

Understanding Lag-Based Feature Learning in Export and Import Trade Forecasting Using Machine Learning

Balasubramanian S.

Research Scholar, Annamalai University, Tamil Nadu, India

Natarajan M.

Assistant Professor / Programmer, Annamalai University, Tamil Nadu, India

Abstract— This paper examines lag-based feature learning in export and import trade forecasting using machine learning. Rather than focusing on forecasting accuracy, the study aims to understand how models learn from historical trade values. Monthly export and import data are converted into a supervised learning framework using lagged features, where past trade values serve as inputs and the next month's trade value is the output. Tree-based machine learning models are applied due to their ability to capture nonlinear patterns and provide feature importance measures. Feature importance is used as the main analytical tool to identify which lagged trade values contribute most to model learning. The results show that recent export values play a dominant role, while selected medium-term import lags also influence learning. In contrast, distant historical lags have limited impact. These findings indicate that trade forecasting models rely on selective temporal dependencies rather than uniformly using all past observations, highlighting the usefulness of feature-based learning analysis in trade forecasting.

Keywords—Trade forecasting, Machine learning, Lagged features, Feature learning, Feature importance, Export and import data, Big data analytics

I. INTRODUCTION AND RELATED WORK

Forecasting export and import trade is important for economic planning, policy formulation, and business decision making. Trade time series show strong temporal dependence, where past trade values influence future outcomes. As a result, historical trade data are commonly used as inputs in forecasting models.

Machine learning methods have gained increasing attention in trade forecasting due to their ability to learn nonlinear patterns from data. XGBoost, introduced by Chen and Guestrin, is an efficient tree-boosting framework that performs well on structured and time-dependent data, including trade series [1]. Breiman earlier proposed the Random Forest algorithm, which established the foundation of ensemble tree learning and demonstrated how collections of decision trees learn selectively from input variables [2]. These tree-based models are widely used in economic forecasting because of their flexibility and strong learning capability.

As machine learning models became more complex, the need to understand how they learn from data also increased. SHAP, proposed by Lundberg and Lee, provides a unified

framework for interpreting model predictions by quantifying feature contributions [3]. This approach was later extended through TreeSHAP, which enables efficient and consistent feature attribution for tree-based models [4]. Altmann and coauthors emphasized that feature importance measures should be interpreted with care and highlighted the importance of relative comparisons among features [5].

In the forecasting literature, Makridakis and colleagues highlighted the growing role of machine learning methods while noting that most studies continue to emphasize predictive accuracy [6]. Several studies have applied machine learning models to international and national trade forecasting and demonstrated their practical relevance [7], [8]. However, these studies largely focus on forecast performance rather than on understanding how historical trade values influence model learning behavior. Recent work by Balasubramanian and Natarajan examined feature learning behavior in trade forecasting and showed that machine learning models rely selectively on historical trade information during the learning process. Nevertheless, the specific role of temporal lag structures in shaping model learning remains underexplored.[9]

Despite the increasing use of machine learning in trade forecasting, limited attention has been given to lag-based feature learning. This study addresses this gap by focusing on how machine learning models learn from lagged export and import values. Feature importance is treated as the main analytical outcome to examine which historical trade inputs contribute most to the learning process, rather than prioritizing accuracy-based evaluation.

II. RESEARCH OBJECTIVES AND QUESTIONS

This section outlines the objectives and research questions of the study. The emphasis is on examining how lag-based features influence the learning process of machine learning models in export and import trade forecasting.

Research Objectives: The objectives of the study are:

- To investigate how machine learning models learn temporal patterns from lagged export and import data.
- To examine the contribution of different lagged trade inputs to the learning process.

- To identify dominant lag structures that influence export and import trade forecasting.
- To analyze feature importance patterns as indicators of lag-based learning behavior.
- To frame trade forecasting as a feature learning problem within a data analytics context rather than a pure accuracy task.

Research Questions: The study seeks to answer the following questions:

- How do machine learning models learn from lagged export and import trade values?
- Which lagged trade features have the strongest influence on model learning?
- Are certain lag lengths consistently important across the learning process?
- How can feature importance results improve understanding of temporal trade patterns?

III. DATA SOURCE AND PREPARATION

The analysis is based on monthly export and import trade data observed over a continuous time period. The dataset represents aggregate trade values and reflects both gradual structural changes and short-term variations in trade activity. Each record corresponds to a single month, which makes the data suitable for temporal analysis and lag-based modeling.

Export and import series are treated as related but distinct time-dependent variables. Before analysis, the data were examined for missing entries and basic inconsistencies. Only minimal preprocessing was applied to maintain the original structure of the trade series. No external economic indicators or derived variables were included. This ensures that the learning behavior observed in the models is driven solely by historical trade information.

For machine learning analysis, the monthly trade series were organized into a supervised learning format. Past export and import values were used as explanatory inputs, while the trade value in the subsequent month was defined as the prediction target. This setup allows the models to learn temporal relationships directly from historical observations.

To reflect realistic forecasting conditions, the dataset was divided into training and testing subsets using a time-aware splitting strategy. The chronological order of observations was preserved to avoid information leakage. This data preparation approach supports transparent analysis of lag-based feature learning in export and import trade forecasting.

IV. LAG DESIGN AND FEATURE FORMATION

The purpose of feature construction in this study is to represent historical trade behavior in a form that allows machine learning models to learn temporal patterns. Since export and import values evolve over time, information from previous months is treated as a source of predictive structure.

The original monthly trade series are reorganized into a learning matrix by shifting past export and import values forward in time. Each observation is described by trade values

from earlier months, while the outcome corresponds to the trade value in the following period. This transformation enables the learning algorithm to associate past trade conditions with future outcomes.

Lagged inputs are created separately for export and import series to preserve their individual temporal characteristics. A range of lag lengths is considered so that both recent trade movements and more persistent effects are represented. Shorter lags reflect immediate market conditions, while longer lags capture delayed responses in trade activity.

The lag construction process introduces missing values at the beginning of the series. These observations are removed to maintain consistency in the learning dataset. No scaling, smoothing, or derived indicators are applied. This choice ensures that feature importance results can be directly interpreted in terms of actual historical trade values.

By structuring the data in this manner, the feature set allows clear examination of which past trade periods are emphasized by the learning process. This design supports the study's objective of understanding lag-based feature learning rather than enhancing forecast accuracy.

V. MODELING FRAMEWORK AND LEARNING STRATEGY

This study treats export and import trade forecasting as a learning problem centered on historical information. The modeling framework is designed to examine how machine learning models absorb and prioritize lagged trade values rather than to optimize forecasting accuracy.

The learning task is formulated as a supervised regression problem. Lagged export and import values from previous months form the input space, while the trade value in the subsequent month is defined as the output. Each observation therefore represents a snapshot of past trade conditions linked to a future outcome.

Tree-based machine learning models are used because they can capture nonlinear relationships and provide direct measures of feature relevance. During training, the models construct decision rules by repeatedly selecting lagged trade inputs that best explain variation in the output. This process naturally reveals which past trade values are emphasized during learning.

Model training follows a time-aware strategy that respects the chronological order of observations. Only past data are used to predict future values, which reflects realistic forecasting conditions. After training, feature importance scores are extracted and analyzed to study lag-based learning behavior. The emphasis is on interpreting these scores to understand temporal feature usage rather than on comparing predictive performance.

VI. RESULTS AND ANALYSIS

This section presents the empirical results of the lag-based feature learning analysis. The emphasis is on understanding how machine learning models utilize lagged export and import values during learning.

Lagged Feature Structure: Table 1 summarizes the lagged input variables constructed from the export and import time series. A total of twelve lagged features were generated,

consisting of six export lags and six import lags. Each feature represents trade information from a specific past month relative to the prediction period.

Feature Name	Variable Type	Lag Length (Months)
Export_lag_1	Export	1
Export_lag_2	Export	2
Export_lag_3	Export	3
Export_lag_4	Export	4
Export_lag_5	Export	5
Export_lag_6	Export	6
Import_lag_1	Import	1
Import_lag_2	Import	2
Import_lag_3	Import	3
Import_lag_4	Import	4
Import_lag_5	Import	5
Import_lag_6	Import	6

Table 1. Lagged Trade Features Used in the Model

This structured input design allows the model to learn temporal relationships directly from historical trade values.

Learning Model Representation : The learning process can be formally expressed as:

$$y(t) = f(x(t-1), x(t-2), \dots, x(t-k)) + e(t) \quad (1)$$

where $y(t)$ denotes the trade value at time t , $x(t-k)$ represents lagged export or import values, $f(\cdot)$ denotes the tree-based learning function, and $e(t)$ captures unexplained variation.

Feature Importance Results: Table 2 presents the ranked feature importance values obtained from the trained machine learning model. These values quantify the relative contribution of each lagged input to the learning process.

Rank	Feature	Importance
1	Export_lag_1	0.321709
2	Import_lag_4	0.283721
3	Import_lag_5	0.192506
4	Export_lag_5	0.057294
5	Import_lag_6	0.031814
6	Export_lag_2	0.030502
7	Import_lag_1	0.025065
8	Import_lag_2	0.024759
9	Import_lag_3	0.018878
10	Export_lag_6	0.005465
11	Export_lag_3	0.004445
12	Export_lag_4	0.003844

Table 2. Feature Importance Ranking of Lagged Trade Inputs

Visual Interpretation of Feature Importance: Figure 1 illustrates the feature importance values reported in Table 2 using a bar chart. The figure highlights the dominance of recent export lags and selected import lags, while also

showing a sharp decline in importance for distant historical lags.

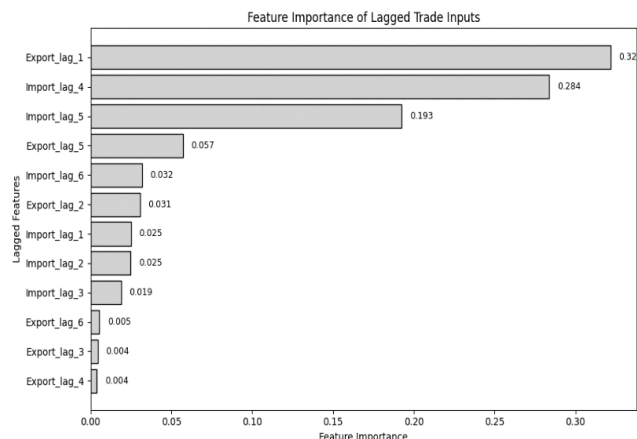


Figure 1. Feature Importance Distribution Across Lagged Trade Inputs

The visual pattern reinforces the tabular results by clearly distinguishing dominant, moderate, and weak lag contributions.

Summary of Findings: The empirical results demonstrate that the machine learning model relies on a selective subset of lagged trade features during learning. Recent export values and certain medium-term import lags play a central role, while older trade information contributes marginally. These findings provide concrete evidence that lag-based feature learning governs the model's behavior rather than uniform use of all historical observations.

VII. DISCUSSION

The objective of this study was to understand how machine learning models learn from lagged export and import trade values. The discussion interprets the feature importance results presented in Table 2 and the visual patterns shown in Figure 1.

The results indicate that recent export values play a dominant role in the learning process. The high importance of Export_lag_1 suggests that short-term export movements contain the most relevant information for predicting future trade values. This finding reflects strong temporal dependence in export series, where recent performance strongly influences near-term outcomes.

At the same time, selected import lags show meaningful influence on learning behavior. Medium-term import lags, particularly those around four to five months, contribute more than several export lags. This pattern suggests delayed interactions between import activity and future trade levels. Such effects may reflect adjustment processes in production, inventory, or supply chains that operate with time delays.

The gradual decline in importance across higher-order lags indicates that distant historical trade values have limited relevance for current forecasting. This confirms that the learning function does not rely uniformly on all past observations. Instead, it selectively emphasizes recent and

medium-term information, which aligns with the lag-based learning structure defined by the model formulation.

The learning model representation clarifies how historical trade values are mapped to future outcomes. The dominance of specific lagged inputs observed in Figure 1 represents the practical realization of this learning function. Feature importance patterns therefore provide direct insight into how the model approximates trade dynamics.

Overall, the discussion shows that feature-based learning analysis offers insights that extend beyond traditional accuracy-focused evaluation. By examining which lagged trade values guide model learning, the study improves transparency and supports more informed interpretation of machine learning-based trade forecasts.

VIII. SCOPE AND BOUNDARIES OF THE STUDY

While the study provides clear insights into how tree-based machine learning models learn from lagged trade values, certain limitations should be acknowledged.

First, the analysis relies exclusively on historical export and import data. No external macroeconomic variables such as exchange rates, inflation, or global demand indicators are included. As a result, the learning behavior examined in this study reflects only the information contained in past trade values and does not account for external economic influences.

Second, the study focuses on a fixed lag structure derived from monthly data. Although multiple lag lengths are included, the analysis does not explore very long lag horizons or adaptive lag selection mechanisms. Different lag configurations may reveal additional learning patterns, particularly in periods of structural change.

Third, the modeling framework is limited to tree-based machine learning models. While these models are well suited for feature importance analysis, the findings may not directly generalize to other model families such as neural networks or linear time series models. The study does not aim to compare learning behavior across different algorithmic classes.

Fourth, feature importance is interpreted as a measure of learning contribution rather than causal impact. Although importance scores indicate which lagged inputs influence model decisions, they do not establish causal relationships between past and future trade values. The results should therefore be understood as reflecting model learning behavior rather than economic causality.

Finally, the analysis is conducted at an aggregate trade level. Disaggregated trade data by sector, commodity, or trading partner may exhibit different temporal learning patterns. The current findings may not fully capture such heterogeneity.

Despite these limitations, the study provides a transparent and focused examination of feature-based learning behavior in trade forecasting. The acknowledged constraints also highlight clear directions for future research, which are discussed in the following section.

IX. CONCLUSION AND FUTURE SCOPE

This study examined lag-based feature learning in export and import trade forecasting using machine learning methods. The focus was on understanding how models learn from

historical trade values rather than on improving forecasting accuracy. By constructing lagged export and import features and analyzing feature importance, the study provided clear evidence of selective temporal learning behavior.

The results show that recent export values play a dominant role in the learning process, while selected medium-term import lags also contribute meaningfully. In contrast, distant historical trade values have limited influence. These findings indicate that machine learning models emphasize specific temporal patterns rather than relying uniformly on all past observations. Feature importance analysis proved effective in revealing how historical trade information is used during learning.

The study contributes a transparent and interpretable perspective to machine learning-based trade forecasting. By treating feature importance as the primary analytical outcome, it enhances understanding of trade dynamics and supports explainability-driven data analytics.

Future research can extend this framework by examining alternative lag lengths and rolling time windows to test the stability of learning patterns. The approach can also be applied separately to import-focused forecasting or expanded to multivariate trade systems involving multiple countries or commodities. Incorporating external economic indicators may further reveal how models balance internal trade history with broader economic signals. Comparative analysis across different machine learning models may also provide deeper insight into variations in lag-based learning behavior.

REFERENCES

- [1] [1] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," arXiv preprint arXiv:1603.02754, 2016. <https://arxiv.org/abs/1603.02754>
- [2] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. <https://link.springer.com/article/10.1023/A:1010933404324>
- [3] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," arXiv preprint arXiv:1705.07874, 2017. <https://arxiv.org/abs/1705.07874>
- [4] S. M. Lundberg, G. Erion, and S. I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," arXiv preprint arXiv:1802.03888, 2018. <https://arxiv.org/abs/1802.03888>
- [5] A. Altmann, L. Tološi, O. Sander, and T. Lengauer, "Permutation Importance: A Corrected Feature Importance Measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010. <https://academic.oup.com/bioinformatics/article/26/10/1340/193348>
- [6] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018. <https://www.sciencedirect.com/science/article/pii/S0169207019301128>
- [7] F. A. Batarseh and E. Gonzalez, "Predicting International Trade Flows Using Machine Learning," arXiv preprint arXiv:1910.03112, 2019. <https://arxiv.org/pdf/1910.03112.pdf>
- [8] H. Jošić and B. Žmuk, "Machine Learning in International Trade Forecasting," *EconStor Working Paper*, 2022. <https://www.econstor.eu/bitstream/10419/318808/1/1835756867.pdf>
- [9] S. Balasubramanian and M. Natarajan, "Feature Learning Behavior in Trade Forecasting: Evidence from Tree-Based Machine Learning Models," *Eng. Technol. Appl. Sci. Res.*, vol. 16, no. 2, pp. 34097–34101, 2026. <https://etasr.com/index.php/ETASR/article/view/17612>