Explainable AI in Demand Forecasting Machine Learning Models

Bharathram Nagaiah

bharathram.nagaiah@gmail.com

Abstract:
Demand forecasting is a crucial element of effective supply chain management, retail planning, and production scheduling. Although models of machine learning (ML) have been reinventing the way demands are predicted with consistently greater accuracy, the complexity can sometimes be a burden to interpret. Such a transparency deficit may impede trust, choices, and adoption in critical business circumstances. XAI methods are applicable in solving this problem because they help in rendering the process of decision-making of complex ML models comprehensible to human stakeholders. In the given article, the authors consider the use of XAI when demand is forecasted, presenting multiple explainability methods: SHAP, LIME, permutation importance, partial dependence plots, and counterfactual explanations. The paper highlights the role of these tools to reveal the motivators behind predictions, enhance user trust, and help users make well-informed decisions.

Keywords: Explainable AI, demand forecasting, SHAP, LIME, time series, XGBoost, random forest, machine learning, interpretability, supply chain.

1.INTRODUCTION

Effective demand forecasting plays a decisive role in planning across industries by allowing an organization to manage inventory, distribute workforce effectively, implement an opportune marketing campaign, and automate the logistics processes. Bad predictions may cause overstocks, stockouts, higher operation expenses, and loss of sales. [1,2]

Machine learning (ML) models have become a potent way to enhance the precision of forecasts with the passage of years. High-performance methods, including ensemble algorithms (e.g., Random Forest, XGBoost), and deep learning networks (e.g., LSTM) have the power to pull complicated patterns and non-linear connections in huge and dynamic databases. Such models are more precise and flexible than simpler statistical models such as ARIMA or exponential smoothing.

The lack of transparency is cited as an issue in the application of such models to practice in the real world, even though these models are successful predictors. Such black-box models offer little or no information about the making of predictions, further restricting both trust and applicability, compared to low-stakes systems and contexts where the key distinguishing attributes of accountability, transparency, and explainability are at a premium. Regulatory agencies, business leaders, and other professionals working in various fields are not interested in accurate numbers only; they need to know how these predictions were made. [3]

Explainable Artificial Intelligence (XAI) appears as a tool to resolve this problem. XAI is a collection of technologies and approaches to making the decision-making of machine learning models easy and transparent to human beings. XAI can help users in decision-making, trust more, and be able to validate their models as it provides both global and local interpretability.

This paper discusses how XAI can be incorporated into the demand forecasting system. It examines how methods, including SHAP, LIME, permutation significance, ab during rely plots, and counterfactual explanations might help give insight into the behavior of the model, boost stakeholder trust, and eventually enlarge the usefulness of the ML-dependent forecast models utilized in business settings. [4]

2.METHODOLOGY

In order to study the role of XAI in demand forecasting, we provided a comparative analysis of various models of machine learning with different levels of complexity and interpretability. We were trying to find out to what extent explainability methods could be utilized to help reveal the internal logic of the model decision-making process and how these understandings could be altered to increase familiarity and trust.

We chose a mix of three kinds of models, covering a continuum of interpretable/predictive capability:

- **Linear Regression**: A simple, interpretable model used as a baseline to compare with more complex approaches. It allows coefficients to be analyzed directly and so it is a natural yardstick of transparency. [5]
- **Random Forest and XGBoost:** The ensemble models are popular due to their accuracy level and also their non-linear relationship features capability and features interaction. However, they lack intrinsic interpretability, thus requiring post-hoc XAI methods. [6]
- **LSTM (Long Short-Term Memory networks)**: A type of recurrent neural network suitable for sequential and time-series data. LSTM models can capture long-term dependencies in demand trends but are considered black-box models due to their complex internal structures. [7]

Each model was trained on a structured retail sales dataset. The dataset included the following features:

- **Product category** (e.g., electronics, groceries)
- **Promotion status** (whether the product was on sale)
- **Holiday indicators** (public holidays and seasonal events)
- **Pricing information**
- **Historical sales data** (time-series format)

To interpret these models, we applied the following XAI techniques:

1. **SHAP (SHapley Additive exPlanations)**: SHAP was used to compute the contribution of each feature to the final output for both individual predictions (local interpretability) and overall model behavior (global interpretability). SHAP values were visualized to highlight the most influential features across different time periods and product categories. [8]
2. **LIME (Local Interpretable Model-agnostic Explanations)**: LIME was used to interpret individual predictions by perturbing the input and training a simple, interpretable model (like linear regression) locally. This assisted in showing why the model provided a specific forecast for a specific case. [9]
3. **Permutation Importance:** This method referred to the shuffling feature column and assessed the difference in the model accuracy. A greater decline in performance indicated a more significant feature. Such an approach served to confirm SHAP findings and gave extra strength to our findings to understand interpretability. [10]
4. **Partial Dependence Plots (PDP):** PDPs were employed to see the impact of the variation in a given feature on the predicted demand, keeping the contribution of the rest of the features at a mean value. This approach was especially effective in describing the non-linear impact of prices or advertising on sales projections. [11]
5. **Counterfactual Explanations:** Counterfactual analysis enabled us to formulate what-if scenarios by slightly changing the features of the input, and in turn, see how differently the predictions would be made. For example, we tested the impact of a 10% price increase and simulated a holiday period to observe their effects on the demand forecast

These explainability methods were introduced via the machine learning ecosystem in Python. The SHAP values were calculated with the help of the shap library, and permutation importance, and PDPs were performed through ELI5 and scikit-learn. LIME explanations were generated using the lime package, while counterfactuals were manually crafted using custom logic guided by domain expertise. We tested the deep learning models, such as LSTM on Keras, to develop the model and SHAP DeepExplainer to interpret.

The result is a holistic framework whereby we incorporated these XAI methods into our forecasting pipeline, and this allows us not only to get an accurate prediction but also to get an explainable insight that you can communicate to the business stakeholders**.** [12]

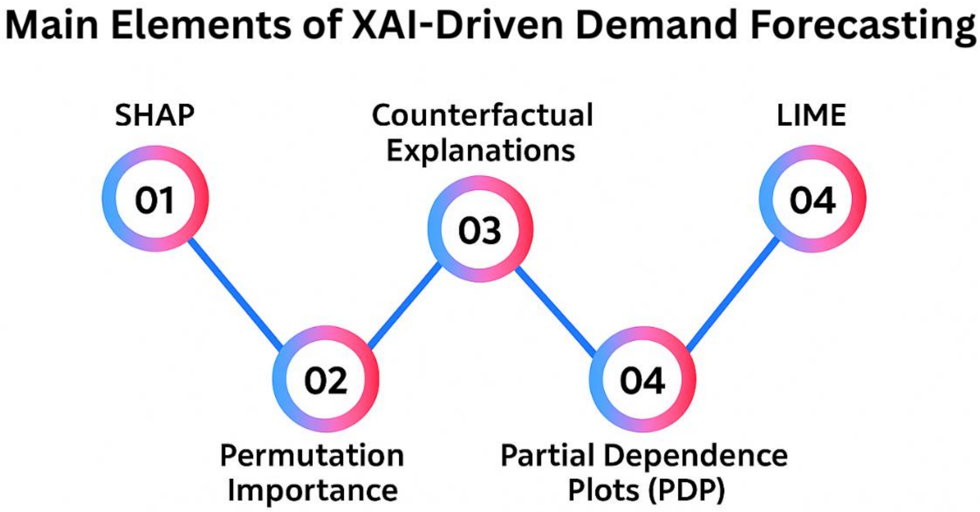

Figure 1. Main XAI-Driven ML Models

When comparing the results of the integration of the Explainable AI (XAI) approaches to demand forecasting models, the results led to meaningful understanding of what machine learning models predict, and that there are also transparency and trustworthiness advantages.
In all the tested models, including Linear Regression, Random Forest, XGBoost, and LSTM, SHAP (SHapley Additive exPlanations) revealed the following three factors as the most influential in predicting the accuracy of future promotions: promotion status, indicators of holidays, and product category. These characteristics had high SHAP values, so they were statistically significant as well as intuitively correct for the retail business. SHAP has been used in ensemble models such as Random Forest and XGBoost to offer a powerful diagnosis tool to business analysts, with single-model versions offering both global interpretations (e.g., average feature influence over all predictions) and local ones (e.g., feature contribution on a given forecast).

LIME (Local Interpretable Model-agnostic Explanations) was used to generate local explanations for individual forecasts. This came in handy, particularly in cases where outliers, which include unforeseen peaks or declines in demand, can be validated with confidence or explored by the stakeholders with confidence. Such localized interpretations were easy to picture and explain to average users, which is why they are perfect in the case of operations teams.

SHAP was complemented with **permutation importance** that measured the loss to the model accuracy rating due to a random shuffle of the single features. The matching of the received results improved the confidence in the interpretability procedure since the model learned the patterns. It also acted as a sanity check that ensured that the model was not maintaining some spurious correlations.
Partial Dependence Plots (PDPs) revealed non-linear interactions—particularly between **price and demand**, where demand dropped sharply beyond certain price thresholds and plateaued within expected discount ranges. This validated the current knowledge on the domain and helped make strategic pricing decisions supported by insights generated through the model.

Users could use counterfactual explanations to run a what-if simulation like the following: What would the forecasted call be at the finish line once we priced it 10 percent more? Or what would have happened had we advertised such a product during a holiday? These simulations empowered planners to model the impact of potential actions, improving strategic forecasting, campaign planning, and inventory control.

In the case of **LSTM models**, which are inherently less transparent due to their sequential deep learning architecture, explainability was more nuanced. Using SHAP's DeepExplainer and **time-distributed feature importance scores**, we identified which temporal features (e.g., sales trends, past demand spikes) influenced the prediction at each time step. While these insights were less granular than those from tree-based models, they still uncovered meaningful patterns. Incorporating **attention mechanisms** further enhanced interpretability, allowing for better traceability of time-dependent signals such as recurring holiday spikes or promotional cycles.

On the whole, such XAI methods not only confirm the correctness of such a model but also change the black-box predictions into working insights, which benefit the business through successful decision-making in terms of clarity and certainty.

3.DISCUSSION

The inclusion of XAI in demand forecasting paths will enable the underserved space between model accuracy and human comprehension. SHAP and LIME are also very helpful to both technical and business users as they come in local and global views. These insights can be used to inform promotional strategies, allocation of resources, and improvement of overall inventory in operational environments.

Besides, explainability enables debugging and model validation by indicating the positions and reasons for the differences between predictions and expectations. It is also significant in regulatory compliance, particularly in the pharmaceutical and financial sectors, where auditability is paramount.

Nevertheless, not everything is going smoothly. XAI also introduces computational cost and, in many cases, it demands skills to be interpreted as well. In addition, not all deep learning models are fully transparent, even when they are capable of processing. Further studies might aim at investigating hybrid models that offer a trade-off between accuracy and interpretation, or even designing new XAI techniques that will be applied to time-series forecasting.

4.CONCLUSION

Explainable AI renders critical capabilities that enable the demand forecasting models to be more transparent, interpretable, and executable. By using SHAP, LIME, and other XAI systems, companies will be able to open the black box of machine learning, gain trust in AI systems, and make sure decisions with more confidence and the use of data. With the ongoing development and wider acceptance of machine learning, the trade-off between accuracy and explainability is likely to become crucial to broadly deploying AI in any application of demand forecasting and beyond.

REFERENCES:

1. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM.
2. Hvitfeldt, E., Pedersen, T. L., & Benesty, M. (2022). *lime: Local Interpretable Model-Agnostic Explanations* (Version 0.5.3) [R package]. CRAN. https://CRAN.R-project.org/package=lime
3. Castel, N., Jarrahi, M. H., & Pomerol, J. C. (2023). Users' trust in black-box machine learning algorithms: Understanding the role of transparency and explainability. *Review of Economics and Management*, *11*(3), 245–262. https://doi.org/10.1108/REGE-06-2022-0100
4. Yvonne. (2025, June 10). *Explainable AI in demand forecasting: How to understand what the model sees*. LinkedIn. https://www.linkedin.com/pulse/explainable-ai-demand-forecasting-how-understand-what-yvonne-74npc
5. Salih, A. M., & Wang, Y. (2024). *Are linear regression models white box and interpretable?* arXiv. https://doi.org/10.48550/arXiv.2407.12177
6. Bassan, S., Provenzano, A., Viale, A., Zanella, M., & Cesa-Bianchi, N. (2025). *What makes an ensemble (un)interpretable?* arXiv. https://arxiv.org/abs/2506.08216
7. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.
8. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4768–4777.

9. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
10. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
11. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.
12. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, *31*(2), 841–887.