



Article

Econometric Modeling of Food Industry Development using Regression and Random Forest: Evidence from Kashkadarya Region, Uzbekistan

Hasan Abduxalimovich Isokov*¹

1. Independent Researcher, Karshi State Technical University, Karshi, Uzbekistan

*Corresponding: h37077866@gmail.com

Abstract: This study examines the development of the food industry in the Kashkadarya region of Uzbekistan by integrating classical regression analysis with the Random Forest (RF) algorithm. The research aims to identify the key economic factors influencing food industry development, evaluate their relationships with production performance, and compare the explanatory and predictive capabilities of conventional econometric and machine learning approaches. Multiple regression analysis is applied to estimate the direction, magnitude, and statistical significance of factor effects, whereas Random Forest is employed to capture potential nonlinear relationships and assess the relative importance of predictors. The performance of the models is evaluated using R^2 , MAE, and RMSE metrics. Using annual data for 2011–2024, the estimated regression model explains approximately 96.8% of the variation in food production ($R^2 = 0.9677$; adjusted $R^2 = 0.9580$; $F = 99.86$). Investment in food production (coefficient = 22.5863, $p = 0.0016$) and the number of food manufacturing enterprises (coefficient = 693.7199, $p = 0.0313$) are statistically significant, whereas employment is not ($p = 0.4507$). In the Random Forest model, employment, investment, and the number of enterprises account for approximately 40.4%, 33.8%, and 25.8% of feature importance, respectively, suggesting nonlinear effects or interactions not captured by the linear specification. The findings provide a methodological basis for identifying the major drivers of regional food industry development, improving forecasting accuracy, and supporting evidence-based decisions aimed at strengthening investment activity, production capacity, and the sustainable development of the food industry in the Kashkadarya region.

Citation: Isokov, H. A. Econometric Modeling of Food Industry Development using Regression and Random Forest: Evidence from Kashkadarya Region, Uzbekistan. American Journal of Economics and Business Management 2026, 9(10), 111-122.

Received: 15th Aug 2026

Revised: 05th Sept 2026

Accepted: 20th Sept 2026

Published: 06th Oct 2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

Keywords: Food Industry, Econometric Modeling, Regression Analysis, Random Forest, Machine Learning, Investment, Employment, Production Volume, Regional Development, Forecasting

1. Introduction

The food industry is one of the strategically important sectors of the regional economy, as it plays a significant role in ensuring food security, creating employment, increasing value added, processing agricultural raw materials, and improving the competitiveness of regional production. In regions with substantial agricultural potential, the sustainable development of food production is closely associated with the efficient use of local resources, investment activity, expansion of production capacities, technological modernization, and the development of entrepreneurship. In this regard, identifying the factors that determine food industry development and quantitatively assessing their effects are important tasks of contemporary economic research.

Kashkadarya region is one of Uzbekistan's major agricultural regions and possesses considerable potential for the production and processing of agricultural products. The availability of agricultural raw materials creates favorable conditions for expanding food processing, increasing the production of higher-value-added products, and strengthening intersectoral linkages between agriculture and manufacturing. At the same time, the effective utilization of this potential requires the expansion of investment, modernization of production facilities, improvement of labor productivity, and development of food-producing enterprises.

Traditional econometric methods are widely used to analyze the relationships between food production and the economic factors affecting its development. In particular, multiple regression analysis makes it possible to estimate the direction and magnitude of the relationship between a dependent variable and a set of explanatory variables. Regression coefficients, statistical significance tests, and goodness-of-fit measures provide a relatively transparent framework for interpreting economic relationships. Therefore, regression models remain important instruments for evidence-based analysis and decision-making in regional economic studies.

At the same time, linear regression requires the functional form of the relationships to be specified in advance and may not fully capture nonlinear effects and interactions among explanatory factors. Machine learning methods, and the Random Forest algorithm in particular, provide a flexible data-driven alternative for detecting such patterns and assessing the predictive importance of explanatory variables [1], [2].

Nevertheless, the application of Random Forest should not be regarded as a replacement for econometric analysis. Regression analysis and Random Forest address partly different research objectives. Regression models are particularly useful for parameter interpretation and statistical inference under their assumptions, whereas Random Forest is primarily oriented toward flexible prediction and the detection of complex patterns in data. Consequently, combining the two approaches can provide a more comprehensive analytical framework for examining regional food industry development.

From this perspective, comparing regression and Random Forest models using the same dataset is of particular methodological interest. Regression analysis can reveal the direction, magnitude, and statistical significance of relationships between food production and selected economic factors. Random Forest can subsequently be used to investigate whether nonlinear structures and interactions improve out-of-sample predictive performance and to assess the relative predictive importance of the explanatory variables. The models can then be compared using common performance indicators such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE).

The contribution of the study lies in integrating the interpretability of regression-based econometric analysis with the flexible predictive capacity of Random Forest within a unified empirical framework for regional food industry analysis. Rather than treating econometric and machine learning approaches as competing methodologies, the study considers their complementary strengths. The resulting framework can contribute to a more comprehensive assessment of food industry development and provide an analytical basis for investment, production, and regional development decisions in Kashkadarya and comparable regions of emerging economies.

Accordingly, the aim of this study is to identify the key economic factors associated with food industry development in the Kashkadarya region and to compare the explanatory and predictive capabilities of multiple regression and Random Forest. The remainder of the paper is organized as follows. The rest of Section 1 reviews the relevant literature; Section 2 describes the data and methods; Section 3 presents the empirical

results; Section 4 discusses the findings and limitations; and Section 5 concludes and formulates recommendations.

Literature Review

The methodological foundations of the present study are based on the integration of traditional econometric modeling, modern machine learning techniques, and contemporary approaches to food industry development. The existing literature demonstrates that econometric and machine learning methods should not necessarily be regarded as competing approaches. Instead, their complementary application can improve both the economic interpretation of relationships and the predictive performance of empirical models.

One of the fundamental methods used in this study is the Random Forest (RF) algorithm introduced by Breiman. Random Forest is an ensemble learning method based on constructing multiple decision trees and aggregating their predictions. Its major advantages include the ability to capture nonlinear relationships and complex interactions among explanatory variables, relative robustness to overfitting compared with individual decision trees, and the possibility of assessing variable importance [1]. These characteristics make RF particularly relevant for modeling food industry development, where production outcomes may depend on complex interactions among investment, employment, enterprise activity, and technological factors.

The increasing availability of large and multidimensional datasets has expanded the methodological boundaries of econometric research. Varian argues that modern data environments create opportunities to complement traditional econometric tools with machine learning techniques, particularly when researchers face nonlinear relationships, numerous potential predictors, and complex interactions [3]. From this perspective, machine learning can be considered an extension of the analytical toolkit available to applied economists rather than a replacement for econometric reasoning.

The distinction between explanation and prediction has also received considerable attention in the economic literature. Kleinberg et al. show that many economic and policy problems contain an important prediction component and that improvements in predictive methods can contribute to better decision-making [4]. This argument is relevant to regional food industry analysis, where identifying influential factors and obtaining reliable predictions of production dynamics represent related but methodologically distinct objectives.

Mullainathan and Spiess provide one of the most important links between econometrics and machine learning. They argue that machine learning methods are particularly effective for prediction problems, whereas traditional econometric methods are generally designed around parameter estimation and statistical inference [5]. Accordingly, the combination of regression analysis and Random Forest used in the present research makes it possible to examine food industry development from two perspectives: the regression model provides interpretable estimates of statistical relationships, while Random Forest provides a flexible framework for detecting nonlinear patterns and improving predictive performance.

Athey and Imbens develop a similar argument in their reviews of applied econometrics and of the machine learning methods most relevant to economists, emphasizing that prediction-oriented algorithms and tools for causal and policy analysis serve different purposes and can be used in a complementary manner [6], [7].

The integration of machine learning and econometrics has become an independent research direction in recent years. Kreinovich, Sriboonchitta, and Yamaka examine various possibilities for applying machine learning techniques to econometric and related problems and demonstrate the potential benefits of combining data-driven algorithms

with conventional quantitative economic analysis [8]. Their work provides a contemporary methodological basis for integrated econometric–machine learning frameworks.

Empirical research also demonstrates the potential advantages of hybrid modeling. Kristjanpoller applies a combined econometric and machine-learning framework to modeling realized volatility in the natural gas market, illustrating that hybrid approaches can exploit information that may not be fully captured by a single modeling technique [9]. Although the empirical field differs from food manufacturing, the methodological principle is directly relevant to the present study: econometric models can be complemented by flexible machine-learning algorithms to analyze complex economic dynamics.

Recent macroeconomic forecasting research provides further evidence regarding the applicability of machine learning to economic data. Kant, Pick, and de Winter investigate machine learning methods in GDP nowcasting and demonstrate the potential of such algorithms for processing complex information and improving short-term economic assessment [10]. Their results reinforce the argument that machine learning techniques, including tree-based methods, can complement traditional quantitative approaches when economic relationships are complex or nonlinear.

Research on the food sector itself also points to the importance of technological upgrading. A review of empirical studies on food process innovation summarizes the current state of evidence and outlines a future research agenda for process innovation in food manufacturing [11], which supports the inclusion of investment and enterprise-level capacity among the determinants of food production.

For the regional context, Khudoyorov investigates the dynamic modeling of sustainable development of small business and entrepreneurship and emphasizes the role of quantitative modeling in identifying development patterns and supporting regional economic decision-making [12]. Although the study focuses on small business and entrepreneurship, its regional modeling perspective is relevant to the analysis of food industry development in Kashkadarya, where entrepreneurial activity, investment, employment, and production capacity are closely interconnected.

Therefore, the present study contributes to the literature by combining multiple regression analysis and Random Forest for the analysis of food industry development in the Kashkadarya region of Uzbekistan. The regression model is used to assess the direction, magnitude, and statistical significance of relationships between food production and its selected determinants, while Random Forest is employed to capture potential nonlinearities and evaluate the predictive importance of the same factors. Comparing the models through common performance measures provides a methodological basis for determining their relative strengths and for developing a more comprehensive framework for regional food industry analysis.

2. Materials and Method

This study employed econometric and machine learning methods to analyze the dynamics of food industry development in the Kashkadarya region and to identify the key economic factors influencing food production. The methodological framework of the study is based on the integration of multiple regression analysis and the Random Forest (RF) algorithm, which makes it possible to combine the explanatory capabilities of traditional econometric methods with the predictive advantages of machine learning.

2.1. Data and variables

At the initial stage of the study, statistical data characterizing the development of the food industry in the Kashkadarya region were collected and systematized. The volume of food production was selected as the dependent variable (Y), while employment in the food industry (X_1), investment in food production (X_2), and the number of legal entities engaged in food manufacturing (X_3) were considered explanatory variables. Based on these data,

the dynamics, growth rates, and general development trends of the regional food industry were analyzed using economic and statistical methods [13].

The analysis covers annual observations for 2011–2024 ($n = 14$). The variables included in the econometric and machine learning models are summarized in Table 1.

Table 1. Variables Used in the Study

Symbol	Variable	Role in the model
Y	Volume of food production	Dependent variable
X ₁	Employment in the food industry	Explanatory variable
X ₂	Investment in food production	Explanatory variable
X ₃	Number of legal entities engaged in food manufacturing	Explanatory variable

2.2. Multiple regression model

At the next stage, a multiple regression model was developed to assess the relationships between food production and the selected economic factors. The model parameters were estimated using the Ordinary Least Squares (OLS) method. The statistical significance of individual coefficients and the overall adequacy of the model were evaluated using the t-test, F-test, coefficient of determination (R^2), and adjusted R^2 .

The empirical specification of the multiple regression model is:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t \quad (1)$$

where t denotes the year, β_0 is the intercept, β_1 , β_2 , and β_3 are the regression coefficients of the explanatory variables, and ε_t is the random error term [13]. Because the model is estimated on observational data, the coefficients are interpreted as conditional statistical associations rather than as causal effects [14].

At the first stage, multiple regression analysis was used to estimate the relationship between food production volume (Y) and the selected explanatory variables: employment in the food industry (X₁), investment in food production (X₂), and the number of legal entities engaged in food manufacturing (X₃). The regression approach makes it possible to determine the direction and magnitude of the conditional relationship between each factor and food production while holding the other included variables constant. The statistical significance of the estimated coefficients should be evaluated using the t-statistics and corresponding probability values, while the overall adequacy of the model is assessed through the F-statistic and R^2 .

2.3. Random Forest algorithm

In the following stage, the Random Forest algorithm was applied to the same set of variables. This method was used to identify possible nonlinear relationships among the indicators and determine the relative predictive importance of employment, investment, and enterprise activity in explaining food industry development [1]. The performance of the regression and Random Forest models was compared using R^2 , Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

A Random Forest regression model averages the predictions of B decision trees, each grown on a bootstrap sample of the data with a random subset of predictors considered at each split [1], [2], [15]:

$$\hat{Y}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2)$$

where $T_b(x)$ is the prediction of the b -th regression tree for the vector of explanatory variables x . The relative importance of each predictor is assessed by its contribution to the predictive performance of the ensemble [1], [15].

2.4. Model evaluation criteria

The predictive accuracy of the models was assessed using the coefficient of determination (R^2), the mean absolute error (MAE), and the root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (5)$$

where Y_t is the observed value of food production, $\hat{Y}_t$ is the predicted value, $\bar{Y}$ is the sample mean, and n is the number of observations. A higher R^2 and lower MAE and RMSE indicate better model performance [15].

The obtained results provide a methodological basis for comparing traditional econometric and Random Forest approaches, identifying the key factors associated with food industry development, and improving the analytical framework for evidence-based decision-making in the food industry of the Kashkadarya region.

3. Results

3.1. Multiple regression estimates

The estimated multiple regression model demonstrates a high overall explanatory capacity, with an R^2 value of 0.9677 and an adjusted R^2 of 0.9580. This indicates that approximately 96.8% of the variation in food production observed within the sample is associated with changes in employment, investment, and the number of food manufacturing enterprises. The overall model is statistically significant, as confirmed by an F-statistic of 99.86.

Among the explanatory variables, investment in food production (X_2) demonstrates the strongest statistical evidence, with a coefficient of 22.5863 and a p-value of 0.0016. The number of food manufacturing enterprises (X_3) is also statistically significant at the 5% level, with a coefficient of 693.7199 and a p-value of 0.0313. Employment (X_1), despite its positive coefficient, is not statistically significant in the multiple regression specification ($p = 0.4507$).

The main regression estimates are summarized in Table 2.

Table 2. Multiple Regression Estimates for Food Production (Y)

Explanatory variable	Coefficient	p-value	Significance (5% level)
Employment in the food industry (X_1)	Positive	0.4507	Not significant
Investment in food production (X_2)	22.5863	0.0016	Significant
Number of food manufacturing enterprises (X_3)	693.7199	0.0313	Significant
$R^2 = 0.9677$; adjusted $R^2 = 0.9580$; F-statistic = 99.86			

3.2. Random Forest feature importance

The Random Forest analysis provides a complementary perspective. The preliminary feature-importance estimates assign approximately 40.4% of predictive importance to employment, 33.8% to investment, and 25.8% to the number of enterprises.

Table 3. Random Forest Feature Importance of the Explanatory Variables

Explanatory variable	Feature importance, %	Rank
Employment in the food industry (X_1)	40.4	1
Investment in food production (X_2)	33.8	2
Number of food manufacturing enterprises (X_3)	25.8	3
Total	100.0	–

4. Discussion

The difference between the OLS and Random Forest results is methodologically important. In particular, employment is not statistically significant in the linear regression model but appears to contain substantial predictive information in the Random Forest model. This may indicate nonlinear relationships or interactions that are not explicitly represented in the linear regression specification.

Consequently, the results demonstrate the value of integrating econometric regression with Random Forest. Regression provides interpretable coefficients and statistical inference, whereas Random Forest provides additional information on nonlinear predictive patterns and variable importance. Their combined application therefore offers a broader analytical framework for assessing food industry development in the Kashkadarya region.

The analysis of food industry development in the Kashkadarya region shows that the sector has experienced significant changes in production volume, investment activity, employment, and the number of operating enterprises. The observed dynamics indicate that food industry growth is determined not by a single factor but by the combined influence of labor resources, capital investment, and entrepreneurial activity. Therefore, the use of both regression analysis and the Random Forest algorithm provides a broader methodological basis for assessing these relationships.

The economic interpretation of the model is particularly important. Employment reflects the labor potential of the industry, while investment represents opportunities for expanding production capacity, introducing modern equipment, and improving technological processes. The number of operating enterprises characterizes the level of entrepreneurial activity and the expansion of the regional production base. Consequently, changes in these indicators may directly or indirectly contribute to the dynamics of food production.

However, the relationship between these factors and food production may not always be linear. For example, an increase in investment may generate a stronger production effect after new capacity becomes operational, while employment growth may produce different outcomes depending on labor productivity and technological conditions. For this reason, the Random Forest algorithm was used as a complementary modeling approach capable of detecting nonlinear relationships and interactions that may not be fully captured by the linear regression specification [1].

Random Forest also makes it possible to assess the relative predictive importance of the explanatory variables. This provides additional information on which indicators contribute most strongly to the model's predictions of food production. In this context, RF feature importance complements rather than replaces regression coefficients: regression coefficients describe conditional statistical relationships, whereas Random Forest importance reflects the contribution of variables to predictive performance.

The comparative assessment of the two models is based on R^2 , MAE, and RMSE. A model with a higher out-of-sample R^2 and lower MAE and RMSE can be considered more accurate from a predictive perspective. Nevertheless, predictive accuracy alone is

insufficient for selecting the most appropriate model. Regression analysis offers greater transparency and economic interpretability, whereas Random Forest provides greater flexibility in capturing complex and nonlinear patterns.

The results therefore support an integrated modeling framework rather than treating regression and Random Forest as mutually exclusive alternatives. Regression analysis can first be used to establish economically interpretable relationships among the variables, after which Random Forest can test whether nonlinearities and interactions provide additional predictive information. This combination strengthens both the explanatory and predictive dimensions of regional food industry analysis.

Thus, the comparison can be summarized as follows:

Table 4. Comparative Characteristics of the Regression Model and Random Forest Algorithm

Criterion	Regression model	Random Forest
Economic interpretation	High	Moderate
Statistical inference	High	Limited
Detection of nonlinear relationships	Limited without explicit specification	High
Interaction effects	Must generally be specified	Captured automatically
Variable assessment	Coefficients, <i>t</i> -tests, <i>p</i> -values	Feature importance
Predictive flexibility	Moderate	High
Main advantage	Explanation	Prediction and nonlinear modeling

Overall, the findings suggest that the development of the Kashkadarya food industry should be analyzed through the interconnected effects of investment, employment, and enterprise activity. The combined use of regression and Random Forest provides a more comprehensive analytical framework for identifying the key drivers of food production and can support evidence-based decisions concerning investment allocation, production capacity, and regional food industry development.

The comparative characteristics presented in Table 4 indicate that regression analysis and Random Forest address different but complementary dimensions of economic modeling. From an econometric perspective, the regression model provides a structured framework for evaluating conditional relationships between food production and its determinants. Random Forest, in contrast, provides additional flexibility by allowing the response of food production to explanatory variables to vary across different regions of the predictor space.

An important methodological issue in food industry modeling is the possibility of nonlinear and threshold effects. The effect of investment on food production, for example, may not be proportional at all levels of investment. At relatively low levels, additional investment may have a limited immediate effect because of insufficient production capacity, infrastructure, or complementary resources. Once a certain level of capital accumulation is reached, however, investment may generate stronger production effects through technological modernization and capacity expansion. Such relationships can be difficult to represent with a conventional linear specification unless the relevant nonlinear terms are explicitly introduced.

A similar argument applies to employment. An increase in the number of employees does not necessarily result in a proportional increase in food production. The economic effect of additional labor depends on productivity, capital intensity, technological

equipment, employee qualifications, and the utilization of existing production capacity. Consequently, Random Forest can be useful for examining whether the relationship between employment and production varies across different economic conditions.

The number of food manufacturing enterprises also deserves particular attention. Growth in the number of enterprises may contribute to higher production through increased competition, entrepreneurial activity, and processing capacity. However, the economic effect depends not only on the number of enterprises but also on their size, technological level, productivity, and access to investment resources. Therefore, a larger number of firms cannot automatically be interpreted as a proportional increase in sectoral output.

Another important aspect of the analysis concerns the relative importance of factors. In the regression model, the importance of explanatory variables is evaluated through estimated coefficients, standard errors, t-statistics, and p-values. In Random Forest, variable importance is evaluated according to the contribution of each predictor to the model's predictive performance. Therefore, the two approaches provide different types of information and should be interpreted jointly rather than directly equated.

This distinction makes it possible to construct a more informative classification of the factors:

Table 5. Economic Interpretation of Factors Used in Regression and Random Forest Models

Factor	Economic role	Regression assessment	Random Forest assessment	Expected contribution
Employment (X_1)	Labor resources and production potential	Coefficient and statistical significance	Predictive importance	Expansion of productive capacity
Investment (X_2)	Capital formation and modernization	Marginal conditional relationship	Nonlinear and predictive importance	Technological upgrading and output growth
Number of enterprises (X_3)	Entrepreneurial and production activity	Conditional relationship with output	Importance and interactions	Expansion of the regional production base

The methodological value of this classification lies in separating economic effect, statistical association, and predictive importance. A variable may have a statistically significant regression coefficient but contribute only modestly to out-of-sample prediction. Conversely, a variable may be highly useful for Random Forest prediction because of nonlinearities and interactions even when its linear regression coefficient is not statistically significant. Such differences should not be interpreted as contradictions; rather, they reflect the different objectives of the two modeling approaches.

From an applied perspective, the combined approach can provide useful information for regional food industry policy. If investment is identified as both statistically relevant in the regression model and highly important in Random Forest prediction, this would provide stronger empirical support for prioritizing capital investment and technological modernization. If employment is important mainly in interaction with investment or enterprise activity, policies aimed solely at increasing employment may be less effective than coordinated measures that simultaneously improve technology, capital availability, and labor productivity.

4.1. Limitations and directions for future research

The empirical results should be interpreted in light of the relatively small annual sample covering 2011–2024. Future research should extend the dataset, where possible, by using longer annual series or quarterly observations and additional indicators of technological modernization, innovation, productivity, exports, and digitalization. Out-of-sample and time-series validation should also be applied to provide a more robust comparison of the predictive performance of regression and Random Forest models.

5. Conclusion

The study examined the development of the food industry in the Kashkadarya region by combining multiple regression analysis with the Random Forest algorithm. The results confirm that the integration of traditional econometric methods and modern data-driven approaches provides a broader analytical framework for identifying the factors associated with regional food production.

The estimated multiple regression model demonstrated a high explanatory capacity, with an R^2 of 0.9677 and an adjusted R^2 of 0.9580. Among the selected factors, investment in food production (X_2) showed the strongest statistical evidence, with a coefficient of 22.5863 and a p-value of 0.0016. The number of food manufacturing enterprises (X_3) was also statistically significant at the 5% level, with a coefficient of 693.7199 and a p-value of 0.0313. Employment (X_1) had a positive coefficient but was not statistically significant in the multiple regression specification.

The Random Forest analysis provided a complementary perspective on the predictive relevance of the selected variables. The preliminary feature-importance results indicated that employment accounted for approximately 40.4%, investment for 33.8%, and the number of enterprises for 25.8% of total feature importance. In particular, the relatively high predictive importance of employment, despite its statistical insignificance in the linear regression model, suggests that its relationship with food production may involve nonlinearities or interactions with other factors. However, Random Forest feature importance should be interpreted as predictive relevance rather than as a causal economic effect.

Based on the findings, several recommendations can be proposed. First, investment activity in the food industry should be strengthened, with particular attention to technological modernization, expansion of processing capacity, energy-efficient equipment, and advanced production technologies. The regression results indicate that investment is one of the most statistically relevant factors associated with food production growth.

Second, policies aimed at increasing the number and productive capacity of food manufacturing enterprises should be supported. Improving access to financing, production infrastructure, modern technologies, logistics, and regional markets can contribute to strengthening the industrial base and increasing value added from locally produced agricultural raw materials.

Third, regional food industry analysis should increasingly combine econometric and Random Forest approaches. Regression models can be used to estimate interpretable statistical relationships and test economic hypotheses, while Random Forest can complement them by detecting nonlinear patterns, interactions, and predictive importance. Such integration can improve the information base available for regional economic decision-making.

Based on the results of the study, the following conclusions were drawn:

1. The development of the food industry in the Kashkadarya region is influenced by the combined effects of employment, investment, and the number of food manufacturing

enterprises. The results confirm the importance of applying a multi-factor approach to the analysis of regional food production.

2. The estimated multiple regression model demonstrated a high explanatory capacity, with an R^2 of 0.9677 and an adjusted R^2 of 0.9580. Thus, approximately 96.8% of the variation in food production within the analyzed sample is explained by the variables included in the model.
3. Investment in food production (X_2) was found to be one of the most statistically relevant factors. Its estimated coefficient was 22.5863, with a p-value of 0.0016, indicating a strong positive conditional association between investment and food production.
4. The number of food manufacturing enterprises (X_3) also showed a statistically significant positive relationship with food production. This finding indicates the importance of expanding entrepreneurial activity and production capacity in the regional food industry.
5. Although employment (X_1) had a positive regression coefficient, it was not statistically significant in the multiple regression model. This suggests that increasing employment alone may not be sufficient to ensure proportional production growth and that labor productivity, technology, and capital availability should also be considered.
6. The Random Forest analysis provided additional information on the predictive relevance of the selected factors. The preliminary feature-importance estimates were approximately 40.4% for employment, 33.8% for investment, and 25.8% for the number of enterprises.
7. The difference between the regression and Random Forest results for employment is particularly important. While X_1 was not statistically significant in the linear regression, it showed relatively high predictive importance in Random Forest. This suggests the possible presence of nonlinear relationships and interactions that require further empirical investigation.

Overall, the study confirms that the development of the food industry in the Kashkadarya region is closely associated with investment activity, employment, and enterprise development. The combined use of regression analysis and Random Forest provides a more comprehensive assessment by integrating economic interpretation with the identification of nonlinear predictive patterns. The findings suggest that increasing investment efficiency, modernizing production, improving labor productivity, and strengthening food manufacturing enterprises are key priorities for the sustainable development of the regional food industry.

REFERENCES

- [1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7.
- [3] H. R. Varian, "Big data: New tricks for econometrics," *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014, doi: 10.1257/jep.28.2.3.
- [4] J. Kleinberg, J. Ludwig, S. Mullainathan, and Z. Obermeyer, "Prediction policy problems," *American Economic Review*, vol. 105, no. 5, pp. 491–495, 2015, doi: 10.1257/aer.p20151023.
- [5] S. Mullainathan and J. Spiess, "Machine learning: An applied econometric approach," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, 2017, doi: 10.1257/jep.31.2.87.
- [6] S. Athey and G. W. Imbens, "The state of applied econometrics: Causality and policy evaluation," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 3–32, 2017, doi: 10.1257/jep.31.2.3.
- [7] S. Athey and G. W. Imbens, "Machine learning methods that economists should know about," *Annual Review of Economics*, vol. 11, pp. 685–725, 2019, doi: 10.1146/annurev-economics-080217-053433.

-
- [8] V. Kreinovich, S. Sriboonchitta, and W. Yamaka, Eds., *Machine Learning for Econometrics and Related Topics*. Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-031-43601-7.
 - [9] W. Kristjanpoller, "A hybrid econometrics and machine learning based modeling of realized volatility of natural gas," *Financial Innovation*, vol. 10, Art. no. 45, 2024, doi: 10.1186/s40854-023-00577-0.
 - [10] D. Kant, A. Pick, and J. de Winter, "Nowcasting GDP using machine learning methods," *AStA Advances in Statistical Analysis*, vol. 109, pp. 1–24, 2025, doi: 10.1007/s10182-024-00515-0.
 - [11] D. L. Naila, F. A. Nandonde, and J. Makindara, "Review of empirical studies on food process innovation: Status quo and future research agenda," *Sustainable Technology and Entrepreneurship*, vol. 3, no. 3, Art. no. 100073, 2024, doi: 10.1016/j.stae.2024.100073.
 - [12] L. N. Khudoyorov, "Dynamic modeling of sustainable development of small business and entrepreneurship," *American Journal of Economics and Business Management*, vol. 8, no. 9, pp. 4685–4694, 2025.
 - [13] D. N. Gujarati and D. C. Porter, *Basic Econometrics*, 5th ed. Boston, MA, USA: McGraw-Hill/Irwin, 2009.
 - [14] J. D. Angrist and J.-S. Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ, USA: Princeton University Press, 2009.
 - [15] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning: With Applications in Python*. Cham, Switzerland: Springer, 2023, doi: 10.1007/978-3-031-38747-0.