Jurnal Penelitian dan Aplikasi Sistem dan Teknik Industri (PASTI)
Vol. XX, No. 1, April 2026, 35-48
p-ISSN 2085-5869/ e-ISSN 2598-4853

Managing Workflow Time Overruns: A Workload-Aware Operational Management Approach Supported by Machine Learning

Rizaldi Mu'min[1]*, Jakfat Haekal[2], Andrian Haro[3], Rhamdalia Fanny Gustaji[4], Joval Ifghaniyafi Farras[5]

[1,3,4)] Program Studi S1 Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Negeri Jakarta
Alamat: Jl. R.Mangun Muka Raya, RT.11/RW.14, Rawamangun, Kec. Pulo Gadung, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia
[2)] Program Studi S1 Teknik Industri, Fakultas Teknik, Universitas Esa Unggul
Alamat: Jalan Arjuna Utara No.9, Duri Kepa, Kebon Jeruk, Jakarta Barat, Daerah Khusus Ibukota Jakarta, Indonesia
[5)] Program Studi D4 Pemasaran Digital, Fakultas Ekonomi dan Bisnis, Universitas Padjajaran
Alamat: Jl. Raya Bandung Sumedang KM.21, Hegarmanah, Kec. Jatinangor, Kabupaten Sumedang, Jawa Barat, Indonesia
Email: rizaldi.mumin@unj.ac.id* , jakfat.haekal@esaunggul.ac.id , andrian.haro@unj.ac.id, rhamdalia_fanny@unj.ac.id , joval.ifghaniyafi@unpad.ac.id

(Received: 31-03-2026; Revised: 30-04-2026; Accepted: 30-04-2026)

Abstrak

Keterlambatan waktu alur kerja bukan sekadar kesalahan peramalan, melainkan masalah pengendalian operasional yang berulang. Ketika waktu penyelesaian aktual melebihi rencana, manajer menghadapi ketidakstabilan jadwal, kehilangan koordinasi, hambatan persetujuan, dan kenaikan biaya layanan. Dengan menggunakan 2.500 observasi tingkat tugas, penelitian ini mengkaji bagaimana analitik berbasis beban kerja dari data workflow rutin dapat meningkatkan pengendalian operasional atas time overrun. Analisis menempatkan time overrun sebagai luaran utama dan menilai apakah variabel seperti jenis tugas, departemen, prioritas, tingkat persetujuan, beban kerja karyawan, estimasi durasi, dan biaya dapat memberikan visibilitas terhadap risiko keterlambatan. Hasil penelitian menunjukkan bahwa data workflow rutin mampu mengindikasikan lokasi akumulasi risiko overrun, terutama pada kualitas estimasi, kondisi beban kerja, kebutuhan persetujuan, dan heterogenitas tugas. Nilai manajerial utama analitik terletak pada penguatan disiplin perencanaan, kalibrasi estimasi, peninjauan beban kerja, dan pemantauan pengecualian.

Kata kunci: *manajemen operasional; time overrun workflow; pengendalian operasional; akurasi perencanaan; keseimbangan beban kerja; tata kelola proses*

Abstract

Workflow time overruns are recurring operational control problems rather than mere forecasting errors. When actual completion times exceed plan, managers face schedule instability, coordination losses, approval bottlenecks, and rising service costs. Using 2,500 task-level observations, this study examines how workload-aware analytics from routine workflow data can improve operational control over time overruns. The analysis treats time overrun as the main outcome and evaluates whether variables such as task type, department, priority, approval level, employee workload, estimated duration, and cost provide useful visibility into overrun risk. The results show that routine workflow data can indicate where overrun exposure tends to accumulate, especially around estimate quality, workload conditions, approval requirements, and task heterogeneity. However, the strongest managerial value of analytics lies less in replacing judgment than in improving planning discipline, estimate calibration, workload review, and exception monitoring. The study therefore reframes workflow overrun analysis as an operational control and process-governance issue.

Keywords: operational management; workflow time overruns; operational control; planning accuracy; workload balancing; process governance

INTRODUCTION

In many organizations, workflow time overruns are not isolated timing deviations but repeated failures of operational control. When planned completion times are missed, managers must revise schedules, absorb coordination costs, protect service levels, and respond to downstream bottlenecks with incomplete visibility into where the next disruption will emerge, a problem also observed in applied studies of production-delay control (Kannan et al., 2022) and remaining-time planning support (Botelho et al., 2026). Controlling overruns is therefore a central operational task rather than a narrow reporting problem.

In the dataset analyzed in this study, the operational performance gap is quantitatively visible. The planned completion time averages 123.91 minutes, whereas the actual completion time averages 180.86 minutes. This creates an average time overrun of 56.95 minutes, or approximately 46% above the planned standard. The median overrun is 57 minutes and the maximum reaches 120 minutes. The pattern is also system-wide rather than isolated: average overruns vary only from 55.04 to 57.87 minutes across process families and from 56.11 to 58.05 minutes across departments. These figures show a measurable gap between expected and actual workflow performance and justify the need for stronger operational control.

Recent advances in workflow analytics and machine learning create new opportunities to support that control task (Weinzierl et al., 2024). Yet the managerial value of analytics does not lie in prediction for its own sake. Its value lies in improving operational visibility, helping managers see where planning standards are weak, where workload concentration raises execution risk, and where approval structures or coordination routines contribute to avoidable slippage—an interpretation that aligns with recent work on predictive process monitoring and process-oriented decision support (Ceravolo et al., 2024). When analytical insight remains interpretable and actionable, it becomes more useful for managerial intervention (Breuker et al., 2016).

Prior research in workflow and process analytics has demonstrated that historical execution data can be used to anticipate outcomes such as remaining time, delay, and process completion (van der Aalst et al., 2011; Maggi et al., 2014). That literature is highly relevant, but it often places the technical prediction problem at the center of the discussion. For operational management, the more important question is slightly different: how can routine workflow data strengthen day-to-day control over time overruns, not merely forecast them? Review and benchmark studies reinforce the importance of this shift because the usefulness of time-related analytics depends heavily on target definition, data structure, and operational context (Tax et al., 2017; Verenich et al., 2019; Teinemaa et al., 2019).

This distinction matters because time overruns are shaped not only by task characteristics, but also by planning accuracy, approval complexity, workload balancing, and execution governance. Resource-aware research shows that workload and inter-case conditions can materially influence completion behavior (Aalikhani et al., 2025), while intervention-oriented studies stress that analytical value depends on whether managers can act on those signals in time (Fahrenkrog-Petersen et al., 2022). An analytics exercise becomes valuable when it clarifies where managerial attention should be directed, whether in estimate calibration, workload allocation, escalation rules, or process redesign.

The state of the art in business process management emphasizes the use of process data to monitor execution, support prediction, and improve managerial decision making (Dumas et al., 2018; Weske, 2019). Predictive analytics research further argues that models become valuable when they improve decisions rather than merely produce statistically accurate forecasts (Shmueli and Koppius, 2011). In workflow prediction, process-mining and deep-learning studies have shown that execution traces can anticipate process behavior, remaining time, and future outcomes (Evermann et al., 2017; Tax et al., 2017; Verenich et al., 2019).

However, much of the existing literature still emphasizes technical prediction performance, while fewer studies translate routine operational variables into practical control routines. This leaves a managerial gap: organizations need to know how ordinary workflow records, including workload, approval level, task type, estimated duration, and cost, can be converted into actionable routines for estimate calibration, workload balancing, approval review, and exception handling.

The present study addresses this issue using a task-level workflow dataset containing 2,500 observations with information on process family, task type, department, assigned employee, priority, approval level, workload, timestamps, planned duration, realized duration, and task-level cost. The analysis treats time overrun as the primary outcome because it captures the extent to which execution departs from plan. Machine learning is used here as a supporting analytical instrument, not as the central contribution of the article.

Accordingly, this study aims to determine how routine workflow data can improve operational control over time overruns and to identify which workload, planning, approval, and task-context variables provide useful managerial visibility. The contribution is practical and operational: the study reframes workflow overrun analysis as a control problem and links the analytical results to routines for planning discipline, workload governance, approval review, and exception-based monitoring.

METHODS

Dataset and Operational Context

The empirical analysis uses a provided workflow dataset consisting of 2,500 observations and 15 original variables spanning the period from 01 January 2024 to 30 December 2024. Each record captures a single operational task and includes process family, task type, priority, department, assigned employee, start and end timestamps, estimated duration, actual duration, approval level, workload indicator, and task-level cost. The dataset is therefore well suited to examining how routine operational descriptors can support visibility over time overruns.

An important structural characteristic is that Workflow_ID and Task_ID are unique at the row level. The file therefore represents one realized task outcome per observation rather than a full multi-event case history. This matters methodologically because prior reviews of predictive process monitoring show that richer event histories often support stronger temporal prediction than single-record snapshots (Verenich et al., 2019; Ceravolo et al., 2024). The present study is therefore designed around the level of operational visibility that routine workflow records can realistically provide.

The primary dependent variable is Time_Overrun_Minutes, defined as Actual_Time_Minutes minus Estimated_Time_Minutes. This measure is more appropriate than a binary delay indicator because it captures the magnitude of deviation from plan rather than collapsing performance into a yes/no outcome. That choice is consistent with prior work showing that target definition strongly shapes the usefulness of predictive process analysis and its managerial interpretation (Verenich et al., 2019; Teinemaa et al., 2019; Ceravolo et al., 2024).

Table 1. Variable Definitions and Analytical Roles

Variable	Type	Description	Analytical role
Process Name	Nominal	Business process family (e.g., invoice approval, customer complaint)	Predictor
Task Type	Nominal	Operational task category	Predictor
Priority Level	Ordinal label	Priority assigned to the task	Predictor
Department	Nominal	Responsible functional unit	Predictor
Assigned Employee ID	Nominal	Assigned employee identifier	Predictor
Task Start Time	Timestamp	Observed start timestamp	Used for temporal feature extraction only
Task End Time	Timestamp	Observed end timestamp	Excluded from prediction to avoid leakage
Estimated Time Minutes	Numeric	Planned task duration in minutes	Predictor
Actual Time Minutes	Numeric	Observed task duration in minutes	Outcome support / excluded as predictor
Delay Flag	Binary	1 if Actual_Time_Minutes > Estimated_Time_Minutes	Supplementary screening outcome
Approval Level	Ordinal label	Approval tier associated with the task	Predictor
Employee Workload	Numeric discrete	Current workload indicator of the assigned employee	Predictor
Cost Per Task	Numeric	Task-level cost estimate	Predictor
Time Overrun Minutes	Numeric	Actual_Time_Minutes - Estimated_Time_Minutes	Primary operational outcome
Time Overrun Ratio	Numeric	Actual_Time_Minutes / Estimated_Time_Minutes	Descriptive derived variable
Start Hour	Numeric discrete	Hour of day derived from Task_Start_Time	Derived predictor
Start DayOfWeek	Nominal	Day of week derived from Task_Start_Time	Derived predictor
Start Month	Nominal	Calendar month derived from Task_Start_Time	Derived predictor
Start Weekend	Binary	1 if task started on weekend	Derived predictor
Cost per Estimated Minute	Numeric	Cost_Per_Task / Estimated_Time_Minutes	Derived predictor

Note. The analytical objective is to evaluate ex ante operational visibility. Accordingly, realized-performance fields that would leak outcome information (Task End Time, Actual Time Minutes, and Delay Flag) were excluded from the predictor set. This follows established practice in predictive workflow studies, which stress the importance of separating information available before completion from information revealed only after the outcome is known (Maggi et al., 2014; Ceravolo et al., 2024).

Data Preparation and Variable Construction

Data preparation followed four steps. First, the timestamp fields were parsed into datetime objects. Second, time consistency was checked by comparing the minute difference between Task_Start_Time and Task_End_Time with the recorded Actual_Time_Minutes. Third, the derived variables Time_Overrun_Minutes and Time_Overrun_Ratio were constructed. Fourth, the dataset was screened for missing values, implausible entries, and duplicate observations, none of which proved materially problematic. The sequence was kept deliberately transparent because data quality, label construction, and temporal consistency are central to the credibility of time-related workflow analysis (Verenich et al., 2019; Ceravolo et al., 2024).

The derived predictors were chosen to preserve managerial interpretability. Start_Hour, Start_DayOfWeek, Start_Month, and Start_Weekend were extracted from Task_Start_Time to capture broad temporal operating conditions, while cost-normalized and plan-normalized fields were created to retain visibility over resource intensity and planning quality. This emphasis on interpretable, workload-relevant signals is consistent with research that values comprehensible predictive support (Breuker et al., 2016) and with resource-aware studies that foreground workload and execution context (Aalikhani et al., 2025).

Categorical variables were encoded using one-hot encoding with unknown categories ignored at scoring time. Numerical variables were standardized for the linear models and passed through without scaling for the tree-based models. These preprocessing choices were intended to balance comparability, robustness, and interpretability rather than to maximize technical complexity, in line with the study's decision-support orientation (Breuker et al., 2016).

Analytical Assessment Strategy

The analytical strategy combined a simple managerial benchmark with a small set of defensible regression models. Linear regression and ridge regression were used as transparent baselines, while random forest and gradient boosting were included to test whether non-linear patterns in the workflow variables could provide additional visibility over overrun risk. The comparison is intentionally pragmatic: the aim is to determine whether routine workflow data add enough incremental information to improve operational control beyond a simple historical average. This pragmatic model framing is consistent with systematic review evidence showing that applied BPM studies draw value from diverse methods when those methods remain tied to decision support and process improvement (Weinzierl et al., 2024; Breuker et al., 2016).

The sample was divided into training and holdout test subsets using an 80/20 split with a fixed random seed. Five-fold cross-validation on the training set was used to assess model stability, and the final comparison relied on out-of-sample mean absolute error (MAE), root mean squared error (RMSE), and R-squared. These metrics were selected because they support managerial interpretation of forecast error in minutes rather than only abstract statistical fit. Prior benchmark work likewise emphasizes that evaluation design should reflect both predictive validity and the operational meaning of the target being assessed (Verenich et al., 2019; Ceravolo et al., 2024).

A supplementary classification exercise was conducted on Delay_Flag using balanced logistic regression and balanced random forest classification. This step was treated as supporting analysis only. In highly imbalanced settings, binary delay labels can obscure the scale of operational deviation and can make performance metrics look stronger than the underlying managerial signal really is, a concern that is also visible in outcome-oriented benchmark studies (Teinemaa et al., 2019; Ceravolo et al., 2024).

RESULT AND DISCUSSION

Descriptive View of Operational Exposure

Table 2 reports the core descriptive profile of the dataset. Estimated task duration averages 123.91 minutes, whereas actual duration averages 180.86 minutes. The mean time overrun is 56.95 minutes with a standard deviation of 36.54, indicating that tasks are delayed on average and that the extent of the overrun varies substantially across observations. Median overrun is 57 minutes, which confirms that positive deviation from plan is not an isolated phenomenon but a routine operational pattern in the data.

Differences across broad operational categories are present but modest in magnitude relative to the overall variance of the target. The range of mean overruns is only 2.83 minutes across process families, 1.95 minutes across departments, and 4.67 minutes across priority levels. This pattern suggests that overrun exposure is not concentrated in one obviously problematic process class. Instead, it appears as a diffuse control issue that cuts across the workflow system.

Workload shows a more nuanced pattern. Figure 2 indicates that mean overrun is not monotonic in the employee workload indicator. Average overrun is lowest around workload level 5 and rises again at the upper end of the scale, while approval level appears to condition that pattern. Read managerially, this suggests that overrun exposure is shaped less by workload volume alone than by the way workload interacts with control intensity and task context, a reading consistent with recent resource-aware findings (Aalikhani et al., 2025).

Table 2. Descriptive Profile of The Main Numerical Variables

Variable	Mean	SD	Min	Median	Max
Estimated Time Minutes	123.91	66.20	10.00	121.00	240.00
Actual Time Minutes	180.86	75.56	7.00	180.00	357.00
Time Overrun Minutes	56.95	36.54	-5.00	57.00	120.00
Time Overrun Ratio	1.80	1.17	0.58	1.46	12.80
Employee Workload	5.52	2.84	1.00	6.00	10.00
Cost Per Task	277.98	130.76	50.32	279.36	499.94
Start Hour	11.34	6.93	0.00	11.00	23.00

Note. The primary overrun measure is reported alongside a ratio-based companion measure to show the proportional deviation between realized and planned duration.

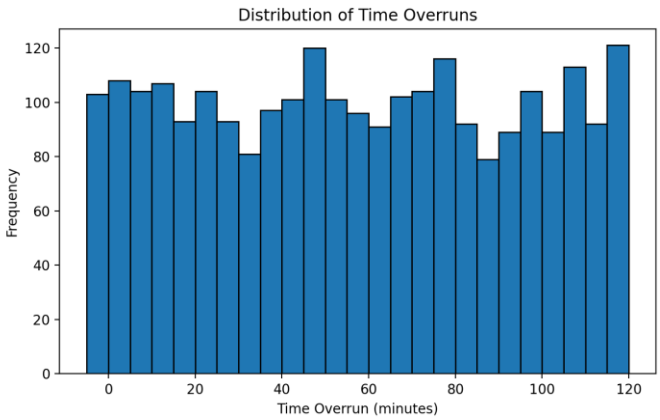

Figure 1. Distribution of Time Overruns

Note. Overrun values are broadly dispersed across the support of the variable rather than concentrated near a narrow band, reinforcing the need for operational monitoring rather than reliance on a single fixed tolerance.

Table 3. Mean Time Overruns Across Major Operational Categories

Group	Category	N	Mean Overrun (min)
Process Name	Customer Complaint	501	57.16
Process Name	HR Onboarding	510	55.04
Process Name	IT Support Ticket	492	57.25
Process Name	Invoice Approval	480	57.87
Process Name	Purchase Order	517	57.51
Task Type	Approval	493	55.78
Task Type	Data Entry	480	55.73
Task Type	Escalation	504	57.19
Task Type	Review	530	60.50
Task Type	Validation	493	55.27
Priority Level	Critical	641	58.07
Priority Level	High	618	57.01
Priority Level	Low	663	58.53
Priority Level	Medium	578	53.85
Department	Customer Service	492	56.22
Department	Finance	504	58.05
Department	HR	502	56.11
Department	IT	497	57.24
Department	Operations	505	57.13
Approval Level	Level 1	826	56.40
Approval Level	Level 2	821	56.78
Approval Level	Level 3	853	57.66

Note. Category-level differences are visible, but they remain small relative to total dispersion in the target, indicating that overrun control is a cross-cutting operational issue rather than a single-category problem.

Figure 2. Mean Time Overrun by Employee Workload and Approval Level

Note. The descriptive relationship between workload and overrun is non-monotonic and varies across approval levels, suggesting that managerial interpretation should focus on conditional combinations of exposure rather than isolated averages.

Comparative Analytical Assessment

Table 4 compares the out-of-sample performance of alternative analytical approaches. Cross-validation on the training sample ranked random forest first, followed closely by gradient boosting, but the differences among models were small. On the holdout test set, random forest was the strongest machine-learning model with a test MAE of 31.56 minutes and RMSE of 36.82 minutes, while gradient boosting performed similarly. The linear models were only marginally weaker. This narrow separation across models is not unusual in workflow prediction, where evaluation outcomes are often shaped as much by data structure and target specification as by algorithm choice (Verenich et al., 2019; Ceravolo et al., 2024).

The central operational comparison, however, is against the simple historical benchmark. The mean baseline achieved a test MAE of 31.01 minutes and RMSE of 36.18 minutes, slightly outperforming every machine-learning model on the primary target. This result is substantively important because it suggests that routine workflow descriptors carry only limited incremental signal about deviation from plan. In operational terms, the main challenge is therefore not merely selecting a more sophisticated algorithm, but improving the planning baseline and the visibility of execution conditions—an interpretation that echoes application studies linking analytics to planning support rather than to model contest alone (Kannan et al., 2022; Botelho et al., 2026).

A secondary analysis on actual task duration clarifies the point. Using the same predictors, gradient boosting achieved a test MAE of 31.82 minutes and an R-squared of 0.736. The contrast is important. Routine workflow descriptors appear more informative for estimating how long work will take than for estimating how much it will exceed plan. This result aligns with broader process-analytics discussions that distinguish between absolute performance estimation and deviation-based control outcomes (Verenich et al., 2019; Ceravolo et al., 2024).

Table 4. Comparative Analytical Assessment for The Primary Overrun Measure

Analytical approach	Test MAE (min)	Test RMSE (min)	Test R-squared
Mean Baseline	31.01	36.18	-0.006
Random Forest	31.56	36.82	-0.042
Gradient Boosting	31.82	37.21	-0.064
Ridge Regression	32.07	38.02	-0.111
Linear Regression	32.23	38.27	-0.126

Note. The historical mean benchmark slightly outperformed every machine-learning model on the primary overrun measure, which suggests that routine workflow descriptors provide only modest incremental signal for deviation from plan.

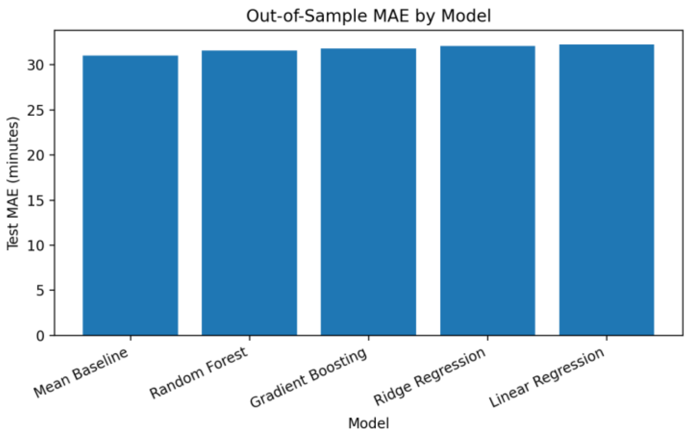

Figure 3. Comparative Test Error Across Analytical Approaches

Note. Error differences are small across approaches, indicating that the managerial issue is less about algorithm selection and more about the information content of the workflow data used to monitor overrun exposure.

Operational signals associated with overrun risk

Table 5 and Figure 4 summarize the relative analytical signal from the best-performing random forest model. The strongest indicators are cost per task, cost intensity relative to planned minutes, planned duration, start hour, employee assignment, and workload. Read operationally, these indicators point to three broad sources of exposure: the quality and resource intensity of the original plan, the assignment and workload context of execution, and broad temporal operating conditions. This pattern is consistent with the workload-aware logic developed in recent resource-sensitive prediction research (Aalikhani et al., 2025).

These signals should be interpreted cautiously. Because overall predictive fit for the primary overrun outcome is weak, the rankings do not establish definitive drivers of overrun. They are better understood as directional cues for managerial attention. That interpretation also follows the broader argument that predictive support is most useful when it remains comprehensible and action-oriented rather than overstated as causal explanation (Breuker et al., 2016).

Table 5. Relative Analytical Signal of Operational Variables in The Best-Performing Random Forest Model

Operational variable	Relative analytical signal
Cost Per Task	0.140
Cost per Estimated Minute	0.114
Estimated Time Minutes	0.114
Start Hour	0.094
Assigned Employee ID	0.078
Employee Workload	0.069
Start Month	0.069
Process Name	0.055
Task Type	0.055

Operational variable	Relative analytical signal
Start DayOfWeek	0.055

Note. Given the weak overall predictive fit for the primary target, these values are interpreted as directional signals rather than conclusive evidence of causal importance.

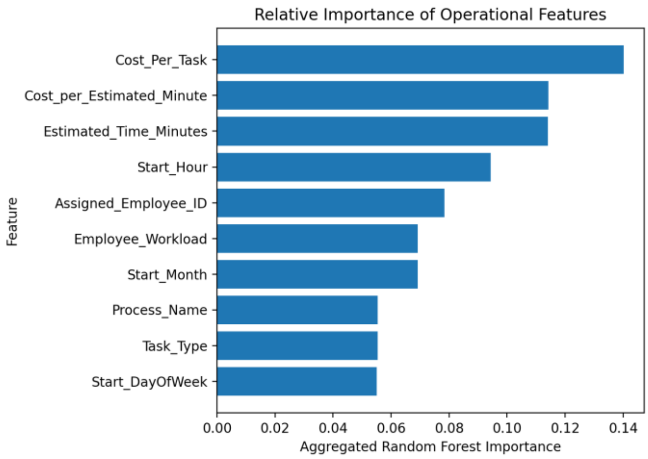

Figure 4. Relative Signal of Operational Variables Associated with Time Overruns
Note. Planning-related, cost-related, and assignment-related variables carry more signal than broad process descriptors, but total predictive signal remains limited on the primary target.

Supplementary Delay-Flag Assessment

The supplementary classification exercise further supports the decision to keep Delay_Flag outside the core contribution of the study. Balanced logistic regression produced precision of 0.949, recall of 0.500, F1-score of 0.655, and ROC-AUC of 0.646, while balanced random forest achieved precision of 0.958, recall of 0.639, F1-score of 0.767, and ROC-AUC of 0.580. Although these values look respectable on selected metrics, the class distribution is highly uneven. As prior benchmark work warns, outcome framing can materially shape how informative such scores really are for operational use (Teinemaa et al., 2019; Ceravolo et al., 2024).

Overall, the results yield a coherent operational synthesis. First, time overruns are systematically positive on average and broadly dispersed. Second, routine workflow descriptors provide useful but limited visibility into where overrun risk accumulates. Third, workload, approval structure, plan quality, and assignment context matter more than broad process labels alone. Finally, analytics appears most useful when directed toward estimate calibration, workload governance, and exception handling rather than toward stand-alone prediction contests, which is consistent with recent intervention-oriented monitoring research (Fahrenkrog-Petersen et al., 2022).

Findings

The findings suggest that workflow time overruns are best understood as operational control issues rather than purely predictive events. Historical workflow data can reveal where exposure tends to accumulate, especially around workload conditions, approval requirements, and planning-related variables, but the managerial meaning of those signals lies in how they inform control routines rather than in model performance alone. This reading is consistent with recent process-science work that treats predictive monitoring as valuable when it supports decision making and intervention rather than isolated forecasting (Ceravolo et al., 2024; Fahrenkrog-Petersen et al., 2022).

The contrast between actual duration and time overrun is especially important. Absolute duration partly reflects task size, which is reasonably captured in this dataset through estimated time and related operational attributes. Time overrun, by contrast, represents deviation from an expected standard. That makes it more sensitive to estimate quality, coordination breakdowns, approval waiting, and local execution shocks that are not fully visible in static workflow descriptors. For operational management, this means that overrun control depends as much on the quality of planning and governance as on the quality of prediction. Prior work on time-related process analysis also suggests that target construction and data granularity strongly shape what can be learned from operational records (Verenich et al., 2019; Ceravolo et al., 2024).

This interpretation extends prior workflow-analytics research in a more managerial direction. The literature on predictive process monitoring shows that data can support earlier intervention (Maggi et al., 2014; Ceravolo et al., 2024), while resource-aware analysis highlights the role of workload and operating context (Aalikhani et al., 2025). The present study supports that broader view by showing that the value of analytics is greatest when it strengthens operational visibility, estimate discipline, workload governance, and approval review rather than when it is framed as a standalone technical solution. That emphasis on actionability also aligns with intervention-oriented monitoring research (Fahrenkrog-Petersen et al., 2022).

The secondary actual-duration results are particularly informative for theory and practice. They show that the same workflow variables which are only weakly informative for overrun control can still be useful for direct duration estimation. This suggests that managers should distinguish between two related but different questions: how long a task is likely to take, and how much it is likely to exceed its plan. The first question is better supported by the current data than the second, which reinforces the importance of improving the planning baseline itself. Application studies in production settings point in the same direction by showing that analytical support is often most valuable when connected to planning reliability and execution control (Kannan et al., 2022; Botelho et al., 2026).

Finally, the delay-flag results caution against reducing operational performance to a binary alarm. Once the target is compressed into delayed versus not delayed, the operational meaning of deviation is partly lost. This is one reason the article treats binary delay classification as supplementary rather than central. That decision is consistent with outcome-oriented benchmark studies showing that target framing can materially alter the managerial usefulness of predictive outputs (Teinemaa et al., 2019).

Managerial Implications

The first managerial implication is that organizations should treat time overruns as a planning-and-control issue, not only as a prediction problem. The most immediate routine is estimate calibration: completion-time standards should be reviewed by task type, department, and approval level so that planned durations better reflect actual operating conditions. This recommendation fits the applied operations literature, which consistently ties analytical value to planning quality and control discipline (Kannan et al., 2022; Botelho et al., 2026).

Second, the results support stronger workload governance. Because overrun exposure varies across workload conditions rather than rising in a simple linear fashion, managers should define workload thresholds that trigger supervisory review before execution reliability deteriorates. Such thresholds can be built into routine workload meetings, staffing reviews, or daily control dashboards. Resource-aware research provides strong support for treating workload as a controllable organizational condition rather than merely a background descriptor (Aalikhani et al., 2025).

Third, approval design deserves closer operational attention. The descriptive patterns suggest that approval requirements interact with workload conditions and can amplify timing instability. Periodic review of layered approvals, especially in high-volume or high-variance tasks, may therefore reduce avoidable delay without weakening governance. More generally, this reflects the argument that analytical insight becomes most useful when translated into concrete intervention routines (Fahrenkrog-Petersen et al., 2022).

Fourth, analytics should be embedded in exception handling rather than treated as a one-off modeling project. Managers can use workflow data to maintain dashboards of high-overrun tasks, identify recurring control points, and escalate cases that combine weak planning standards with adverse workload conditions. This kind of use is closely aligned with recent calls to connect predictive support, process visibility, and managerial action (Ceravolo et al., 2024; Fahrenkrog-Petersen et al., 2022).

Finally, a practical roadmap emerges. In the short term, organizations can use existing workflow data for estimate recalibration, workload monitoring, and exception-based review. In the medium term, they should capture richer execution data on approval waiting, queueing, reassignment, and interruption. Over time, the objective should be to build a governance system in which analytical visibility, planning discipline, and process review reinforce one another rather than operate as separate initiatives (Weinzierl et al., 2024; Ceravolo et al., 2024).

Conclusion

This study shows that workload-aware analytics from routine workflow data can improve operational control by making time-overrun exposure more visible. The evidence from 2,500 task-level observations indicates that planned completion time averaged 123.91 minutes, while actual completion time averaged 180.86 minutes, producing a mean overrun of 56.95 minutes. This confirms that the gap between planned and actual workflow performance is systematic and managerially significant.

The objective of identifying operational signals associated with overrun risk is addressed through the finding that planning quality, cost intensity, employee workload, approval level, assignment context, and temporal operating conditions provide useful but limited visibility. Therefore, the practical value of machine learning in this study is not algorithmic replacement of managerial judgment, but support for estimate calibration, workload governance, approval review, and exception-based monitoring.

Limitations

Several limitations should be acknowledged. First, the study is based on a single supplied dataset and does not establish cross-organizational generalizability. Second, the data consist of one task outcome per row rather than richer longitudinal execution records, which limits visibility over queueing, handoffs, approval waiting, reassignment, and other process conditions that may influence overrun behavior. Third, the design is predictive rather than causal and therefore cannot determine whether the observed associations reflect underlying mechanisms or correlated operating conditions. Each of these limitations echoes broader concerns in the predictive process literature regarding data granularity, target construction, and external validity (Verenich et al., 2019; Ceravolo et al., 2024).

A further limitation concerns data provenance and scope. The file is internally coherent and suitable for analysis, but the article cannot independently verify whether it originates from a live organizational system or a constructed workflow scenario. This does not invalidate the empirical exercise, but it does caution against strong claims about external validity or organization-specific policy design.

Future research should therefore move in three related directions. The first is richer operational data capture, including approval waiting times, queue conditions, reassignment events, and other execution-context variables that speak directly to control. The second is stronger integration between analytics and managerial routines, for example through dashboard design, escalation triggers, and intervention protocols. The third is broader validation across organizations and workflow settings so that time-overrun control can be studied not only as a modeling task, but as an evolving capability in operational management. These directions are consistent with current calls for richer process data, stronger intervention logic, and closer links between predictive support and action (Ceravolo et al., 2024; Fahrenkrog-Petersen et al., 2022).

REFERENCES

Aalikhani, R., Fathian, M., & Rasouli, M. R. (2025). Enhancing remaining time prediction in business processes by considering system-level and resource-level inter-case features. Software and Systems Modeling, 24(5), 1515-1547. https://doi.org/10.1007/s10270-025-01267-4

Botelho, J. G. S., Choueiri, A. C., Pécora Júnior, J. E., & Santos, E. A. P. (2026). A process mining and machine learning based approach for remaining time prediction of production orders. Journal of Intelligent Manufacturing. https://doi.org/10.1007/s10845-026-02792-9

Breuker, D., Matzner, M., Delfmann, P., & Becker, J. (2016). Comprehensible predictive models for business processes. MIS Quarterly, 40(4), 1009-1034.

Ceravolo, P., Comuzzi, M., De Weerdt, J., Di Francescomarino, C., & Maggi, F. M. (2024). Predictive process monitoring: Concepts, challenges, and future research directions. Process Science, 1, 2. https://doi.org/10.1007/s44311-024-00002-4

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Springer.

Evermann, J., Rehse, J. R., & Fettke, P. (2017). Predicting process behaviour using deep learning. Decision Support Systems, 100, 129-140. https://doi.org/10.1016/j.dss.2017.04.003

Fahrenkrog-Petersen, S. A., Tax, N., Teinemaa, I., Dumas, M., de Leoni, M., Maggi, F. M., & Weidlich, M. (2022). Fire now, fire later: alarm-based systems for prescriptive process monitoring. Knowledge and Information Systems, 64, 559-587. https://doi.org/10.1007/s10115-021-01633-w

Kannan, R., Abdul Halim, H. A., Ramakrishnan, K., Ismail, S., & Wijaya, D. R. (2022). Machine learning approach for predicting production delays: a quarry company case study. Journal of Big Data, 9, 94. https://doi.org/10.1186/s40537-022-00644-w

Maggi, F. M., Di Francescomarino, C., Dumas, M., & Ghidini, C. (2014). Predictive monitoring of business processes. In Advanced Information Systems Engineering (Lecture Notes in Computer Science, Vol. 8484, pp. 457-472). Springer. https://doi.org/10.1007/978-3-319-07881-6_31

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. MIS Quarterly, 35(3), 553-572.

Tax, N., Verenich, I., La Rosa, M., & Dumas, M. (2017). Predictive business process monitoring with LSTM neural networks. In Advanced Information Systems Engineering (Lecture Notes in Computer Science, Vol. 10253, pp. 477-492). Springer. https://doi.org/10.1007/978-3-319-59536-8_30

Teinemaa, I., Dumas, M., La Rosa, M., & Maggi, F. M. (2019). Outcome-oriented predictive process monitoring: Review and benchmark. ACM Transactions on Knowledge Discovery from Data, 13(2), 17:1-17:57. https://doi.org/10.1145/3301300

van der Aalst, W. M. P., Schonenberg, M. H., & Song, M. (2011). Time prediction based on process mining. Information Systems, 36(2), 450-475. https://doi.org/10.1016/j.is.2010.09.001

Verenich, I., Dumas, M., La Rosa, M., Maggi, F. M., & Teinemaa, I. (2019). Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring. ACM Transactions on Intelligent Systems and Technology, 10(4), 1-34. https://doi.org/10.1145/3331449

Weinzierl, S., Zilker, S., Dunzer, S., & Matzner, M. (2024). Machine learning in business process management: A systematic literature review. Expert Systems with Applications, 253, 124181. https://doi.org/10.1016/j.eswa.2024.124181

Weske, M. (2019). Business process management: Concepts, languages, architectures (3rd ed.). Springer.