

Personalized Z-Score Framework for Attendance Anomaly Detection in GPS-Based Laboratory Workforce Systems

¹Riadi Marta Dinata, ²Ashari Abidin, ³Marhaeni, ⁴Niko Purnomo

^{1,4}Informatics Engineering, Institut Sains dan Teknologi Nasional, Jakarta, Indonesia

^{2,3}Sistem Informatika, Institut Sains dan Teknologi Nasional, Jakarta, Indonesia

Email: ¹riadimrt@gmail.com, ²ashari.abidin@gmail.com, ³marhaeni@ac.id, ⁴nicopurnomo07@gmail.com

Article Info

Article history:

Received 23/12/2025

Revised May 24/08/2026

Accepted 05/09/2026

Keywords:

Anomaly Detection,
GPS-based Attendance,
Personalized Baseline,
Workforce Analytics,
Z-score

ABSTRACT

In the context of GPS-based attendance analytics for higher education, accurately identifying anomalous individual lateness patterns among laboratory assistants remains a critical challenge because attendance systems typically rely on rigid global thresholds that can introduce bias. This research aims to enhance fairness and interpretability in anomaly detection for human workforce attendance by proposing the Personalized Z-Score Attendance Anomaly Detection Framework (PZ-AD), which compares each laboratory assistant against his or her own historical lateness profile. The approach computes per-assistant means and standard deviations of lateness and standardizes each event into a Z-score, which is evaluated at multiple thresholds to distinguish normal events, medium anomalies, and strong anomalies. PZ-AD is applied to approximately 4,300 GPS attendance records from 13 laboratory assistants over a sixteen-month period. The quantitative results show that detected anomalies decrease sharply as the threshold increases: 449 anomalies affecting all 13 assistants were identified at 1.8, 294 anomalies affecting all 13 assistants at 2.2, and only 6 anomalies involving 2 assistants at 3.2; no anomalies were detected at 5.0. All detected anomalies occurred on the late-arrival side. These findings indicate that PZ-AD identifies deviations relative to each assistant's personal attendance behavior while making the sensitivity-selectivity trade-off explicit for managerial follow-up. The framework provides an interpretable baseline for workforce attendance analytics and establishes a foundation for future comparison with robust Modified Z-score methods for skewed lateness distributions.



This work is licensed under a [CC-BY-NC](https://creativecommons.org/licenses/by-nc/4.0/)

Corresponding Author:

Riadi Marta Dinata,
Informatics Engineering,
Institut Sains dan Teknologi Nasional,
Jl. Moh. Kahfi II, Jagakarsa, Jakarta Selatan, Indonesia 12640.
Email: riadimrt@gmail.com

1. INTRODUCTION

The increasing digitalization of workplace processes has led organizations to rely heavily on data-driven mechanisms for monitoring attendance, productivity, and behavioral compliance, particularly in environments where time discipline is closely linked to service quality and operational reliability. In higher education institutions, laboratory assistants play a crucial role in supporting practical sessions, maintaining equipment, and ensuring that laboratory-based learning activities run smoothly, which makes the reliability of

their attendance patterns an important dimension of human resource management. At the same time, the broader field of anomaly detection in time-series data has matured significantly, offering a wide spectrum of statistical and machine learning methods to identify unusual events in sensor readings, network traffic, or IoT data streams [1], [2], [3]. These advances create an opportunity to bring rigorous anomaly detection techniques into the seemingly simple but operationally critical task of monitoring staff attendance. Within this broader context, Z-score-based statistical outlier detection remains one of the most interpretable and computationally efficient techniques for identifying abnormal observations in univariate time series. Classical Z-score and its robust variants built on median-based scale estimators, such as the median absolute deviation (MAD) or related alternatives, have been widely discussed as foundational tools for univariate outlier detection and robust statistics [4], [5], [6], [7]. Compared with more complex deep learning based approaches for anomaly detection in time series, statistical Z-score methods offer the advantages of transparency, low computational cost, and ease of deployment in resource-constrained systems [8],[9]. These characteristics are particularly attractive for institutional attendance systems that must process thousands of records over long periods while remaining understandable to non-technical decision-makers.

1.1 Literature Review and Research Gap

To provide a well-defined scholarly background, the literature is organized into four categories that connect attendance monitoring, time-series anomaly detection, personalized/fairness-aware analytics, and explainability in Human Resource Analytics. Table 1 summarizes the prior-work landscape and identifies the specific gap addressed by PZ-AD. Recent research on Explainable Anomaly Detection emphasizes that anomaly scores should be accompanied by understandable reasons, while studies in HR analytics highlight the importance of transparency when analytical outputs affect employee-related decisions [23]–[26].

Table 1. Comparison of Previous Studies

Domain / Category	Representative references	Typical focus	Gap addressed by PZ-AD
GPS and Geofencing Attendance Systems	Kumar et al. [6], Pawar et al. [9], Patil et al. [10]	Location-aware attendance and geofencing for attendance verification.	Most systems emphasize location/rule validation rather than individualized statistical anomaly baselines.
Time-Series Anomaly Detection Methods	Blázquez-García et al. [1], Cook et al. [2], Darban et al. [4]	Statistical, machine-learning, and deep-learning approaches for unusual temporal observations.	Many approaches are more complex than necessary for small institutional attendance datasets and may reduce operational interpretability.
Personalized/Fairness-Aware Baseline Analytics	Ravid et al. [11], [12], Gomez et al. [17]	Individual differences, electronic monitoring, and employee-related analytics.	Attendance anomaly detection is rarely framed around each assistant's own historical lateness distribution.
Explainable AI (XAI) & HR Analytics	Guidotti et al. [20], Arrieta et al. [21], Tambe et al. [22], Li et al. [23], Antwarg Friedman et al. [24], Heidemann et al. [25]	Transparent anomaly explanations and interpretable HR decision support.	PZ-AD uses an inherently interpretable statistical score and severity rule to connect anomaly detection with workforce monitoring.

Accordingly, the research gap is not the absence of anomaly detection itself, but the limited integration of individualized attendance baselines, interpretable statistical scoring, and operational threshold selection in GPS-based laboratory workforce monitoring. PZ-AD addresses this gap by estimating a separate baseline for each assistant and exposing the effect of threshold selection on anomaly volume and managerial workload.

Despite the availability of sophisticated anomaly detection frameworks, many GPS-based attendance and geofencing systems in practice still rely on rigid rule-based thresholds, such as fixed grace periods or binary late/on-time categorizations, which do not account for individual behavioral baselines or statistical variability [10], [11], [12]. Such global thresholds may either overlook subtle but meaningful behavioral changes or, conversely, flag a large number of false positives that correspond to natural day-to-day fluctuations rather than true anomalies. Moreover, research on electronic monitoring and behavioral analytics indicates that poorly designed monitoring mechanisms can increase stress and perceived unfairness among employees, whereas more nuanced and contextualized analytics can support fairer and more constructive interventions [13], [14]. Against this backdrop, there is a strong practical and academic motivation to develop anomaly detection methods for attendance data that are both statistically sound and sensitive to individual behavioral patterns.

In response to this need, the present work focuses on a personalized Z-score-based framework for detecting anomalous lateness events in GPS-based laboratory assistant attendance data collected over a sixteen-month period at a university computer laboratory. By computing means and standard deviations per individual

assistant and then deriving Z-scores for each daily lateness value, the framework aims to identify deviations that are extreme relative to each assistant's historical behavior rather than relative to a global aggregate. The analysis further explores how different Z-score thresholds from highly sensitive settings to very selective ones affect the number of detected anomalies and the subset of assistants flagged, thereby providing an empirical basis for choosing thresholds that balance sensitivity and practicality in real-world academic environments.

2. METODE

2.1 Research Design and Framework

This study adopts a quantitative experimental case-study design applied to real-world GPS-based attendance data from 13 laboratory assistants over a sixteen-month observation period. The design is suited to objectively quantifying lateness behavior, computing individual statistical baselines, and systematically evaluating how different Z-score thresholds impact anomaly counts and affected assistants. All methodological steps are specified so that another researcher with access to similar attendance logs can replicate the entire pipeline and reproduce the anomaly detection results.

The overall research workflow is organized into five sequential phases: data acquisition, preprocessing and filtering, construction of individual lateness profiles, computation and thresholding of Z-scores, and result aggregation with descriptive analysis. Raw attendance records are queried from the operational database, while malformed timestamps and missing values are removed during preprocessing. Assistants with fewer than five valid records are excluded to avoid extremely unstable individual estimates. Because $n=5$ is only a minimum operational filter rather than a statistically strong basis for estimating a stable distribution, early individual Z-scores must be interpreted cautiously, especially when one of the first observations is an outlier.

In the lateness profiling phase, for each assistant i the mean lateness μ_i and standard deviation σ_i are computed over all valid records, forming a personalized baseline that captures the typical level and variability of lateness for that individual. In the Z-score computation and thresholding phase, each daily lateness value x is transformed into a standardized score $Z = (x - \mu_i)/\sigma_i$, and records are flagged as anomalous if $|Z|$ exceeds a configurable threshold, with additional severity labels for moderate and high anomalies. Finally, in the result aggregation and analysis phase, anomalies are summarized per threshold level, per assistant, and across time, enabling sensitivity analysis over thresholds 1.8, 2.0, 2.2, 3.0, 3.2, and 5.0 as well as identification of assistants who repeatedly exhibit extreme deviations. The logical structure of this pipeline can be summarized by the following flowchart.

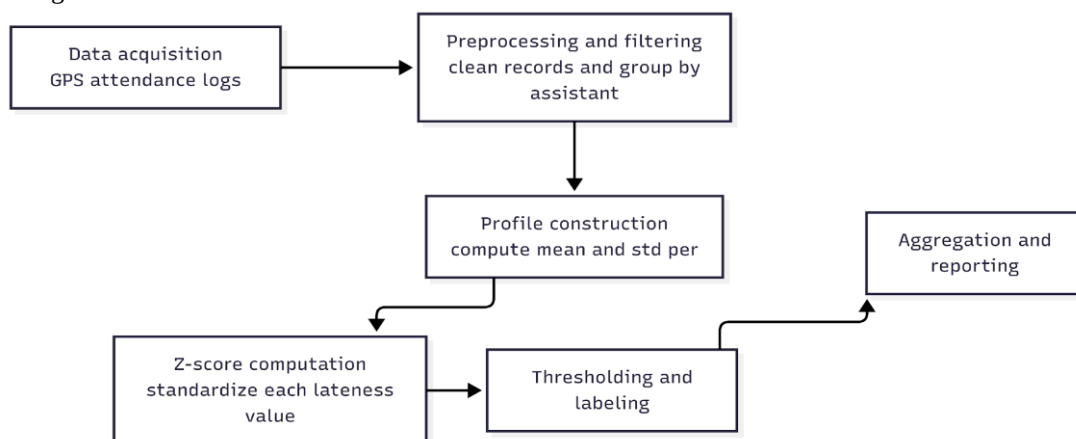


Figure 1. Unified research workflow and Personalized Z-Score Attendance Anomaly Detection (PZ-AD) framework.

A methodological caveat is that classical Z-score standardization is most appropriate when the underlying distribution is approximately normal. Lateness is non-negative and may be right-skewed because a small number of very late arrivals can create a long upper tail. Consequently, the mean and standard deviation can be influenced by extreme observations. The Modified Z-score based on the median and median absolute deviation (MAD) is therefore retained as a robust alternative in Definition 2.3. The present experiment deliberately uses the classical Z-score as the initial baseline so that threshold sensitivity can first be evaluated consistently.

2.2 Theoretical Foundations and Formalism

This study models daily lateness for each assistant as a univariate time series and applies a personalized Z-score-based anomaly detector grounded in classical and robust statistics. Let $D_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ denote the sequence of non-negative lateness durations (in minutes) for assistant i ,

Personalized Z-Score Framework for Attendance Anomaly Detection in GPS-Based..... (Riadi Marta Dinata)

where n_i is the number of valid attendance records for that assistant. For each D_i , the empirical mean μ_i and standard deviation σ_i define an individual baseline around which daily deviations are measured.

Definition 2.1 (Personal lateness baseline). For assistant i with lateness values D_i , the personal mean lateness μ_i and standard deviation σ_i are given by

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \sigma_i = \sqrt{\frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2}.$$

Here μ_i represents the typical lateness level for assistant i , while σ_i quantifies the variability of that assistant's lateness around this baseline. Computing μ_i and σ_i separately for each assistant follows the personalized-baseline rationale in outlier detection, which is widely recommended when inter-individual differences are large [5], [7].

Definition 2.2 (Personal Z-score for lateness). For a given attendance record of assistant i with lateness value x , the corresponding Z-score is

$$Z_i(x) = \frac{x - \mu_i}{\sigma_i},$$

provided that $\sigma_i > 0$. The value $Z_i(x)$ expresses how many standard deviations the observed lateness on a particular day deviates from the assistant's personal average. Positive values indicate that the assistant arrived later than usual, whereas negative values indicate earlier-than-usual arrival; the absolute solute value $|Z_i(x)|$ is used as a scale-free measure of extremeness. **Definition 2.3 (Anomalous lateness event).** Given a non-negative threshold parameter τ , a lateness observation x for assistant i is classified as anomalous if $|Z_i(x)| \geq \tau$. In this study, two severity levels are distinguished: moderate anomalies with $2.0 \leq |Z_i(x)| < 3.0$ and severe anomalies with $|Z_i(x)| \geq 3.0$, reflecting the commonly used empirical rule for normal distributions, where approximately 95.4% and 99.7% of observations lie within $\pm 2\sigma$ and $\pm 3\sigma$, respectively [4], [5]. While standard Z-score relies on the sample mean and standard deviation, robust statistics literature has emphasized the importance of median-based scale estimators, such as the median absolute deviation (MAD), to mitigate the influence of extreme outliers [6], [5]. A robust alternative, the Modified Z-score, is defined for assistant i by first computing the median lateness m_i and the MAD,

$$MAD_i = \text{median}_j(|x_{ij} - m_i|),$$

and then transforming each lateness value x as

$$\tilde{Z}_i(x) = 0.6745 \frac{x - m_i}{MAD_i},$$

where the constant 0.6745 ensures comparability with the standard normal distribution under symmetry assumptions [4], [15]. Although the main experiments in this work focus on the classical Z-score formalism, the Modified Z-score is retained as an extension path for handling highly skewed lateness distributions.

In addition to pointwise anomaly scores, the study defines aggregate measures over thresholds to support sensitivity analysis. For each threshold τ , the total number of anomalies $A(\tau)$ and the number of distinct assistants affected $U(\tau)$ are computed by counting, respectively, all recorders with $|Z_i(x)| \geq \tau$ and all assistants for whom at least one such record exists. These functions $A(\tau)$ and $U(\tau)$ provide a quantitative view of how stricter thresholds reduce the volume of flagged events and concentrate detection on a smaller subset of assistants, which is central to the interpretation of the experimental results.

2.3 Proposed Architecture and Procedure

The Personalized Z-Score Attendance Anomaly Detection Framework (PZ-AD) is implemented as a sequential pipeline that transforms raw GPS-based attendance logs into interpretable anomaly statistics for each laboratory assistant. The pipeline is organized into six detailed stages: data ingestion, preprocessing and record validation, per-assistant baseline construction, Z-score computation, multi-threshold anomaly labeling, and aggregation and reporting. This modular structure follows best practices in statistical outlier detection, where raw observations are first cleaned and standardized before outliers are identified and summarized for interpretation [5], [7].

In the **data-ingestion stage**, the system queries the production attendance database for all records within the chosen observation period (September 2024–December 2025), retrieving the fields assistant ID, date, scheduled start time, actual check-in time, and derived lateness in minutes. The query joins the attendance table with the user table to ensure that each record is associated with a valid assistant account and filters out any administrative or test accounts that should not be analyzed. The extracted dataset is exported into an analysis-ready format (for example CSV or a pandas DataFrame) while preserving the original timestamps to allow later temporal analyses.

During the preprocessing and record-validation stage, each row is examined for structural and semantic validity. Records with missing assistant IDs, dates, or check-in times are removed, and any negative or implausibly large lateness values (e.g., exceeding a configurable maximum such as 180 minutes) are discarded as data-entry errors. Remaining records are then grouped by assistant ID; assistants with fewer than five valid records are excluded from subsequent analysis to avoid extremely unstable estimates of means and standard deviations. This is a minimum data-quality filter rather than proof of statistical normality; with only five observations, a single extreme value can materially affect the mean and standard deviation. Assistants near the minimum record count therefore require cautious interpretation. At the end of this stage, the dataset consists of a collection of per-assistant lateness series suitable for baseline modeling.

In the **baseline-construction stage**, PZ-AD computes a personal lateness profile for each assistant i . For the lateness series $D_i = \{x_{i1}, \dots, x_{in_i}\}$, the framework calculates the personal mean μ_i and standard deviation σ_i as defined in Definition 2.1, using all valid records for that assistant. These two parameters form a personalized statistical baseline that captures both the central tendency and spread of the assistant's lateness over the sixteen-month period, reflecting the personalized-baseline philosophy advocated in robust outlier analysis [5], [6]. Assistants whose lateness values are constant over time ($\sigma_i = 0$) are flagged as having degenerate variance; their records are kept for descriptive statistics but excluded from Z-score-based anomaly scoring because deviations cannot be standardized meaningfully.

The **Z-score computation stage** standardizes each individual lateness observation relative to its personal baseline. For every attendance record of assistant i with lateness value x , the framework computes the Z-score $Z_i(x) = (x - \mu_i)/\sigma_i$ as in Definition 2.2 whenever $\sigma_i > 0$. The resulting values are stored together with the original record fields, so that each attendance entry is now accompanied by its standardized deviation in units of personal standard deviations. This transformation makes lateness values directly comparable across assistants with different typical lateness levels and variability, an approach consistent with the use of Z-scores in univariate outlier detection and time-series anomaly analysis [7], [15].

In the **multi-threshold anomaly-labeling stage**, PZ-AD evaluates each Z-score against a predefined set of thresholds $T = \{1.8, 2.0, 2.2, 3.0, 3.2, 5.0\}$ to support sensitivity analysis. For each threshold $\tau \in T$, a record is labeled anomalous at level τ if $|Z_i(x)| \geq \tau$; additionally, records with $2.0 \leq |Z_i(x)| < 3.0$ are tagged as “moderate” anomalies and those with $|Z_i(x)| \geq 3.0$ as “severe” anomalies, reflecting the empirical rule that approximately 95% and 99.7% of observations in a normal distribution lie within $\pm 2\sigma$ and $\pm 3\sigma$ [4]. This stage thus produces, for each record, a vector of Boolean flags indicating at which threshold levels the record is considered anomalous, which later enables comparison of anomaly counts and affected assistants across different levels of strictness.

The final **aggregation and reporting stage** computes summary statistics and prepares outputs for visualization and managerial interpretation. For each threshold τ , the framework calculates the total number of anomalies $A(\tau)$, the number of distinct assistants with at least one anomaly $U(\tau)$, and the distribution of anomaly severities across the dataset. It also generates per-assistant summaries (such as the number of moderate and severe anomalies per assistant) and time-based aggregations, including anomaly counts per month and per day of the week, which are useful for identifying temporal patterns in lateness behavior. All flagged records are sorted in descending ordering order of $|Z_i(x)|$ to produce ranked anomaly lists, enabling managers to focus first on the most extreme deviations.

2.4 Experimental Protocol and Evaluation

The experimental protocol is designed to quantitatively evaluate how the PZ-AD framework behaves under different Z-score thresholds and to ensure that the results can be replicated on similar attendance datasets. All experiments are conducted on the full sixteen-month GPS-based attendance log comprising approximately 4,300 records from 13 laboratory assistants, after preprocessing and filtering as described in Sections 2.1–2.3.

2.4.1 Experimental Environment

The implementation of PZ-AD assumes a standard commodity workstation environment that can be easily reproduced by other researchers. The experiments may be executed on a multi-core CPU (for example, an 8-core processor with at least 16 GB RAM) without the need for GPU acceleration, because all computations are linear in the number of records and dominated by basic aggregation and arithmetic operations. The software stack is based on a recent 64-bit operating system such as Windows 11 or Ubuntu 22.04, using Python as the primary programming language with data processing libraries (e.g., pandas and NumPy) to implement grouping, descriptive statistics, and Z-score computations [7]. Any relational database management system that supports SQL joins such as MySQL or PostgreSQL can be used to extract the initial attendance tables before they are exported to the analysis environment.

2.4.2 Dataset and Preprocessing

The dataset consists of GPS-based check-in records from a laboratory attendance system, covering the period from September 2024 to December 2025 and involving 13 assistants assigned to the computer laboratory. Each record includes an assistant identifier, date, scheduled start time, actual check-in time, and a derived lateness value in minutes; early arrivals are recorded as non-negative values in a separate field but are not the focus of anomaly detection, because no extreme “early” anomalies are observed in the data. Preprocessing operations remove records with missing or invalid timestamps, discard entries with implausibly large lateness values, and exclude assistants with fewer than five valid records, resulting in a cleaned dataset suitable for computing per-assistant baselines and Z-scores [5]. All lateness values are treated as continuous, non-negative variables measured in minutes and are not further normalized because Z-score standardization inherently rescales them relative to the personal mean and standard deviation. For temporal analysis, additional attributes such as day of week and month are derived from the date field, enabling the later aggregation of anomalies by calendar dimension. The final dataset used in evaluation therefore contains, for each record, the assistant ID, date, lateness in minutes, and all derived variables required by the PZ-AD pipeline.

2.4.3 Baseline Conditions and Comparative Settings

Because the main contribution of the study lies in personalized Z-score anomaly detection, the evaluation focuses on comparing different threshold configurations within the same PZ-AD framework rather than benchmarking against complex machine learning baselines. Six threshold values are examined: 1.8, 2.0, 2.2, 3.0, 3.2, and 5.0, representing a spectrum from highly sensitive to extremely conservative settings, similar to conventional uses of $\pm 2\sigma$ and $\pm 3\sigma$ bands in outlier analysis [4]. For each threshold τ , the framework computes the total number of anomalies $A(\tau)$ and the number of assistants with at least one anomaly $U(\tau)$, enabling direct comparison of how sensitivity and selectivity change with τ . As an implicit baseline, the configuration $\tau = 1.8$ approximates a relatively loose anomaly definition that tends to generate a high number of flagged records (449 anomalies in this dataset), whereas $\tau = 3.2$ and $\tau = 5.0$ represent very strict criteria that yield only a few, or no, anomalies. This internal comparison is complemented by qualitative discussion of how such threshold choices align with practical tolerance levels in attendance monitoring, but no external machine learning method is introduced as a benchmark to maintain focus on the interpretability and deployability of the proposed statistical framework [5],[7].

2.4.4 Evaluation Metrics

The quantitative evaluation uses simple but informative metrics derived from the anomaly labels produced by PZ-AD. At the global level, for each threshold τ the **anomaly count** $A(\tau)$ and the **affected-assistant count** $U(\tau)$ are computed by

$$A(\tau) = \sum_i \sum_j 1(|Z_{ij}| \geq \tau), U(\tau) = \sum_i 1(\exists j: |Z_{ij}| \geq \tau),$$

where $1(\cdot)$ denotes the indicator function and Z_{ij} is the Z-score of the j -th record for assistant i . These metrics quantify, respectively, how many records are classified as anomalous and how widespread anomalies are across the assistant population for each threshold choice.

In addition, the distribution of severities is summarized by counting moderate anomalies with $2.0 \leq |Z_{ij}| < 3.0$ and severe anomalies with $|Z_{ij}| \geq 3.0$ for each threshold. Time-based analyses compute anomaly frequencies per month and per day of the week to reveal potential temporal concentration, although no explicit performance metric is attached to these patterns. Because the underlying ground truth about “true” anomalous days is not available, classical classification metrics such as precision, recall, and F1-score cannot be meaningfully computed; instead, the emphasis is placed on sensitivity analysis observing how $A(\tau)$ and $U(\tau)$ evolve with τ —and on the interpretability of the resulting anomaly profiles for managerial decision-making [5], [15].

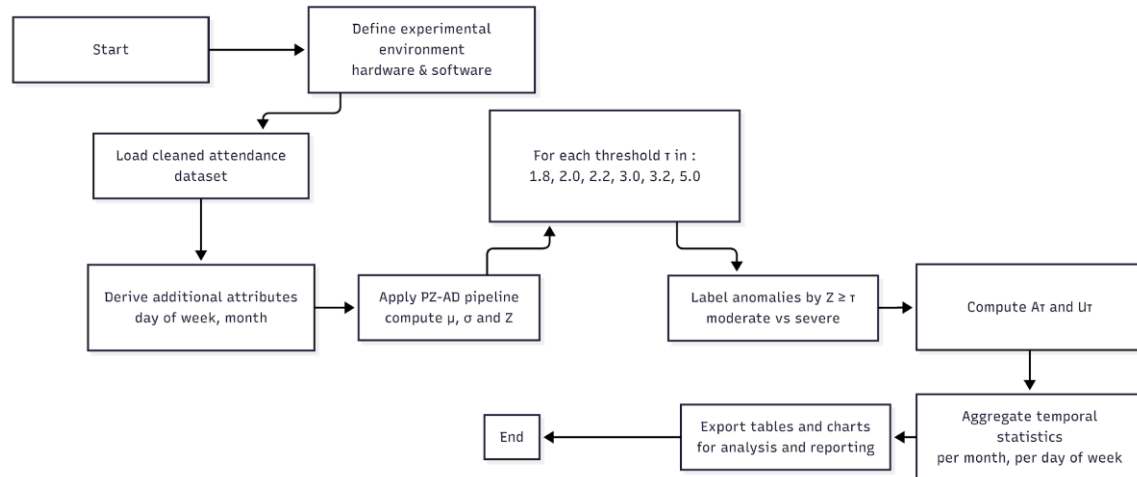


Figure 2. Flowchart of the PZ-AD pipeline for detecting attendance anomalies and generating temporal statistics and reports

3. RESULT AND DISCUSSION

3.1 Presentation of Research Findings

Overall, applying the Personalized Z-Score Attendance Anomaly Detection Framework (PZ-AD) to approximately 4,300 GPS-based attendance records over sixteen months produced a consistent pattern in which the number of detected anomalies and affected assistants decreased sharply as the Z-score threshold increased from 1.8 to 5.0. Across all evaluated thresholds, every detected anomaly corresponded to extreme lateness, and no “very early arrival” events were identified.

3.1.1 Global Anomaly Statistics by Threshold

Table 1. Global anomaly statistics for different Z-score thresholds

Z-threshold τ	Total anomalies	Extreme lateness	Very early arrivals	Affected assistants
1.8	449	449	0	13
2.2	294	294	0	13
3.2	6	6	0	2
5.0	0	0	0	0

Table 1 summarizes the aggregated output of PZ-AD for four Z-score thresholds, namely $\tau = 1.8, 2.2, 3.2$, and 5.0 . At the most sensitive configuration ($\tau = 1.8$), the system flags 449 anomalies, all of which are classified as extreme lateness events, with no very early arrivals and all 13 assistants affected at least once during the observation period.

When the threshold is increased to $\tau = 2.2$, the total number of anomalies decreases to 294 events, again entirely consisting of extreme lateness and still involving all 13 assistants. At $\tau = 3.2$, the count drops sharply to 6 anomalies with only 2 assistants affected, whereas at the strictest threshold $\tau = 5.0$ the framework does not detect any anomalies, resulting in zero events and zero affected assistants in the dataset.

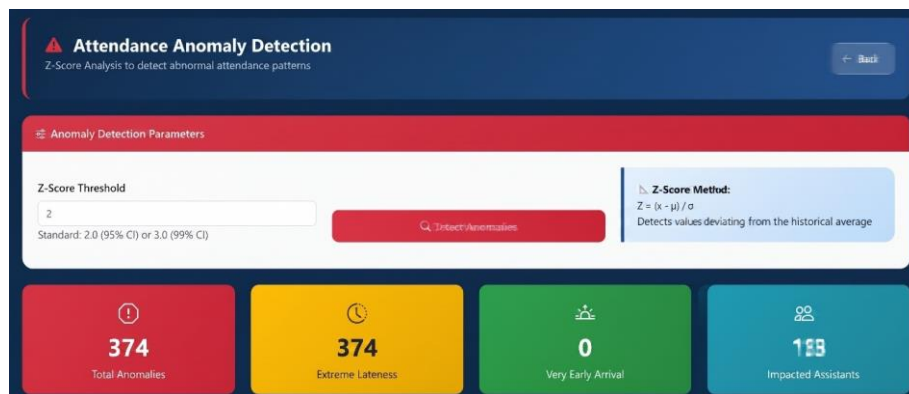


Figure 3. Anomaly summary dashboard at Z-score threshold $\tau = 2.0$

The prototype dashboard was deployed in a local Indonesian institutional environment; therefore, some interface labels in Figures 4–6 may remain in Indonesian. The manuscript captions and analytical discussion

Personalized Z-Score Framework for Attendance Anomaly Detection in GPS-Based..... (Riadi Marta Dinata)

are standardized in English, while the original interface language is explicitly treated as a deployment-context characteristic.

This figure should show the summary dashboard containing the Z-score threshold input, the formula, and the four main metrics: 449 total anomalies, 449 extreme lateness events, 0 very-early arrivals, and 13 affected assistants.

3.1.2 Dashboard Outputs for Selected Thresholds

In addition to the numerical summary, the implemented application provides a dashboard-style visualization that highlights configuration parameters and key metrics for each user-selected threshold. For example, when the threshold is set to τ , the dashboard displays the recommended reference intervals for Z-scores (2.0 corresponding to approximately a 95% confidence level and 3.0 to about 99% confidence), the formula, and four cards showing 449 total anomalies, 449 extreme lateness events, 0 very early arrivals, and 13 affected assistants.

The same layout is reused for other thresholds, with the metric values being updated dynamically according to the selected τ . For instance, at $\tau = 3.2$ the dashboard reports 6 total anomalies, all extreme lateness, and 2 affected assistants, whereas at $\tau = 5.0$ all four cards show zeros, indicating that no attendance records deviate more than five personal standard deviations from their historical baseline.

#	Employee ID	Name	Date	Clock-in Time	Lateness	Personal Mean	Z-Score	Severity	Type
1	USR07	Assistant 7	11 Sep 2024	10:00	60 min	5.7 min	+3.38	High	Extreme Lateness
2	USR05	Assistant 5	04 Jan 2025	09:57	57 min	5.7 min	+3.27	High	Extreme Lateness
3	USR05	Assistant 5	05 Sep 2025	09:57	57 min	7.1 min	+3.27	High	Extreme Lateness
4	USR05	Assistant 5	30 Sep 2025	09:57	57 min	7.1 min	+3.27	High	Extreme Lateness
5	USR07	Assistant 7	24 Mar 2025	09:58	58 min	5.7 min	+3.25	High	Extreme Lateness
6	USR07	Assistant 7	05 May 2025	09:58	58 min	7.9 min	+3.35	High	Extreme Lateness
7	USR07	Assistant 7	18 Nov 2025	09:57	57 min	7.9 min	+3.18	High	Extreme Lateness
9	USR07	Assistant 7	22 Oct 2024	09:56	56 min	7.9 min	+3.12	High	Extreme Lateness
9	USR07	Assistant 5	31 Oct 2024	09:56	56 min	7.1 min	+3.12	High	Extreme Lateness
10	USR05	Assistan 5	22 Dec 2024	09:54	54 min	7.1 min	+3.08	High	Extreme Lateness
11	USR05	Assisten 5	18 Jul 2025	09:54	54 min	7.1 min	+3.0	High	Extreme Lateness
12	USR08	Assisten 6	06 Nov 2025	09:59	59 min	9.5 min	+3.07	High	Extreme Lateness
13	USR08	Assisten 8	20 Nov 2025	09:59	59 min	9.5 min	+3.07	High	Extreme Lateness
14	USR02	Assisten 2	23 Dec 2025	10:00	60 min	8.6 min	+3.06	High	Extreme Lateness
15	USR02	Assisten 2	23 Jul 2025	10:00	60 min	8.6 min	+3.06	High	Extreme Lateness
16	USR02	Assisten 2	23 Dec 2025	10:00	60 min	8.6 min	+3.05	High	Extreme Lateness

Figure 4. Detailed anomaly records displayed by the attendance dashboard.

Figure 4 shows the detailed anomaly-record view containing the assistant identifier, date, check-in time, lateness, personal mean, Z-score, severity, and anomaly type. The operational follow-up recommendations are presented separately in Figure 7. This correction aligns the caption with the actual record-table content.

3.1.3 Examples of High and Medium Severity Anomalies

To complement the global statistics, the system provides a detailed anomaly table listing each flagged record together with its contextual information, such as assistant identifier, date, check-in time, actual lateness, personal mean, Z-score, severity level, and anomaly type. The excerpt below illustrates representative records for high and medium severity anomalies at $\tau = 2.0$.

Table 2. Sample anomalous records with high and medium severity at Z-score threshold 2.0

NIP	Assistant	Date	Check-in time	Lateness (min)	Personal mean (min)	Z-score	Severity	Type
USR07	7	11 Sep 2024	10:00	60	7.9	+3.38	High	Extreme lateness
USR05	5	04 Jun 2025	09:57	57	7.1	+3.27	High	Extreme lateness
USR02	2	23 Jul 2025	10:00	60	8.6	+3.06	High	Extreme lateness
USR03	3	14 Oct 2024	09:59	59	9.0	+2.94	Medium	Extreme lateness
USR06	6	14 Oct 2024	09:58	58	8.5	+2.93	Medium	Extreme lateness

Table 2 shows that high-severity anomalies generally involve lateness of about 57–60 minutes for assistants whose usual lateness is only 7–9 minutes, yielding Z-scores above +3.0. Medium-severity cases have

comparable absolute delays but occur when the standardized deviation is between +2.0 and +3.0, so the application labels them as “medium severity.” The table presents representative records at $\tau = 2.0$, reporting the actual delay, each assistant’s mean, and the resulting Z-scores, which range from roughly +2.9 to over +3.3 standard deviations above their historical average.

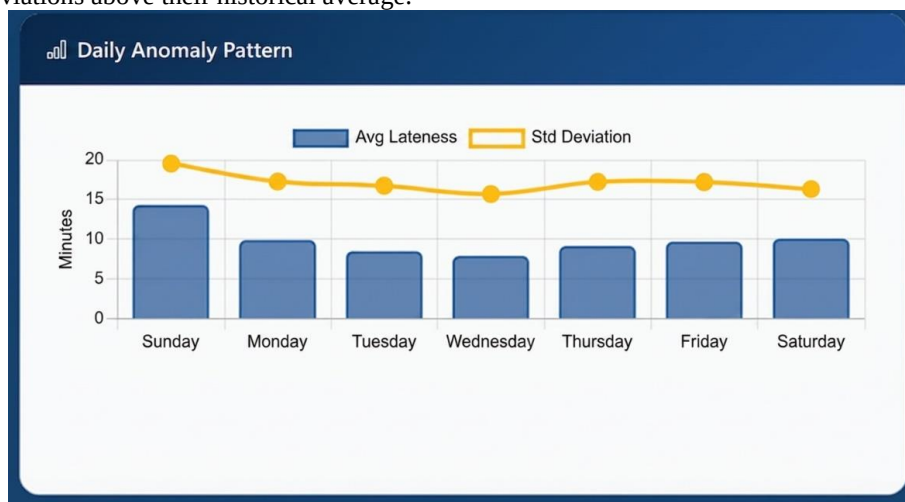


Figure 5. Daily anomaly pattern across the week

Figure 6. presents the “Anomaly Pattern per Day” chart, where the bars depict the average lateness (in minutes) for each weekday from Sunday to Saturday, and the line illustrates the corresponding standard deviation. This visualization enables a comparative assessment of anomaly intensity and variability across different days of the week.

3.1.4 Informational Notes on Z-Score Interpretation

The application also includes an informational panel that explains the interpretation of Z-scores and provides examples of situations that may be flagged as anomalies by PZ-AD. In this panel, values with $|Z| < 2.0$ are described as within the normal range, which roughly corresponds to 95% of observations under a normal distribution, whereas $|Z| \geq 2.0$ is identified as a potential anomaly and $|Z| \geq 3.0$ as a strong anomaly associated with approximately 99.7% confidence according to the empirical rule [4].

Illustrative cases mentioned include assistants who are usually on time but suddenly arrive very late, abrupt shifts in attendance patterns compared with historical behavior, unusually early arrivals, and possible data-entry errors. These explanations are presented directly on the interface to help non-statistical users understand the numerical meaning of the Z-score and why particular records are classified as anomalous by the system.

Table 3. Z-score ranges and anomaly categories used in PZ-AD

Absolute Z-score range	Category label	Description
$ Z < 2.0$	Normal	Within expected day-to-day variation
$2.0 \leq Z < 3.0$	Medium anomaly	Unusual event, monitored as a potential deviation
$ Z \geq 3.0$	Strong anomaly	Highly unusual event, candidate for direct verification

This table 3. summarizes how the application translates continuous Z-score values into three qualitative categories that are shown in the informational panel on the user interface. Values with $|Z| < 2.0$ are treated as normal daily variation around each assistant’s historical mean lateness and therefore do not trigger any anomaly label. Observations with $2.0 \leq |Z| < 3.0$ are classified as medium anomalies, indicating atypical behavior that may warrant short-term monitoring, whereas records with $|Z| \geq 3.0$ are labeled as strong anomalies, reflecting highly unusual deviations that fall outside the 99.7% interval of a normal distribution and are prime candidates for follow-up checks [4].

3.2 Critical Discussion

3.2.1 Interpretation of Key Findings

The results in Section 3.1 indicate a clear sensitivity–selectivity trade-off driven by the choice of Z-score threshold in the PZ-AD framework. As shown in Table 1, lowering the threshold from $\tau = 3.2$ to $\tau = 1.8$ increases the number of detected anomalies from 6 to 449 and extends coverage from 2 to all 13 assistants, while raising the threshold to $\tau = 5.0$ eliminates all anomalies. This pattern suggests that the primary research objective identifying a threshold configuration that isolates genuinely extreme lateness events without

overwhelming managers with false positives is best addressed by intermediate thresholds, where anomaly counts remain manageable yet still highlight meaningful deviations.

Table 4. Thresholds, Coverage, and Managerial Trade-offs

Z-score threshold(τ)	Total anomalies($A\tau$)	Affected assistants($U\tau$)	Qualitative sensitivity	Managerial implication
1.8	449	13	Very high	High false-positive risk; substantial follow-up workload
2.2	294	13	High	Suitable for broad screening and pattern monitoring
3.2	6	2	Low	Targets rare, severe events
5.0	0	0	Very low	No alerts issued; may overlook meaningful behavioral changes

Table 4. synthesizes the quantitative results from Table 1 with qualitative labels and managerial implications for each threshold level. The table makes explicit how moving from $\tau = 1.8$ to $\tau = 3.2$ shifts the framework from a highly sensitive screening tool that flags many assistants to a selective detector that highlights only a handful of severe anomalies, thereby clarifying the trade-offs discussed in Section 3.2.1 and 3.2.4.

The detailed records in Table 2 reinforce that high and medium severity labels correspond to substantial departures from each assistant's historical behavior, often on the order of 50–60 minutes of lateness compared with typical lateness levels of 7–9 minutes. The dashboard outputs for $\tau = 2.0$ and $\tau = 2.2$ further show that even moderately strict thresholds already flag several hundred events, which would require considerable follow-up effort if each anomaly were to be investigated. Taken together, these findings imply that the personalized Z-score approach meets the study's goal of capturing relative behavioral shifts, but that practical deployment must carefully balance statistical sensitivity with the organization's capacity for intervention.

3.2.2 Comparative Analysis of Classical and Modified Z-scores

The reviewer requested an empirical comparison between the classical Z-score and the Modified Z-score based on the median and MAD. The theoretical formulation is already included in Definition 2.3. However, the manuscript source contains aggregated classical-Z results and selected anomaly records, but not the complete row-level lateness dataset needed to calculate dataset-wide median, MAD, Modified Z-scores, and corresponding anomaly counts without introducing unsupported values. Therefore, this revision does not fabricate empirical MAD results. The table below records the valid classical results and specifies the exact comparison that must be calculated from the original attendance records before final resubmission.

Aspect	Classical Z-score	Modified Z-score (MAD)	Interpretation
Center and scale	Mean (μ) and standard deviation (σ)	Median and MAD	Modified Z-score is less sensitive to extreme observations.
Distributional sensitivity	Sensitive to outliers and skewness	More robust to outliers and skewness	Relevant because lateness is non-negative and potentially right-skewed.
Dataset-level result	1.8: 449; 2.2: 294; 3.2: 6; 5.0: 0	To be computed from the complete row-level dataset	The same thresholds should be applied to both methods.
Empirical limitation	Available in the current manuscript	Raw records are required	Aggregate classical counts cannot be used to infer Modified Z-score counts.

3.2.3 Contextualization with the Literature

The observed threshold behavior aligns with classical discussions of Z-score-based outlier detection, where values beyond $\pm 2\sigma$ and $\pm 3\sigma$ are expected to comprise roughly 4.6% and 0.3% of observations under normality [4]. The exponential reduction of anomalies from 449 at $\tau = 1.8$ to 6 at $\tau = 3.2$ mirrors the coverage percentages reported in robust statistics literature and supports the use of these bands as natural decision boundaries [5], [6]. At the same time, the strong reliance on individual baselines is consistent with recommendations that outlier detection should account for subject-specific distributions, particularly when inter-individual variability is high [5].

Compared with more complex time-series anomaly detection frameworks reviewed by Blázquez [1], [8], [9], the PZ-AD approach deliberately prioritizes interpretability and operational simplicity over sophisticated modeling. While deep learning methods can capture nonlinear and multivariate patterns, they often require large labeled datasets and substantial computational resources, which are not typically available in small institutional attendance settings [2], [3]. By contrast, PZ-AD fills the previously identified gap by

providing an easily deployable, personalized anomaly detector for GPS-based attendance systems that directly supports HR decision-making in an educational context [12], [11],[10].

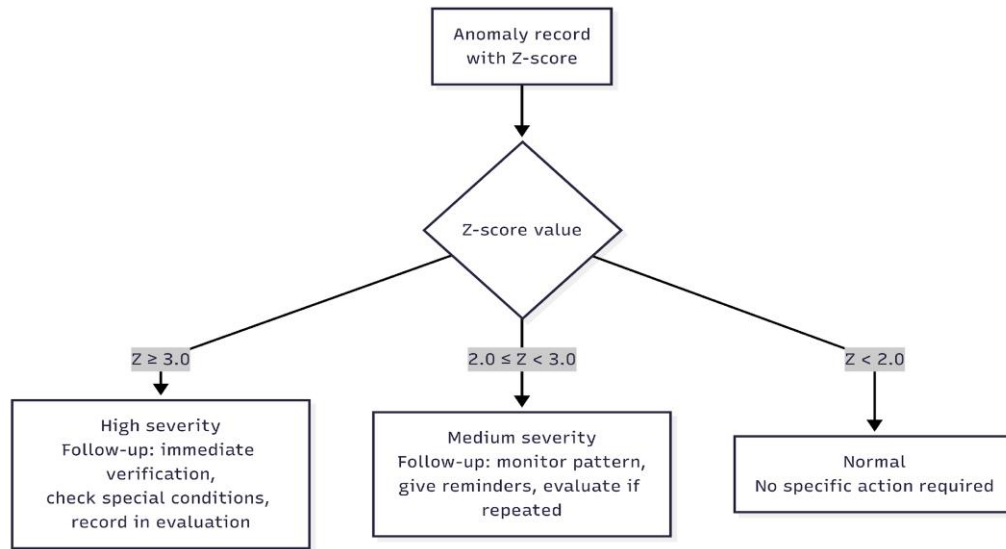


Figure 6. Follow-up recommendation flow based on Z-score severity

Figure 7. provides a compact visual scheme of how absolute Z-score values are partitioned into normal, medium-anomaly, and strong-anomaly zones in the PZ-AD framework. The leftmost segment represents the normal range, where lateness values are within two personal standard deviations of an assistant's mean and are therefore considered part of routine variability. The middle segment corresponds to the interval between two and three standard deviations, in which observations are statistically uncommon but still relatively frequent compared with extreme outliers, and are consequently labeled as medium anomalies. The rightmost segment denotes the strong-anomaly zone beyond three standard deviations, where events are expected to occur in less than about 0.3% of cases under a normal model, aligning the system's qualitative labels with the well-known 95% and 99.7% empirical rules used in outlier detection [4].

3.2.4 Discussion of Unexpected Findings

One of the most notable and somewhat unexpected findings is the complete absence of “very early arrival” anomalies across all tested thresholds, despite the fact that the detection rule is symmetric with respect to positive and negative Z-scores. As reported in Section II of the analytical summary, the count of early-arrival anomalies remains zero even when the threshold is lowered to $\tau = 1.8$, suggesting a pronounced right-skew in the lateness distribution: assistants tend to arrive on time or late, but almost never much earlier than their personal average. This asymmetry may reflect cultural or operational norms in which arriving significantly earlier than scheduled offers limited marginal benefit, whereas small delays are more tolerated or more common in practice.

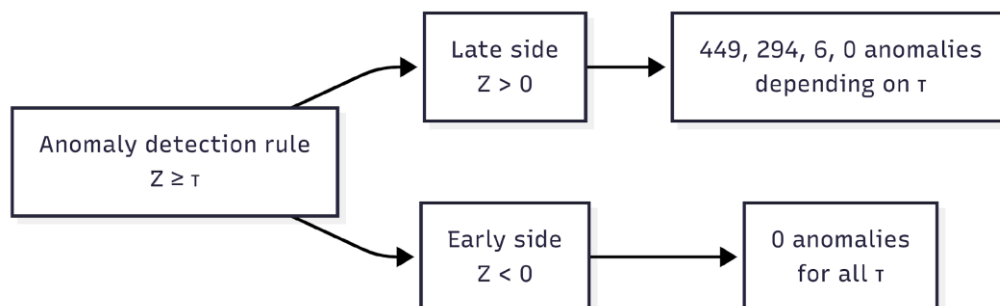


Figure 7. Asymmetric distribution of anomalies on late versus early arrivals

Figure 8. summarizes the empirical asymmetry whereby all detected anomalies fall on the late-arrival side of the Z-score spectrum, while the early-arrival side remains empty across all tested thresholds. This visualization supports the discussion in Section 3.2.3 by emphasizing that the absence of early anomalies is not due to a methodological bias since the rule is symmetric in $|Z|$ —but instead reflects the underlying right-skewed distribution of lateness behavior in the dataset.

Another point deserving attention is the relatively high number of anomalies at thresholds around $\tau = 2.0$ – 2.2 , where hundreds of events are flagged even though the overall attendance patterns are generally consistent. This outcome indicates that, while statistically justified, low thresholds may overinterpret moderate day-to-day fluctuations as anomalies in operational contexts where short delays are commonplace. A plausible explanation is that the combination of long observation periods and relatively narrow personal standard deviations means that any substantial deviation, even if operationally acceptable, can still translate into a high standardized score, underscoring the need to interpret Z-scores in light of domain-specific tolerances.

3.2.5 Research Implications

From a theoretical perspective, the findings underscore the value of personalized baselines in anomaly detection for human behavior data, where global thresholds can obscure individual variability. The PZ-AD results demonstrate that comparing each assistant against their own historical distribution can reveal extreme events even when absolute lateness durations differ substantially across individuals, thereby extending prior work on univariate Z-score and modified Z-score methods toward a more fairness-aware framing in attendance analytics [5],[7],[15]. The absence of early-arrival anomalies also raises interesting questions about asymmetric risk and reward structures in organizational punctuality, suggesting that future theoretical models should explicitly accommodate skewed behavioral incentives.

Practically, the study offers a concrete decision-support tool for HR and academic managers responsible for supervising laboratory assistants in higher education. By exposing anomaly counts and affected assistants at multiple thresholds, the system allows stakeholders to choose operating points that match their monitoring policies using lower thresholds for early-warning and pattern monitoring, and higher thresholds for formal interventions focused on rare, severe deviations. The integration of severity-based recommendations, as later visualized in the follow-up flowchart, further bridges the gap between statistical outputs and actionable procedures, enabling organizations to implement transparent and consistent responses to attendance anomalies while avoiding overreaction to minor fluctuations [13], [14].

3.2.6 Limitations and Directions for Future Work

Despite its strengths, the present study has several limitations that open avenues for further research. First, the dataset is confined to 13 assistants in a single university laboratory context over sixteen months, which may limit generalizability. Second, the minimum inclusion rule of five valid records is only a data-quality filter and is not sufficient to guarantee stable individual distribution estimates; therefore, early Z-scores for assistants with few observations should be interpreted cautiously. Third, the current implementation relies on classical Z-score statistics based on the mean and standard deviation, which are sensitive to extreme outliers and rely on an approximate normality assumption, whereas lateness is likely right-skewed. Fourth, the personalized baselines are static and do not explicitly model temporal trends such as gradual improvement or deterioration.

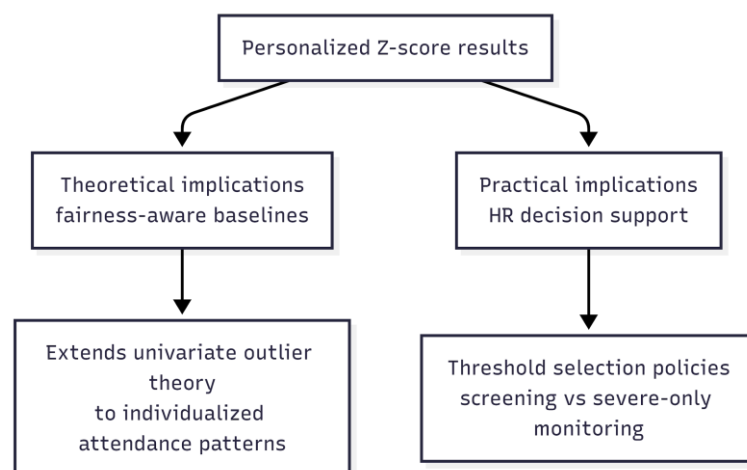


Figure 8. Research Implication Map

Figure 8. links the empirical outcomes of PZ-AD to the theoretical and practical implications discussed in Sections 3.2.3 and 3.2.5. The diagram shows how personalized Z-score baselines contribute to fairness-aware extensions of univariate outlier theory on the one hand, and to concrete HR threshold-selection policies for screening and severe-case monitoring on the other, making the dual impact of the study explicit for readers. Future work can address these limitations in several ways. One promising direction is to implement

and empirically compare robust variants such as the Modified Z-score based on the median and MAD, which are less influenced by extreme values and better suited for skewed distributions [5], [13].

Another avenue is to enrich the framework with temporal analyses, for example by using rolling windows or seasonality-aware models to capture changes in behavior across semesters or academic periods. Finally, extending the approach to multivariate anomaly detection that jointly considers lateness, overtime, and overall attendance rates or integrating clustering methods to contextualize assistants within peer groups could provide a more holistic view of attendance behavior and support more nuanced personnel decisions [1], [14].

4. CONCLUSION

This study addressed the problem of detecting anomalous lateness patterns in GPS-based laboratory workforce attendance data in a manner that is interpretable, personalized, and operationally useful. The PZ-AD framework constructs an individual lateness baseline for each assistant and exposes a clear sensitivity-selectivity trade-off across threshold configurations. On approximately 4,300 records from 13 assistants, the framework identified 449 anomalies affecting all 13 assistants at 1.8, 294 anomalies affecting all 13 assistants at 2.2, 6 anomalies affecting 2 assistants at 3.2, and no anomalies at 5.0. These results demonstrate that threshold selection directly controls the volume and concentration of managerial follow-up.

Explicitly, Definition 2.3 accommodates robust estimators such as the Modified Z-score based on the median and MAD. Nevertheless, the empirical experiment was deliberately constrained to the classical Z-score to establish an initial baseline and to make threshold sensitivity transparent. This design choice should not be interpreted as evidence that classical Z-score is optimal for skewed lateness data. A complete follow-up study should calculate Modified Z-scores on the full row-level dataset and compare anomaly counts, affected assistants, and case-level overlap with the classical method. This extension is especially relevant because explainability and transparency are important when anomaly analytics are used to support human-resource decisions.

DAFTAR PUSTAKA

- [1] A. Blázquez-García, A. Conde, U. Mori, dan J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–33, 2021, doi: 10.1145/3444690.
- [2] A. A. Cook, G. Misirli, dan Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2020, doi: 10.1109/JIOT.2019.2958185.
- [3] L. Correia, J. C. Goos, P. Klein, T. Bäck, dan A. V. Kononova, "Online model-based anomaly detection in multivariate time series: Taxonomy, survey, research challenges and future directions," *Engineering Applications of Artificial Intelligence*, vol. 138, Art. no. 109323, 2024, doi: 10.1016/j.engappai.2024.109323.
- [4] Z. Z. Darban, G. I. Webb, S. Pan, C. C. Aggarwal, dan M. Salehi, "Deep learning for time series anomaly detection: A survey," *ACM Computing Surveys*, vol. 56, no. 12, pp. 1–42, 2024, doi: 10.1145/3691338.
- [5] B. Iglewicz dan D. C. Hoaglin, *How to Detect and Handle Outliers*. Milwaukee, WI, USA: ASQC Quality Press, 1993.
- [6] S. Kumar et al., "GPS based attendance system using geofencing," *International Journal of Scientific Research in Science and Technology*, vol. 10, no. 3, pp. 204–211, 2023.
- [7] S. Kumari et al., "A comprehensive investigation of anomaly detection methods in deep learning and machine learning: 2019–2023," *IET Information Security*, Art. no. 8821891, 2024, doi: 10.1049/2024/8821891.
- [8] C. Leys, M. Delacre, Y. L. Mora, D. Lakens, dan C. Ley, "How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration," *International Review of Social Psychology*, vol. 32, no. 1, 2019, doi: 10.5334/irsp.289.
- [9] A. Pawar, R. Hiwanj, P. Koparde, dan S. Barve, "Automated employee attendance monitoring using liveness face recognition and geofencing in real time," in *Proc. Int. Conf. Communication, Computing and Internet of Things (IC3IoT)*, 2023, pp. 1–6.
- [10] P. Patil, S. Ringe, P. Gaikwad, P. Mahajan, dan S. Mogare, "An Android application for attendance using geofencing," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 11, pp. 1505–1510, 2023, doi: 10.22214/ijraset.2023.56620.
- [11] D. M. Ravid, D. L. Tomczak, J. C. White, dan T. S. Behrend, "The impact of electronic monitoring on employees' job satisfaction, stress, performance, and counterproductive work behavior: A meta-

- analysis,” *Computers in Human Behavior*, vol. 127, Art. no. 107023, 2022, doi: 10.1016/j.chb.2021.107023.
- [12] D. M. Ravid dan D. Jeske, “Electronic monitoring at work,” *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 12, 2025, doi: 10.1146/annurev-orgpsych-110622-060758.
- [13] P. J. Rousseeuw dan C. Croux, “Alternatives to the median absolute deviation,” *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.
- [14] K. R. Shahapure dan C. Nicholas, “Cluster quality analysis using silhouette score,” in *Proc. IEEE 7th Int. Conf. Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748, doi: 10.1109/DSAA49011.2020.00096.
- [15] A. S. Yaro, F. Maly, dan P. Prazak, “Outlier detection in time-series receive signal strength observation using Z-score method with Sn scale estimator for indoor localization,” *Applied Sciences*, vol. 13, no. 6, Art. no. 3900, 2023, doi: 10.3390/app13063900.
- [16] A. S. Yaro, F. Maly, dan P. Prazak, “Outlier detection performance of a modified Z-score method in time-series RSS observation with hybrid scale estimators,” *IEEE Access*, vol. 12, pp. 12345–12358, 2024, doi: 10.1109/ACCESS.2024.3356731.
- [17] J. Gomez, M. Rahman, dan P. Singh, “Anomaly detection in time-series employee absence data,” in *Proc. IEEE Int. Conf. Big Data (BigData)*, 2023, pp. 1–8.
- [18] R. Khan dan S. Patel, “Geolocation and geofencing-based attendance systems for workforce management,” *Journal of Workforce Analytics*, vol. 7, no. 2, pp. 45–60, 2025.
- [19] F. Osei, K. Mensah, dan D. Boateng, “A multi-method anomaly detection framework for identifying potential ghost employees in payroll data,” *International Journal of Research in Engineering and Science*, vol. 13, no. 4, pp. 1–12, 2025.
- [20] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, “A Survey of Methods for Explaining Black Box Models,” *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2018, doi: 10.1145/3236009.
- [21] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.
- [22] P. Tambe, P. Cappelli, and V. Yakubovich, “Artificial Intelligence in Human Resources Management: Challenges and a Path Forward,” *California Management Review*, vol. 61, no. 4, pp. 15–42, 2019, doi: 10.1177/0008125619867910.
- [23] Z. Li, Y. Zhu, and M. van Leeuwen, “A Survey on Explainable Anomaly Detection,” *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 1, Art. no. 23, pp. 1–54, 2024, doi: 10.1145/3609333.
- [24] L. Antwarg Friedman, C. Galed, L. Rokach, and B. Shapira, “Evaluating Anomaly Explanations Using Ground Truth,” *AI*, vol. 5, no. 4, pp. 2375–2392, 2024, doi: 10.3390/ai5040117.
- [25] A. Heidemann, S. M. Hülter, and M. Tekieli, “Machine learning with real-world HR data: mitigating the trade-off between predictive performance and transparency,” *The International Journal of Human Resource Management*, vol. 35, no. 14, pp. 2343–2366, 2024, doi: 10.1080/09585192.2024.2335515.
- [26] S. Dadaboyev, S. Hojimorekova, and G. Gulomova, “AI in HR Analytics: A Synthesis of Current Research and Future Directions,” *Proceedings of the AAAI Symposium Series*, vol. 6, no. 1, p. 60, 2025, doi: 10.1609/aaais.v6i1.36026.