

A Lightweight Decision Support System for Real-Time Predictive Process Monitoring

Narendra B O¹, Dr.Usha G R²,Dr.Sreenivasa B R³

¹PG Student, Department of Artificial Intelligence and Data Science, Bapuji Institute of Engineering and Technology, Davangere

² Associate Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere

³Professor & PG Coordinator, Department of Artificial Intelligence and Data Science, Bapuji Institute of Engineering and Technology, Davangere

Abstract—Despite the fact that predictive process monitoring (PPM) is vital for keeping tabs on normal enterprise operations, getting from a hard and fast of regulations' theoretical correctness to placing it into workout isn't always any easy feat. no matter the truth that deep reading architectures that depend closely on computation had been the focal point of recent studies, process managers want gear which may be easy to recognize and use so that you can interfere in real-time. To fill this void, the authors of this article present an interactive DSS. The device, which is built as a surrender-to-surrender internet framework, takes raw data from event logs and makes use of dynamic trace extraction to create visualizations of usage distributions and computing case durations in real-time. a framework-based totally automated eliminate detection device is used for anomaly manipulate. This mechanism flags deviations based totally on imply-period thresholds. the use of a mild-weight dual-model structure in conjunction with prefix-based totally trace encoding, the predictive module guarantees fast inference without the want for specialised hardware. On one hand, a concurrent Linear Regression model estimates the last steps to case of completion, even as on the other, a Logistic Regression classifier is used to forecast the following probabilistic interest in an on-foot trace. With a baseline accuracy of sixty-two% and the capability to provide process stakeholders fast, actionable insights, the predictive classifier is evaluated on a multi-interest event log that includes the subsequent states: begin, examine, Approve, entire, Reject, and Revise. The counselled framework provides a without problem implementable answer to the issues of proactive workflow optimization and subsequent-satisfactory-movement recommendation via the use of placing an emphasis on computational overall performance and consumer-centric layout in preference to mathematical complexity.

Index Terms—Predictive Process Monitoring, Process Mining, Decision Support Systems, Logistic Regression, Delay Detection, Event Log Analysis.

I. INTRODUCTION

Timestamped occasion facts are continuously being generated across various corporate information structures as a result of the fast digital transformation of modern-day corporations. the sphere of technique mining has grown in significance as a way of coming across, tracking, and enhancing real-global workflows by way of gaining beneficial information from these occasion logs. As this subject develops, Predictive technique tracking (PPM) is a step within the proper route, marking an exchange in emphasis from looking lower back at tactics to planning in advance. the use of PPM, you will count on what a jogging technique instance will do subsequent, calculate how plenty longer it's going to take to finish the modern venture, and notice uncommon delays before they motive SLA breaches.

modern-day, modern-day techniques for predictive technique analytics use intricate deep mastering architectures such as CNNs, Transformer-based totally interest fashions, and lengthy brief-term memory (LSTM) networks. even though these superior algorithms are fantastic at capturing non-linear manage flows and feature excellent forecast accuracy on benchmark datasets, there's a massive impediment to the use of them in commercial settings. because these fashions are frequently opaque and function as "black packing containers," stakeholders

have a hard time understanding what is going into generating a sure forecast. additionally, for training and inference, they need large pc assets. there's a sizeable barrier to realistic, every day adoption for operational technique managers because of the heavyweight fashions' lack of interpretability and the architectural difficulties of embedding them into close to-real-time operational dashboards. more and more, stakeholders are looking for solutions which can make predictions, but also install quickly, are light-weight, and offer visual feedback proper away.

This work provides a light-weight, interactive decision help machine (DSS) designed for real-time predictive technique analytics to bridge the distance between algorithmic complexity and realistic viability. The advised paradigm puts an emphasis on cease-to-cess operational software program engineering in place of optimizing algorithmic complexity in a way that compromises deploy ability. built as an interactive net interface, the machine takes raw facts from occasion logs and uses it to run hint extraction in real time, calculate non-stop case intervals, and robotically discover instances which might be delayed according to statistical threshold guidelines.

The machine uses a simplified predictive module that uses traditional, computationally green device mastering fashions to offer fast inference without the requirement for specialized hardware. The technique uses prefix-based totally hint encoding to transform sequences of historic events into function vectors of described period. a Logistic Regression classifier anticipates the subsequent probabilistic movement in a jogging hint the use of these vectors, and a Linear Regression version estimates the ultimate steps to finish the case the use of these vectors. customers are capable of input modern occasion sequences and right away gain actionable hints the use of this twin-version method.

This paper's principal contributions start with the implementation of an interactive visualization framework. This framework acts as an analytical dashboard that captures hobby distributions and case-stage duration monitoring, permitting for fast, automated visual exam of occasion logs. the use of this as a foundation, the machine consists of automated postpone detection the use of a dynamic anomaly detection technique that determines the average duration of beyond traces and highlights continual times that surpass this barrier. Our new light-weight

predictive structure is a twin-version device mastering pipeline, and it powers these insights. This structure makes use of prefix-based totally label encoding together with Logistic Regression and Linear Regression to attain fast inference with minimal computational overhead. ultimately, the framework bridges the distance between facts technological know-how and technique management by way of imparting real-time decision help through a prediction interface that is reachable to customers. This interface converts encoded version outputs into insights that technique stakeholders can act on in close to-real-time. What follows is the define for the relaxation of the item. part II discusses relevant research on decision-help structures and predictive technique tracking. The mathematical components of the prediction fashions, function engineering, and occasion log preprocessing are all included in segment III, which outlines the proposed technique. The experimental results and performance opinions are offered in segment IV, collectively with the machine layout. The work is concluded and ability destiny research guidelines are mentioned in segment V.

II. LITRATURE REVIEW

a move far from reactive system discovery and towards proactive, dynamic forecasting in business workflows is symbolized via the emergence of Predictive system monitoring (PPM). the existing frontier locations an emphasis on version explainability, consumer accept as true with, and realistic operational deployment, in contrast to early studies that focused extraordinarily correct but opaque deep gaining knowledge of algorithms. To determine wherein PPM is now and wherein vital studies desires exist, this overview compiles primary difficulties, superior prediction strategies, and interactive choice-guide structures.

a cohesive taxonomy is needed inside the fragmented environment that has advanced as a result of the quick increase of Predictive system monitoring (PPM) during the last decade. Ceravolo, Comuzzi, De Weerd, Di Francescomarino, and Maggi tackle this problem. it's miles difficult to standardize technique or stumble on modern studies gaps because of this dispersion, as stated via the authors. although complex deep gaining knowledge of fashions performs admirably whilst prefix encoding, the examine

observed that they have got difficulty with explainability and real-time operational deployment whilst compared to indirect, bucketed strategies. After undertaking a thorough survey of PPM researchers, the authors had been capable of validate their proposed 9 examine avenues, 3 popular studies issues, and unified taxonomy of PPM ideas. as a way to address noisy, real-international event logs, the examine recommends that the community priorities integrating area-information-primarily based bucketing, standardizing assessment criteria, and making fashions greater explainable to cease users [1].

the subject of reliability in machine gaining knowledge of fashions for system prediction is mentioned in Galanti et al. (2023). They draw attention to the truth that, regardless of predictive fashions' outstanding accuracy, they're frequently run as "black packing containers," main stakeholders to distrust and reject predictive recommendation whilst they're not able to recognize their common sense. finding that system actors want prescriptive analytics in addition to predictive analytics and that explaining those predictions to users significantly reduces consumer distrust turned into the authors' primary finding. as a way to deal with this, they created a framework that uses Shapley Values to hyperlink sure system continuations with expected KPIs. They had been capable of correctly encompass this framework into the commercial IBM system Mining Suite as a module. The authors endorse expanding explainability to situations regarding the simultaneous modeling and explanation of several interacting items as an area for future studies [2].

software development workflows are the primary emphasis of Dorado et al. (2025), who point out that whilst traditional DevOps metrics, such the ones observed inside the DORA framework, offer wide perspectives of performance, they don't choose up on granular system dynamics or CI/CD pipeline bottlenecks. Their examine proves that system mining techniques may be immediately applied to software engineering via changing unprocessed DevOps facts like builds, pull requests, and commits into observable event logs. So, to create those logs, the authors supplied "CodeSight," a comprehensive device that draws facts from GitHub. Pull Request (PR) decision times and closing date compliance had been expected with high precision and F1 scores via CodeSight using an LSTM (long brief-term reminiscence) neural

network version applied to the logs. the subsequent steps on this assignment are to enhance the incorporation of machine gaining knowledge of for proactive assignment management and to amplify predictive modeling to manipulate bigger, pass-repository workflows [three].

With an emphasis at the shift from static system discovery to dynamic, predictive monitoring, Khan, Ghose, Dam, and Syed (2023) offer a thorough assessment synthesising the approaches wherein machine gaining knowledge of enhances corporate approaches. Reinforcement gaining knowledge of and disbursed privateness-retaining mining are two of the most popular superior methods inside the examine. those allow decentralized version training and dynamic choice guide without sacrificing facts privateness throughout silos. modern prediction algorithms had been categorised via the authors in line with their algorithmic approach, and their performance turned into compared using enterprise-standard event log benchmarks. They pressure that integrating advanced information-centric fashions and pass-silo system mining are two of the most vital regions for future studies [four].

the problem of preserving track of complex and numerous real-international business approaches for execution inconsistencies is tackled via Tariq, Charles, McClean, McChesney, and Taylor. whilst traditional system mining methods do a terrific activity of presenting descriptive facts, the authors point out that modern predictive techniques don't offer enough facts on how a real-lifestyles system example compares to a perfect goal system version. as a way to deal with this, the studies give a framework for on-line system prediction that use a heuristic mining set of rules to find beginning fashions and conducts conformance evaluation to extract elements associated with manipulate-go with the flow and hint alignment. using a Fisher discriminant ratio (FDR) ranking, we may additionally assume the last system outcome via isolating surprisingly contributing features via mapping out hint fluctuations throughout detailed temporal cut-offs (ranging from 25% to 100%). Their architecture turned into tested using various algorithms, including Naive Bayes, Logistic Regression, Random Forest, and guide Vector machine, and it turned into evaluated in opposition to real-international patron diagnostic logs from a telecom's employer. Outperforming graph-simplest

baseline prediction techniques inside the literature, the experimental findings confirmed high accuracy, keep in mind, and F-degree scores, making them a greater correct predictor of very last system success or failure [5].

consumer accept as true with and operational attractiveness in predictive system monitoring systems are addressed in Galanti, de Leoni, Marazzi, Bottazzi, Delsante, and Folli's (2021) work. Even whilst contemporary algorithms produce fairly precise predictions, humans involved inside the system are understandably hesitant to accept as true with automatic predictions if the device fails to offer an obvious reason behind why a case is being flagged. To fill this want, the authors created a cloud-primarily based prediction module that is each ready-to-use and anchored in cooperative recreation idea; it generates nearby and worldwide factors using SHapley Additive factors (SHAP). Catboost, a high-performance gradient boosting choice tree framework, replaced traditional LSTM architectures inside the framework, significantly cutting version training times in half of to a few quarters whilst preserving accuracy ranges unchanged. thru an interactive analytics dashboard, the device visualizes system "influencers" and efficiently integrates with the commercial IBM system Mining Suite. It maps event houses to associated time and price deviations compared to a favored common baseline. via making use of them to a dataset of banking system events with greater than 730,000 data, they proven that even complex backend machine gaining knowledge of my additionally offer without difficulty comprehensible insights without the want for area stakeholders to possess extra technical knowledge [6].

although expected accuracy continues to be vital, the real benefit of system mining is in its real use, as the today's trends in PPM display. To put in force proactive, facts-driven choice guide in real-time enterprise settings, modern frameworks must strike a compromise among computational efficiency and human-in-the-loop explainability.

III. PROPOSED METHODOLOGY

The proposed framework employs a modular architectural approach to predictive process monitoring, designed to balance computational efficiency with operational utility. By moving away

from resource-intensive deep learning models, the methodology prioritizes rapid, interpretable inference suitable for real-time decision support in enterprise environments. The pipeline is structured into four distinct phases: event log preprocessing and trace extraction, threshold-based delay detection, feature engineering through prefix-based encoding, and the training of a dual-model predictive architecture. This systematic workflow ensures that raw, unstructured process data is transformed into actionable, near-real-time insights, as illustrated in the system architecture diagram below.

3.1 Dataset Description

a based event log dataset mimicking actual-global workflow executions in a banking-style placing is used for the experimental evaluation. The event log big.csv document statistics all elements of commercial enterprise process instances, along with their beginning and ending factors, and includes critical fields like Case identification, Timestamp, and hobby name. The workflow is formed by separate hobby routes, some of which might be standard approval techniques (begin → review → Approve → complete) even as others are extra worried and targeted on rework (begin → review → Reject → Revise → Approve → complete). The predictive model can inform the distinction among easy, efficient instances and those with bottlenecks like rejection and revision that extend the case time because of these pathways.

Datset Summary Table

Dataset Attribute	Description
Dataset Name	event_log_large.csv
Data Format	Comma-Separated Values (CSV)
Key Attributes	Case ID, Timestamp, Activity Name
Workflow Activities	Start, Review, Approve, Complete, Reject, Revise
Application Domain	Banking/Finance (Simulated)

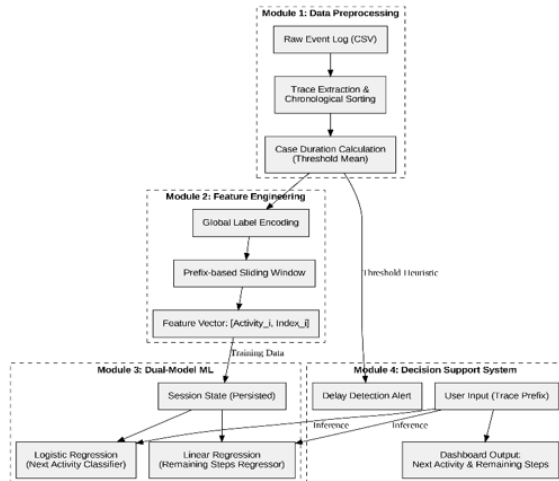


Fig1: Methodology

3.2 Event Log Preprocessing and Trace Extraction

the usage of a dependent occasion log that incorporates executions of approaches from the past, the predictive framework may build its basis. in this definition, each occasion e in the occasion log L is a tuple $e = (c, an, t)$ that represents the case identity c , the pastime an , and the timestamp t . The uncooked occasion log is sorted chronologically by using c and t at some point of the preprocessing segment to ensure temporal integrity. Ordered pastime traces are created by using grouping the occasions in line with their case IDs. With a case c , a hint T_c is described as $T_c = \{a_1, a_2, \dots, a_n\}$. the whole length D_c is decided for each example c by using subtracting the minimum length (t_c) from the most length (t_c). To decide a dynamic, postpone threshold θ , we common the time of all past cases that have been finalized and mark a case as "delayed" if $D_c > \theta$.

3.3 Feature Engineering and Sequence Encoding

- a global label encoder is used to transform express hobby labels to integer representations $v \in \mathbb{Z}$ on the way to get sequential hint facts equipped for machine mastering. From steps $i=1$ to n , sliding windows extract capabilities X and targets y , y_time for each encoded hint T .unit one:
- an X -vector, in which $X_i = [v_{(i-1)}, i]$, representing the modern index and the formerly encoded hobby, makes up the input.type aim (y): The aim is the fee so as to be encoded for the subsequent step: you could write $y_i = v_i$.

- The aim of the regression is the number of duties left to do: $y_time, i = n - i$.

3.4 Dual-Model Predictive Architecture

- a global label encoder is used to transform express hobby labels to integer representations $v \in \mathbb{Z}$ on the way to get sequential hint facts equipped for machine mastering. From steps $i=1$ to n , sliding windows extract capabilities X and targets y , y_time for each encoded hint T .unit one:
- an X -vector, in which $X_i = [v_{(i-1)}, i]$, representing the modern index and the formerly encoded hobby, makes up the input.type aim (y): The aim is the fee so as to be encoded for the subsequent step: you could write $y_i = v_i$.
- The aim of the regression is the number of duties left to do: $y_time, i = n - i$.

IV. REAL-TIME INFERENCE

The trained models are persisted within the application's session state. Through the interactive dashboard, stakeholders input an ongoing, partial trace. The system validates this input against historical prefixes, transforms it via the label encoder, and passes the feature vector to the dual-model pipeline. In doing so, it bridges the space between facts technology and process management and presents process stakeholders with on the spot, actionable records.

The cautioned structure proves that deep gaining knowledge of's computational complexity makes no sense for predictive process monitoring to produce useful effects. The methodology offers a computationally efficient, interpretable, and robust alternative to contemporary industry standards through designing a modular pipeline that movements from raw event log enter to threshold-based bottleneck discovery and then to dual-model system gaining knowledge of inference. ultimately, this methodical process takes care of the essential operational requirement for decision assist within the close to actual-time. thru the usage of skilled fashions inside an interactive Streamlit display, this methodology permits stakeholders to make facts-pushed interventions with minimum infrastructure overhead. It bridges the space between static process analytics and dynamic, proactive process management.

V. RESULTS AND DISCUSSION

The proposed predictive framework changed into evaluated the usage of the event_log_large.csv dataset, which captures workflow activities in a simulated banking-fashion approval procedure. The assessment aimed to assess the machine's efficacy in pastime distribution evaluation, delay detection, and actual-time prediction of future procedure states.

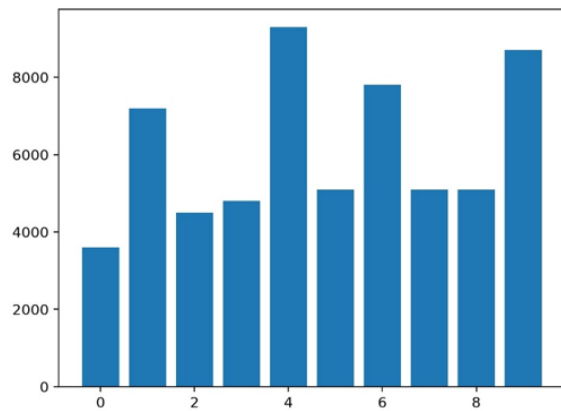


Fig2: Case Duration Analysis

The visualization, depicted in Fig 2, illustrates the total processing time in seconds for ten distinct process cases (case_id 0 through 9). The analysis indicates significant variance in case durations, with some instances exhibiting substantially longer processing times than others. As documented in the project logs, these duration calculations were derived directly from the temporal attributes specifically the start and end timestamps captured within the event logs.

This performance variability is a key indicator of process health. The higher duration peaks observed in the chart correlate with cases that involve rework activities, such as Reject and Revise, confirming that these deviations from the standard approval path are the primary drivers of workflow bottlenecks. For process stakeholders, this visualization serves as a primary diagnostic tool; by instantly identifying high-latency cases that exceed the established operational threshold, managers can prioritize these "at-risk" workflows for resource reallocation or expedited handling. Consequently, the Case Duration Analysis is essential for bridging the gap between passive performance reporting and active, data-driven process optimization.

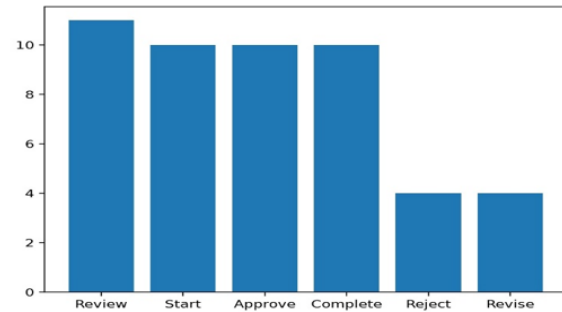


Fig3: Activity Distribution Analysis

The visualization presented in Fig. 3 illustrates the frequency count of each distinct activity identified in the processed event log. The results show that the activities Review, Start, Approve, and Complete maintain high, consistent frequency counts, suggesting these steps form the core sequence of the standard business process. Conversely, Reject and Revise activities exhibit a significantly lower frequency, occurring only in a subset of the total process cases.

This distribution pattern provides critical insight into the structural health of the workflow. The high frequency of Start, Review, Approve, and Complete confirms that the system effectively processes the majority of cases through the intended "straight-through" path. The presence of Reject and Revise activities at lower frequencies identifies these steps as deviations from the standard path, which often represent rework or exception handling. From an operational standpoint, this analysis is vital because these low-frequency activities are identified as the primary sources of process bottlenecks and increased completion times. By highlighting these frequent rework points, the dashboard enables process managers to pinpoint specific stages where the workflow is most prone to inefficiency, facilitating targeted interventions to optimize process throughput.

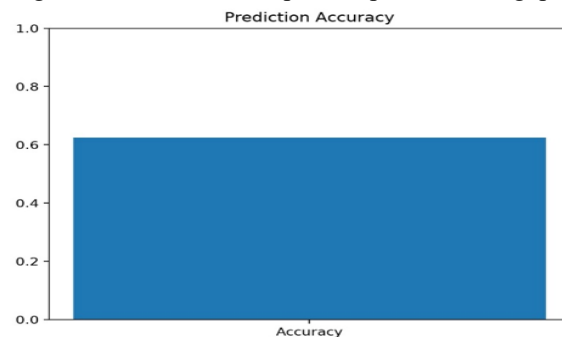


Fig4: Prediction Accuracy

The visualization presented in Fig. 4 displays the overall classification accuracy achieved by the machine learning model during the evaluation phase. The model achieved a prediction accuracy of approximately 62.5%. This metric represents the system's ability to correctly identify the next activity in a process sequence based on historical training data. An accuracy of 62.5% demonstrates that the framework effectively captures significant control-flow patterns within the business process using a lightweight, dual-model architecture. While deep learning models, such as LSTMs, are capable of higher predictive precision, they often require extensive computational resources and large-scale datasets. By contrast, this result validates that Logistic and Linear Regression models provide a practical and efficient trade-off, offering sufficient predictive utility for real-time process monitoring with minimal infrastructure overhead.

This functionality allows managers to gain near-real-time insights, such as predicting the next activity (Approve) and estimating remaining steps (3) for active cases, effectively supporting proactive workflow management and informed decision-making.

VI. CONCLUSION AND FUTURE ENHANCEMENT

The Predictive Process Mining Dashboard successfully bridges the gap between static process analytics and dynamic, proactive process management. By integrating lightweight machine learning models specifically Logistic Regression for activity classification and Linear Regression for remaining steps estimation with a Streamlit-based interactive interface, the system provides a scalable solution for near-real-time workflow monitoring. Experimental results demonstrate that the framework effectively identifies process bottlenecks, flags delayed cases through duration thresholding, and provides actionable insights that help organizations optimize operational efficiency and improve decision-making. To enhance the dashboard's capabilities and robustness, future iterations could incorporate advanced modeling through deep learning architectures such as LSTM or Bi-LSTM networks to capture more complex temporal dependencies in event sequences. Furthermore, the system could be

expanded to support industry-standard XES event logs, enabling integration with a wider range of enterprise software systems. Future developments may also involve implementing real-time event streaming and IoT-enabled process monitoring to provide truly live, automated decision-support systems, alongside integrating advanced conformance checking to compare current workflow executions against idealized process models, and leveraging cloud-based deployment for improved accessibility and scalability.

REFERENCES

- [1] Ceravolo, P., Comuzzi, M., De Weerd, J., Di Francescomarino, C., and Maggi, F. M., "Predictive process monitoring: Concepts, challenges, and future research directions," *Process Science*, vol. 1, no. 1, p. 2, 2024.
- [2] Galanti, R., de Leoni, M., Monaro, M., Navarin, N., Marazzi, A., Di Stasi, B., and Maldera, S., "An explainable decision support system for predictive process analytics," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105904, 2023.
- [3] Dorado, A., Folguez, I., Martín, S., Martín, G., Porto, Á., Ramos, A., and Wallace, J., "A process mining-based system for the analysis and prediction of software development workflows," *arXiv preprint*, arXiv:2510.25935, 2025.
- [4] Khan, A., Ghose, A., Dam, H., and Syed, A., "Advances in process optimization: A comprehensive survey of process mining, predictive process monitoring, and process-aware recommender systems," *arXiv preprint*, arXiv:2301, 2023.
- [5] Delsante, M., "Integration of an explainable predictive process monitoring system into IBM Process Mining Suite," 2021.
- [6] Tax, N., Verenich, I., La Rosa, M., and Dumas, M., "Predictive business process monitoring with LSTM neural networks," in *Proc. Int. Conf. Advanced Information Systems Engineering (CAiSE)*, Springer, 2017, pp. 477–492.
- [7] Evermann, J., Rehse, J.-R., and Fettke, P., "Predictive business process monitoring with structured and unstructured data," in *Proc. Business Process Management (BPM)*, Springer, 2016, pp. 401–417.

- [8] Taymouri, F., La Rosa, M., Erfani, S., Bozorgi, Z. D., and Verenich, I., "Predictive business process monitoring via generative adversarial nets: The case of next event prediction," in *Proc. Business Process Management (BPM)*, Springer, 2020, pp. 237–256.
- [9] Bukhsh, Z. A., Saeed, A., and Dijkman, R. M., "ProcessTransformer: Predictive business process monitoring with transformer network," *arXiv preprint*, arXiv:2104.00721, 2021.
- [10] Kratsch, J., Manderscheid, J., Röglinger, M., and Seyfried, J., "Machine learning in business process monitoring: A comparison of deep learning and classical approaches used for outcome prediction," *Business & Information Systems Engineering*, vol. 63, pp. 261–276, 2021.
- [11] Pasquadibisceglie, V., Appice, A., Castellano, G., Malerba, D., and Modugno, G., "ORANGE: Outcome-oriented predictive process monitoring based on image encoding and CNNs," *IEEE Access*, vol. 8, pp. 184073–184086, 2020.
- [12] Nguyen, A., Chatterjee, S., Weinzierl, S., Schwinn, L., Matzner, M., and Eskofier, B., "Time matters: Time-aware LSTMs for predictive business process monitoring," in *Proc. ICPM Workshops*, Springer, 2021, pp. 112–123.
- [13] Kahani, S., Pourmasoumi, A., and Beheshti, H. A., "HAM-Net: Predictive business process monitoring with a hierarchical attention mechanism," *Knowledge-Based Systems*, vol. 236, p. 107722, 2022.
- [14] Lin, L., Wen, L., and Wang, J., "MM-Pred: A deep predictive model for multi-attribute event sequence," in *Proc. SIAM Int. Conf. Data Mining*, 2019, pp. 118–126.
- [15] Harane, N., and Rathi, S., "Comprehensive survey on deep learning approaches in predictive business process monitoring," in *Intelligent Systems Reference Library*, Springer, 2020, pp. 115–128.
- [16] Di Francescomarino, C., and Ghidini, C., "Predictive process monitoring," in *Process Mining Handbook*, Springer, 2022, pp. 320–346.
- [17] Weytjens, H., and De Weerd, J., "Creating unbiased public benchmark datasets with data leakage prevention for predictive process monitoring," in *Proc. Business Process Management Workshops*, Springer, 2021, pp. 18–29.
- [18] Verenich, I., Dumas, M., La Rosa, M., Maggi, F. M., and Teinemaa, I., "Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 4, pp. 1–34, 2019.
- [19] Teinemaa, I., Dumas, M., La Rosa, M., and Maggi, F. M., "Outcome-oriented predictive process monitoring: Review and benchmark," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 2, pp. 1–57, 2019.