Journal of Agricultural Science and Technology (JAST), 28(5)
In Press, Pre-Proof Version

Applying Machine Learning to Strategic Market Prioritization in the Dairy Export Sector

Morteza Majidian[1], and Esmaeil Pishbahar[1*]

Abstract

This study presents an innovative framework for optimizing Iran's dairy exports by integrating machine learning and multi-criteria decision-making techniques. Utilizing a comprehensive 20-year dataset spanning 2003 to 2022, sourced from credible international databases, four machine learning models, Bagging Regression, CatBoost Regression, Gradient Boosted Regression, and Extreme Gradient Boosting, were employed to forecast dairy export values. The CatBoost Regression model demonstrated superior predictive accuracy, achieving a coefficient of determination of 0.93. To enhance interpretability, SHAP analysis was applied, revealing population, economic size, and trade potential as the most influential factors driving export performance. Concurrently, the TOPSIS method was used to prioritize potential export markets based on economic and trade-related criteria, identifying Turkey, Iraq, and Pakistan as the top destinations due to their proximity, market demand, and trade compatibility. This dual approach combines predictive analytics with strategic market ranking, offering actionable insights for policymakers and exporters aiming to bolster Iran's non-oil economy. The findings highlight the critical role of regional markets and trade infrastructure in enhancing dairy export competitiveness. By leveraging advanced analytics, this research supports sustainable agricultural development and economic diversification in Iran, addressing the vulnerability of its oil-dependent economy. The methodology and results provide a robust foundation for future export strategies, emphasizing the synergy between data-driven forecasting and systematic decision-making in agricultural trade optimization.

***Keywords**:* Dairy Exports, Machine Learning, Multi-Criteria Decision-Making, Predictive Analytics, Trade Optimization.

JEL Classification: F14, Q17, L11, Q13, L13.

1. Introduction

Iran's economy heavily relies on oil and gas revenues, possessing 10% of global reserves and ranking as OPEC's second-largest producer after Saudi Arabia (Farzanegan & Markwardt, 2009). This dependency renders it vulnerable to oil price fluctuations, often leading to balance-

[1] Department of Agricultural Economics, Faculty of Agriculture, College of Agriculture and Natural Resources, University of Tehran, Karaj, Islamic Republic of Iran.
Corresponding's author; e-mail: esmaeil.pishbahar@ut.ac.ir

of-payments deficits. Diversifying the economy through non-oil exports, such as dairy products, is critical to mitigate this risk. Dairy exports offer a promising avenue to enhance foreign exchange earnings and reduce oil reliance, given Iran's robust agricultural and livestock sectors. Currently, food and agricultural products, notably dairy, account for about 8% of Iran's non-oil exports, highlighting their role in economic diversification.

Developing agricultural export markets requires a deep understanding of domestic production capabilities and global demand. Meeting foreign consumer expectations is vital for sustainable export profitability. Iran's agricultural sector, particularly dairy, holds significant export potential due to its comparative advantages in production. Enhancing dairy product quality and diversity could position Iran as a leading exporter. Between 2003 and 2022, dairy export values fluctuated significantly (Fig. 1), rising from 2006 to a peak of $718 million in 2016, followed by a decline during 2018–2020, likely due to global economic conditions and the COVID-19 pandemic, before recovering to $628 million by 2022. This trend underscores the need for strategic planning to bolster dairy exports and support economic resilience.

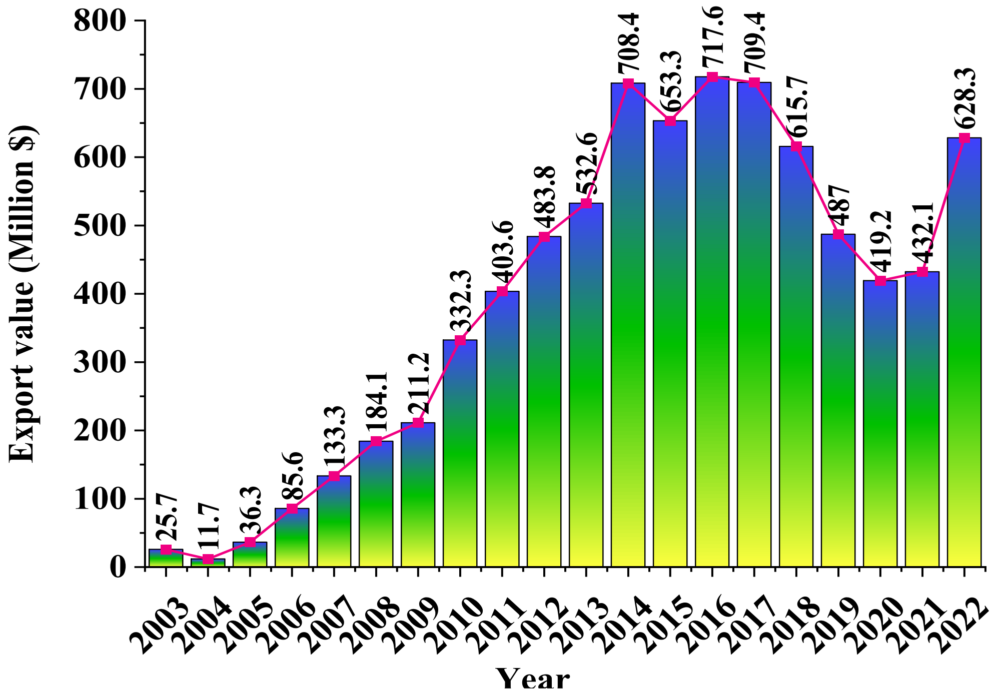

Figure 1. Export value of Iran's dairy products during the period 2003–2022 (million USD) (Source: ITC, 2024).

Effective enhancement of Iran's dairy sector hinges on understanding market performance drivers. Sound agricultural policies, informed by domestic capabilities and global dynamics, are essential for sustainable growth. Misguided policies, however, could hinder economic progress and societal welfare. Accurate predictive models are thus crucial for identifying viable export markets. This study employs the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), a multi-criteria decision-making (MCDM) method, to rank export destinations systematically. TOPSIS, valued for its simplicity, selects optimal options by measuring proximity to an ideal solution, making it effective for market prioritization (Madanchian & Taherdoost, 2023).

Traditional forecasting methods, like time series and econometric models, have long been used for export prediction. However, data-driven approaches, such as artificial intelligence (AI) and machine learning (ML), offer superior precision (Guo et al., 2024). AI mimics human cognitive functions, while ML, a subset, excels at modeling complex phenomena with advanced algorithms (Tasneem et al., 2024). These techniques often surpass human analytical capacity, increasingly replacing traditional methods across sectors (Y. Wang et al., 2024). For dairy export forecasting, ensemble ML methods like Random Forest and Gradient Boosting enhance accuracy by integrating multiple models. Typically, 70–80% of pre-processed data trains these models, with 20–30% reserved for evaluation, ensuring reliable predictions for strategic planning (Binkhonain & Zhao, 2019).

While tree-based ML algorithms provide interpretability via feature importance scores, measures like gain can be inconsistent. Explainable AI, particularly SHAP (Shapley Additive Explanations), addresses this by using game theory to quantify each feature's contribution to predictions, improving transparency (Ghafarian Nia et al., 2025b; H. Wang et al., 2024). In this study, SHAP elucidates factors driving dairy export performance, enhancing model interpretability.

2. Literature Review

Table 1 synthesizes prior research on dairy export modeling across diverse geographical contexts, including Ukraine, Ireland, New Zealand, Russia's Altai region, Latvia, and the European Union (Popko et al., 2024; Cele et al., 2022; Luo & Bano, 2020; Belyaev et al., 2019; Du et al., 2006; Bojnec & Fertő, 2014). These studies differ substantially in their choice of predictors, ranging from macroeconomic indicators such as GDP and trade competitiveness metrics to demographic factors like population size and consumption patterns, as well as sector-specific production and market data. Correspondingly, the applied analytical approaches span traditional econometric techniques (e.g., multiple linear regression, hazard rate modeling, gravity analysis) and advanced computational methods (e.g., artificial neural networks). The reported outputs encompass competitiveness rankings, export potential assessments, survival analysis of trade relationships, and demand forecasting. This comparative synthesis highlights both methodological and contextual diversity in the field and offers a foundation for identifying the research gaps addressed in the present study.

Table 1. Overview of dairy export modeling approaches and methods used in previous studies.

Country	Input Variables (Predictors)	Output Variables (Target Criteria)	Methods and Models Used	Sources
Ukraine	Population, Population Growth Rate, Per Capita GDP, Total Dairy Imports, Per Capita Dairy Consumption, Trade and Economic Relations with Ukraine, Political Status	Ranking of Importing Countries for Domestic Dairy Products and a Comprehensive Evaluation of the Potential Attractiveness for Exporting Countries	Multiple Linear Regression (MLR) and Numerical Taxonomy Analysis	Popko et al., 2024
Ireland	Productivity Indicators, Normalized Revealed Comparative Advantage (NRCA), Trade Competitiveness Metrics	Farm-level Competitiveness and Net Export Share, NRCA at Export Level	Productivity Indicators, NRCA, and Statistical Methods	Cele et al., 2022
New Zealand	Dairy Export Data (1989-2017), Supply, Demand Factors, and Gravity Drivers	Duration and Survival Patterns of New Zealand Dairy Export Relations, Impact of Tariff and Non-Tariff Barriers on Hazard Rate	Hazard Rate Models and Gravity Analysis	Luo & Bano, 2020
Russia	Production and Export Data from the Altai Region, Dairy Market Analysis	Evaluation of Export Potential and Key Factors Affecting It; Suggestions for Export Improvement and Processing Industry Development	Production-Consumption Balance Method, Statistical and Economic Analysis	Belyaev et al., 2019
European Union	RCA Index, Markov Transition Probability Matrix, Kaplan-Meier Estimation, Wilcoxon Test	RCA Index, HS-6 Level Trade Data (2000-2011), Dairy Chain Product Classification into Four Groups	RCA Index, Markov Model, Kaplan-Meier, and Wilcoxon Test	Bojnec & Fertő, 2014
Latvia	Domestic Market Data: Production Levels, Consumption, Prices, Market Share, Consumer Behavior, Domestic Demand	Domestic Dairy Demand Forecast and Its Impact on Export Capacity and Trade Balance	ANN	Du et al., 2006

The comparative analysis in Table 2 underscores the methodological and contextual distinctiveness of the present study relative to prior dairy export research. None of the six benchmark studies combine advanced machine learning with multi-criteria decision-making techniques, nor do they integrate explainable AI to elucidate the relative importance of predictive factors. Furthermore, systematic market prioritization and a focus on Iran's dairy export sector are absent from the existing literature. Addressing these gaps, the present research develops a state-of-the-art ML framework leveraging a 20-year dataset (2003–2022) to forecast export values, applies TOPSIS for structured market ranking, and employs SHAP to ensure interpretability and transparency. This integrated approach delivers both high-precision forecasts and actionable strategic insights, offering policymakers a robust evidence base for targeted interventions aimed at diversifying and strengthening Iran's dairy export portfolio.

Table 2. Comparative analysis of key methodological features and research gaps in dairy export studies.

ML Models	MCDM (TOPSIS)	Explainable AI (SHAP)	Market Prioritization	Iran Focus	Reference
✗	✗	✗	✓	✗	Popko et al., 2024
✗	✗	✗	✗	✗	Cele et al., 2022
✗	✗	✗	✗	✗	Luo & Bano, 2020
✗	✗	✗	✗	✗	Belyaev et al., 2019
✗	✗	✗	✗	✗	Bojnec & Fertő, 2014
✓	✗	✗	✗	✗	Du et al., 2006
✓	✓	✓	✓	✓	Present Study

Figure 2 provides an overview of the research workflow, starting from data collection (ITC and World Bank, 2003–2022) and data preparation, followed by the application of machine learning models (BGR, CBR, GBR, XGB) and SHAP-based interpretation. The process continues with market prioritization using entropy-weighted TOPSIS and culminates in policy insights. This integrated framework ensures accurate prediction, transparent interpretation, and practical recommendations for strengthening Iran's dairy export competitiveness.

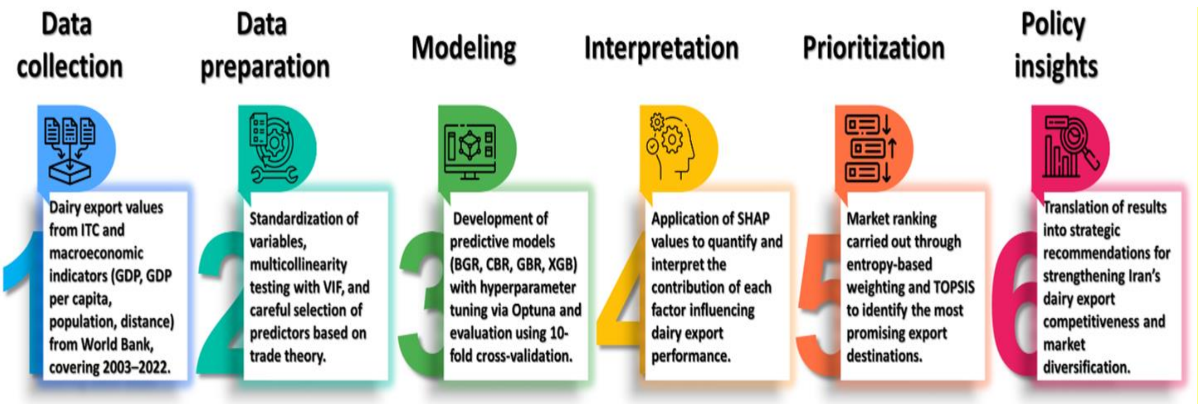

Figure 2. Overview of the research workflow: from data collection to market prioritization.

3. Research Methodology

3.1. Data Compilation

Dairy exports are vital to Iran's economy and global trade. This study uses reliable data to predict Iran's dairy export values accurately, identifying opportunities and challenges for expansion. Data include dairy export values from the International Trade Center and economic-demographic variables, geographical distance, per capita income, GDP, and population of importing countries, from the World Bank, covering 2003–2022. Excel spreadsheets calculated trade potential indices, economic size, structural differences, and TOPSIS weights for target countries. These variables were selected for their significant influence on export trends:

distance affects transportation costs and competitiveness, population indicates market demand, and per capita income and GDP reflect purchasing power and economic conditions of importing nations.

3.2. Feature Selection

Feature selection improves model accuracy and interpretability by reducing complexity and overfitting. Guided by trade theory and empirical evidence, this study focuses on key demographic, economic, and trade-related factors shaping Iran's dairy exports. These variables align with the Intercountries Trade Force (ITF) model (Rasoulinezhad & Jabalameli, 2019), which refines the gravity trade framework through a dynamic gravity index to better capture real-world trade patterns.

a) Population: Indicates future demand; larger populations suggest greater import needs.

b) Geographical Distance: Shorter distances reduce transportation costs, increasing trade likelihood (inverse used in ranking) (Tiits et al., 2024a).

c) Per Capita GDP: Reflects consumer purchasing power, driving demand (Tiits et al., 2024a).

d) Per Capita Income: Captures income effects on dairy demand, reducing variable redundancy (Tiits et al., 2024b).

e) Economic Structure Difference Index: Calculated via Equation 1, higher similarity boosts trade potential (Majidian & Dourandish, 2024):

$$\mathbf{DISSIZE} = \mathbf{Ln}\left[\mathbf{1} - \left(\frac{\mathbf{GDP_{it}}}{\mathbf{GDP_{it}} + \mathbf{GDP_{jt}}}\right)^{2} - \left(\frac{\mathbf{GDP_{jt}}}{\mathbf{GDP_{it}} + \mathbf{GDP_{jt}}}\right)^{2}\right] \qquad (1)$$

f) Domestic Demand: Calculated as the sum of Iran's domestic production and imports, this variable captures the total available dairy supply, reflecting internal consumption needs and market size.

g) Economic Size: Product of exporting and importing countries' GDPs, influencing export volume (Antonucci & Manzocchi, 2006; Kahouli & Maktouf, 2015).

h) Trade Potential: Assessed via cosine similarity (Equation 2), measuring export-import pattern alignment (Dolphin et al., 2021):

$$\mathbf{Cos_{ij}} = \frac{\sum_{k=1}^{n} \mathbf{E_{ik}} \cdot \mathbf{M_{jk}}}{\sqrt{\sum_{k=1}^{n} \mathbf{E_{ik}}^{2} \cdot \sum_{k=1}^{n} \mathbf{M_{jk}}^{2}}} f \qquad (2)$$

Where E_{ik} and M_{jk} are exports and imports of product k. Values near 1 indicate high trade potential.

3.3. Data Preprocessing

To ensure model reliability, data were preprocessed using Z-score normalization, standardizing variables to a mean of zero and a standard deviation of one. This prevents scale-related bias, optimizing convergence and interpretability in regression and tree-based models. The standardized dataset was input for machine learning, enhancing predictive accuracy and minimizing biases from scale disparities (Kern et al., 2020).

3.4. Machine Learning Models

Four machine learning models, Bagging Regression (BGR), CatBoost Regression (CBR), Extreme Gradient Boosting (XGBoost), and Gradient Boosted Regression (GBR), were used to predict dairy export values.

BGR: Combines multiple decision trees, averaging predictions (Equation 3) to reduce variance (Rodriguez-Galiano et al., 2015):

$$\hat{y}_i = \frac{1}{B}\sum_{b=1}^{B} f_b(X_i) \qquad (3)$$

Where B is the number of base models. $f_b(X_i)$ is the prediction of base model b for sample i. This process results in a more stable model that can perform better on new data.

CBR: Employs **gradient**-boosted trees (Equation 4), enhancing weak learners iteratively:

$$y = f(x) = \sum_{i=1}^{n} a_i h_i(x) \qquad (4)$$

Where y represents the predicted value, and x denotes the input features. The resulting function f(x) is a linear combination of the base functions $h_i(x)$. The coefficients a_i Determine the importance of the base functions in this linear combination (Ghafarian Nia et al., 2025c).

XGBoost: Uses regularization and parallel processing, minimizing loss (Equation 5):

$$L(\emptyset) = \sum_{i=1}^{n} L(\hat{y}_i, y_i) + \sum_{k=1}^{n} \Omega(f_k) \qquad (5)$$

Where n is the number of training samples and y_i is the actual value of sample i. $\hat{y}_i$ Is the predicted value of sample i. K is the number of trees in the model, and f_k is the k-th decision tree. $\Omega(f_k)$ is the regularization term that penalizes the complexity of the model to prevent overfitting (Sun et al., 2024).

GBR: Iteratively corrects errors (Equation 6), excelling with nonlinear data:

$$\boldsymbol{f_n(x_t) = \sum_{i=1}^{n} f_i(x_t)} \qquad (6)$$

Where f_i represents the prediction of the weak learner, and n is the total number of weak learners in the ensemble (Hongliang et al., 2024).

Implemented in Python, models were optimized using Optuna for hyperparameters and 10-fold cross-validation to prevent overfitting. Performance was evaluated with MAE, RMSE, and R^2 metrics (Ghafarian Nia et al., 2025a).

3.5. Feature Importance Analysis

SHAP analysis, rooted in game theory, quantified each feature's contribution to predictions, improving transparency over traditional methods. It reveals nonlinear relationships and interactions, visualized via summary and dependence plots, fostering trust in AI-driven trade strategies (Hou et al., 2024).

3.6. Multi-Criteria Decision Analysis

To rank export destinations based on multiple economic and trade variables, this study implemented the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). This multi-criteria decision-making (MCDM) framework evaluates each alternative by comparing its proximity to an ideal scenario representing the best possible performance and a counter-ideal representing the poorest performance. The weighting of indicators was determined objectively using the entropy method, ensuring that the importance of each variable reflects the variability in the dataset (Chen & Hwang, 1992). The key steps of the TOPSIS method in this study are as follows:

A) Formulation of the Decision Matrix: A decision matrix X was established containing m countries and n indicators:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{a_{11}} & \cdots & \boldsymbol{a_{1n}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a_{m1}} & \cdots & \boldsymbol{a_{mn}} \end{bmatrix} \quad \textit{Where } \ i = 1, \ldots, m \textit{ and } j = 1, \ldots, n \qquad (7)$$

B) Normalization: Vector normalization was applied to standardize different units of measurement:

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{k=1}^{m}(a_{kj})^2}} \quad (8)$$

C) Entropy-Based Weighting and Weighted Normalized Matrix: The entropy method was used to calculate each indicator's weight Wj ,producing the weighted normalized matrix V (Dehdasht et al., 2020):

$$V_{ij} = \begin{bmatrix} w_1 r_{11} & \cdots & w_n r_{1n} \\ \vdots & \ddots & \vdots \\ w_1 r_{m1} & \cdots & w_n r_{mn} \end{bmatrix} \quad (9)$$

D) Identification of Ideal and Anti-Ideal Solutions: The positive ideal (A^+) and negative ideal (A^-) were defined as:

$$\begin{aligned} A^+ &= \{Max(V_{ij}|j\varepsilon J), Min(V_{ij}|j\varepsilon J')\} \\ A^- &= \{Min(V_{ij}|j\varepsilon J), Max(V_{ij}|j\varepsilon J')\} \end{aligned} \quad (10)$$

Where J denotes benefit-type criteria and J′ denotes cost-type criteria.

E) Distance Measures: Euclidean distances from each alternative to A^+ and A^- were calculated:

$$\begin{aligned} S_i^+ &= \sqrt{\sum_{j=1}^{n}(V_{ij} - A_j^+)^2} \\ S_i^- &= \sqrt{\sum_{j=1}^{n}(V_{ij} - A_j^-)^2} \end{aligned} \quad (11)$$

F) Closeness Coefficient: The relative closeness to the ideal solution was computed as:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^+} \quad (12)$$

The resulting C_i^* values range between 0 and 1, where a value closer to 1 indicates a higher level of attractiveness for the respective export market.

Recent studies have examined the persistence and determinants of agri-food export competitiveness using revealed comparative advantage (RCA) and its normalized or symmetric variants, often applying duration analysis to assess stability over time (e.g., Balassa, 1965; Bojnec & Fertő, 2017; 2018a; 2016; 2018b; Majidian et al., 2022). These works highlight that factors such as economic development, trade costs, export diversification, product differentiation, and macroeconomic shocks (e.g., economic crises) can significantly influence the survival or fragility of comparative advantage. Building on this literature, Iran's dairy export potential was assessed using RCA and RSCA indices for competitiveness and TOPSIS for market ranking based on criteria such as income, GDP, distance, and trade potential. The TOPSIS method was selected for its capacity to integrate multiple indicators with varying units and optimization directions into a single continuous "closeness-to-ideal" index, enabling transparent and consistent market ranking. Its compatibility with entropy-based objective weighting eliminates subjective bias, while its computational simplicity and interpretability make it suitable for policy-oriented decision-making. Entropy weighting ensured that criterion importance reflected the variability in the dataset, offering robust, evidence-based insights for export strategy optimization. Like other MCDA approaches, TOPSIS operates under certain assumptions, such as fixed criteria weights and independence among indicators. While these do not undermine its applicability, they highlight the importance of periodic updates and complementary analyses to ensure the rankings remain aligned with evolving market conditions.

4. Result

4.1. Descriptive Analysis of the Collected Dataset

A statistical analysis of Iran's dairy export dataset to neighboring countries was conducted, with key parameters (mean, standard deviation, skewness, kurtosis, etc.) summarized in Table 3. Economic size ranges from $\$4.31\times10^{20}$ to $\$1.51\times10^{25}$ (mean: $\$1.29\times10^{23}$, SD: $\$9.74\times10^{23}$), with high skewness (15.27) and kurtosis (235.35) indicating extreme values. GDP per capita spans $\$1.99\times10^{2}$ to $\$9.11\times10^{13}$ (mean: $\$3.79\times10^{11}$, SD: $\$5.87\times10^{12}$), reflecting economic disparities. Population varies from 7.49×10^{5} to 2.36×10^{8} (mean: 3.16×10^{7}, SD: 5.58×10^{7}), with skewness (2.44) suggesting lower-value concentration. The dissimilarity index (-3.84 to -0.69, mean: -1.54) and import demand (1.91×10^{3} to 2.69×10^{6} tons, mean: 3.46×10^{5}) show structural and demand diversity. Distance ranges from 899 to 5160 km (mean: 1600 km), impacting costs, while the similarity index (Cosij) (0.06–0.73, mean: 0.39) and export values ($\$0$–$\$5.37\times10^{5}$, mean: $\$3.04\times10^{4}$) indicate trade variability. Iran's domestic demand for dairy products ranged from approximately 6.56×10^{6} to 9.14×10^{6} tons (mean: 7.91×10^{6}, SD: 7.06×10^{5}), with relatively low skewness (-0.31) indicating a balanced distribution around the mean. This stability in domestic consumption reflects a consistent internal market that can influence export capacity and policy decisions. This varied dataset supports machine learning models for predicting exports and informing trade policy.

Table 3. Descriptive statistics of input/output parameters.

Variables	Mean	Standard deviation	Variance	Skewness	Kurtosis	Minimum	Median	Maximum
Economic Size ($)	1.29×10^{23}	9.74×10^{23}	9.49×10^{47}	15.27	235.35	4.31×10^{20}	2.75×10^{22}	1.51×10^{25}
GDP per capita	3.79×10^{11}	5.87×10^{12}	3.44×10^{25}	15.49	240	1.99×10^{2}	7.89×10^{3}	9.11×10^{13}
Population	3.16×10^{7}	5.58×10^{7}	3.12×10^{15}	2.44	5.01	7.49×10^{5}	6.02×10^{6}	2.36×10^{8}
Dissize Index	-1.54	0.81	0.65	-0.77	-0.62	-3.84	-1.28	-0.69
Import Demand (ton)	3.46×10^{5}	4.99×10^{5}	2.49×10^{11}	3.04	10.05	1.91×10^{3}	1.72×10^{5}	2.69×10^{6}
Distance (km)	1.60×10^{3}	5.55×10^{2}	3.09×10^{5}	1.51	5.59	8.99×10^{2}	1.50×10^{3}	5.16×10^{3}
Cos_{ij}	0.39	0.16	0.03	0.07	-0.94	0.06	0.38	0.73
Export Values	3.04×10^{4}	8.99×10^{4}	8.09×10^{9}	4.10	17.31	0.00	1.65×10^{3}	5.37×10^{5}
Iran Demand (Ton)	7.91×10^{6}	7.06×10^{5}	4.99×10^{11}	-0.31	2.27	6.56×10^{6}	7968685	9.14×10^{6}

To ensure the robustness of the empirical model, multicollinearity among the explanatory variables was examined using the Variance Inflation Factor (VIF) test (Table 4). All VIF values were well below the accepted threshold of 10 (and the more conservative threshold of 5) (Johnston et al., 2018), with a mean VIF of 1.48, confirming the absence of problematic collinearity. In particular, the VIF values for population (1.53) and GDP per capita (1.04) indicate that these variables capture distinct aspects of market characteristics, overall size, and per capita economic strength, without redundancy, justifying their simultaneous inclusion in the model.

Table 4. Variance Inflation Factor (VIF) results for model predictors (dimensionless).

Variable	VIF	1/VIF
Dissize Index	2.00	0.499755
Import Demand	1.77	0.563733
Cosij	1.77	0.564177
Population	1.53	0.652589
Distance	1.45	0.688023
Economic Size	1.23	0.814749
Iran Demand	1.05	0.951279
GDP Per Capita	1.04	0.961270
Mean VIF	**1.48**	

Spearman's correlation (Fig. 2) showed that Import Demand strongly correlates with Cosij (trade potential) ($r = 0.50$), indicating that cultural proximity boosts demand. Dissize Index (economic disparity) is positively correlated with Distance ($r = 0.44$) and Import Demand ($r = 0.49$), underlining infrastructure needs. Economic Mass is linked to Iran Demand ($r = 0.25$) and Dissize Index ($r = 0.82$). Moreover, Cosij is associated with Dissize Index ($r = 0.27$) and Export Values ($r = 0.29$), highlighting trade cooperation's role (Fig. 3). PCA (Fig. 4) showed export values (0.59) and population (0.65) driving PC3 and PC2, with distance (-0.52) hindering trade, reinforcing demand and proximity effects.

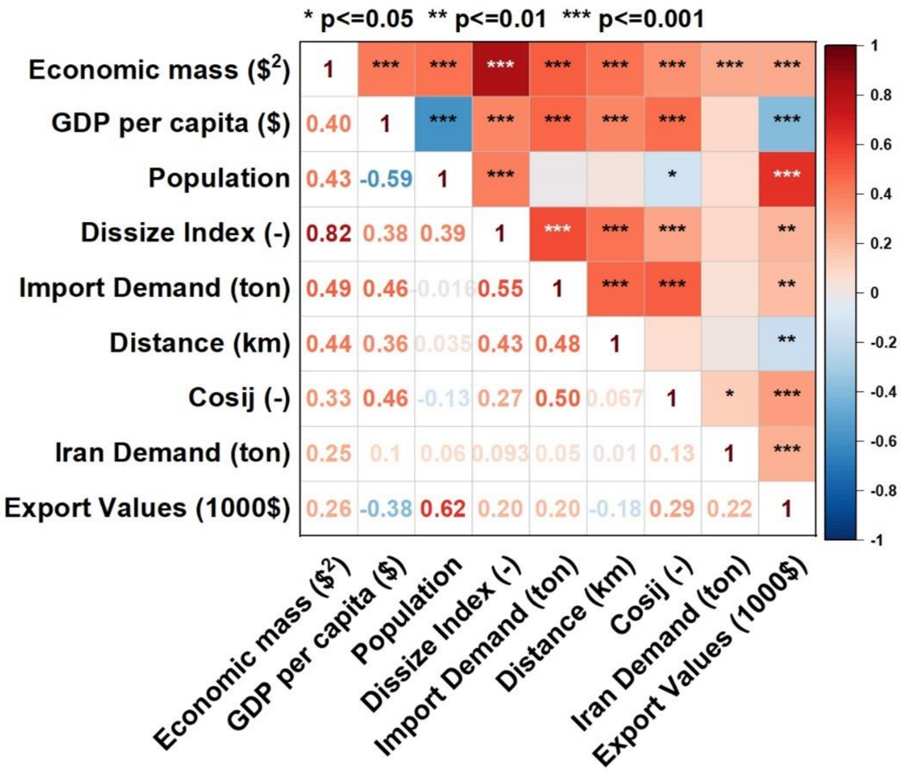

Figure 2. Spearman's rank correlation coefficients (lower triangular matrix) and associated p-values (upper triangular matrix) for all input variables.

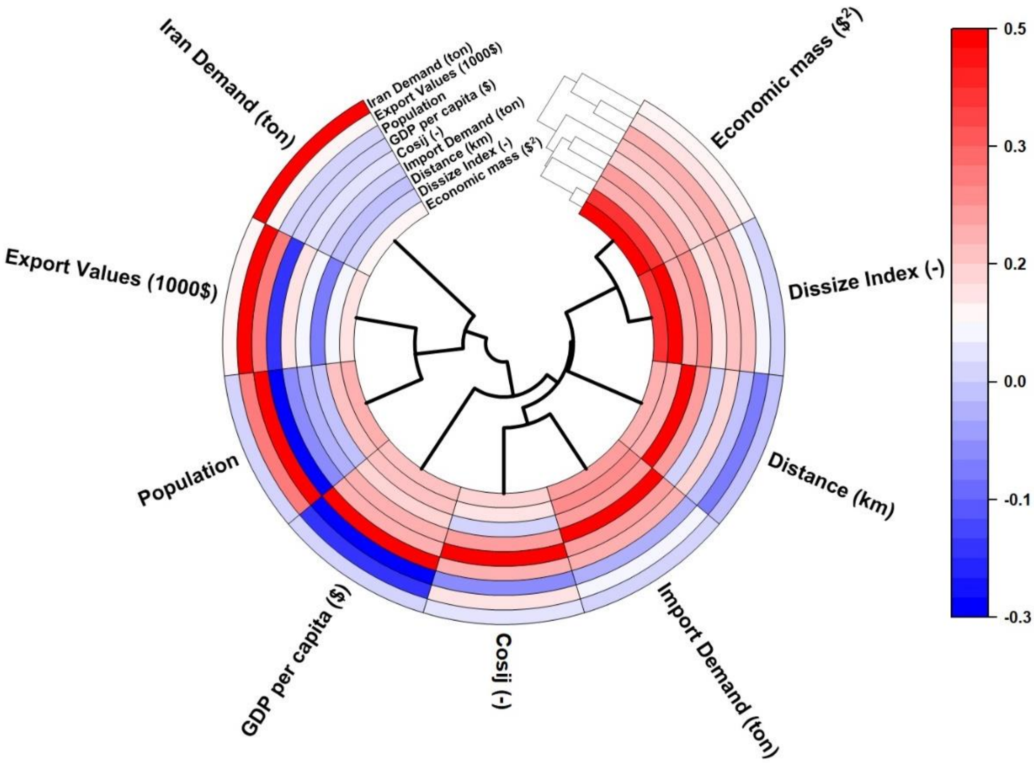

Figure 3. Heatmap with hierarchical clustering dendrogram displaying all combinations of analyzed descriptors based on the Spearman correlation matrix.

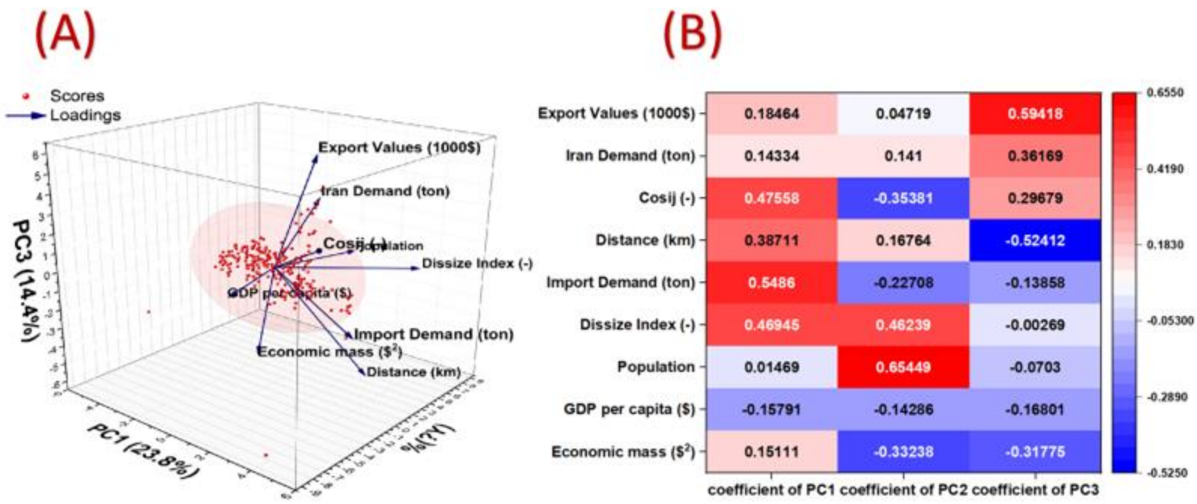

Figure 4. Principal component analysis (PCA) of the collected database: (A) 3D scatter plot illustrating the impact of input descriptors and output targets across the entire dataset; (B) coefficients of each descriptor for the top three principal components.

4.2. Developing a Machine Learning Model

Four models, BGR, CBR, GBR, and XGB, were trained with Optuna-optimized hyperparameters and 10-fold cross-validation, assessed via R^2, RMSE, and MAE (Table 5). Training R^2 ranged from 0.96–0.99, but testing showed CBR leading (R^2=0.93) and GBR lagging (R^2=0.88). CBR's low errors (RMSE: 2032.97, MAE: 1374.15 in training; 5492.4, 3682.4 in testing) outperformed BGR (RMSE: 7785.41, MAE: 3425.67 in testing), indicating less overfitting. Boosting models (CBR, XGB) excel in capturing nonlinear trade patterns (Table 6: CBR hyperparameters).

Table 5. Average statistical parameters from k-fold cross-validation for predicting dairy exports to target countries.

		Model type	
		Training step	**Testing step**
		Export Value (%)	Export Value (%)
R^2	BGR	0.96	0.84
	CBR	0.99	0.93
	GBR	0.98	0.88
	XGB	0.98	0.92
RMSE	BGR	7785.41	17417.6
	CBR	2032.97	5492.4
	GBR	31.07	75.6
	XGB	333.74	737
MAE	BGR	3425.67	7618.3
	CBR	1374.15	3682.4
	GBR	28.46	65.3
	XGB	274.41	534.8

Table 6. Hyperparameters of the developed machine learning models for predicting target dairy export markets for each output variable.

Parameter	Value
iterations	392
depth	5
learning_rate	0.052
l2_leaf_reg	1.381

4.3. Feature Importance Analysis

SHAP analysis enhanced model transparency, assigning values to features (Fig. 5). Population (mean SHAP: 32,086.86), economic mass, and trade potential dominated predictions, aligning with demand and trade theories. GDP per capita and dissimilarity had weaker effects, suggesting wealthier nations may favor local dairy. Distance showed mixed impacts, balancing cost barriers and proximity benefits. SHAP interactions indicated structural differences reduce feasibility (Fig. 6: economic indicators 53.19%, trade factors 5.92%, geographical and distance factors 38.29%, dissimilarity 2.60%).

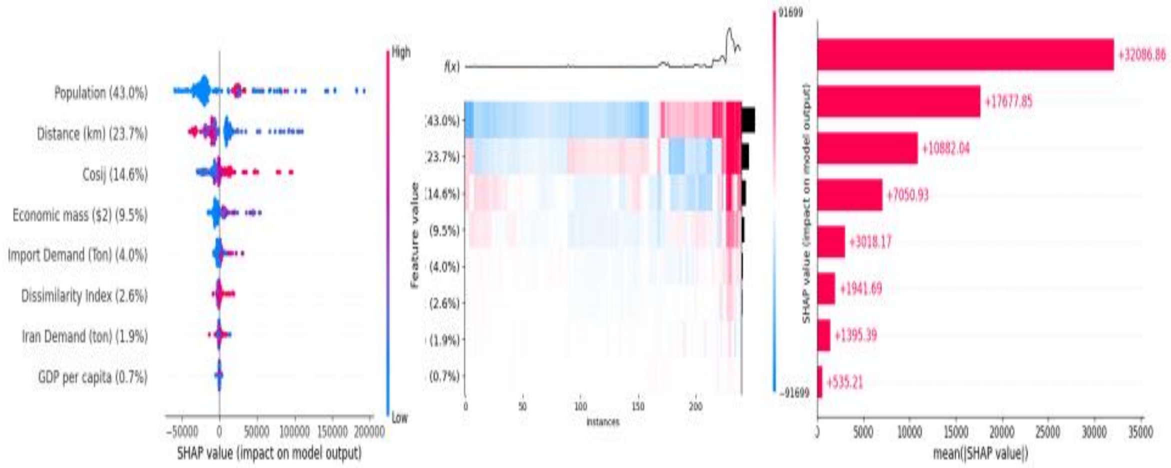

Figure 5. Analysis of feature importance: Beeswarm plot (left), heatmap of SHAP values (middle), bar plot of mean absolute SHAP values (right).

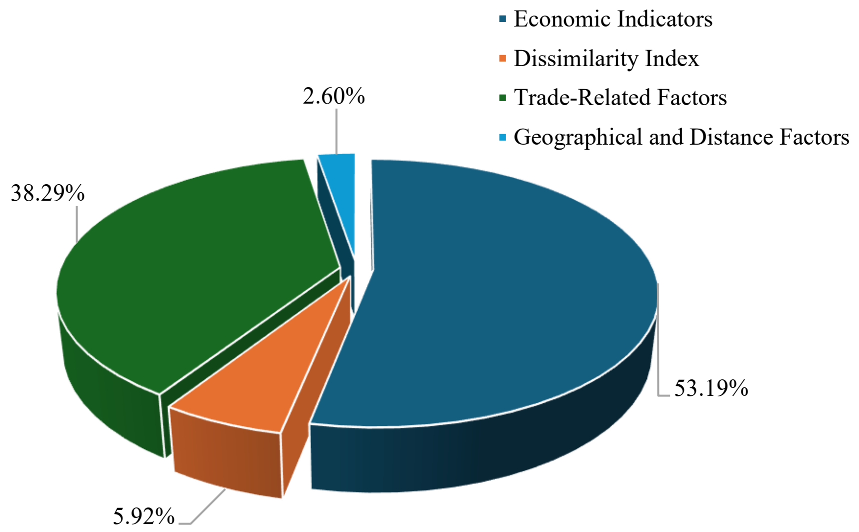

Figure 6. Proportional contribution (%) of feature categories to Iran's dairy export predictions based on SHAP analysis.

4.5. Multi-Criteria Decision Analysis Results

TOPSIS ranked Turkey (0.1481), Iraq (0.1303), and Pakistan (0.1301) as top markets due to proximity, trade ties, and demand (Table 7). UAE, Qatar, and Kuwait followed, leveraging purchasing power despite smaller sizes, while Oman, Turkmenistan, and others ranked lower due to economic and logistical constraints. RSCA (Fig. 7) showed dairy competitiveness rising since 2013, unlike volatile cereal products, suggesting targeted interventions for weaker sectors. These results guide resource allocation and trade strategies.

Table 7. Results of prioritizing target countries for Iran's dairy product exports.

Priority based on TOPSIS	Average weighted importance coefficient	Average rank of importing countries from Iran	Country
1	0.1481	1.4	Turkiye
2	0.1303	2.6	Iraq
3	0.1301	2.6	Pakistan
4	0.1166	4.1	United Arab Emirates
5	0.1032	5.0	Qatar
6	0.0849	5.9	Kuwait
7	0.0541	8.0	Oman
8	0.0555	8.1	Turkmenistan
9	0.0538	8.4	Bahrain
10	0.0458	10.2	Afghanistan
11	0.0423	10.6	Azerbaijan
12	0.0352	11.5	Armenia

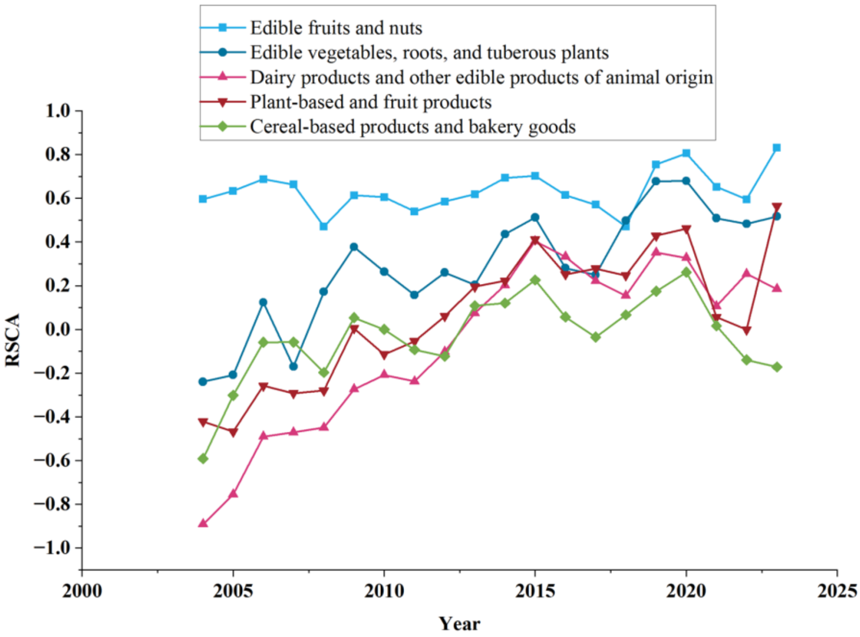

Figure 7. Revealed Symmetric Comparative Advantage (RSCA) index for the export of five agricultural tariff categories of Iran (2000–2023).

5. Discussion

In addition to market prioritization, the composition of Iran's dairy exports reflects both product durability and market-specific demand. According to International Trade Center (Trade Map) data for HS code 0406 (cheese and curd), the majority of Iran's cheese exports are destined for Iraq, with only a comparatively small share shipped to Türkiye. These exported cheeses predominantly include processed and semi-hard varieties such as feta and processed cheese, which are well-suited to Iraq's consumption habits and distribution channels. Türkiye, while geographically close and a significant dairy producer itself, imports only limited quantities of Iranian cheese; instead, its import profile is more diversified and competitive, reducing Iran's market share (ITC, 2024). Long shelf-life products such as whole and skimmed milk powders, sweetened condensed milk, evaporated milk, and ultra-high temperature (UHT) milk (HS code 0402) also account for a substantial portion of Iran's dairy exports. Their extended storage stability, ranging from 8–9 months for UHT milk to over a year for powders, supports long-distance trade and aligns with the logistical and food safety requirements of importing countries (Karlsson et al., 2019).

6. Conclusions and Future Directions

This study developed a robust framework for predicting Iran's dairy exports by integrating machine learning (ML) and multi-criteria decision-making (MCDM) methods. Using a 20-year dataset (2003–2022), four ML models, BGR, CBR, GBR, and XGBoost, forecasted export values, with CBR excelling (R^2 = 0.93). SHAP analysis clarified feature importance, identifying population, distance, and trade potential as key drivers. TOPSIS ranked Turkey, Iraq, and Pakistan as top export markets, emphasizing their strategic value.

Findings highlight distance as a critical factor, with greater distances increasing costs and reducing trade, consistent with trade theories. Proximity favors neighbors like Turkey and Iraq over distant markets, suggesting a regional focus for Iran's dairy strategy. Trade potential, measured by cosine similarity, and population size further shape export feasibility, urging focus on compatible, high-population markets like Pakistan. Policymakers should reduce transportation costs via improved logistics (e.g., cold chains), secure trade agreements with key partners, and promote Iranian dairy quality through marketing and digital platforms. Export subsidies could boost long-term competitiveness.

This research aligns with prior studies (e.g., Ukraine, New Zealand, Latvia) using traditional models but advances the field by combining ML with market prioritization for greater precision. Limitations include omitting consumer preferences, geopolitical risks, and trade policies, alongside reliance on basic ML rather than deep learning, and exclusion of dynamic economic shocks.

Limitations and Future Research: This study does not directly account for certain qualitative determinants of trade, such as consumer preferences, political risk, and detailed non-tariff barriers, due to limited long-term comparable data and potential endogeneity issues. While these effects are partially addressed through high-dimensional fixed effects, future research could extend this framework by incorporating survey-based preference indices, established political risk measures, and comprehensive non-tariff barrier datasets (e.g., UNCTAD TRAINS). Additionally, expanding datasets, integrating sentiment analysis for qualitative factors, and adopting deep learning approaches could better capture complex trade patterns. Implementing real-time policy analysis and multi-objective optimization would further refine decision-making, ultimately strengthening Iran's dairy export strategy and enhancing economic resilience.

References

Antonucci, D., and Manzocchi, S. (2006). Does Turkey have a special trade relation with the EU? *Economic Systems*, **30**(2), 157–169. DOI: 10.1016/j.ecosys.2005.10.003.

Balassa, B. (1965). Trade liberalisation and “revealed” comparative advantage 1. The Manchester School, 33(2), 99-123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x.

Belyaev, V., Butakova, M., and Goryaninskaya, O. (2019). Export of dairy products of the Altai Region: Opportunities, prospects, and problems. *IOP Conference Series: Earth and Environmental Science*, **395**(1), 012100. DOI: 10.1088/1755-1315/395/1/012100.

Binkhonain, M., & Zhao, L. (2019). A review of machine learning algorithms for identification and classification of non-functional requirements. Expert Systems with Applications: X, 1, 100001. https://doi.org/10.1016/j.eswax.2019.100001.

Bojnec, S., & Ferto, I. (2016). Drivers of the duration of grain competitiveness in European Union countries. Journal of Agricultural Science and Technology. 2016, 18(5), 1173-1185.

Bojnec, Š., & Fertő, I. (2017). The duration of global agri-food export competitiveness. British Food Journal, 119(6), 1378-1393. DOI: 10.1108/BFJ-07-2016-0302.

Bojnec, S., & Ferto, I. (2018a). Drivers of the duration of comparative advantage in the European Union's agri-food exports. Agricultural Economics/Zemědělská Ekonomika, 64(2). DOI: 10.17221/173/2016-AGRICECON.

Bojnec, S., & Ferto, I. (2018b). Economic crisis and the fragility of comparative advantage in European agriculture. German Journal of Agricultural Economics (GJAE), 67(3), 147-159.

Bojnec, Š., and Fertő, I. (2014). Export competitiveness of dairy products on global markets: The case of the European Union countries. *Journal of Dairy Science*, **97**(10), 6151–6163. DOI: 10.3168/jds.2013-7711.

Cele, L. P., Hennessy, T., and Thorne, F. (2022). Evaluating farm and export competitiveness of the Irish dairy industry: Post-quota analysis. *Competitiveness Review: An International Business Journal*, **32**(7), 1–20. DOI: 10.1108/CR-11-2020-0136.

Chen, S.-J., and Hwang, C.-L. (1992). Fuzzy multiple attribute decision making methods. Pages 289–486 in *Fuzzy Multiple Attribute Decision Making Methods*. DOI: 10.1007/978-3-642-46768-4_5.

Dehdasht, G., Ferwati, M. S., Zin, R. M., & Abidin, N. Z. (2020). A hybrid approach using entropy and TOPSIS to select key drivers for a successful and sustainable lean construction implementation. PLOS ONE, 15(2), e0228746. https://doi.org/10.1371/journal.pone.0228746.

Dolphin, R., Smyth, B., Xu, Y., & Dong, R. (2021). Measuring Financial Time Series Similarity with a View to Identifying Profitable Stock Market Opportunities (pp. 64–78). https://doi.org/10.1007/978-3-030-86957-1_5.

Du, C., Swigonová, Z., and Messing, J. (2006). Retrotranspositions in orthologous regions of closely related grass species. *BMC Evolutionary Biology*, **6**, 62. DOI: 10.1186/1471-2148-6-62.

Farzanegan, M. R., & Markwardt, G. (2009). The effects of oil price shocks on the Iranian economy. Energy Economics, 31(1), 134–151. https://doi.org/10.1016/j.eneco.2008.09.003.

Ghafarian Nia, S. A., G., Korkzan, M. M., Seyedalikhani, S. A., Javaheri, P. M., Ahmadi, M., Tabatabaei, M., & Aghbashlo, M. (2025a). Accelerating renewable ammonia production via machine learning-driven photocatalysis. Renewable Energy, 124153. https://doi.org/10.1016/j.renene.2025.124153.

Ghafarian Nia, S. A., Shahbeik, H., Ahmadi, M., Golvirdizadeh, M., Shafizadeh, A., Javaheri, P. M., ... & Aghbashlo, M. (2025b). Machine learning optimization of catalytic CO_2 hydrogenation to dimethyl ether for clean fuel production. Journal of Environmental Chemical Engineering, 118064. https://doi.org/10.1016/j.jece.2025.118064.

Ghafarian Nia, S. A., Shahbeik, H., Shafizadeh, A., Rafiee, S., Hosseinzadeh-Bandbafha, H., Kiehbadroudinezhad, M., Sheikh Ahmad Tajuddin, S. A. F., Tabatabaei, M., & Aghbashlo, M. (2025c). Machine learning-driven optimization for sustainable CO2-to-methanol conversion through catalytic hydrogenation. Energy Conversion and Management, 325, 119373. https://doi.org/10.1016/j.enconman.2024.119373.

Guo, W., Shafizadeh, A., Shahbeik, H., Rafiee, S., Motamedi, S., Ghafarian Nia, S. A., Nadian, M. H., Li, F., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2024). Machine learning for predicting catalytic ammonia decomposition: An approach for catalyst design and performance prediction. Journal of Energy Storage, 89, 111688. https://doi.org/10.1016/j.est.2024.111688.

Hongliang, G., Zhiyao, Z., Ahmadianfar, I., Escorcia-Gutierrez, J., Aljehane, N. O., and Li, C. (2024). Multi-step influenza forecasting through singular value decomposition and kernel ridge regression with MARCOS-guided gradient-based optimization. *Computers in Biology and Medicine*, **169**, 107888. DOI: 10.1016/j.compbiomed.2023.107888.

Hou, T.-Y., Tseng, Y.-C., & Yuan, C. W. (Tina). (2024). Is this AI sexist? The effects of a biased AI's anthropomorphic appearance and explainability on users' bias perceptions and trust. International Journal of Information Management, 76, 102775. https://doi.org/10.1016/j.ijinfomgt.2024.102775.

International Trade Center. (2024). Trade Map Home. *Trade Map*. URL: https://www.trademap.org/Index.aspx.

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Quality & quantity, 52(4), 1957-1976.

Kahouli, B., and Maktouf, S. (2015). The determinants of FDI and the impact of the economic crisis on the implementation of RTAs: A static and dynamic gravity model. *International Business Review*, **24**(3), 518–529. DOI: 10.1016/j.ibusrev.2014.10.009.

Karlsson, M. A., Langton, M., Innings, F., Malmgren, B., Höjer, A., Wikström, M., & Lundh, Å. (2019). Changes in stability and shelf-life of ultra-high temperature treated milk during long-term storage at different temperatures. Heliyon, 5(9). https://doi.org/10.1016/j.heliyon.2019.e02431.

Kern, R., Al-Ubaidi, T., Sabol, V., Krebs, S., Khodachenko, M., & Scherf, M. (2020). Astro- and Geoinformatics – Visually Guided Classification of Time Series Data. In Knowledge Discovery in Big Data from Astronomy and Earth Observation (pp. 267–282). Elsevier. https://doi.org/10.1016/B978-0-12-819154-5.00025-4.

Luo, Y., and Bano, S. (2020). Modelling New Zealand dairy products: Evidence on export survival and duration. *Australian Journal of Agricultural and Resource Economics*, **64**(3), 605–631. DOI: 10.1111/1467-8489.12372.

Madanchian, M., & Taherdoost, H. (2023). A comprehensive guide to the TOPSIS method for multi-criteria decision making. Sustainable Social Development, 1(1). https://doi.org/10.54517/ssd.v1i1.2220.

Majidian, M., and Dourandish, A. (2024). Prioritizing Iran's saffron target markets based on market competition indices. *Journal of Agricultural Economics and Development*, **38**(2), 177–194. DOI: 10.22067/jead 2024.87254.1258.

Majidian, M., Roshani, E., & Zeinali, Y. Z. (2022). Iran's export competitiveness in the supply of pistachio products in target markets, *NeuroQuantology*, 20:(11), 1062–1072.

Popko, O., Malchyk, M., Tolchanova, Z., Salamakha, O., and Illin, V. (2024). Logistic model for forecasting the growth of Ukrainian dairy exports. *BIO Web of Conferences*, **114**, 01029. DOI: 10.1051/bioconf/202411401029.

Rasoulinezhad, E., & Jabalameli, F. (2019). Russia-EU gas game analysis: evidence from a new proposed trade model. Environmental Science and Pollution Research, 26(24), 24482-24488. https://doi.org/10.1007/s11356-019-05681-2.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees, and support vector machines. Ore Geology Reviews, 71, 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.

Sun, M., Yang, J., Yang, C., Wang, W., Wang, X., & Li, H. (2024). Research on prediction of PPV in open-pit mine used RUN-XGBoost model. Heliyon, 10(7), e28246. https://doi.org/10.1016/j.heliyon.2024.e28246.

Tasneem, S., Ageeli, A. A., Alamier, W. M., Hasan, N., & Goodarzi, M. (2024). Development of machine learning-based models for describing processes in a continuous solar-driven biomass gasifier. International Journal of Hydrogen Energy, 52, 718–738. https://doi.org/10.1016/j.ijhydene.2023.08.043.

Tiits, M., Kalvet, T., Ounoughi, C., and Ben Yahia, S. (2024a). Relatedness and product complexity meet gravity models of international trade. *Journal of Open Innovation: Technology, Market, and Complexity*, **10**(2), 100288. DOI: 10.1016/j.joitmc.2024.100288.

Tiits, M., Kalvet, T., Ounoughi, C., and Ben Yahia, S. (2024b). Relatedness and product complexity meet gravity models of international trade. *Journal of Open Innovation: Technology, Market, and Complexity*, **10**(2), 100288. DOI: 10.1016/j.joitmc.2024.100288.

Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. Journal of Big Data, 11(1), 44. https://doi.org/10.1186/s40537-024-00905-w.

Wang, Y., Shahbeik, H., Moradi, A., Rafiee, S., Shafizadeh, A., Khoshnevisan, B., Ghafarian Nia, S. A., Nadian, M. H., Li, M., Pan, J., Tabatabaei, M., & Aghbashlo, M. (2024). Predictive modeling for hydrogen storage in functionalized carbonaceous nanomaterials using machine learning. Journal of Energy Storage, 97, 112914. https://doi.org/10.1016/j.est.2024.112914.

World Bank. (2024). World Bank Data. *World Bank*.

کاربرد یادگیری ماشین در اولویت‌بندی راهبردی بازار در بخش صادرات لبنیات

مرتضی مجیدیان، و اسماعیل پیش‌بهار

چکیده

این مطالعه چارچوبی نوآورانه برای بهینه‌سازی صادرات لبنیات ایران از طریق تلفیق یادگیری ماشین و تکنیک‌های تصمیم‌گیری چندمعیاره ارائه می‌دهد. با بهره‌گیری از یک پایگاه داده جامع ۲۰ ساله (۲۰۰۳ تا ۲۰۲۲) برگرفته از منابع معتبر بین‌المللی، چهار مدل یادگیری ماشین شامل رگرسیون بگینگ، رگرسیونCatBoost ، رگرسیون گرادیان تقویتی و XGBoost برای پیش‌بینی مقادیر صادرات لبنیات به‌کار گرفته شد. در میان این مدل‌ها، رگرسیون CatBoost بالاترین دقت پیش‌بینی را نشان داد و ضریب تعیین 0.93 را به دست آورد. برای افزایش قابلیت تفسیر نتایج، تحلیل SHAP به‌کار گرفته شد که نشان داد جمعیت، اندازه اقتصادی و ظرفیت تجاری مهم‌ترین عوامل مؤثر بر عملکرد صادرات هستند. هم‌زمان، روش TOPSIS برای اولویت‌بندی بازارهای بالقوه صادراتی بر اساس معیارهای اقتصادی و تجاری استفاده شد و کشورهای ترکیه، عراق و پاکستان به دلیل نزدیکی جغرافیایی، تقاضای بازار و سازگاری تجاری به‌عنوان مقاصد اصلی شناسایی شدند. این رویکرد دوگانه با ترکیب تحلیل‌های پیش‌بینانه و رتبه‌بندی راهبردی بازار، بینش‌های کاربردی برای سیاست‌گذاران و صادرکنندگانی که در پی تقویت اقتصاد غیرنفتی ایران هستند، فراهم می‌کند. یافته‌ها بر نقش حیاتی بازارهای منطقه‌ای و زیرساخت‌های تجاری در ارتقای رقابت‌پذیری صادرات لبنیات تأکید دارند. این پژوهش با بهره‌گیری از تحلیل‌های پیشرفته، از توسعه پایدار کشاورزی و تنوع‌بخشی اقتصادی در ایران پشتیبانی می‌کند و به آسیب‌پذیری اقتصاد نفت‌محور کشور پاسخ می‌دهد. روش‌شناسی و نتایج این مطالعه بنیانی مستحکم برای راهبردهای آینده صادرات فراهم می‌آورد و هم‌افزایی میان پیش‌بینی داده‌محور و تصمیم‌گیری نظام‌مند در بهینه‌سازی تجارت کشاورزی را برجسته می‌سازد.