ASSESSING FOOD INDUSTRY DEVELOPMENT USING MACHINE LEARNING METHODS: THE CASE OF KASHKADARYA REGION

Isoqov Hasan Abduxalimovich
Independent Researcher, Karshi State Technical University, Uzbekistan
DOI:https://doi.org/10.5281/zenodo.24052248

Abstract. This article investigates the modeling of food manufacturing industry development using machine learning algorithms, with Kashkadarya Region as a case study. Economic, investment, technological, and production-related factors influencing the industry's innovative development are systematized, with particular attention to assessing their relationships with the outcome indicator. The study envisages the application of statistical analysis, data preprocessing, machine learning, and comparative evaluation. The potential of machine learning algorithms to identify complex nonlinear relationships, assess the relative importance of factors, and select a model with high predictive accuracy is examined.

Keywords: food manufacturing industry, innovative development, machine learning, modeling, Random Forest, XGBoost, forecasting, innovation factors, Kashkadarya Region.

INTRODUCTION

Rapid advances in digital technologies are expanding opportunities to apply new methods of data analysis, trend identification, and forecasting across economic sectors. Machine learning algorithms, particularly in the processing of large economic and statistical datasets, are becoming effective complements to traditional analytical methods. These approaches can reveal complex relationships among economic indicators, identify important factors, and potentially improve forecast accuracy.

The food manufacturing industry is an important component of the regional economy. Its sustainable development contributes to supplying the population with quality food, processing agricultural raw materials, creating value added, increasing employment, and supporting regional economic growth. Identifying the factors associated with industry development, assessing their relative influence, and anticipating future trends are therefore important scientific and practical tasks.

Kashkadarya Region possesses considerable resource potential for agricultural production and processing. Development of food manufacturing in the region is closely associated with efficient use of local raw materials, modernization of existing production capacity, investment attraction, higher labor productivity, and improved storage and logistics infrastructure. The interactions among these factors shape the pace and direction of industry development.

Food manufacturing development is a multifactor economic process. Output may be associated with investment, fixed capital, labor resources, agricultural raw materials, the number of enterprises, production costs, technological capacity, and other indicators. These relationships are not necessarily linear: the importance of some factors may change after certain thresholds are reached or under different conditions. This creates a need for flexible modeling methods.

Machine learning algorithms offer new opportunities to study such complex economic relationships. They can learn patterns from available data without imposing a rigid functional form in advance. Random

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Forest, Gradient Boosting, and XGBoost can account for linear and nonlinear relationships, assess the relative importance of predictors, and support the development of forecasting models.

An important feature of the machine learning approach is the possibility of comparing different models using consistent criteria. Metrics such as R^2, MAE, and RMSE can help identify the algorithm that best predicts the economic process under investigation. Assessing feature importance can also identify indicators that provide the most predictive information and help prioritize areas for further investigation and management attention.

RESEARCH METHODOLOGY AND METHODS

The study adopted a comprehensive methodological approach to identifying patterns in the development of the food manufacturing industry in Kashkadarya Region, examining its principal associated factors, and developing predictive models using machine learning algorithms. The approach combined conventional statistical analysis with modern machine learning methods.

The information base comprised statistical indicators of food production in Kashkadarya Region and economic, investment, production, labor, and resource indicators potentially associated with its development. Indicators were selected with attention to economic relevance, their relationship to food manufacturing, data availability, continuity, and comparability over time.

The research was organized into several sequential stages. First, the outcome indicator describing the development of food manufacturing and a set of potentially relevant predictors were identified. These included investment, fixed capital, labor resources, the number of enterprises, local raw-material volumes, labor productivity, technological capacity, and other relevant economic indicators.

Next, the initial data were prepared for machine learning. Missing values, outliers, and differences in measurement scales were examined. Normalization or standardization was considered where necessary to address differences in the scales of indicators.

Descriptive statistics and correlation analysis were used for preliminary data exploration. These methods helped identify initial associations between the outcome and explanatory variables, examine relationships among indicators, and describe the principal trends. Particular attention was given to strongly correlated or economically overlapping indicators.

Random Forest, Gradient Boosting, and XGBoost were considered for modeling the development of the food manufacturing industry. They were selected because of their ability to capture complex and nonlinear as well as linear relationships among economic indicators. Considering several algorithms makes it possible to compare their performance and identify a model suited to the available data.

Random Forest aggregates the predictions of multiple decision trees, thereby reducing sensitivity to the random variation of individual trees. It can identify complex patterns in multivariate data and evaluate the relative predictive importance of explanatory factors.

Gradient Boosting constructs models sequentially, with each successive stage seeking to reduce errors from the preceding model. This feature can help capture complex relationships among economic indicators and improve predictive performance.

XGBoost is an optimized implementation of gradient boosting designed to improve predictive performance and generalization. Its mechanisms for controlling model complexity are important for reducing the risk of overfitting.

High predictive accuracy alone was not regarded as sufficient when considering the application of machine learning to economic data. Consistency with economic reasoning, stability on new observations,

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

and practical interpretability of identified predictors were also considered. This helps reduce the risk of selecting a model that fits the data well mathematically but lacks adequate economic justification.

RESULTS

Statistical data covering 2011–2025 were examined to investigate the development of food manufacturing in Kashkadarya Region using machine learning approaches. The outcome variable was the volume of food production (Y). The explanatory variables were the number of workers employed in the food industry (X1), investment in the sector (X2), and the number of legal entities engaged in food manufacturing (X3).

The analysis indicated a long-term upward trend in food production in Kashkadarya Region, with particularly marked growth in recent years. According to the official statistics cited in the source manuscript, food production in the region reached UZS 10.2 trillion in 2025 (Statistika.uz).

Preliminary analysis suggested relationships among Y, X1, X2, and X3. However, conventional correlation analysis primarily describes linear associations, whereas the relationships of investment, employment, and enterprise numbers with food production may be nonlinear. Accordingly, Random Forest, Gradient Boosting, and XGBoost were proposed for subsequent analysis.

In the proposed machine learning models, food production volume (Y) serves as the target variable, while employment (X1), investment (X2), and the number of legal entities (X3) serve as explanatory features. Comparing algorithms using R^2, MAE, and RMSE would help identify a model that more accurately describes food manufacturing development in Kashkadarya Region.

The availability of only 15 annual observations for 2011–2025 is an important limitation for machine learning. Complex models can overfit such small datasets. Therefore, validation methods that preserve temporal order are preferable to a simple random training/test split.

Including 2025 data enables a more up-to-date assessment of recent trends in Kashkadarya Region's food manufacturing industry. The reported production volume of UZS 10.2 trillion in 2025 underscores the sector's regional economic importance (Statistika.uz). Machine learning could further support comprehensive assessment of the predictive relevance of employment, investment, and the number of enterprises, identification of priority factors, and forecasting of future development indicators.

DISCUSSION

The analysis suggests that food manufacturing development in Kashkadarya Region is closely associated with employment, investment activity, and the number of producing business entities. The 2011–2025 data indicate a long-term growth trend alongside fluctuations in output or growth rates in particular years. This supports the relevance of studying food manufacturing not only through linear trends but also through methods capable of accounting for more complex relationships.

The substantial increase in food production during the period under review suggests a growing role for the industry in the regional economy. Particularly rapid growth in recent years may be associated with expanding productive capacity, greater investment activity, and business development. However, changes in price levels can also contribute to increases in nominal output; nominal production value should therefore not be treated as a sole measure of real production efficiency.

An additional advantage of machine learning is its potential to assess the relative importance of selected predictors. Model-based analysis could identify which of X1, X2, and X3 contributes the most information to forecasts of food production. Such findings could inform the prioritization of areas for further policy analysis and resource allocation.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Nevertheless, the annual frequency and limited number of observations require caution in evaluating machine learning models. With a small sample, complex algorithms may fit the observed data very closely while performing poorly on new observations. Model selection should therefore consider MAE and RMSE alongside R^2 and employ time-aware validation procedures.

Overall, the analysis supports considering machine learning algorithms alongside conventional statistical methods when assessing food manufacturing development. Statistical analysis describes trends and associations, while machine learning offers tools for exploring complex relationships, assessing feature importance, and potentially improving forecasts. Together, these approaches provide a methodological basis for more informed regional management decisions.

Conclusions and Recommendations

The study indicated that food manufacturing development in Kashkadarya Region is associated with labor resources, investment activity, and the number of enterprises engaged in production. Examination of the 2011–2025 data revealed a long-term growth trend as well as fluctuations in output during particular periods.

According to the analysis presented, food production was positively associated with employment, investment volume, and the number of legal entities. The number of producing legal entities and investment activity were highlighted as particularly closely related to industry development. These observations suggest the importance of investment and expanded business production capacity alongside the availability of labor resources.

The findings support considering machine learning algorithms in addition to conventional statistical analysis when examining food manufacturing development. Random Forest, Gradient Boosting, and XGBoost can expand the ability to identify complex nonlinear relationships, assess the relative predictive importance of factors, and forecast future production indicators.

The following practical recommendations are proposed:

1. Increase investment in food manufacturing. Prioritize investment in creating new production capacity, modernizing existing enterprises, and introducing energy- and resource-efficient equipment.

2. Encourage the activity of producing enterprises. Support the establishment of new food manufacturing businesses, expand the production capacity of existing small enterprises, and improve their access to financial resources.

3. Prioritize higher labor productivity. In addition to increasing employment, develop a skilled workforce capable of adopting modern technologies and strengthen employees' digital and technological competencies.

4. Expand the deep processing of local raw materials. Encourage the processing of agricultural products grown in the region into finished food products with higher value added rather than selling them primarily as raw materials.

5. Establish a machine learning-based monitoring and forecasting system. Regularly update datasets on food production, investment, employment, enterprise numbers, and other economic indicators, and use them to develop short- and medium-term forecasts.

Applying machine learning algorithms in this process could improve the use of available statistical data, help identify predictors relevant to industry development, and support management decisions with forecasting evidence. In this respect, machine learning may serve as a modern methodological tool for evaluating the prospective development of food manufacturing in Kashkadarya Region.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

REFERENCES

1. Khudoyorov, L. N. (2025). Dynamic Modeling of Sustainable Development of Small Business and Entrepreneurship. American Journal of Economics and Business Management, 8(9), 4685–4694. (Decision of the Presidium of the Higher Attestation Commission, No. 360, August 28, 2024).
2. Xudoyorov, L. N. (2025). Evaluation of the Quality and Significance of Econometric Models in Small Business and Entrepreneurship Development According to Selected Criteria. Digital Economy Scientific Electronic Journal, (12), 1631–1639. (Decision of the Presidium of the Higher Attestation Commission, No. 347, December 29, 2023). https://infocom.uz/magazine-article-detail/971
3. Hassoun, A. (Ed.). (2025). Food Industry 4.0 and Food Sustainability. Cham: Springer Nature Switzerland AG. 286 pp. https://doi.org/10.1007/978-3-032-06527-8
4. Hassoun, A. (2025). Food Sustainability 4.0: Harnessing Fourth Industrial Revolution Technologies for Sustainable Food Systems. Discover Food, 5, Article 171. https://doi.org/10.1007/s44187-025-00461-z

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.