Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 278 (2026) 898–905

Procedia
Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2025

Predicting Order Activity Sequence Using Contextual Process Mining

Diogo A. Barbeiro[a], Ricardo F. G. Martinho[b,*], Carlos J. R. Ferreira[a]

[a]*ESTG, Polytechnic University of Leiria, Apartado 4133, 2411-901 Leiria, Portugal.*
[b]*INESCC-DL, ESTG, Polytechnic University of Leiria, Portugal, Apartado 4133, 2411-901 Leiria, Portugal.*

Abstract

Logistics processes depend heavily on changing conditions and making accurate forecasts is becoming more and more important to preventing delays and/or predict risks. Predictive Process Monitoring has advanced through deep learning and process-mining approaches, yet current methods often lack interpretability and lose accuracy when context varies. Recent research shows that contextual factors can improve predictions, but their integration into transparent, model-driven frameworks remains limited.

This article presents a context-aware predictive approach that filters historical event logs by important attributes, discovers process models with the Inductive Miner, and predicts future activities and timestamps using token-based replay and polynomial regression. Experiments with real logistics data show that incorporating context reduces prediction errors, while the use of process mining ensures an interpretable and operationally practical forecasting solution for logistics environments.

Keywords: Process Mining; Predictive Monitoring; Spatio-temporal Analysis; Usability Testing; Decision Support Systems.

1. Introduction

Predictive Process Monitoring (PPM) uses process mining and machine learning to forecast how running process instances will evolve based on historical event data [1]. Typical prediction goals include identifying the next activity,

---

* Corresponding author.
E-mail address: ricardo.martinho@ipleiria.pt

1877-0509

estimating the remaining time, or assessing the likelihood of specific outcomes [2]. By combining process discovery with predictive modeling, PPM allows organizations to anticipate deviations and act before performance issues occur.

In logistics, processes are long and often influenced by external conditions. Maintaining accurate forecasts is important because even small delays can spread through the chain, leading to missed delivery windows or service-level penalties. When an order is at risk of falling behind, early detection is often the only chance to intervene before the situation escalates. Predictive monitoring offers a data-driven way to detect potential delays early and to improve resource planning [3]. Real-time forecasts help managers make proactive rather than reactive decisions. For example, predicting the remaining time of an order or identifying cases likely to miss delivery targets can support better routing and dispatching strategies [4].

Previous studies have shown that process-mining-based forecasting improves transparency and coordination in complex operations [6]. In this work, we apply predictive process mining to large-scale order fulfillment data. Using the Inductive Miner from the PM4Py framework [7], we derive process models from event logs and enrich them with contextual attributes such as route, transport mode, and system workload. This combination of process mining and contextual filtering enables more accurate activity and timestamp predictions under varying operational conditions.

2. Related Work

Predictive Process Monitoring (PPM) combines process mining and machine learning to forecast the future behavior of ongoing processes based on historical event logs [8, 9]. Unlike traditional analysis, which focuses only on completed cases, PPM predicts outcomes such as the next activity, remaining time, or the probability of delay [1].

Early studies relied on explicit process models, such as Petri nets or transition systems, to simulate how running cases would continue. As event data became more available, research shifted toward data-driven methods that learn predictive patterns directly from logs. A key milestone was the work of Tax et al., who used Long Short-Term Memory (LSTM) neural networks to predict next activities and remaining times [1]. While deep learning models offer high accuracy, they often lack interpretability, which is essential for business contexts requiring transparent and auditable results [10].

To maintain interpretability while improving predictive capability, hybrid and model-driven methods remain relevant. The Inductive Miner algorithm [7] is widely used for process discovery because it generates structured models that capture realistic control-flow behavior and tolerate noise.

Recent studies emphasize the importance of context in improving prediction accuracy. [11] showed that filtering event logs by workload, route, and transport type reduces variability and enhances precision in order-fulfillment scenarios. Similarly, [12] demonstrated that context-aware monitoring improves performance in smart manufacturing. More recently, a context-enriched framework was proposed that integrates workload and environmental data to improve stability under dynamic conditions [13].

Overall, the literature confirms that combining contextual information with interpretable process models increases both accuracy and practical value in predictive process monitoring, particularly in domains such as logistics and manufacturing.

3. Methodology

This study followed the Design Science Research Methodology (DSRM) [14]. DSRM provides a structured process for developing and evaluating artifacts that address real organizational problems. In this work, it guided the design of a predictive process monitoring component for a logistics information system. Figure 1 summarizes the six DSRM activities and how they were applied in this work.

The research began with the identification of the problem: existing predictive models often lack interpretability and accuracy when forecasting delivery delays and activity sequences in real-time logistics operations [10]. The main objective was to create an interpretable model capable of delivering reliable, low-latency predictions suitable for industrial use.

During the design and development phase, a prediction engine was implemented using polynomial regression trained on event logs. Contextual filtering and optimization techniques were incorporated to improve prediction accuracy and scalability, following approaches similar to those described by Campos et al. [15]. The system was demonstrated and validated using historical event data from our case study logistics processes and through real-time simulations that compared predicted versus actual delivery outcomes.

Evaluation focused on predictive accuracy, runtime performance, and usability. Metrics such as mean absolute error (MAE) were used to assess quantitative performance, while feedback from logistics stakeholders supported qualitative validation. Finally, results were communicated through this research paper.

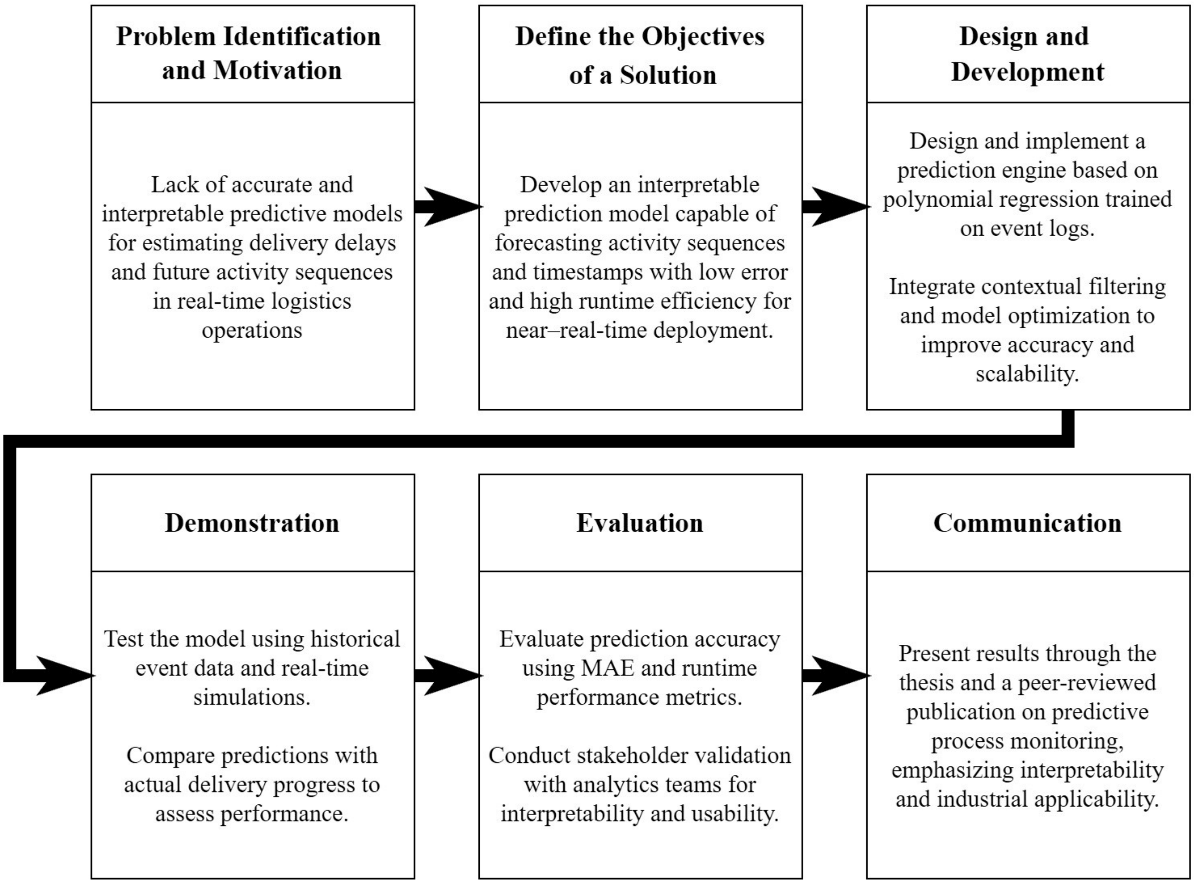


Fig. 1. Dsrm followed for the development of the artifact.

4. Activity and Timestamp Prediction

This section describes the method used to predict the remaining activities and timestamps of ongoing logistics orders. The approach combines process mining with contextual modeling to produce data-driven forecasts.

Before model construction, event logs were preprocessed to ensure consistent timestamps and to handle missing or incomplete entries.

4.1. Contextual Filtering

To improve prediction accuracy, the historical event log is filtered to include only cases that share similar contextual attributes with the running order. This ensures that predictions are generated from behavior observed under comparable operational conditions. The following context features are used:

- Origin and destination locations;
- Transportation mode (air or ground) and shipping company;
- Hour of the day, derived from the order creation timestamp;
- Season, derived from the order month;
- System workload, measured by the number of active orders at the time of the event.

This filtering step narrows the data to a smaller subset of relevant cases, which are then used to build context specific predictive model.

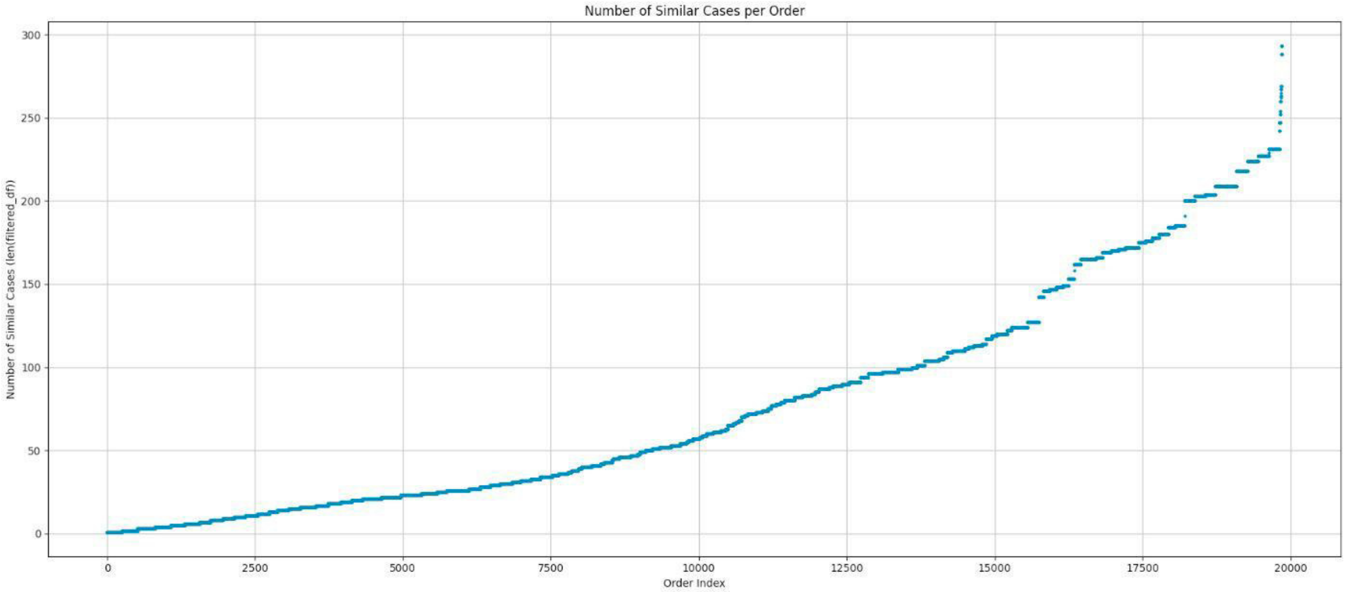


Fig. 2. Number of similar cases used for prediction model discovery.

4.2. Process Model Discovery

The Inductive Miner algorithm from the PM4Py library [5] is applied to the filtered event log to discover a structured process tree. The resulting model captures the control flow of order handling, including sequences, parallel activities, and decision points.

As shown in Figure 2, the number of similar cases per order ranges from 0 to 300, with most cases having fewer than 150. Sparse subsets make process discovery challenging: when too few examples exist, the discovered model may represent irregular or exceptional paths rather than the dominant process behavior. This can lead to fragmented process trees, characterized by multiple weakly connected branches and frequent early splits.

Limited case coverage also reduces the stability of statistical measures such as activity frequency and transition duration. Consequently, while the discovered models remain informative, their reliability for prediction and conformance analysis is constrained by data sparsity.

4.3. Token Replay for Predictive Path Simulation

After discovering a process model, the current trace of an ongoing order is aligned using token-based replay. The algorithm identifies the case's current position in the model and simulates future behavior based on the most frequent transitions in similar historical cases. The last observed activity acts as an anchor, and predictions consider only the past cases that contained this activity.

4.4. Prediction Strategies

The system supports two prediction modes:

Optimistic Path Strategy. In this mode, predictions are constrained to a best-case scenario. To simulate a successful delivery, the system may adjust the order of predicted activities so that the trace follows a path leading to a successful outcome.

Model-Based Strategy. In this mode, the Inductive Miner discovers a process model from historical data, and token-based replay aligns the ongoing case with that model. The system then simulates the most likely future activities and estimates their timestamps using a context-aware duration model.

4.5. Timestamp Estimation via Polynomial Regression

For each predicted transition $A_i \rightarrow A_{i+1}$, the expected duration Δt is estimated using a polynomial regression model on the system workload w:

$$\Delta t = f(w) = a_0 + a_1 w + a_2 w^2 + \ldots + a_n w^n$$

The timestamp of the next activity is computed as:

$$T(A_{i+1}) = T(A_i) + \Delta t$$

If workload is not used as a contextual feature, the algorithm defaults to the mean observed duration between activities within the filtered log:

$$T(A_{i+1}) = T(A_i) + \mathrm{E}[\Delta t \mid A_i \rightarrow A_{i+1}, \mathrm{Context}]$$

The algorithm alternates between predicting the next activity and estimating its timestamp until an end marking is reached. The resulting output is a complete predicted trace containing both control flow and timing information.

4.6. System Integration and Output Visualization

Predicted and completed activities are combined and exported as a JSON structure. The frontend retrieves this data through a REST API and displays it chronologically. Completed activities appear in blue, while predicted activities are shown in gray with a green marker (Figure 3). When few similar cases are available, the system allows selective disabling of filters to ensure a predictive model can still be generated.

5. Results and Discussion

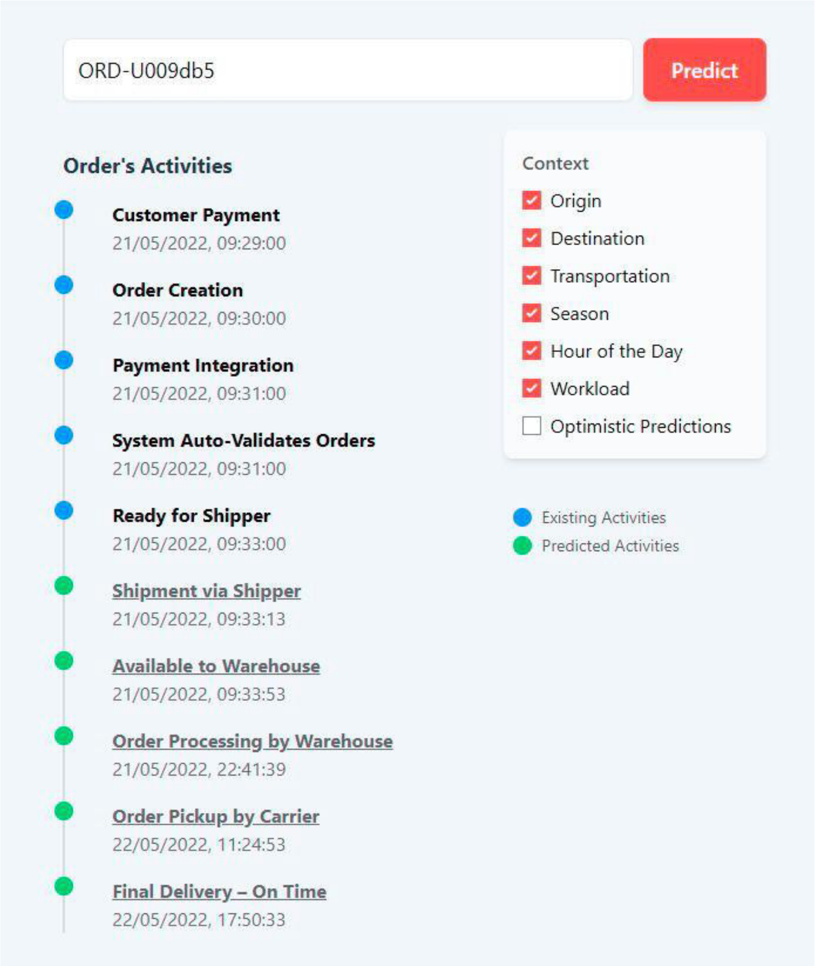
To evaluate the proposed approach, predictions were generated assuming that each order had already reached its fourth activity. This setting reflects real operational conditions, where forecasting occurs after part of the process has been completed. Conditioning the model on partial case progress allows a realistic assessment of predictive performance during active deliveries.

5.1. Evaluation Metrics

Two main metrics were used to measure prediction accuracy: Mean Absolute Error (MAE) in hours and in ratio form. The first quantifies the average deviation between predicted and actual transition durations, expressed in hours, it provides an interpretable measure of accuracy in operationally meaningful units of time [9, 1]. The second normalizes the absolute error by dividing it by the actual duration, it allows fair comparison between short activities (e.g., scans) and long-running ones (e.g., transportation) [8]. Additionally, a *count* value indicates how many times each transition occurred in the dataset, reflecting the reliability of each result.

$$1\,\overset{n}{X} \qquad\qquad 1\,\overset{n}{X}\,|\hat{t}_i - t_i|$$

$$\mathrm{MAE}_{\mathrm{hours}} = \_ \quad |\hat{t}_i - t_i| \qquad \mathrm{MAE}_{\mathrm{ratio}} = \_ \quad \underline{\qquad}$$
$$n \;\; n \; t_i \; i=1 \quad i=1$$



ORD-U009db5

Predict

Order's Activities

● **Customer Payment**
21/05/2022, 09:29:00

● **Order Creation**
21/05/2022, 09:30:00

● **Payment Integration**
21/05/2022, 09:31:00

● **System Auto-Validates Orders**
21/05/2022, 09:31:00

● **Ready for Shipper**
21/05/2022, 09:33:00

● Shipment via Shipper
21/05/2022, 09:33:13

● Available to Warehouse
21/05/2022, 09:33:53

● Order Processing by Warehouse
21/05/2022, 22:41:39

● Order Pickup by Carrier
22/05/2022, 11:24:53

● Final Delivery – On Time
22/05/2022, 17:50:33

Context
☑ Origin
☑ Destination
☑ Transportation
☑ Season
☑ Hour of the Day
☑ Workload
☐ Optimistic Predictions

● Existing Activities
● Predicted Activities

Fig. 3. Predicted and completed activities for order ORD-U009db5.

5.2. Relationship Between Accuracy and Transition Frequency

As Figure 4 shows, transitions that occur more frequently consistently show lower errors, confirming that a larger number of examples improves model stability. Low-frequency transitions exhibit greater variability, which is expected because rare transitions are harder to model accurately. A clear inverse trend between frequency and MAE exists, indicating that data availability is a major factor in predictive performance.

Two major outliers were detected, with errors of 552.04 and 2107.92, both associated with transitions that occurred only twice. Given such limited training data, these extreme errors are expected. A smaller outlier of 166.8 hours is also present. To improve visibility, these three outliers have been removed from the Figure.

5.3. Impact of Contextual Modeling

Table 1 compares predicted transition times obtained from a baseline model without contextual filtering against those from a context-aware model, alongside the actual observed durations. The inclusion of context variables such as route, workload, and transport type consistently reduces prediction errors.

For early automated transitions (e.g., *Customer Payment* → *Order Creation*), the context-aware model predicts near-zero durations, matching their instantaneous nature. The strongest improvements occur in transitions affected by operational factors. For example, *Available at Warehouse → Warehouse Processing* improves from 200.9h to 3.63h, closely aligning with the observed mean of 3.63h.

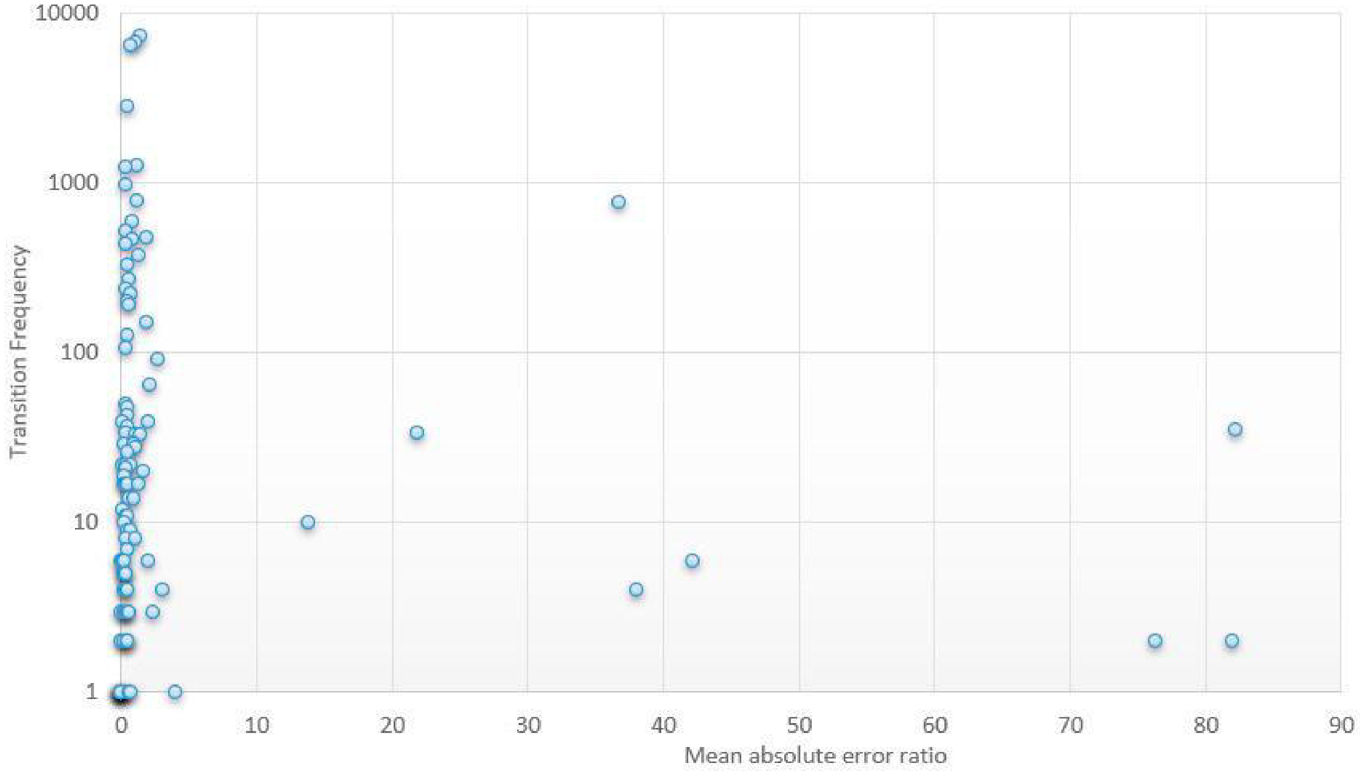


Fig. 4. Relationship between Mean Absolute Error and Transition Frequency.

Table 1. Transition time comparison between baseline (no context) and context-aware models.

Activity Transition	No Context (h)	With Context (h)	Actual (h)
Customer Payment → Order Creation	0.393	0.017	0.017
Order Creation → Payment Integration	0.251	0.017	0.017
Payment Integration → Automatic Validation	1.092	0.000	0.000
Automatic Validation → Ready for Shipper	2.803	0.017	0.017
Ready for Shipper → Shipment Dispatch	0.057	0.006	0.000
Shipment Dispatch → Available at Warehouse	4.453	0.011	0.017
Available at Warehouse → Warehouse Processing	200.901	3.628	3.633
Warehouse Processing → Carrier Pickup	34.863	3.394	3.233
Carrier Pickup → Final On-Time Delivery	394.375	6.228	7.267

5.4. Discussion

The results confirm that integrating contextual information significantly enhances predictive performance in process mining applications. Filtering event logs by relevant features reduces variability in duration estimates and improves both absolute and relative accuracy. Frequent transitions yield lower errors due to greater statistical support, while rare transitions show higher variance, reflecting data sparsity.

Transitions with high observation counts consistently yield lower prediction errors, indicating that larger context specific datasets provide more stable estimates. However, the variability observed among infrequent transitions highlights the limitations of purely frequency-based evaluation. Some low-frequency transitions achieve low error when their process behavior is stable, suggesting that process regularity and context relevance are also critical factors.

6. Conclusion

This work presented a predictive framework that combines process mining and data-driven modeling to forecast future activities and timestamps in logistics processes. By integrating contextual filtering, process model discovery with Inductive Miner, and token-based replay, the system can predict both the next activities and their expected completion times for ongoing orders. This enables proactive detection of potential delays and supports more efficient operational planning.

The results show that prediction accuracy depends strongly on the availability of similar historical cases. When sufficient comparable data exist, the model produces stable and realistic forecasts. However, when few similar cases are found, discovered models become fragmented and less reliable. Despite these limitations, contextual filtering significantly improves prediction quality by reducing noise and focusing on relevant subsets of data.

Timestamp prediction experiments confirmed these findings. Context-aware models produced much closer estimates to actual transition times than baseline models without context, especially for complex or human-dependent activities.

Some limitations remain. The approach relies on classical process mining and regression methods, which ensure interpretability but limit predictive power compared to deep learning techniques, also data sparsity reduces model robustness.

Future work should explore hybrid approaches that combine interpretable process models with neural network predictors. Expanding the dataset and performing real-time validation in production environments would improve generalization and reliability. Overall, the proposed framework contributes to more transparent, context-aware, and proactive decision-making in large-scale logistics operations.

References

[1] Tax, Niek, Ilya Verenich, Marcello La Rosa, and Marlon Dumas. (2017) “Predictive business process monitoring with LSTM neural networks.” In *Lecture Notes in Computer Science*, Vol. 10253, pp. 477–492. Springer. doi:10.1007/978-3-319-59536-8 30.

[2] Marquez-Chamorro, Alfonso Eduardo, Manuel Resinas, and Antonio Ruiz-Cortes. (2018) “Predictive Monitoring of Business Processes: A Survey.” *IEEE Transactions on Services Computing* 11(6). doi:10.1109/TSC.2017.2772256.

[3] Petropoulos, Fotios, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K. Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J. Bessa, Jakub Bijak, John E. Boylan, and others. (2022) “Forecasting: theory and practice.” *International Journal of Forecasting*. Elsevier. doi:10.1016/j.ijforecast.2021.11.001.

[4] Gunnarsson, Bjorn Rafn, Seppe K. L. M. Vanden Broucke, and Jochen De Weerdt. (2019) “Predictive Process Monitoring in Operational¨ Logistics: A Case Study in Aviation.” Unpublished manuscript.

[5] Berti, Alessandro, Sebastiaan J. van Zelst, and Wil M. P. van der Aalst. (2020) “PM4Py: A Process Mining Library for Python.” In *Business Process Management Workshops*, pp. 169–179. Springer. doi:10.1007/978-3-030-66498-5 13.

[6] Badakhshan, Peyman, Bastian Wurm, Thomas Grisold, Jerome Geyer-Klingeberg, Jan Mendling, and Jan vom Brocke. (2022) “Creating Business Value with Process Mining.” *The Journal of Strategic Information Systems* 31: 101745. Elsevier. doi:10.1016/j.jsis.2022.101745.

[7] Berti, Alessandro, Sebastiaan van Zelst, and Daniel Schuster. (2023) “PM4Py: A process mining library for Python.” *Software Impacts* 17: 100556. Elsevier B.V. doi:10.1016/j.simpa.2023.100556.

[8] Di Francescomarino, Chiara, Marlon Dumas, Alessandro Federici, Fabrizio Maria Maggi, and Ivana Teinemaa. (2017) “Predictive process monitoring methods: Which one suits me best?” *Business Process Management Journal*. Emerald Publishing Limited.

[9] Teinemaa, Ivana, Marlon Dumas, Marcello La Rosa, and Fabrizio Maria Maggi. (2019) “Outcome-oriented predictive process monitoring: Review and benchmark.” *ACM Transactions on Knowledge Discovery from Data (TKDD)* 13 (2): 1–57. ACM.

[10] Elkhawaga, Ghada, Mervat Abuelkheir, and Manfred Reichert. (2022) “Explainability of Predictive Process Monitoring Results: Can You See My Data Issues?” *arXiv preprint arXiv:2202.08041*.

[11] Intayoad, Patipat and Jorg Becker. (2018) “Contextual Factors for Predictive Monitoring of Order Fulfillment Times.” In¨ *Lecture Notes in Business Information Processing*, Vol. 308. Springer.

[12] Becker, Jorg, Patipat Intayoad, and Tobias Matzner. (2018) “Context-Aware Predictive Process Monitoring in Smart Manufacturing.” In¨ *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp. 1303–1312.

[13] Al-Jebrni, Abdullah, Florian Fessler, and Marlon Dumas. (2023) “Context-Aware Predictive Process Monitoring Under Dynamic Workload Conditions.” *Information Systems*, 118: 103199. Elsevier.

[14] Hevner, Alan R., Salvatore T. March, Jinsoo Park, and Sudha Ram. (2004) “Design Science in Information Systems Research.” *MIS Quarterly* 28 (1): 75–105. JSTOR. doi:10.2307/25148625.

[15] Campos, and [Co-authors, if applicable]. (2025) “Predictive Process Monitoring for Logistics: Context-Aware Timestamp and Activity Prediction.” In *Proceedings of the DCE 2025 Conference*. To appear.