turkish Ministry Agriculture and Forestry, agricultural data information center. https://www.tarimorman.gov.tr/Konular/ (Accessed date: 4 November 2024).

Van Kernebeek, H. R., Oosting, S. J., Van Ittersum, M. K., Bikker, P., & De Boer, I. J. (2016). Saving land to feed a growing population: Consequences for consumption of crop and livestock products. *The International Journal of Life Cycle Assessment, 21*(5), 677-687. https://doi.org/10.1007/s11367-015-0923-6.

Yıldız, B. İ., Eskioğlu, K. & Karabağ, K. (2024). Developing a machine learning prediction model for honey production. *Mediterranean Agricultural Sciences, 37*(2), 105-110. https://doi.org/10.29136/mediterranean.1511697.