



UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

Demand Forecasting Tool Driving the Digital Twin of a Perishable Food Process

This is the peer reviewed version of the following article:

Original

Demand Forecasting Tool Driving the Digital Twin of a Perishable Food Process / Lucantoni, L., Croci, S., Mazzuto, G., Ciarapica, F.E., Bevilacqua, M., Perenzoni, S. - In: IEEE/CAA JOURNAL OF AUTOMATICA SINICA. - ISSN 2329-9274. - 12:11(2025), pp. 2356-2358. [10.1109/JAS.2025.125591]

Availability:

This version is available at: 11566/362352 since: 2026-09-02T14:35:29Z

Publisher:

Published

DOI:10.1109/JAS.2025.125591

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.

This item was downloaded from IRIS Università Politecnica delle Marche (<https://iris.univpm.it>). When citing, please refer to the published version.

Publisher copyright:

IEEE - Postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. To access the final edited and published work see 10.1109/JAS.2025.125591

(Article begins on next page)

Demand Forecasting Tool Driving the Digital Twin of a Perishable Food Process

Laura Lucantoni, Stefano Croci, Giovanni Mazzuto, Filippo Emanuele Ciarapica, Maurizio Bevilacqua, and Severino Perenzoni

Dear Editor,

The food industry emphasizes improving demand forecasting to align production with consumer needs and reduce waste. This letter thus presents a study that integrates artificial intelligence (AI) and digital twin (DT) technologies to enhance decision-making and efficiency in food production. A data-driven DT was implemented in an Italian company for Raspberry production planning, based on a daily demand forecasting tool powered by a dynamic extreme gradient boosting (XGBoost) algorithm. The model achieved a mean absolute percentage error (MAPE) of 16.37% with 1.69 average of absolute extra working hours (AEW) and a tracking signal (TS) range of $[-1.9, +4.3]$. The food processing industry faces significant challenges in achieving sustainable food production, with food waste presenting economic, social, and environmental concerns [1]. Reducing waste aligns with the United Nation's 17 Sustainable Development Goals. However, the food sector struggles with limited data availability and inadequate analytical tools, which hinder effective decision-making for sustainability [2]. Recent studies emphasize the importance of real-time visibility, data analytics, and remote control in addressing these challenges [3]. Emerging technologies, such as AI and DT, are gaining prominence in demand forecasting and perishable food management within the food industry [4], including application in supply chain and operations management, as discussed by Dolgui and Ivanov [3]. However, the integration of AI-based demand forecasting with DT systems remains underexplored in the food sector. To address this gap, the present study developed a data-driven DT system for raspberry production based on a dynamic XGBoost algorithm. This approach optimizes production by efficiently responding to demand fluctuations, improving resource efficiency, and reducing waste. The methodology was validated through 150 tests across various models and database structures, addressing the company's need to rapidly adapt production processes due to the perishable nature of raspberries.

Related work

A systematic search was conducted in Scopus, Web of Science, and Google Scholar using the search string "Digital Twin" AND "Demand Forecasting" AND "Food Industry". The search was limited to titles, abstracts, and keywords, resulting in the identification of approximately 20 articles. Among these, 14 studies were published in 2023, indicating significant growth and a shift toward recognizing the applications of DT in the food sector, particularly for process improvement and demand forecasting [5]. Following a thorough screening process, which included reviewing abstracts, only 7 studies effectively integrated DT with predictive analytics. The majority of the remaining studies focused on digitalizing agriculture or supply chain management (SCM), addressing areas such as logistics efficiency, inventory management, and food security and sustainability [6]. For example, [7] optimized SCM, while [8] applied machine learning for shelf-life prediction and preventive maintenance.

However, three key gaps were identified in the food sector regarding DT applications: 1) When combined with predictive analytics, DTs primarily focus on production scheduling and shelf-life management, with limited attention to demand forecasting; 2) There is limited integration of AI tools with DTs, which are mostly used for monitoring internal conditions (e.g., crop growth or energy use); and 3) The absence of a fully developed demand-driven DT, i.e., a system directly powered by robust demand forecasting tools. To address these limitations, the current study integrates AI-based demand forecasting with DT models. By employing a gradient-boosting approach, the proposed system facilitates demand-driven production planning, thereby enhancing the role of DTs in the food industry.

Research approach

The implementation of a DT requires a systematic approach, divided into key phases to ensure the project's success. The first phase involves defining the scope of the DT, during which objectives and purposes are clarified. This step is critical for establishing a clear vision of the system and its operational boundaries, providing a roadmap for subsequent phases. For Sant'Orsola, a company specializing in small fruit processing, the DT is designed to optimize production activities by integrating information such as sales data, inventory levels, and production line status. Using AI, the DT accurately forecasts customer demand, leveraging these predictions to plan and monitor production activities in real-time, adjusting operations to improve efficiency as needed.

The second phase focuses on analyzing the physical asset. The company operates a two-level facility where materials are received, classified, and stored before being processed on four production lines: two automated and two manual. Each line produces 300 units per hour, with the option to extend shifts to increase production capacity. Demand forecasts, received by 11 am, guide the configuration of the production lines, while orders are confirmed and adjusted, if necessary, in the afternoon.

The third phase regards system modelling and simulation [9]. The DT was developed using Discrete Event Simulations implemented in Python with the Simpy library. Production elements, such as pallets, crates, and operators, are modelled as entities with specific attributes, allowing the DT to replicate real workflows accurately. The simulation includes input and output stations, quality and weight checks, and the internal transport system. This digital representation enables business process analysis, inefficiency identification, and operations optimization.

The final phase focuses on DT intelligence, linking the physical system with the digital model. The DT uses demand forecasting (Fig. 1) and operational data to set production parameters, configure lines, and allocate resources optimally. This integration facilitates real-time system monitoring, driving continuous improvement in the production process.

Data analytics

- The analysis identified key features of raspberry demand, including summer peaks, intra-seasonal fluctuations (e.g., higher demand in May), and weekly trends.
- Saturday showed the lowest sales (619–849 units), while Thursday and Friday the highest ones (up to 3,769 units).
- Missing data on prices and promotions complicated the analysis of anomalies, such as the 27% increase in year X.

To address the current forecasting error of 109.17%, three main KPIs were calculated: AEW (± 1 hour per line, 1200 units/day), MAPE $\leq 20\%$, and TS within $[-6, +6]$. Models were also assessed for their ability to capture weekly seasonality and their computational efficiency.

Table 1. Extraction of Significant Short-Term Tests Results

# Test	Algorithm	AEW	MAPE _x	TS
11-CAL	MLP (Dec X-1)	1,28	18 (23)	$[-9,8; 11]$
4-CAL	GBT (Dec X-1)	0,78	12 (21)	$[6,5; 8,8]$
73-CAL	GBT (Jan–May X)	1,50	16 (22)	$[-13,8; 2,1]$
74-CAL	DGBT (Jan–May X)	1,46	17 (23)	$[-8,2; 4,9]$
80-CAL	DGBT (Mar–May X)	1,49	19 (25)	$[-5,8; 6,4]$

# Test	Algorithm	AEW	MAPE _x	TS
43-DOTY	DGBT (Mar–May X)	1,51	15 (20)	[–6,2; 5,7]
116-DOTY	DGBT–No ordered (Mar–May X)	1,41	16 (28)	[–5; 7,2]
104-WDA	DGBT (Mar–May X)	1,51	13 (17)	[–8; 2]
111-WDA	DGBT–Continuous standardization (Mar–May X)	1,31	12 (16)	[–5; 5]
142-WKB	DXGBT–Continuous standardization (Mar–May X)	1,29	15 (19)	[0,7; 8,2]
144-WKB	DXGBT–Continuous standardization, no ordered (Mar–May X)	1,56	14 (18)	[–0,9; 3,7]
146-WKB	DXGBT–Continuous standardization, no ordered (Mar–May X)	1,69	16 (21)	[–1,9; 4,3]
150-WKB	DXGBT–Continuous standardization, no ordered (Mar–May X)	1,61	15 (19)	[–0,9; 4,3]
137-WKB	XGBT–Continuous standardization, incremental training, no ordered (Mar–May X)	1,89	22 (34)	[1; 22]

AEW (Average Extra Working Hours) = $\frac{\sum_j \left| \frac{Actual_j - Forecast_j}{Hourly\ Units\ Produced_j} \right|}{Number\ of\ tests\ conducted}$, indicating the absolute average of extra working hours per observation.

MAPE_x – Max Mean Absolute Percentage Value

$MMAPE_x$ (Max Mean Absolute Percentage Value) = $\max \left(\left| \frac{Actual - Forecast}{Actual} \right| \times 100 \right)$, indicating the maximum value among the absolute percentage errors for all observations (the average in parentheses)

TS (Tracking Signal) = $\frac{\sum (Actual - Forecast)}{MAD}$, indicating the direction and magnitude of forecast bias

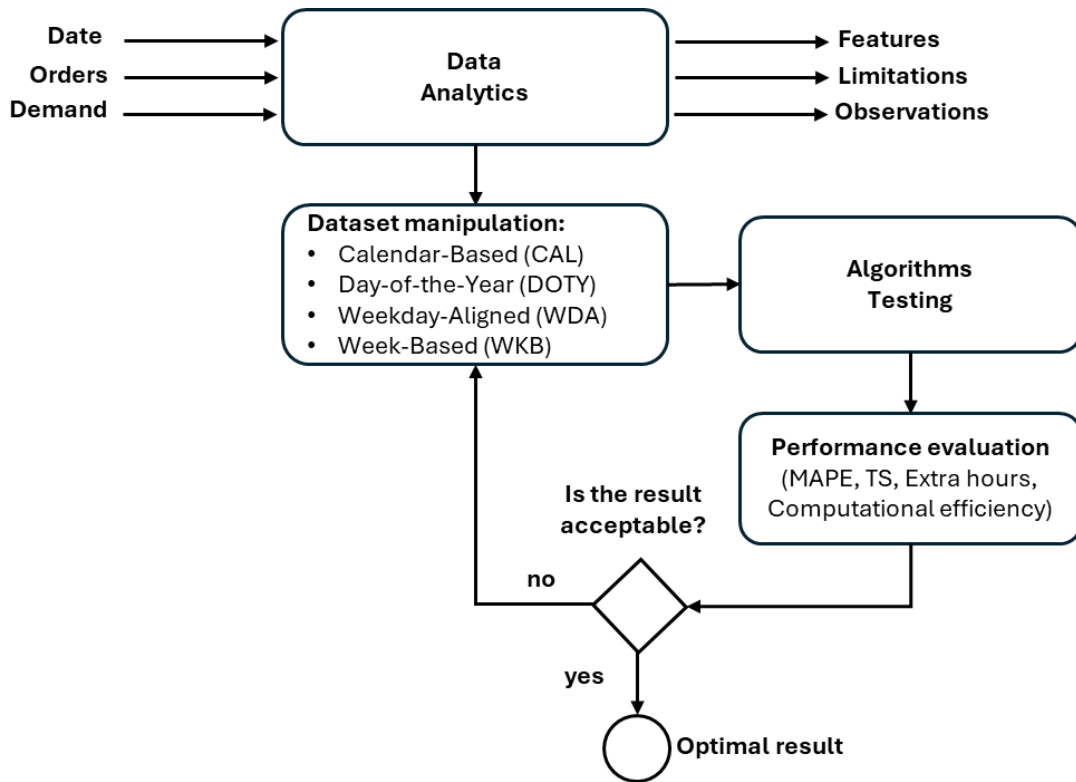


Fig. 1. End-costumers demand forecasting.

Datasets manipulation

- The calendar-based (CAL) database organizes data by date.
- The day-of-the-year (DOTY) database aligns data by the day of the year, allowing for comparison of demand patterns across years.
- The weekday-aligned database (WDA) emphasizes weekly seasonality by grouping data by weekdays within each week of the year.
- The week-based database (WKB) includes data from the preceding six days and introduces the average amount variable, reflecting historical sales trends for the same weekday over the past three years.

Each database was tested with moving averages, lag features, and binary variables.

Algorithm testing and model optimization

The preliminary model selection involved evaluating statistical, neural network, and tree-based models for demand forecasting of $X-1$, using the CAL and DOTY databases. Linear regression failed to capture non-linear trends, while SARIMA lacked incremental training capabilities and struggled without exogenous data, such as pricing. Holt-winters performed poorly, achieving a 27% MAPE and unstable TS. The long short-term model, although effective for capturing long-term trends, performed poorly on weekly seasonality and incurred high computational costs (30 minutes of training time). The gradient boosted tree (GBT) outperformed the alternatives, achieving a MAPE of 12.33%, 0.78 extra working hours, and stable TS (± 6.49), reducing forecast errors from 39.19% to 1.57% compared to the existing system (Table 1). For year X , errors increased due to missing promotional data, but dynamic training reduced short-term gaps from 12.6% to 5.3%, although long-term underestimation and MAPE degradation persisted. In the DOTY Database, GBT's heavy reliance on the "Ordered" variable (46%) led to underestimations, particularly on anomalies such as March 4th, when the order volume was abnormally low (74% error). Errors propagated to subsequent forecasts, but dynamic training (with the

algorithm acronym D) allowed for self-correction within two iterations (Fig. 2). Feature engineering and deseasonalization slightly improved accuracy (March 4th improved by 30%). Still, errors persisted on Saturdays (reduced from 93% to 40%). Removing the orders improved TS, indicating its non-essential role and reducing the risk of overfitting.

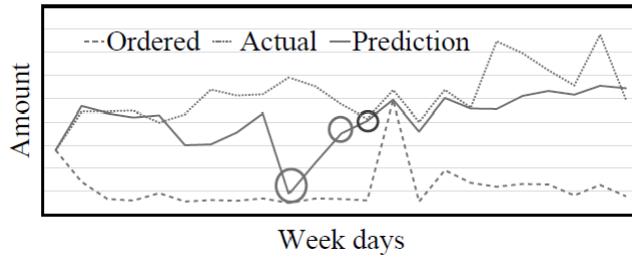


Fig. 2. Anomalies management of the dynamic GBT model.

The WDA database captured weekly seasonality, reducing shortterm errors, although TS dropped to -8 , reflecting long-term underestimation. Structural complexity, particularly with leap years, remained challenging. To address this limitation, the WKB database showed further improvements, with a MAPE of 15% and 1.29 extra working hours, identifying “day of the week” as the key predictor (41%). Removing the variable of orders reduced TS error and simplified operations. Incremental training favored XGBoost over Light-GBM for faster convergence. The dynamic XGBoost (DXGBT) model, incorporating continuous standardization and the pseudo huber error loss function (Test 146–Fig. 3), achieved the best results, also proving reliable for long-term forecasting, with an MAPE of 19% and balanced TS range of $[-4.2; +5.67]$.

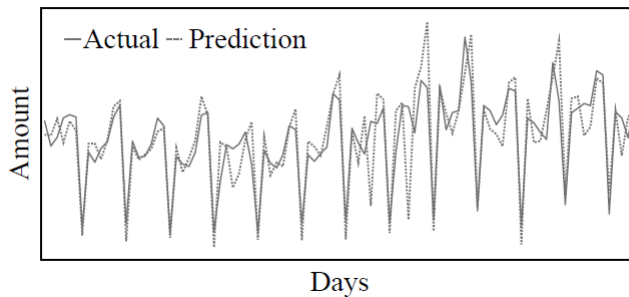


Fig. 3. Best predictive model results

The model’s average forecasted demand of 2 880 units closely aligned with the actual demand of 2394 units, showcasing its reliability. Producing 2880 units required 9.6 hours. Efficient production strategies, such as activating two lines, effectively managed deviations, minimizing extra working hours and optimizing operational efficiency. In cases of maximum overestimation (3874 units vs. 2836 units, +3.46 hours), production remained efficient. Two lines completed the forecasted demand in 6.5 hours, while actual demand required only 5 hours, leaving 3 hours of downtime per line for quality control and preparation, thus ensuring operational efficiency.

Conclusions

This study developed a DT integrated with a DXGBoost model to enhance demand forecasting in raspberry production, achieving an 85.9% reduction in forecasting error over three months, capturing weekly and seasonal trends, despite missing pricing and promotional data. The week-based database with continuous standardization improved MAPE, TS, and operational efficiency, effectively aligning production schedules with actual demand. The final implementation of the whole system is scheduled

for January 2025. Future enhancements will focus on integrating promotional data, expanding the model to encompass additional products, and evolving the DT into a collaborative system with real-time synchronization across the supply chain.

References

- [1] V. Amicarelli and C. Bux, "Food waste in Italian households during the Covid-19 pandemic: A self-reporting approach," *Food Secur.*, vol. 13, no. 1, pp. 25–37, Feb. 2021,
- [2] S. Abbas, H. Munir, and Y. Ahmad, "Integrating ECO-Labeling and green advertising in achieving Sustainable Development Goal 12," *Business Strategy & Development*, vol. 7, no. 2, p. e378, Jun. 2024.
- [3] A. Dolgui and D. Ivanov, "Internet of behaviors: Conceptual model, practical and theoretical implications for supply chain and operations management," *Int. J. Prod. Res.*, pp. 1–8, Jul. 2024, doi: 10.1080/00207543.
- [4] H. Ding, J. Tian, W. Yu, et al., "The application of artificial intelligence and big data in the food industry," *Foods*, vol. 12, no. 24, p. 4511, Dec. 2023.
- [5] P. Maheshwari, S. Kamble, A. Belhadi, V. Mani, and A. Pundir, "Digital twin implementation for performance improvement in process industries- A case study of food processing company," *Int. J. Prod. Res.*, vol. 61, no. 23, pp. 8343–8365, Dec. 2023,
- [6] R. Sirazetdinov, R. Sabitov, G. Smirnova, S. Sabitov, and A. Eponeshnikov, "Building a digital twin of agricultural production based on mathematical modeling of sustainability and development processes," *Adv. Syst. Sci. Appl.*, vol. 22, no. 4, pp. 116–125, 2022.
- [7] T. Y. Melesse, M. Bollo, V. Di Pasquale, and S. Riemma, "Digital twin for inventory planning of fresh produce," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 2743–2748, 2022.
- [8] T. Eppinger, G. Longwell, P. Mas, K. Goodheart, U. Badiali, and R. Aglave, "Increase food production efficiency using the executable digital twin (xDT)," *Chem. Eng. Trans.*, vol. 87, pp. 37–42, 2021.
- [9] S. Croci, G. Mazzuto, M. Ortenzi, F. E. Ciarapica, M. Bevilacqua, and G. Osler, "A digital twin modeling for the production line optimized management in the soft and deformable food sector," in *Proc. IEEE Int. Conf. Engineering, Technology, and Innovation*, Jun. 2024, pp. 1–8