

Mining procedural non-compliance in enterprise asset management data: a CRISP-DM study with association rules on IBM Maximo work orders

Burman Bagaskara ^{1*}, Aris Puji Widodo ², Indra Waspada ²

¹*Master of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Semarang, Indonesia*

²*Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia*

Abstract Enterprise asset management platforms accumulate millions of maintenance transactions, yet the way users actually populate those records is almost never audited after go-live. This paper treats procedural non-compliance as a mineable object and reports a complete CRISP-DM study on the work order module of an IBM Maximo installation run by a rail transport company. Seven relational tables were extracted from IBM DB2. After filtering to corrective (CM) and preventive (PM) work, 12,368 orders remained, 5,790 corrective and 6,578 preventive. Completeness and integrity were profiled first; child tables were then aggregated to order level and recoded into a binary transaction matrix carrying eleven items, two contextual and nine describing distinct forms of procedural failure. Apriori and FP-Growth were executed over four scenarios at support 0.10, confidence 0.70 and lift above 1.0, under a two-step ablation designed to neutralise items that are almost always present. Every reported rule was additionally submitted to a one-sided Fisher exact test, corrected for multiple testing across all 155 tested rules, and each syndrome was re-estimated in a multivariable logistic model so that the items are adjusted for one another, and both thresholds were varied over a sixteen-cell grid. The two algorithms returned numerically identical rule sets, which establishes robustness, while Apriori ran 23 to 98 times faster on this low-dimensional item space; because every absolute time lies below one second, that ordering is reported as a property of small binary matrices rather than as a general ranking of the two algorithms. Two behavioural syndromes were isolated. Corrective work fails on timing: 47.0% of CM orders close within two minutes and 98.0% of those carry an earlier execution stamp, at lift 1.350. Preventive work fails on scheduling: overlapping labour combined with backdating implies late completion at 96.8% confidence and lift 1.759. Both syndromes survive Benjamini-Hochberg correction at $q < 0.001$, retain adjusted odds ratios of 49.3 and 69.7 in the logistic models, and reappear at every one of the sixteen threshold combinations tested, where they are also the only rules common to all cells. The extracted rules were converted into platform validation designs, a blueprint checklist and a rule engine that states each finding in natural language.

Keywords Association Rule Mining, CRISP-DM, Data Quality, Enterprise Asset Management, Apriori, FP-Growth

AMS 2010 subject classifications 62-07, 68T09, 90B25

DOI: 10.19139/soic-2310-5070-4473

1. Introduction

Organisations that operate large fleets of physical assets depend on enterprise asset management (EAM) software to schedule maintenance, record what was done and account for the condition of every item they own. EAM is frequently filed under enterprise resource planning, although the two are not interchangeable: EAM carries its own maintenance semantics and its own failure taxonomies, which is why dedicated case evidence remains valuable [1]. Rail operators make the point concretely. A recent review of railway maintenance digitalisation traces how asset registers, condition monitoring and work management have merged into one decision layer [2], and the same

*Correspondence to: Burman Bagaskara (Email: bagasgibran@gmail.com). Master of Information Systems, School of Postgraduate Studies, Universitas Diponegoro, Jl. Imam Bardjo SH No. 5, Semarang, Central Java, Indonesia (50241).

convergence now drives interest in predictive and prescriptive strategies [3, 4]. Whether any of those strategies can be trusted depends on a question that is rarely asked out loud: how good is the historical record they are trained on?

Most platforms are not maintained by their owners alone. Accredited consultancies deliver configuration changes and resolve incidents under a service contract that allocates a fixed annual quota of hours. The arrangement contains an awkward asymmetry. Nothing happens until the client asks. If no ticket arrives the quota expires, the client absorbs the cost, and the consultant records little activity against its own targets. Studies of asset management practice describe the resulting risk exposure in general terms [5, 6], and work on post-implementation behaviour identifies the same pattern from the other side, where adaptation is triggered by user demand rather than by evidence that the system has drifted [7]. Metro operators carry an additional constraint, since maintenance decisions translate directly into service penalties for passengers [8].

A second problem is quieter. Once a system goes live, what users type into it is seldom inspected again. Acceptance testing checks functionality, not the longitudinal behaviour of the people entering records, so defects accumulate without producing any visible failure until a report is questioned. Data quality research supplies the vocabulary: completeness and freedom from error are the dimensions that matter most in transactional systems [9], and recent surveys have widened the set with governance and semantic criteria that map onto enterprise software [10, 11, 12]. What that literature does not offer is a way of deciding which defects deserve attention first.

1.1. Related work

Attempts to raise EAM quality through analytics fall into three groups. Oyoo built a collaboration-based framework for automatic data validation during implementation, arguing that validation belongs in the delivery method rather than in later clean-up [13]. Santos and colleagues showed that natural language processing can absorb part of the manual burden in work order handling [14]. Lu and co-authors mined maintenance records to trace why reliability data turns out inaccurate and how that should reshape maintenance management [15]. These three studies share a boundary. Oyoo's framework is preventive and therefore silent about installations that are already live; the text mining pipeline of Santos and colleagues presumes narrative fields dense enough to parse, which the present dataset does not have; and Lu and co-authors stop once a cause has been named, without converting the diagnosis into a control that a delivery team can switch on. The gap this paper addresses is the step none of them takes, namely turning a measured behavioural pattern into a configurable platform rule. Within this journal, the same distance between raw enterprise records and usable decision support has been identified as the central obstacle to business intelligence adoption [42], and quality has been argued to be measurable as a set of explicit dataset attributes rather than treated as a precondition [43].

Association rule mining recurs across all three groups. A 2024 review catalogues its role in maintenance and reliability [16]; facility management researchers pair it with large language models to anticipate requests [17]; process researchers use it to flag and explain anomalous behaviour inside running processes [18, 19]. Applications well outside maintenance, from aircraft event analysis [20] to geochemical pattern recognition [21], confirm that the technique travels wherever labels are unavailable and behaviour has to be inferred from co-occurrence. What these applications have in common is that the input is already curated; none of them is asked to work on a matrix in which the majority of a field is empty by construction.

Process mining is the closest methodological alternative and deserves to be named rather than left implicit. It reconstructs the control flow actually followed by cases and measures how far each deviates from a reference model [40], which makes it the stronger instrument when the question concerns ordering, throughput or conformance to a prescribed path. It is not the stronger instrument here, for three reasons that follow from the data rather than from preference. First, the event log needed for control-flow discovery is exactly the artefact whose reliability is in doubt: this study establishes that more than 95% of corrective status changes are backdated and that 63.4% of closures occur within two minutes, so timestamps cannot be trusted to order events. Second, several of the eleven items are not events at all; a missing attachment and a missing failure report have no position in a trace. Third, the object of interest is co-occurrence across heterogeneous evidence types rather than sequence. Association rules impose no ordering assumption and therefore remain valid on precisely the data whose ordering is unreliable. The two approaches are complementary, and §4 returns to what process mining could add once timestamp integrity has been repaired.

The process discipline comes from CRISP-DM. Two decades after release it remains the reference model for data mining projects [22], and it has been adapted repeatedly: to regulated financial services [23], to production data engineering in small-batch manufacturing [24], and to agile delivery, which compensates for its waterfall bias [25]. Extensions such as CASP-DM add explicit context handling [26]. Reviews continue to track where the model is used and how it evolves [27, 28], and it has recently been combined with process mining to assess simulation models [29]. One theme is constant across these studies: the business understanding phase, not the algorithm, decides whether anyone uses the output [30].

1.2. Contributions

The three strands rarely meet. EAM improvement research produces validation frameworks and text mining pipelines but stops short of a repeatable process; association rule studies demonstrate technique on data that is already clean; CRISP-DM studies demonstrate process on domains other than asset maintenance. To the best of the authors' knowledge, no published work runs a full CRISP-DM cycle over EAM work order transactions in order to audit how users behave, and then carries the resulting rules into a deployable artefact.

It is worth separating two kinds of novelty before listing what follows. Apriori, FP-Growth, CRISP-DM and Fisher's exact test are established and are not claimed as contributions; no algorithm is proposed or modified here. The contribution is an integrated, deployment-oriented framework for discovering procedural non-compliance in EAM work order data: an encoding that converts documentation gaps into mineable items, an ablation that prevents saturated items from masking process structure, a validation layer that separates statistical from practical significance, and a translation of the surviving rules into platform controls. The framework is what transfers between installations; the particular support and confidence values reported below are properties of this operating context and are not claimed to be constants. On that basis the paper makes five contributions.

1. Completeness and integrity are quantified across seven tables of a live Maximo installation, separating a sound administrative backbone from severe sparsity in the fields that describe physical execution.
2. Procedural non-compliance is encoded as a binary transaction matrix of eleven items, so that missing documentation becomes a feature rather than a defect to be imputed away, and rules are mined separately for corrective and preventive work.
3. Apriori and FP-Growth are compared on this matrix. They agree on every rule, yet Apriori is between 23 and 98 times faster, an ordering that is reported here as a property of low-dimensional binary matrices and explained in terms of candidate-space size, not as a general ranking.
4. The rule base is validated three ways rather than one: every rule is corrected for multiple testing, each syndrome is re-estimated in a multivariable logistic model that adjusts each item for the others, and the pairwise correlation structure is reported so that the value added by multi-item rules over simpler statistics can be judged directly.
5. The rule base is converted into three working artefacts: configurable platform validations, a blueprint checklist for new engagements, and a rule engine that renders each rule as a natural-language recommendation.

1.3. Organisation

Section 2 sets out the research framework and then works through the six CRISP-DM phases in order, one subsection each. Section 3 reports the outcome of the same six phases in the same order, so that every design decision can be traced to the result it produced. Section 4 interprets the findings against prior work and states the limitations. Section 5 concludes.

2. Methodology

2.1. Research framework

The study follows the six phases of CRISP-DM as originally specified [31], with one deviation that is documented rather than hidden: the loop from evaluation back to data preparation was traversed once, after the first modelling run produced rules that merely restated a definition. Figure 1 shows the framework as instantiated here, with the concrete activity of each phase written into the diagram.

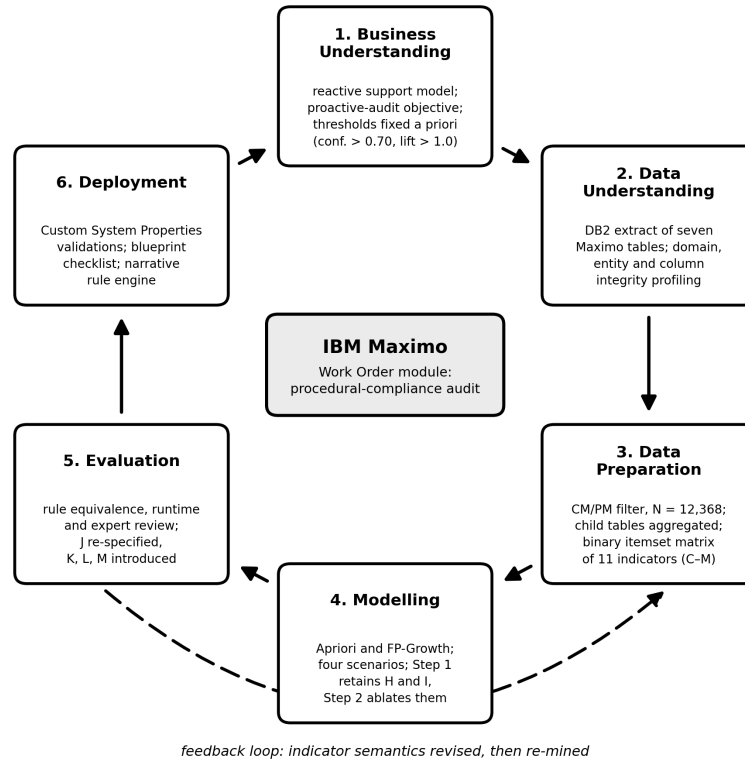


Figure 1. CRISP-DM framework instantiated for the Maximo work order audit.

The object of study is the work order module of IBM Maximo version 9.0 [32], currently the leading product in the EAM software category [33], operated by a rail transport company and supported by an accredited implementation partner. The first author has worked on this account as a Maximo consultant since 2021, which provided both authorised database access and the domain knowledge needed to read operational meaning into the records.

Maximo was selected for three reasons that also bear on transferability. It holds the largest share of the enterprise asset management category [33], so findings expressed in its schema reach the widest installed base. Its work order model is table-stable: WORKORDER, WOSTATUS, WORKLOG, DOCLINKS, LABTRANS, MULTIASSETLOCCI and WOACTIVITY are present in every deployment, which is what allows the encoding of §2.4 to be reproduced without bespoke mapping. And its validation layer is exposed through configuration rather than source code, which is what makes the deployment artefacts of §2.7 feasible at all. A platform whose controls required core modification would not support the same route from rule to control.

All analysis ran in Python 3.11.16 with pandas 2.2.3 and mlxtend 0.23.4, on a single machine with an Apple M3 processor, eight cores and 8 GB of memory, under macOS 26.6.1. The deployment prototype was written in Node.js 20. Runtimes reported in Table 7 are means over thirty repeated executions, each preceded by three warm-up executions that are discarded, on an idle machine connected to mains power so that no frequency scaling applies. Timing is taken with `time.perf_counter` around frequent itemset generation and rule extraction together, with the transaction matrix already resident in memory so that file reading does not enter the measurement. Standard deviations are reported alongside the means, since at this scale the dispersion matters as much as the central value.

Because the phases are the organising principle of the paper as well as of the project, Table 1 maps each of them to the section in which its design is described, the section in which its outcome is reported, and the artefact it produced. The remainder of Section 2 follows the same order.

Table 1. The six CRISP-DM phases mapped to the structure of this paper.

Phase	Design	Outcome	Principal output
1. Business understanding	§2.2	§3.1	Objectives and acceptance thresholds
2. Data understanding	§2.3	§3.2	Completeness and integrity profile
3. Data preparation	§2.4	§3.3	Binary matrix of eleven items
4. Modelling	§2.5	§3.4	Rule sets from Apriori and FP-Growth
5. Evaluation	§2.6	§3.5	Validated rules, significance, sensitivity
6. Deployment	§2.7	§3.6	Validation package, checklist, rule engine

2.2. Phase 1: business understanding

Objectives were written in the language the consultancy already uses. The overarching aim was to replace demand-driven support with evidence-driven support, and five sub-goals followed: find the errors that recur, describe the behaviour that produces them, design an audit that runs without being asked, reduce dependence on client initiative, and point consulting effort at defects that actually cost something.

Acceptance thresholds were fixed in this phase, before any query was written. That ordering matters, because it removes the temptation to tune criteria to whatever the data happens to yield. Rules had to exceed 70% confidence and a lift above 1.0. Recommendations had to be judged implementable by the delivery team. The model had to separate corrective from preventive work rather than let one dominate. Extraction had to be cheap enough to repeat each reporting period.

Confidence and lift alone proved insufficient once the first results arrived, so a practical-significance criterion was added and is stated here rather than applied silently. A rule is called actionable only if it satisfies four conditions at once: support of at least 0.10, so the control it implies would touch a tenth of the population and is worth building; lift of at least 1.25, which excludes associations that are real but barely distinguishable from the base rate; an odds ratio of at least 3, the effect size belonging to the significance test; and a consequent that a platform control can actually block. Statistical significance is a necessary condition in what follows, never a sufficient one. §3.5 shows why the distinction is not academic: three rules in this study are significant at conventional levels while failing every one of the four practical conditions.

2.3. Phase 2: data understanding, source and quality metrics

Data were exported as CSV from IBM DB2 through Cloud Pak for Data. Seven tables were retrieved: WORKORDER as the header, plus MULTIASSETLOCCI, WOSTATUS, WORKLOG, DOCLINKS, LABTRANS and WOACTIVITY. Only attributes visible on the user interface were kept. The restriction is deliberate, since the research question concerns what people see and type, not what the platform computes in the background. The window spans two to three years, long enough to describe settled behaviour and long enough to leave behind the unstable period that follows go-live.

Quality was measured with the completeness and free-of-error metrics of Lee and colleagues [9]. Completeness is defined in (1), where n_{inc} counts incomplete items and n counts all items.

$$\text{Completeness} = 1 - \frac{n_{\text{inc}}}{n} \quad (1)$$

Three dimensions were evaluated: domain integrity, comparing required fields against available fields; entity integrity, the share of rows carrying a non-null primary key; and column integrity, the share of columns holding valid values. Missing values were not dropped. In this study the gap is the observation, a position consistent with recent treatments of missingness as a property of the governance process rather than an accident of measurement [10, 12]. The contrast with inferential modelling is deliberate and worth marking. Where the goal is estimation, missingness is a nuisance to be neutralised by full-information likelihood or by deletion, and the choice among those methods materially changes the estimates [44]. Where the goal is auditing behaviour, the same missingness is the measurement: an imputed work log would erase the very act the study is trying to observe. Treating dataset properties as quantities to be measured in their own right rather than as preconditions to be satisfied follows the same logic [43].

2.4. Phase 3: data preparation, encoding procedural non-compliance

Selection kept records where WOCLASS equals WORKORDER, STATUS is not CAN, and WORKTYPE is either CM or PM. That left 12,368 work orders, 5,790 corrective and 6,578 preventive. Cancelled transactions were discarded because they describe neither completed work nor completed paperwork, so including them would inflate anomaly counts without any operational referent. Long free-text columns were also set aside: column integrity reached only 1.97% in WOACTIVITY and 46.53% in WORKORDER, far below what text mining requires [17]. Preparation is not a preliminary chore in work of this kind, since decisions taken here propagate directly into rule semantics [34].

Association rule mining accepts categorical items only, so every continuous, timestamp and textual input had to be recoded. Child tables were first collapsed to work order level, which prevents the row multiplication a naive join would cause. WORKLOG and DOCLINKS became record counts. LABTRANS produced total hours and, more usefully, a count of technician intervals that intersect intervals booked elsewhere. MULTIASSETLOCCI yielded an additional-asset count, and WOSTATUS the exact INPRG and COMP timestamps. Aggregates were merged into the header with left outer joins on WONUM, WORKORDERID and SITEID, expressed as DB2 common table expressions. Figure 2 traces the whole path. Each work order then became a short itemset labelled with single letters from C to M, the compact basket representation that association analysis expects [35].

The eleven items were not chosen opportunistically, and the selection rule is stated here because it determines what the study can and cannot see. A candidate had to satisfy three conditions. It had to be derivable from tables present in every Maximo deployment, which is what makes the encoding portable. It had to be computable from fields the user can see and edit, since the research question concerns behaviour rather than platform internals. And it had to correspond to a procedural obligation that the consulting team could name independently of the data, so that a rule containing it would be interpretable as a breach rather than as a statistical curiosity. Candidates were discarded when they failed one of these. Repair cost and downtime were dropped because they are populated by a finance interface rather than by the technician. Craft and labour rate were dropped for the same reason. Free-text sentiment on work logs was dropped because column integrity of 30.06% leaves too little text to score. Asset criticality was dropped because it is inherited from the asset master and is therefore a property of the asset rather than of the transaction. Priority survived as a contextual item precisely because it is set per work order by a person. One candidate was revised rather than discarded: the failure report item, whose first specification restated the definition of corrective work and produced rules at confidence 1.000, was rewritten during the single iteration back to this phase described in §2.6.

The encoding itself is mechanical once the aggregates exist. Every anomaly item is the indicator of a predicate over one work order, evaluated after the child tables have been collapsed:

```
for each work order  $w$  in the filtered header set:
   $\text{agg} \leftarrow \text{collapse}(\text{children of } w)$            % counts, extremes, overlap flag
   $E(w) \leftarrow 1$  if  $\text{agg.actstart} < \text{agg.ts.inprg}$  else 0
```

```

F(w) ← 1 if agg.ts_comp - agg.ts_inprg < 120 s else 0
G(w) ← 1 if any labour interval of w intersects an interval
      booked by the same technician on another order else 0
H(w) ← 1 if agg.worklog_count = 0 else 0
I(w) ← 1 if agg.doclink_count = 0 else 0
... and correspondingly for C, D, J, K, L, M
emit the row (C, D, ..., M) into the matrix of worktype(w)

```

No branch of this procedure writes an imputed value, and no branch discards a row for being sparse. A work order with every field empty is a valid transaction carrying many items, which is the intended behaviour.

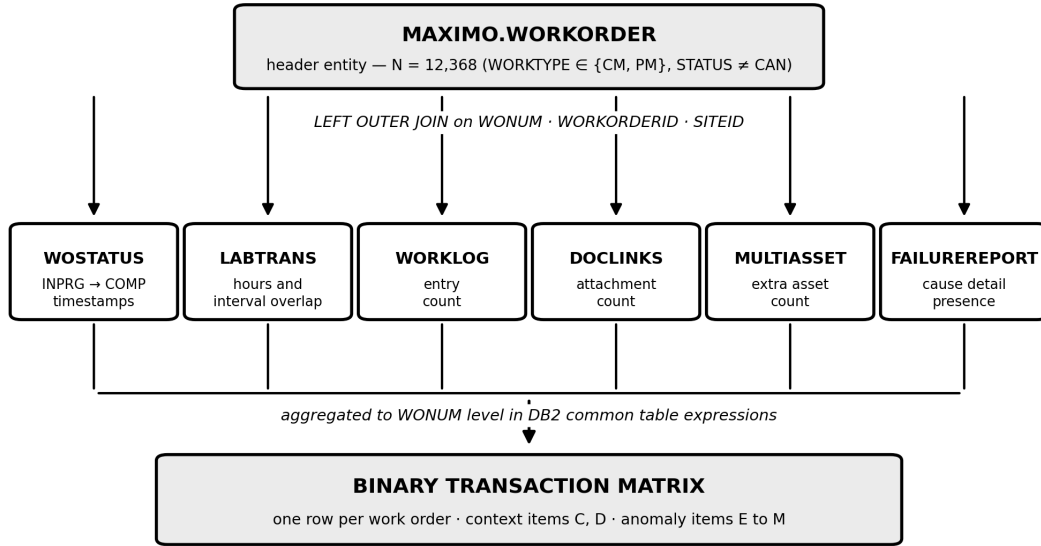


Figure 2. Aggregation of Maximo child tables into the binary transaction matrix.

2.5. Phase 4: modelling, association rules and the ablation design

The goal is discovery rather than prediction, which places the work in the unsupervised branch of data mining [35] and makes association rules the natural instrument. Let X and Y be disjoint itemsets. Support counts how often the two occur together, confidence gives the conditional probability of Y once X holds, and lift compares observed co-occurrence with what independence would predict, so any value above one signals a real association [15, 16].

$$\text{Support}(X \cap Y) = P(X \cap Y) \quad (2)$$

$$\text{Confidence}(Y | X) = \frac{\text{Support}(X \cap Y)}{\text{Support}(X)} \quad (3)$$

$$\text{Lift}(Y | X) = \frac{\text{Confidence}(Y | X)}{\text{Support}(Y)} \quad (4)$$

Apriori and FP-Growth were both implemented and run on identical matrices. The first expands frequent itemsets iteratively under the downward-closure property; the second compresses the database into a frequent pattern tree and skips candidate generation altogether. Hybrid designs combining the two have recently been proposed for retail data [36], which makes a direct comparison on a very different data shape worth reporting. Thresholds were held at support 0.10, confidence 0.70 and lift above 1.0.

Four scenarios were formed by crossing work type with an ablation factor. The reason for ablating is mechanical rather than aesthetic. A binary item present in nearly every transaction saturates the frequent itemsets and crowds out weaker patterns, an effect referred to here as the sponge effect. Step 1 keeps the documentation items H and I; Step 2 removes them.

Ablation is not the only possible response to saturation, and the alternatives were tried rather than dismissed. Raising the support floor does not help, because a saturated item has high support by definition and is removed last, not first; the sensitivity grid in §3.5 confirms this, since the Step 1 rules persist to support 0.20. Post hoc filtering on lift removes the offending rules from the output but not the offending items from the search, so the itemsets that would have carried process structure are still never generated. Removing the items themselves is the only intervention that changes what the algorithm can find. Which items to remove was then settled empirically rather than by assumption: H and I were ablated singly as well as jointly, and Table 8 reports all four outcomes. The result justifies removing both. Each work type is saturated by a different item, so removing only one leaves the other free to dominate that matrix, and removing both is what yields a single encoding usable across scenarios instead of two work-type-specific ones. Splitting by work type follows the same logic, because scheduled preventive rounds and unplanned corrective repairs are different activities and the larger class would otherwise set the tone for both.

2.6. Phase 5: evaluation, statistical validation and sensitivity

Support, confidence and lift describe a rule; none of them establishes that it differs from chance. Every rule carried into the results was therefore also submitted to a one-sided Fisher exact test on the 2×2 contingency table formed by presence or absence of the antecedent against presence or absence of the consequent. Fisher was preferred to a chi-square approximation because several cells become small once the rarer items are involved, and the exact test stays valid there. The odds ratio is reported next to the p -value because it is the effect size belonging to that test, whereas lift is a descriptive ratio. The test is one-sided because the hypothesis is directional. What is at issue is whether two procedural failures co-occur *more* often than independence predicts, which is the same direction already imposed by the lift > 1 filter; a two-sided test would additionally flag negative associations, and a rule stating that one breach makes another less likely is outside the scope of an audit designed to identify controls worth enforcing.

Testing many rules inflates the family-wise error rate, so the number of tests is stated and corrected rather than left implicit. Every rule clearing support and confidence with positive lift in the Step 1 and Step 2 scenarios was tested, giving $m = 155$ tests in total. The Benjamini-Hochberg procedure was applied across all 155 at once [39], controlling the false discovery rate rather than the family-wise rate, which is the appropriate choice when the tested hypotheses are neither independent nor few. Adjusted q -values are reported next to the raw p -values throughout. Because several p -values underflow double precision, the tail probability is accumulated in log space and reported as $\log_{10} p$ where relevant. The stricter Bonferroni bound is also reported in §3.5 for comparison.

A third check addresses a limitation intrinsic to association rules rather than to thresholds. A rule reports a marginal co-occurrence, so an item may appear in a rule only because it travels with a third item that carries the real association. Each syndrome was therefore re-estimated as a multivariable logistic regression in which the consequent is the outcome and every other item enters simultaneously as a predictor, fitted by iteratively reweighted least squares. Items whose prevalence in that matrix falls outside $[0.01, 0.99]$ were excluded before fitting, since a near-constant regressor produces separation and uninterpretable intervals; the excluded items are listed with the results. Adjusted odds ratios and Wald intervals from these models are reported beside the unadjusted Fisher odds ratios, so that the reader can see directly which associations survive mutual adjustment and which do not. The models are a complementary check, not a replacement: logistic regression estimates the effect of each item on one nominated outcome, whereas the rule search makes no such nomination and is what surfaced the syndromes in the first place.

A second check answers a fair objection to any threshold-based method, namely that the findings might only reflect where the cut-offs were placed. Mining was repeated over a grid of four support values (0.05, 0.10, 0.15, 0.20) crossed with four confidence values (0.60, 0.70, 0.80, 0.90), sixteen combinations per work type. The grid was fixed before the runs and was not extended afterwards. Its endpoints bracket the operating threshold on both axes: 0.05 is the lowest support at which a control would still touch enough of the population to be worth configuring,

0.20 the highest at which any three-item combination survives in this data, while 0.60 and 0.90 span the range within which a confidence figure remains defensible to a delivery team. Lift was held at the same floor of 1.0 throughout, since it is a filter on direction rather than on strength and relaxing it would admit associations weaker than the base rate. For each cell the number of surviving rules was recorded together with the identity of every rule in it, not merely the count, so that composition can be compared across cells; the intersection over all sixteen cells is reported as the core rule set, together with whether the syndrome-defining rule for that work type still appeared.

Evaluation had a technical half and a business half. Technically, the algorithms were compared on rule equivalence and runtime, and the models were re-run after the item set was corrected. The first pass exposed a flaw: the failure report item was specified so loosely that it repeated the definition of corrective work and returned rules at confidence 1.000. Its query was rewritten, and three further items covering status bypass, late completion and stagnation were added. On the business side each surviving rule was put to the consulting team and judged on whether it described something they recognised and could act on.

2.7. Phase 6: design of interventions and deployment plan

Deployment was planned along three lines, each tied to a rule that survives evaluation: a configurable validation package for the platform, a compliance checklist to be confirmed at the blueprint stage of new implementations, and an analytical prototype that renders rules in natural language. What each of them contains, and how far each has been realised, is reported in §3.6.

This phase produced designs and one working prototype. Nothing described here has been switched on in the client production environment, and the section title says so rather than leaving the reader to infer it. Because the artefacts are designs, the phase also had to specify how their effect would be established, and that specification is part of the output rather than deferred to future work. Effectiveness would be measured as a before-and-after comparison on the same encoding used in this study: the eleven indicators recomputed monthly, with the twenty-four months analysed here as the baseline, and the primary endpoint the prevalence of the two syndrome antecedents, instant closure on corrective work and labour overlap on preventive work. Because the validation package is delivered through per-client system properties, controls can be enabled for one site while a comparable site continues unchanged, which yields a controlled rather than merely sequential comparison. Three quantities have to be watched for displacement rather than improvement: work orders migrating to work types where the control does not apply, technician hours moving from booked intervals into unbooked ones, and closure timestamps clustering just beyond whatever threshold the control enforces. A control that only relocates a behaviour would show a falling anomaly rate on its own indicator and a rising one on its neighbour, which is why the whole indicator set is tracked rather than the targeted item alone. The rule engine would be assessed separately and qualitatively, by putting a fixed set of rendered rules to maintenance managers and delivery consultants and recording whether each reader can state the implied action without further explanation.

3. Results

3.1. Phase 1: business understanding

The contractual asymmetry described in Section 1 is confirmed by the account record itself. Service hours are consumed only when the client opens a ticket, and the consultant holds no standing mandate to inspect the system. Framing the audit as a data mining problem changes that position, because a rule with measurable support, lift and a significance test behind it is an argument that can be put to a client without waiting for a fault to appear. The acceptance criteria agreed in this phase are what the remainder of this section is measured against.

3.2. Phase 2: data understanding, data quality profile

Four different counts appear in this paper, and separating them at the outset avoids the impression of inconsistency. Table 2 maps each to its scope. The profiling tables below describe the raw extract; the rules described later all rest on the mining population of 12,368.

Table 2. Scope of each population figure used in this paper.

Figure	Count	Scope
Raw extract	12,339	Work order headers as returned by the query, before any modelling filter (Table 3)
Mining population	12,368	Assembled per work type after cancelled records are removed: 5,790 CM + 6,578 PM. Carries every rule reported below
Completed subset	8,090	Orders that reached COMP, used for the documentation ratios of Table 4
Status transitions	—	Figure 3 counts transitions, not orders, so one order contributes to both the INPRG and the COMP bars

Column integrity across the seven tables splits the module cleanly in two, as Table 3 shows. Everything administrative is complete. Identifiers, work type, site, status and change metadata all reach 100%, and asset and location mapping reach 99.55%, which is enough to guarantee that joins between header and children will not silently drop records. Everything describing physical execution is not. Actual start is present in 71.10% of records and actual finish in 66.29%, so a third of the population never receives a closing timestamp. Target dates appear in barely 53%. Failure code sits at 46.34%, unremarkable on its own because classification is expected only on corrective work, although it does limit the scenarios in which the attribute can be used.

Table 3. Column integrity of selected attributes in the Maximo work order module.

Table	Attribute	Records	Populated	Integrity (%)
WORKORDER	WONUM / WORKTYPE / STATUS	12,339	12,339	100.00
WORKORDER	ASSETNUM / LOCATION	12,339	12,218	99.55
WORKORDER	DESCRIPTION	12,339	12,187	99.30
WORKORDER	PERSONGROUP	12,339	10,377	84.55
WORKORDER	ACTSTART	12,339	8,726	71.10
WORKORDER	ACTFINISH	12,339	8,136	66.29
WORKORDER	TARGSTARTDATE	12,339	6,524	53.16
WORKORDER	FAILURECODE	12,339	5,687	46.34
WOSTATUS	STATUS / CHANGEDATE	50,220	50,220	100.00
LABTRANS	STARTDATETIME	16,137	16,086	99.68
MULTIASSETLOCCI	ASSETNUM / LOCATION	13,851	13,837	99.90
WOACTIVITY	WOSEQUENCE	51,745	45,701	88.32
WOACTIVITY	LONGDESCRIPTION	51,745	1,020	1.97
WORKLOG	DESCRIPTION	2,502	752	30.06

Qualitative documentation is scarcer still. Attachments cover about 4.2% of work orders and chronological logs about 20.4%. Table 4 breaks this down by work type, and the asymmetry it reveals is operational rather than statistical. Corrective work carries 2,390 log entries against 4,304 completed orders, a ratio of 0.56; put plainly, roughly half of all breakdown repairs are closed without any written account of what was found. Preventive work carries 33 entries against 3,786 orders, a ratio of 0.01, which says that routine maintenance has been reduced to a status change. Attachments repeat the pattern from the evidentiary side: 97.97% of corrective and 94.29% of preventive orders close with no photograph, inspection sheet or vendor report on file.

Two exploratory checks explain why these gaps had to shape the modelling. Figure 3 asks whether status changes are posted while the work is happening. They are not. Backdated entry exceeds 95% on corrective work and 83% on preventive work, so field execution consistently runs ahead of the system. Instant closure, meaning an INPRG to COMP transition inside two minutes, accounts for 63.4% of completed corrective transactions against 0.05% of

Table 4. Documentation usage by work type for completed work orders.

Work type	Total WO	Total work logs	Log ratio	WO with attachment
CM	4,304	2,390	0.56	2.03%
PM	3,786	33	0.01	5.71%

preventive ones. One consequence follows immediately: status timestamps cannot stand in for work duration, and any duration measure has to be built from the difference between actual finish and actual start.

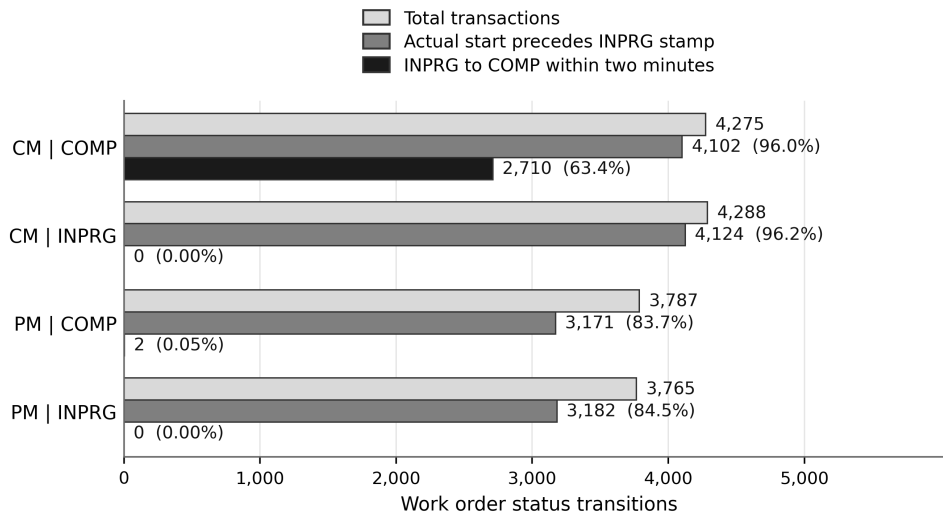


Figure 3. Real-time behaviour of work order status transitions by work type. Percentages are shares of the total transactions on the same row; the two anomaly measures are subsets of that total rather than additive components of it.

Figure 4 turns to labour reporting. Close to 70% of preventive and 27% of corrective orders contain technician intervals that intersect intervals booked on other orders, which means roughly half of all recorded hours are counted twice or otherwise distorted. Raw duration is therefore useless as a continuous predictor unless it is first time-weighted, and that finding is what justified modelling overlap as a flag instead of an hour count. Similar conclusions have been reached elsewhere: inaccurate maintenance records usually originate in reporting habits rather than in system design [15].

3.3. Phase 3: data preparation, item construction

Table 5 lists the eleven items that came out of the recoding. Two describe context (C and D) and nine describe procedural failure. The design decision worth stating is that weaknesses in the data were promoted to features rather than repaired or deleted: the absence of a log became `ERR_NO_WORKLOG`, the absence of an attachment became `ERR_NO_ATTACHMENT`. Nothing was imputed. This keeps the analysis honest, because an imputed work log would have hidden precisely the behaviour under investigation.

Table 6 records how often each item occurs in the two mining matrices. This is the base rate that every later lift value is measured against, so it belongs in the paper rather than in an appendix. The two work types read almost as mirror images. Corrective work is dominated by high priority at 76.99%, backdating at 72.57% and instant closure at 48.03%, while late completion never occurs at all because corrective orders in this installation carry no

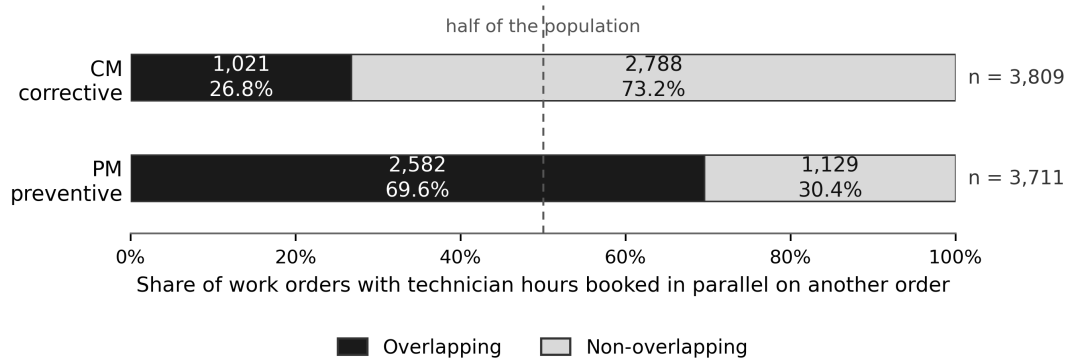


Figure 4. Overlapping versus non-overlapping actual labour records by work type. An order is counted as overlapping when at least one technician interval booked against it intersects an interval booked by the same technician on another order.

Table 5. Binary encoding of contextual and anomaly items.

Code	Derived item	Operational meaning
C	IS_HIGH_PRIORITY	Emergency or high priority (scale 1 or 2)
D	IS_MULTI_ASSET	Covers more than one asset or location
E	ERR_BACKDATED	ACTSTART precedes the INPRG status timestamp
F	ERR_INSTANT_CLOSED	INPRG to COMP transition within two minutes
G	ERR_LABOR_OVERLAP	Technician hours intersect across two work orders
H	ERR_NO_WORKLOG	No repair narrative recorded in WORKLOG
I	ERR_NO_ATTACHMENT	No document, checklist or photograph in DOCLINKS
J	ERR_NO_FAILURE_REPORT	Failure code present but failure report empty
K	ERR_STATUS_BYPASS	Status jumps to COMP without APPR or INPRG
L	ERR_LATE_COMPLETION	ACTFINISH exceeds TARGCOMPDATE
M	ERR_STAGNANT_WO	Over 30 days between APPR and INPRG

meaningful target date. Preventive work inverts the picture: instant closure is effectively absent at 0.03%, but late completion reaches 55.03% and missing work logs 99.48%. Several items are too rare in one or both matrices to enter any frequent itemset at the chosen support. They are kept in the table because their scarcity is itself a result, since it says that multi-asset routing, status bypass and stagnation are not what is wrong here.

3.4. Phase 4: modelling, Apriori compared with FP-Growth

Table 7 sets the two algorithms side by side. Rule counts match exactly in all four scenarios, and so do the support, confidence and lift values, which is the strongest available evidence that the rule base belongs to the data and not to the search strategy. Runtime is another matter. Apriori finished every scenario in under a hundredth of a second and beat FP-Growth by factors between 23 and 98. The absolute magnitudes deserve to be read before the ratios: the slowest single measurement in the study is 0.189 seconds and the fastest is 0.0011 seconds. At that scale a ratio is easy to quote and easy to over-read, which is why means over thirty runs and their standard deviations are given rather than single timings.

That ordering runs against the usual expectation and deserves an explanation rather than a footnote. FP-Growth normally wins, and the benchmarks establishing this are drawn from retail-scale baskets with thousands of distinct items [36, 37]. Its advantage comes from skipping candidate generation, which only pays for itself when the candidate space is large. Here the item space holds eleven binary attributes. Apriori's bit-vector operations are trivially cheap at that scale, whereas FP-Growth still pays the fixed price of building and walking a tree in memory.

Table 6. Prevalence of each item in the two mining matrices.

Code	Item	CM ($n = 5,790$)	PM ($n = 6,578$)
C	IS_HIGH_PRIORITY	76.99%	0.06%
D	IS_MULTI_ASSET	0.00%	0.53%
E	ERR_BACKDATED	72.57%	49.68%
F	ERR_INSTANT_CLOSED	48.03%	0.03%
G	ERR_LABOR_OVERLAP	18.39%	39.86%
H	ERR_NO_WORKLOG	57.82%	99.48%
I	ERR_NO_ATTACHMENT	98.00%	94.36%
J	ERR_NO_FAILURE_REPORT	0.09%	8.67%
K	ERR_STATUS_BYPASS	0.47%	0.50%
L	ERR_LATE_COMPLETION	0.00%	55.03%
M	ERR_STAGNANT_WO	0.97%	0.11%

Table 7. Rule count and execution time of Apriori and FP-Growth by scenario. Times are means over thirty runs; standard deviations in parentheses.

Scenario	Rules	Apriori (s)	FP-Growth (s)	Speed-up
CM, with attachment and log	90	0.0018 (0.0001)	0.1766 (0.0041)	98×
PM, with attachment and log	94	0.0017 (0.0001)	0.1367 (0.0029)	82×
CM, without attachment and log	20	0.0013 (0.0001)	0.0539 (0.0009)	41×
PM, without attachment and log	7	0.0011 (0.0001)	0.0266 (0.0004)	23×

The claim is bounded accordingly. What this result supports is that on a binary matrix of roughly ten items and roughly ten thousand transactions, the tree-construction overhead of FP-Growth is not recovered, so a practitioner auditing compliance data of this shape should not assume the usual ranking holds. What it does not support is any general statement about the two algorithms, and the ratios should not be extrapolated beyond this order of magnitude: at sub-second scale the measured factor is sensitive to library version, memory behaviour and machine load, and it is the direction of the difference rather than its size that is stable. The reason FP-Growth is nevertheless the common default is that the benchmarks defining practice come from retail baskets, where the candidate space is large enough for the overhead to be repaid many times over. The practical guidance is to choose by matrix dimensionality rather than by reputation, a consideration that applied studies seldom make explicit [16, 38].

3.5. Phase 5: evaluation, rules under the ablation design

Before the rules themselves, Table 8 settles which items had to be ablated. Removing H alone and removing I alone were run as separate scenarios alongside the joint removal. The two work types behave differently, and that difference is the justification. On corrective work the attachment item is the sponge: removing I alone cuts the rule set from 90 to 27, while removing H alone leaves 61. On preventive work the relationship inverts, because there it is the work log that is near-constant at 99.48%; removing H alone yields the clean process rule $\{E, G\} \Rightarrow \{L\}$, whereas removing I alone leaves the top rule still carrying H in its consequent. Neither single ablation therefore works for both matrices. Removing both is what produces one encoding that behaves consistently across work types, and it also produces the smallest rule sets, 20 and 7, in which every surviving rule is a process rule rather than a documentation tautology.

Across the four scenarios the two algorithms returned 211 rules that clear support and confidence, of which 155 also carry positive lift. Many are redundant supersets that restate a shorter rule at slightly lower support. A rule was therefore discarded whenever another rule with the same consequent and a strictly smaller antecedent already achieved equal or higher lift, which leaves 35, 30, 18 and 7 rules in the four scenarios. Table 9 reports the strongest

Table 8. Effect of alternative ablation choices at support 0.10 and confidence 0.70.

Work type	Ablation	Items	Rules	Lift > 1	Highest-lift rule
CM	none (Step 1)	9	90	63	$\{C, E, G\} \Rightarrow \{F\}$, lift 1.502
CM	drop H only	8	61	39	$\{C, E, G\} \Rightarrow \{F\}$, lift 1.502
CM	drop I only	8	27	23	$\{C, E, G\} \Rightarrow \{F\}$, lift 1.502
CM	drop H and I (Step 2)	7	20	18	$\{C, E, G\} \Rightarrow \{F\}$, lift 1.502
PM	none (Step 1)	11	94	67	$\{E, G\} \Rightarrow \{H, L\}$, lift 1.767
PM	drop H only	10	26	19	$\{E, G\} \Rightarrow \{L\}$, lift 1.759
PM	drop I only	10	28	26	$\{E, G\} \Rightarrow \{H, L\}$, lift 1.767
PM	drop H and I (Step 2)	9	7	7	$\{E, G\} \Rightarrow \{L\}$, lift 1.759

of those by lift, together with the Fisher test for each. Step 1 keeps the documentation items and demonstrates exactly the problem that motivated the ablation. On corrective work the leading rule states that a missing log implies a missing attachment, at support 0.578 and confidence 0.999. Since the attachment item is set in roughly 95% of transactions it behaves as a near-constant, absorbs the remaining structure and leaves lift at 1.020, which masks the process variables entirely. Preventive work is saturated differently: where technicians book overlapping hours and post backdated entries, the log is omitted in 39.8% of transactions at confidence 0.998.

Table 9. Strongest non-redundant rules per scenario, with one-sided Fisher exact tests. q is the Benjamini-Hochberg adjusted p -value over all $m = 155$ tested rules; the last column marks whether the rule meets the practical-significance criterion of §2.2.

Scenario	Rule	Support	Conf.	Lift	Odds ratio	p	q	Act.
Step 1, CM	$\{H\} \Rightarrow \{I\}$	0.578	0.999	1.020	54.10	< 0.001	< 0.001	no
Step 1, PM	$\{G, E\} \Rightarrow \{H\}$	0.315	0.999	1.004	4.78	0.002	0.002	no
Step 1, PM	$\{G\} \Rightarrow \{H\}$	0.398	0.998	1.003	3.11	0.005	0.006	no
Step 2, CM	$\{C, E, G\} \Rightarrow \{F\}$	0.102	0.721	1.502	3.28	< 0.001	< 0.001	yes
Step 2, CM	$\{C, G\} \Rightarrow \{F\}$	0.102	0.707	1.473	3.05	< 0.001	< 0.001	yes
Step 2, CM	$\{C, F\} \Rightarrow \{E\}$	0.374	0.981	1.352	39.09	< 0.001	< 0.001	yes
Step 2, CM	$\{F\} \Rightarrow \{E\}$	0.470	0.980	1.350	49.50	< 0.001	< 0.001	yes
Step 2, PM	$\{E, G\} \Rightarrow \{L\}$	0.305	0.968	1.759	54.46	< 0.001	< 0.001	yes
Step 2, PM	$\{E\} \Rightarrow \{L\}$	0.471	0.948	1.723	97.33	< 0.001	< 0.001	yes
Step 2, PM	$\{L\} \Rightarrow \{E\}$	0.471	0.856	1.723	97.33	< 0.001	< 0.001	yes
Step 2, PM	$\{G\} \Rightarrow \{L\}$	0.349	0.876	1.593	14.13	< 0.001	< 0.001	yes

Removing H and I brings two separate syndromes into view, summarised in Figure 5. Corrective maintenance breaks on administrative timing. Instant closure implies backdated entry at support 0.470, confidence 0.980, lift 1.350 and an odds ratio of 49.50; nearly half of all corrective orders are opened and closed inside a two-minute window, and almost all of them carry an execution stamp earlier than the status change. The pattern is consistent with sequence rather than falsification: the repair is carried out in the field and the record is created and closed afterwards in one administrative gesture. That reading was put to the consulting team and matches what they describe, but it is an interpretation of a co-occurrence, not something the data establishes, and at least three other mechanisms would produce the same signature. Mobile access may be unavailable at the work face, so entry is necessarily deferred. Status changes may be posted in batches at shift handover by a coordinator rather than by the technician. Or the platform may itself default the execution stamp when a status is set late. Distinguishing these requires interface logs and shift rosters that the present extract does not contain, and §4 treats the question as open.

What complicates the picture is what happens when priority is added. Confidence rises to 0.981 and lift to 1.352 over 37.4% of the population, so the association is stronger, not weaker, on urgent work. The usual expectation is

that priority escalation improves traceability; here the marginal association points the other way. The multivariable model in Table 11 qualifies this, and the qualification matters enough to state immediately: once instant closure and labour overlap are held constant, high priority is associated with slightly *lower* odds of backdating, at an adjusted odds ratio of 0.67. Priority therefore raises backdating through instant closure rather than independently of it, which is a different statement from the one the marginal rule alone would support. The highest lift in this scenario belongs to a three-item antecedent, where high priority, backdating and labour overlap together imply instant closure at lift 1.502. Its support is lower at 0.102, but it describes the same practice observed at its most concentrated.

(a) Corrective maintenance: administrative lag

(b) Preventive maintenance: schedule slippage

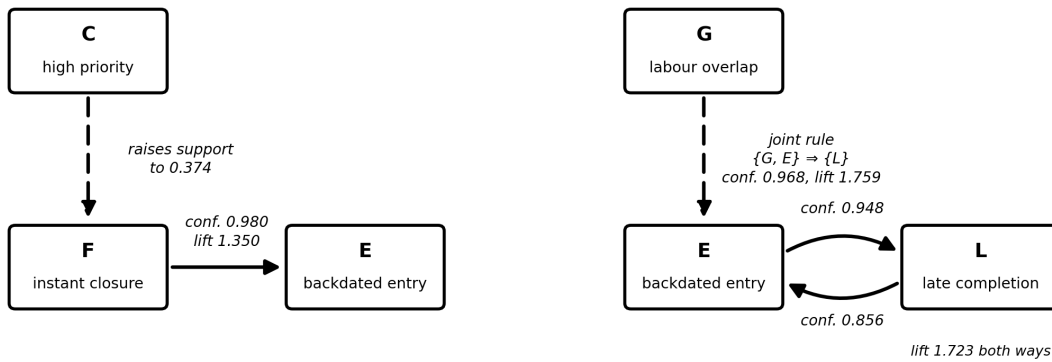


Figure 5. Rule map of the two behavioural syndromes isolated in Step 2.

Preventive maintenance breaks on time instead. Backdating and late completion form a two-way association at support 0.471, confidence 0.948 in one direction and 0.856 in the other, lift 1.723 both ways and an odds ratio of 97.33, the strongest effect anywhere in the study. Support is necessarily identical in both directions, since it counts the same co-occurrence. The strongest single rule joins labour overlap to backdating and predicts schedule slippage across 30.5% of the preventive population at confidence 0.968 and lift 1.759. One reading consistent with all three items is that scheduled rounds accumulate and are then cleared together near the end of a reporting period, which would leave overlapping bookings, retrospective dates and missed targets as joint traces of a single working practice. The consulting team recognises this pattern, and it is offered here as the most plausible of several hypotheses rather than as an established mechanism. Association rules identify co-occurrence; they carry no information about intent, and nothing in this data speaks to why a round was cleared late. Competing explanations that would generate the same three items include preventive schedules generated on a calendar that does not match crew availability, so slippage is structural rather than behavioural; a system configuration in which the target date is populated at generation and never revised; delayed synchronisation between a mobile client and the server; and supervisor-level batch entry on behalf of several technicians. Workaround research in information systems treats exactly this kind of adaptation as a rational response to a mismatch between prescribed procedure and working conditions rather than as non-compliance in the disciplinary sense [41], which is the frame adopted here. The two-way structure is informative on its own account: because the association holds at similar strength in both directions, a simple one-way causal model is inadequate, and the two items are better read as joint symptoms of an upstream process than as cause and effect.

Once its logic was corrected, the failure report item stopped producing trivial rules at confidence 1.000 and the process variables surfaced on their own. That single iteration back to data preparation did more for rule quality than any parameter change, which supports what the adaptation literature has argued about CRISP-DM: the binding constraint is usually feature semantics, not algorithm choice [23, 28].

Every rule in Table 9 is significant under the Fisher test. The correction for multiple testing does not change that. Of the $m = 155$ rules tested, 144 remain significant at a false discovery rate of $q < 0.05$ and 137 at $q < 0.01$; under the far stricter Bonferroni bound, 129 still clear 0.05. The eleven rules in Table 9 all survive, and the four syndrome rules survive by an enormous margin: the adjusted q for $\{E\} \Rightarrow \{L\}$ has $\log_{10} q = -1048.8$ and for $\{F\} \Rightarrow \{E\}$, $\log_{10} q = -451.3$. Correction was never going to overturn effects of this size in a sample of this width, but reporting it removes the objection rather than leaving it standing.

Three rules deserve a comment, and they are the reason the practical-significance criterion of §2.2 was fixed in advance. All three are Step 1 rules involving the saturated documentation items. Their lift is barely above unity, at 1.003 to 1.020, yet the test still rejects independence, and the adjusted q still clears 0.01, because the sample is large. This is precisely the divergence a saturated variable produces: the association is real but almost useless, since an item present in 99.48% of transactions can hardly fail to co-occur with anything. Under the four practical conditions all three fail on lift and two of them also fail on odds ratio, so none is marked actionable in Table 9. Had significance alone governed the selection, the paper would have carried them forward as findings. The Step 2 rules behave in the opposite way, combining negligible q -values with odds ratios between 3.28 and 97.33, and it is those that carry operational meaning. The multiple-testing correction thus strengthens the ablation argument rather than weakening it: it shows that the rules the ablation removes were never removed for being statistically fragile, but for being operationally empty.

Two questions remain that association rules cannot answer about themselves. The first is whether a simpler statistic would have found the same thing. Table 10 reports the pairwise ϕ coefficients among the items that enter the syndrome rules. On preventive work a correlation analysis performs respectably: backdating and late completion correlate at 0.79, labour overlap and late completion at 0.53, so an analyst inspecting the matrix would have noticed both edges. What the matrix cannot express is the joint rule. The conjunction of backdating and labour overlap predicts late completion at lift 1.759, higher than either item alone, and a correlation matrix has no representation for a two-item antecedent. On corrective work the gap is starker still. High priority correlates with instant closure at only 0.05, which is close enough to zero that any correlation-driven screen would discard the item outright, yet high priority appears in the highest-lift rule in the entire corrective matrix, $\{C, E, G\} \Rightarrow \{F\}$ at lift 1.502. Its contribution is conditional, visible only in combination with backdating and labour overlap, and it is exactly the kind of structure that motivates rule mining over pairwise screening.

Table 10. Pairwise ϕ coefficients among the items entering the syndrome rules. The corresponding rule lifts are given for comparison.

Work type	Item pair	ϕ	Lift of the rule containing the pair
CM	E backdating, F instant closure	0.55	1.350 for $\{F\} \Rightarrow \{E\}$
CM	E backdating, G labour overlap	0.26	—
CM	F instant closure, G labour overlap	0.19	—
CM	C high priority, F instant closure	0.05	1.502 for $\{C, E, G\} \Rightarrow \{F\}$
PM	E backdating, L late completion	0.79	1.723 for $\{E\} \Rightarrow \{L\}$
PM	G labour overlap, L late completion	0.53	1.593 for $\{G\} \Rightarrow \{L\}$
PM	E backdating, G labour overlap	0.48	1.759 for $\{E, G\} \Rightarrow \{L\}$

The second question is whether each item earns its place once the others are held constant. Table 11 reports the three logistic models. Both syndromes survive adjustment with their effect sizes essentially intact. On preventive work, backdating retains an adjusted odds ratio of 69.7 and labour overlap 7.1 for late completion, with the model accounting for 60% of the null deviance; the two items therefore contribute independently rather than one standing in for the other, which is what the joint rule $\{E, G\} \Rightarrow \{L\}$ asserts. On corrective work, instant closure retains an adjusted odds ratio of 49.3 for backdating, almost unchanged from the unadjusted Fisher value of 49.50, so that association is not carried by a third item.

The models also correct one reading that the marginal rules would have supported. High priority enters the backdating model with an adjusted odds ratio of 0.67, below unity, even though adding it to the rule raised confidence from 0.980 to 0.981. The explanation is visible in the third model: high priority raises the odds of instant closure by a factor of 1.45, and instant closure in turn raises the odds of backdating by a factor of 47.7. Priority acts on backdating through instant closure, and once that path is blocked by conditioning, its direct association reverses sign. Nothing in the syndrome depends on this, but it disciplines the language: the correct statement is that urgent work is closed instantly more often, and instantly closed work is backdated, not that urgency independently degrades record-keeping. It also illustrates the general point that a marginal rule and an adjusted coefficient answer different questions, which is why both are reported.

Table 11. Multivariable logistic models. Adjusted odds ratios with 95% Wald intervals. Items with prevalence outside [0.01, 0.99] in the relevant matrix were excluded before fitting and are listed beneath each model.

Model	Predictor	Adj. OR	95% CI	<i>p</i>
PM: outcome L, late completion <i>n</i> = 6,578, McFadden R^2 = 0.599	E backdating	69.74	57.60–84.44	< 0.001
	G labour overlap	7.09	5.87–8.57	< 0.001
	I no attachment	0.16	0.07–0.39	< 0.001
	J no failure report	2.30	1.77–2.99	< 0.001
CM: outcome E, backdating <i>n</i> = 5,790, McFadden R^2 = 0.449	F instant closure	49.33	37.27–65.28	< 0.001
	G labour overlap	9.08	6.05–13.63	< 0.001
	C high priority	0.67	0.56–0.81	< 0.001
	H no worklog	0.11	0.09–0.14	< 0.001
	I no attachment	0.08	0.01–0.58	0.013
CM: outcome F, instant closure <i>n</i> = 5,790, McFadden R^2 = 0.265	E backdating	47.73	36.12–63.07	< 0.001
	C high priority	1.45	1.26–1.68	< 0.001
	G labour overlap	1.37	1.17–1.60	< 0.001
	I no attachment	1.69	1.16–2.47	0.007
	H no worklog	1.00	0.88–1.14	0.96

Excluded, PM models: C (0.06%), D (0.53%), F (0.03%), H (99.48%), K (0.50%), M (0.11%).

Excluded, CM models: D (0.00%), J (0.09%), K (0.47%), L (0.00%), M (0.97%).

One limit of these models has to be stated with them. They adjust each binary item for the other binary items, which is the only adjustment the extract supports. They cannot adjust for site, asset class, technician, crew or shift, because those attributes were not carried into the transaction matrix. An association between labour overlap and late completion could still arise because particular teams or asset classes generate both, and nothing here excludes that. §4 treats this as the principal open confound.

Table 12 reports the sensitivity grid. Two things stand out. Neither syndrome-defining rule disappears anywhere in the grid: the corrective rule and the preventive rule survive all sixteen combinations, so the headline findings are not an artefact of the chosen cut-offs. The corrective matrix is more sensitive to support than the preventive one, dropping from 18 rules to 5 as support rises from 0.10 to 0.20 at confidence 0.70, whereas the preventive matrix holds at seven rules throughout. That asymmetry follows from Table 6: the preventive items that matter all have prevalence near or above 0.40, so raising the support floor removes nothing, while the corrective matrix contains several mid-frequency combinations that fall away.

Counting surviving rules is a weak form of stability evidence, so rule composition was compared cell by cell as well. Across the sixteen corrective cells, 23 distinct rules appear at least once and exactly two appear in all sixteen: $\{F\} \Rightarrow \{E\}$ and $\{C, F\} \Rightarrow \{E\}$. Across the sixteen preventive cells, 12 distinct rules appear at least once and again exactly two appear in every cell: $\{E, G\} \Rightarrow \{L\}$ and $\{E\} \Rightarrow \{L\}$. The core sets are therefore not merely non-empty, they are precisely the syndrome-defining rules and their immediate variants, while everything else in the grid is threshold-dependent. The pattern of loss is also orderly rather than arbitrary. Raising confidence removes

rules from the low-confidence end in sequence, and raising support removes the three-item antecedents first, since those sit near support 0.102 in the corrective matrix and fall out between 0.10 and 0.15. Nothing enters the rule set as thresholds move; rules only leave. That monotonicity is what one expects if the underlying structure is stable and the thresholds are simply cutting it at different depths.

Table 12. Rules with positive lift under varying thresholds. Y marks survival of the syndrome rule, $\{F\} \Rightarrow \{E\}$ for CM and $\{E, G\} \Rightarrow \{L\}$ for PM.

Support	CM, confidence				PM, confidence			
	0.60	0.70	0.80	0.90	0.60	0.70	0.80	0.90
0.05	23 Y	18 Y	9 Y	6 Y	12 Y	7 Y	5 Y	2 Y
0.10	23 Y	18 Y	9 Y	6 Y	11 Y	7 Y	5 Y	2 Y
0.15	8 Y	6 Y	3 Y	3 Y	11 Y	7 Y	5 Y	2 Y
0.20	7 Y	5 Y	2 Y	2 Y	11 Y	7 Y	5 Y	2 Y

3.6. Phase 6: designed interventions and deployment plan

One point should be stated plainly before the artefacts are described. None of them is yet running in the client production environment. The deployment phase of this study produced designs and a working prototype, not an installed change, and the sequencing is deliberate: the mining result is what tells the delivery team which validations are worth building, so the configuration package is assembled first and applied at the next implementation window rather than retro-fitted piecemeal. What the study delivers at this stage is an evidence base for moving the service model from reactive to proactive, together with a specification of the controls that the move would consist of and of the measurements that would establish whether it worked. The shift itself is a goal, not a result. Claims about reduced anomaly rates, recovered consulting hours or improved record quality are hypotheses generated by this study, and they remain untested until the package described below is enabled and the endpoints set out in §2.7 are collected.

Three artefacts carry the findings towards practice. The first is a validation package delivered through the platform's custom system properties, so individual controls can be switched on per client without touching source code. The rule base dictates what those controls should be: lock the completion action until a failure report exists on corrective work, reject status transitions shorter than two minutes, and refuse labour records whose intervals collide with an existing booking. Each control follows directly from a rule in Table 9 and is scoped by a measured share of the population, so the delivery team can weigh the cost of building it against the volume of transactions it would touch. The second is a checklist confirmed during the blueprint phase of new implementations, which pushes the same knowledge upstream so that a defect discovered here in year three is prevented elsewhere in month one. The third, shown in Figure 6, is a prototype that accepts an extract, exposes support and confidence as editable parameters and returns each rule as a sentence with a recommendation attached. Rendering analytical output in language managers already use is a recognised condition for adoption in operational reporting [45].

The translation from rule to sentence follows one aim rather than an elaborate scheme: a maintenance manager should be able to read a finding without knowing what support, confidence or lift mean. Concretely, the template names each item in operational language and keeps the item code in parentheses, states the association as a likelihood that the same work order also carries the consequent, expresses support as the share of analysed transactions in which the pattern occurs, and expresses lift as a comparison with chance, so that a lift of 1.294 is read as an association 1.29 times stronger than chance alone would produce. The odds ratio and the p -value are not carried into the rendered sentence, which keeps it short enough to be read in full. Each sentence closes with the recommended control rather than with the finding. Figure 6 shows one rendered rule in its entirety.

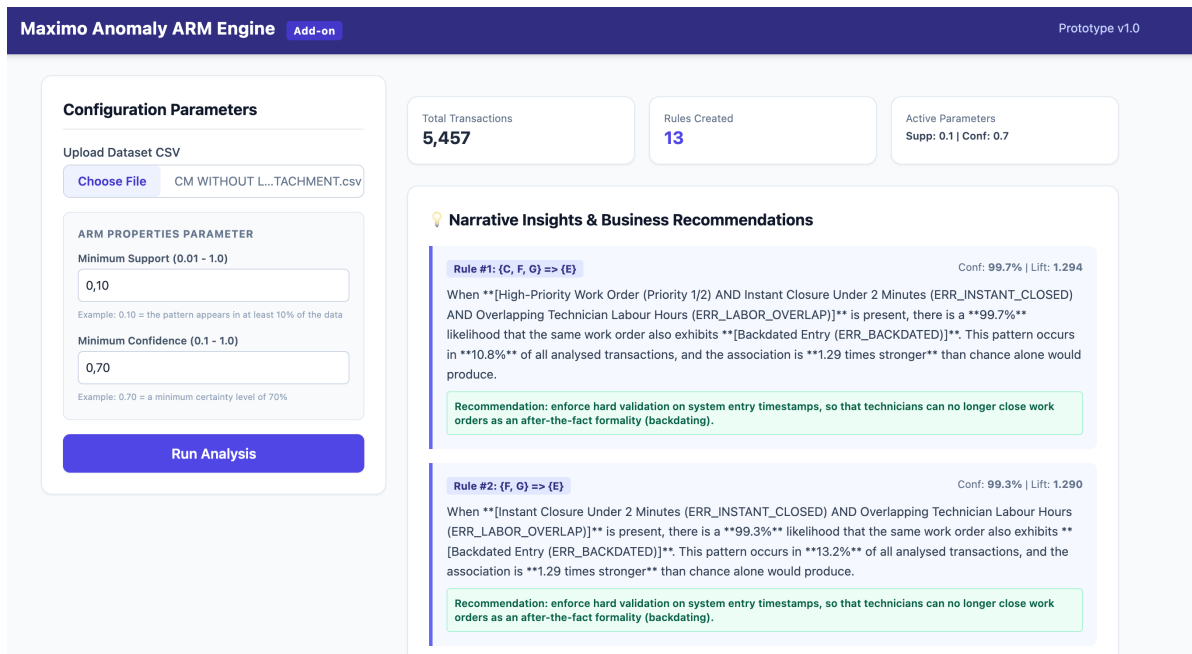


Figure 6. Prototype rule engine rendering association rules as natural-language recommendations.

4. Discussion

Set against the objectives, the results answer a question that descriptive statistics could not. Profiling showed where defects live. It could not say why they survive. The rule base does: defects are not scattered across users and assets at random, they are the predictable residue of two working practices. Corrective work suffers administrative lag and instant closure. Preventive work suffers schedule slippage and inflated labour. Different conditions need different treatments, and a blanket instruction to improve data entry would have missed both.

The contribution extends earlier work in two directions. Lu and colleagues traced inaccurate reliability data to its causes but stopped at diagnosis [15]; the rules here become configuration. Oyoo argued that validation belongs in the delivery method [13]; this study supplies the evidence for choosing which validations are worth enforcing and estimates how much of the population each one would touch. Text-based approaches [14, 17] presume narrative fields that this dataset does not have, with log description integrity at 30.06% and long descriptions below 2%. Under that level of sparsity, structural binary items are the appropriate instrument. The contrast is worth stating in its own right, because it marks a boundary condition for text mining in maintenance data.

Where the findings can be compared with published work, they agree. Reviews of association rules in maintenance report that the technique earns its place when labels are absent and behaviour has to be inferred [16], which is the situation here. Process research uses rule learning in the same way, to explain deviations and grade their severity [18, 19]. The one divergence, the runtime ordering of the two algorithms, is a matter of dimensionality rather than disagreement. On the process side the outcome is consistent with adaptation studies in finance [23] and manufacturing [24]: CRISP-DM needed no structural change, only honest use of its feedback loops.

4.1. Why the syndromes might exist

Naming a syndrome is not the same as explaining it, and the analysis so far supports the first but not the second. Three families of explanation would each produce the observed items, and they carry different remedies, which is why choosing between them matters operationally rather than only intellectually.

The first is procedural mismatch. If mobile access is unavailable at the work face, or if the prescribed sequence requires a status change at a moment when the technician's hands are occupied, then deferred entry is not a deviation but the only way to complete the work at all. Workaround theory treats adaptations of this kind as rational responses to a gap between prescribed procedure and working conditions, and predicts that they persist precisely because they are locally efficient [41]. If this is the mechanism, a hard validation that blocks late entry will not improve the record; it will push the work outside the system entirely.

The second is measurement pressure. Preventive completion rates are reported against target dates, so clearing a backlog near a period boundary improves the reported figure whether or not the rounds were performed on schedule. This explanation predicts temporal clustering of preventive closures near period ends, which is testable but not from the present extract, since the transaction matrix carries no dates.

The third is configuration rather than behaviour. If preventive schedules are generated on a fixed calendar that ignores crew availability, slippage is structural and no amount of validation will remove it. This is the explanation the consulting team considered least likely, but it is also the cheapest to check, since it requires only a comparison of generated schedules against roster capacity.

These are hypotheses, offered as the reading most consistent with the evidence rather than as findings. What the data does establish is narrower and still useful: the failures are not randomly distributed across users and assets, they cluster into two reproducible configurations, and each configuration points at a different part of the operation.

4.2. Limitations

Six limitations bound the claims, and the first is the one that bounds all the others.

The study covers one installation in one industry. Specific support, confidence and lift values should be read as characteristic of this operating context, not as constants, and the paper makes no claim that they transfer. What is claimed to transfer is the framework: the encoding depends only on tables present in every Maximo deployment, the ablation procedure is defined by prevalence rather than by these particular items, and the practical-significance criterion is a template into which another organisation would substitute its own thresholds. Transfer is conditional even so. The encoding requires that status history be retained, that labour be booked by interval rather than by total, and that the installation be past the unstable period following go-live. An organisation that books labour as a daily total cannot compute the overlap item at all, and one that purges status history cannot compute backdating. Replication should therefore begin by verifying those three conditions rather than by assuming portability.

Second, the deployment artefacts are designs and one prototype. No post-intervention measurement exists, and the effectiveness of the controls is an argument from evidence rather than a demonstrated outcome. The measurement plan in §2.7 states what would settle the question.

Third, the analysis measures whether documentation exists, not whether it is any good. A work log holding one meaningless character scores as compliant. This matters more than it may appear: the study is framed partly as a data quality investigation, and completeness is only one of the dimensions that framing implies. Accuracy, consistency and informativeness are all outside what a presence indicator can see, so the correct reading of the documentation findings is that they establish a lower bound on the problem. The true rate of unusable documentation can only be higher than the rate of absent documentation.

Fourth, rules capture association. The behavioural readings above are interpretations validated by domain experts, not experimentally established causes, although the Fisher tests and the multivariable models do exclude chance and mutual confounding among the encoded items as explanations for the co-occurrences themselves.

Fifth, and following directly from the fourth, contextual confounders remain unaddressed. Site, asset class, technician, crew, shift, work order age and organisational unit were not carried into the transaction matrix, so none of them could enter the logistic models. The association between labour overlap and late completion could still be generated by particular crews or asset classes that produce both characteristics, and the present design cannot exclude that. This is the single most consequential open question in the study, and it is addressable without new instrumentation, since the attributes exist in the source tables and were simply not extracted. A stratified re-analysis is the natural next step.

Sixth, the sixteen-cell grid establishes stability across a defensible range of thresholds, not across all possible parameterisations, and lift was held at a fixed floor throughout rather than varied as a third axis.

Each limitation points somewhere concrete: stratified re-analysis by asset class and crew, temporal validation in which rules are derived on an earlier window and tested on a later one, replication across clients, language and image analysis over logs and attachments, and the controlled before-and-after measurement once the validation package goes live. Process mining belongs on that list as well. It is not usable on this extract, for the reasons given in § 1.1, but the relationship is sequential rather than competitive: the controls designed here are aimed precisely at the timestamp integrity that process mining requires, so an installation that enforces them for a year would generate an event log on which conformance checking becomes meaningful. Rule mining diagnoses why the log cannot yet be trusted; process mining is what the log becomes useful for once it can.

5. Conclusion

A complete CRISP-DM cycle turned a governance problem in an enterprise asset management system into an analytical result that a consultancy can act on, and association rule mining over a compact binary matrix proved sufficient to expose behaviour that neither routine reporting nor conventional auditing had made visible. Profiling 12,368 IBM Maximo work orders showed a sound administrative backbone alongside severe sparsity in execution and documentation fields, with work logs on roughly a fifth and attachments on roughly a twenty-fifth of transactions; encoding those gaps as items instead of discarding them allowed nine forms of procedural failure to be mined directly across 5,790 corrective and 6,578 preventive orders. Apriori and FP-Growth returned identical rule sets, which establishes robustness, while Apriori ran 23 to 98 times faster on an eleven-item space, a result reported for matrices of this shape rather than as a general ranking, and one that suggests algorithm choice in compliance auditing should follow matrix dimensionality rather than convention. The two-step ablation separated two operational syndromes: corrective maintenance is governed by administrative lag, where instant closure implies backdated entry across 47.0% of the population at 98.0% confidence and the effect strengthens on high-priority work, whereas preventive maintenance is governed by schedule slippage, where labour overlap combined with backdating implies late completion at 96.8% confidence and lift 1.759. Every reported rule survives a one-sided Fisher exact test and Benjamini-Hochberg correction across all 155 tested rules; both syndromes retain their effect sizes in multivariable logistic models that adjust each item for the others, and both are the only rules present in every one of the sixteen threshold combinations tested. Three rules that pass the significance test while failing a pre-registered practical-significance criterion are reported as such, since the divergence between the two is itself the reason the ablation was necessary. The rules were carried into three designed interventions: a configurable validation package, a blueprint checklist for new engagements and a natural-language rule engine. None is yet installed, so the contribution at this stage is the evidence for moving the consultant and client relationship from reactive ticket handling towards proactive governance, together with a specification of the controls that move would consist of and of the before-and-after measurement that would test whether it worked. The generalisable output is the framework rather than the numbers: an encoding that turns documentation gaps into mineable items, a prevalence-driven ablation, and a validation layer that separates statistical from practical significance. Stratifying the analysis by asset class and crew, deriving rules on an earlier window and testing them on a later one, replicating the encoding across clients, applying language and image analysis to judge the substance of logs and attachments, and measuring whether the deployed validations actually reduce anomaly rates are the natural next steps.

Acknowledgement

The authors thank the School of Postgraduate Studies, Universitas Diponegoro, and the consulting and maintenance management teams who granted access to the system and validated the operational reading of the results. A generative artificial intelligence assistant was used for language editing and for reorganising the first author's thesis material into the structure required by this journal. Research design, data extraction, query construction, feature engineering, model execution, interpretation and conclusions are the work of the authors, and no figure, table, dataset or numerical result was produced by such a tool.

Conflict of Interest

The first author is employed as a consultant by the implementation partner whose service model is examined in this study. The second and third authors declare no competing interests. Every rule reported here was derived from transaction data rather than from consultant judgement, and the academic supervision of the second and third authors was exercised over the full analysis.

Data Availability

The work order extracts underlying this study contain operational records belonging to the system owner and cannot be released publicly. The SQL used to build the binary indicator matrix, together with an anonymised sample of that matrix, can be made available to the editor and to reviewers on request. The analysis code implementing the itemset search, the Fisher tests with Benjamini-Hochberg correction, the logistic models and the threshold grid is available from the corresponding author, and runs on the anonymised matrix without modification.

REFERENCES

1. C. Singh, and S. Pekkola, *Reasons for customizing packaged enterprise systems: a case study on an enterprise asset management system*, Journal of Information Technology Case and Application Research, vol. 25, no. 2, pp. 159–186, 2023.
2. M. Rodríguez-Hernández, A. Crespo-Márquez, A. Sánchez-Herguedas, and V. González-Prida, *Digitalization as an enabler in railway maintenance: a review from the International Union of Railways asset management framework perspective*, Infrastructures, vol. 10, no. 4, p. 96, 2025.
3. I. Hector, and R. Panjanathan, *Predictive maintenance in Industry 4.0: a survey of planning models and machine learning techniques*, PeerJ Computer Science, vol. 10, p. e2016, 2024.
4. E. L. O. Marques, R. E. C. Lucas, and J. M. Borges, *Digital transformation in industrial maintenance: an overview of Industry 4.0 technologies, their advantages and challenges*, International Journal of Quality and Reliability Management, vol. 43, no. 7, pp. 1938–1953, 2026.
5. G. R. Akhmiyeva, and B.-M. A. Azhiev, *Problems and strategies of enterprise asset management*, Ekonomika i Upravlenie: Problemy, Resheniya, vol. 5/8, no. 146, pp. 248–253, 2024.
6. B. Liu, Z. Yang, J. Chen, and J. Li, *Enterprise asset management risks and management and control strategies*, International Journal of Business Management and Economics and Trade, vol. 5, no. 1, 2024.
7. Z. T. Butarbutar, P. W. Handayani, R. R. Suryono, and W. S. Wibowo, *Systematic literature review of critical success factors on enterprise resource planning post implementation*, Cogent Business and Management, vol. 10, no. 3, 2023.
8. A. Consilvio, G. Vignola, P. López Arévalo, F. Gallo, M. Borinato, and C. Crovetto, *A data-driven prioritisation framework to mitigate maintenance impact on passengers during metro line operation*, European Transport Research Review, vol. 16, no. 1, p. 6, 2024.
9. Y. W. Lee, L. L. Pipino, J. D. Funk, and R. Y. Wang, *Journey to Data Quality*, The MIT Press, Cambridge, MA, 2006.
10. R. Miller, H. Whelan, M. Chrusasik, D. Whittaker, P. Duncan, and J. Gregório, *A framework for current and new data quality dimensions: an overview*, Data, vol. 9, no. 12, p. 151, 2024.
11. S. Aggarwal, *Overview of data quality: examining the dimensions, antecedents, and impacts of data quality*, Journal of the Knowledge Economy, vol. 15, no. 1, pp. 1159–1178, 2024.
12. V. Papastergios, L. Ehrlinger, and A. Gounaris, *Unfolding data quality dimensions in practice: a survey*, ACM Journal of Data and Information Quality, 2025.
13. K. Oyoo, *Collaboration-based automatic data validation framework for enterprise asset management*, Advances in Intelligent Systems and Computing, pp. 639–657, 2020.
14. P. Santos, N. Datia, M. P. M. Pato, J. Sobral, N. Gomes, N. Leitao, and M. R. Ferreira, *NLP for enterprise asset management: an emerging paradigm*, Proceedings of the 27th International Conference on Information Visualisation, pp. 238–243, 2023.
15. Y. J. Lu, W. C. Lee, and C. H. Wang, *Using data mining technology to explore causes of inaccurate reliability data and suggestions for maintenance management*, Journal of Loss Prevention in the Process Industries, vol. 83, p. 105063, 2023.
16. R. G. N. Paiva, C. A. V. Cavalcante, and P. Do, *Applying association rules in the maintenance and reliability of physical systems: a review*, Computers and Industrial Engineering, vol. 194, p. 110332, 2024.
17. M. Lowin, *A text-based predictive maintenance approach for facility management requests utilizing association rule mining and large language models*, Machine Learning and Knowledge Extraction, vol. 6, no. 1, pp. 233–258, 2024.

18. G. Mao, X. Fang, and K. Lu, *Business process anomaly behavior detection method based on multiperspective association rules*, Concurrency and Computation: Practice and Experience, vol. 37, p. e70094, 2025.
19. Y. Shi, N. Zhang, X. Song, H. Li, and Q. Zhu, *Novel approach for industrial process anomaly detection based on process mining*, Journal of Process Control, vol. 136, p. 103165, 2024.
20. H. Chen, M. Yang, and X. Tang, *Association rule mining of aircraft event causes based on the Apriori algorithm*, Scientific Reports, vol. 14, p. 13440, 2024.
21. M. Hajihosseini, A. Maghsoudi, and R. Ghezelbash, *Geochemical anomaly detection and pattern recognition: a combined study of the Apriori algorithm, principal component analysis, and spectral clustering*, Minerals, vol. 14, no. 12, p. 1202, 2024.
22. C. Schroer, F. Kruse, and J. M. Gomez, *A systematic literature review on applying CRISP-DM process model*, Procedia Computer Science, vol. 181, pp. 526–534, 2021.
23. V. Plotnikova, M. Dumas, and F. Milani, *Applying the CRISP-DM data mining process in the financial services industry: elicitation of adaptation requirements*, Data and Knowledge Engineering, vol. 139, p. 102013, 2022.
24. J. Brzozowska, J. Pizon, G. Baytikenova, A. Gola, A. Zakimova, and K. Piotrowska, *Data engineering in CRISP-DM process production data: case study*, Applied Computer Science, vol. 19, no. 3, 2023.
25. F. E. Hamdani, I. A. Q. Quintero, M. Enjolras, M. Camargo, D. Monticolo, and C. Lelong, *Agile supply chain analytic approach: a case study combining agile and CRISP-DM in an end-to-end supply chain*, Supply Chain Forum: An International Journal, vol. 25, no. 1, pp. 96–110, 2024.
26. F. Martinez-Plumed, L. Contreras-Ochando, C. Ferri, P. Flach, J. Hernandez-Orallo, M. Kull, N. Lachiche, and M. J. Ramirez-Quintana, *CASP-DM: context aware standard process for data mining*, arXiv preprint arXiv:1709.09003, 2017.
27. S. Peker, and Ö. Kart, *Transactional data-based customer segmentation applying CRISP-DM methodology: a systematic review*, Journal of Data, Information and Management, vol. 5, no. 1-2, pp. 1–21, 2023.
28. A. M. Shimaoka, R. C. Ferreira, and A. Goldman, *The evolution of CRISP-DM for data science: methods, processes and frameworks*, SBC Reviews on Computer Science, vol. 4, no. 1, pp. 28–43, 2024.
29. R. H. Bemthuis, R. R. Govers, and A. Asadi, *A CRISP-DM-based methodology for assessing agent-based simulation models using process mining*, Journal of Simulation, pp. 1–22, 2025.
30. M. Pohl et al., *The linkage to business goals in data science projects*, Proceedings of the Australasian Conference on Information Systems, paper 60, 2023.
31. P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0: Step-by-step Data Mining Guide*, SPSS Inc., 2000.
32. IBM Corporation, *IBM Maximo Asset Management*, White Paper, Armonk, NY, 2019.
33. Gartner Peer Insights, *Enterprise asset management software reviews 2024*, Available at <https://www.gartner.com/reviews/market/enterprise-asset-management-software>, accessed 1 August 2026.
34. R. Nisbet, K. McCormick, and G. Miner, *A data preparation cookbook*, Handbook of Statistical Analysis, Elsevier, pp. 125–134, 2025.
35. M. Atif, *Data mining*, International Journal of Communication and Information Technology, vol. 3, no. 1, pp. 37–40, 2022.
36. A. Sreelakshmi, N. Padhy, and M. K. Senapaty, *Strategic retail decision-making: a hybrid Apriori-FP-Growth algorithm for efficient association rule discovery*, Lecture Notes in Networks and Systems, pp. 117–130, 2025.
37. T. Li, F. Liu, X. Chen, and C. Ma, *Web log mining techniques to optimize Apriori association rule algorithm in sports data information management*, Scientific Reports, vol. 14, p. 24099, 2024.
38. S. Wu, *An association rule-based approach for frequent item mining of multi-stage access data*, Discover Computing, vol. 28, no. 1, p. 139, 2025.
39. Y. Benjamini, and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, Journal of the Royal Statistical Society, Series B, vol. 57, no. 1, pp. 289–300, 1995.
40. W. M. P. van der Aalst, *Process Mining: Data Science in Action*, 2nd ed., Springer, Berlin, 2016.
41. S. Alter, *Theory of workarounds*, Communications of the Association for Information Systems, vol. 34, pp. 1041–1066, 2014.
42. A. A. M. Abid Ali, *A data-driven decision support framework for business intelligence systems leveraging BERT-enhanced LLMs*, Statistics, Optimization and Information Computing, vol. 16, no. 3, pp. 2404–2425, 2026.
43. Z. Banou, S. Elfilali, E. H. Benlahmar, and F.-Z. Alaoui, *Unveiling quality-factor metrics to optimize data collection: a comprehensive framework for Arabic sentiment analysis*, Statistics, Optimization and Information Computing, vol. 16, no. 3, pp. 1933–1954, 2026.
44. S. A. Jumaa, and O. S. Ibrahim, *Addressing missing values in structural equation modeling with application to digital transformation survey data and its role in improving the quality of healthcare in Mosul*, Statistics, Optimization and Information Computing, vol. 16, no. 3, pp. 2807–2825, 2026.
45. K. G. Sanabria-Lizarraga, B. Carballo-Mendivil, A. Arellano-González, and A. Bueno-Solano, *Business intelligence for agricultural foreign trade: design and application of a Power BI dashboard*, Sustainability, vol. 16, no. 21, p. 9576, 2024.