

Article

An Integrated Structural Equation Modelling and Machine Learning Framework for Measurement Scale Evaluation—Application to Voluntary Turnover Intentions

Marcin Nowak ^{1,*}  and Robert Zajkowski ²¹ Faculty of Engineering Management, Poznan University of Technology, 60-965 Poznan, Poland² Department of Banking and Financial Markets, Faculty of Economics, Maria Curie-Skłodowska University in Lublin, 20-031 Lublin, Poland; robert.zajkowski@mail.umcs.pl

* Correspondence: marcin.nowak@put.poznan.pl

Abstract

There is an increasing demand for robust methodologies to rigorously evaluate the psychometric properties of measurement scales used in quantitative research across various scientific disciplines. This article proposes an integrative method that combines structural equation modelling (SEM) with machine learning (ML) to jointly assess model fit and predictive accuracy, limitations often addressed separately in traditional approaches. Using a measurement scale for voluntary employee turnover intention, the method demonstrates clear improvements: RMSEA decreased from 0.073 to 0.065, and classifier accuracy slightly increased from 0.862 to 0.863 after removing three redundant items. Compared to standalone SEM or ML, the integrated framework yields a shorter, better-fitting scale without compromising predictive power. For practitioners, this method enables the creation of more efficient, theoretically grounded, and predictive tools, facilitating faster and more accurate assessments in organisational settings. To this end, this study employs Covariance-Based SEM (CB-SEM) in conjunction with classifiers such as naive Bayes, linear and nonlinear support vector machines, decision trees, k-nearest neighbours, and logistic regression.

Keywords: machine learning; structural equation models; voluntary employee turnover intentions



Received: 7 July 2025

Revised: 31 July 2025

Accepted: 11 August 2025

Published: 13 August 2025

Citation: Nowak, M.; Zajkowski, R. An Integrated Structural Equation Modelling and Machine Learning Framework for Measurement Scale Evaluation—Application to Voluntary Turnover Intentions. *AppliedMath* **2025**, *5*, 105. <https://doi.org/10.3390/appliedmath5030105>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Contemporary research in the field of work psychology and organisational management increasingly emphasises the significant role of accurate and reliable measurement of psychometric variables, such as voluntary turnover intentions. Measurement scales of this kind play a crucial role not only in modelling the mechanisms of organisational behaviour but also in predicting personnel phenomena that directly impact the functioning of enterprises [1–3]. Due to the substantial costs associated with employee turnover, developing tools that allow for its early detection and the explanation of predictive factors remains a problem of high applied value [4]. Despite the availability of various measurement scales, many of them are tested without simultaneously considering the quality of structural model fit and their predictive effectiveness, which limits their usefulness in practical applications.

While numerous studies have utilised either structural equation modelling (SEM) or machine learning (ML) methods to assess psychometric instruments, these approaches are typically applied in isolation, which limits their capacity to address theoretical model

fit and predictive accuracy simultaneously. Traditional SEM procedures often emphasise model fit indices such as RMSEA or CFI but do not evaluate how individual items contribute to out-of-sample prediction performance [5,6]. Conversely, ML models are optimised for classification or regression accuracy but lack theoretical grounding in latent construct measurement [7]. This methodological separation creates a significant gap: current psychometric validation frameworks fail to integrate construct validity with predictive utility in a unified approach. From the perspective of psychometric theory, the validation of measurement instruments relies on establishing both construct validity and criterion-related validity. SEM has traditionally been used to assess the internal structure of scales (e.g., dimensionality, factor loadings), reflecting construct validity, while ML techniques contribute primarily to assessing external validity through predictive performance. The proposed integration aligns with the contemporary understanding of validity as a unitary but multifaceted construct, where internal and external aspects should be addressed simultaneously. By incorporating item-level SEM diagnostics with predictive accuracy metrics from ML, the method operationalises a psychometrically coherent framework that honours both theoretical model specification and empirical utility [8,9].

Recent studies have highlighted the potential of combining SEM and ML, but no standardised or replicable methodology has yet emerged for doing so in scale refinement [10,11]. Addressing this gap, the present study proposes an integrative SEM-ML framework for psychometric scale evaluation that accounts for theoretical validity and predictive effectiveness.

The integration of structural equation modelling (SEM) and machine learning (ML) remains underexplored despite their complementary strengths. SEM provides a robust framework for confirming theoretical constructs and quantifying latent relationships, while ML offers strong predictive capabilities based on complex patterns in data. However, the lack of methodological integration means that researchers must often choose between theoretical validation (SEM) and predictive performance (ML), thus missing opportunities to leverage both. The proposed method bridges this divide by creating a unified analytic pipeline where SEM identifies valid construct structures and ML tests their utility in real-world predictions. This dual functionality is particularly valuable in applied fields like organisational research, where both theoretical rigour and predictive accuracy are essential [12].

This methodological gap defines the aim of the present article, which is to develop an integrated method for evaluating psychometric scales that combines theoretical validation with an assessment of predictive effectiveness. The approach proposed in this article integrates structural equation modelling (SEM) with machine learning (ML), allowing for simultaneous analysis of the scale's fit to the theoretical concept and its utility in case classification. To achieve the stated goal, the covariance-based SEM method was employed (with maximum likelihood as the parameter estimation method), alongside the following machine learning algorithms: naive Bayes, linear and nonlinear support vector machines, decision trees, k-nearest neighbours, and logistic regression.

Such integration places this study at the core of applied mathematics, as it merges optimisation techniques, parameter estimation, and algorithmic learning to solve real-world empirical problems in an organisational context [5,6,13]. The theoretical contribution of this study lies in extending the psychometric validation framework through a mathematically formalised integration of SEM and ML. By demonstrating how predictive metrics and structural model diagnostics can be used in tandem for item selection and scale refinement, this study offers a novel methodological model for balancing construct coherence with empirical performance. This approach challenges the conventional dichotomy between theory-driven and data-driven methods in measurement development, proposing instead

a unified paradigm that can be generalised to various domains where psychometric scale evaluation is required. Thus, the article contributes to the growing interest in using applied mathematics tools in analysing social and psychometric data, offering a novel approach to constructing and testing research instruments. To illustrate the unified pipeline, Figure 1 presents a block diagram of the proposed SEM–ML integration method.

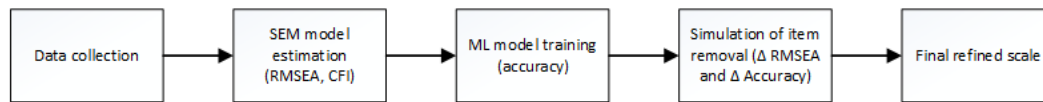


Figure 1. Block diagram of the integrative SEM–ML workflow.

This flowchart summarises the six main stages of our method. First, raw survey data are collected and preprocessed. Next, a covariance-based SEM is fitted to obtain traditional fit indices (e.g., RMSEA, CFI), while, in parallel, an ML classifier is trained to measure predictive performance (accuracy). In the third stage, each item is successively removed, and both Δ RMSEA and Δ Accuracy are computed in simulation runs. The fourth step applies a joint optimisation criterion—retaining only those items whose removal does not substantially worsen fit or prediction, thus balancing theoretical coherence with empirical utility. Finally, the selected subset of items yields a more parsimonious scale, optimised simultaneously for model structure and predictive goals.

The Literature Review section presents the essence of the issues related to using measurement scales for voluntary turnover intentions and explains the core and mathematical formalisation of structural equation modelling (SEM) and machine learning. The Methodology section outlines the procedure of the proposed method for evaluating measurement scales, integrating structural equation modelling with machine learning. The Results chapter presents the implementation of the method using the example of evaluating a measurement scale for employee voluntary turnover intentions.

2. Literature Review

2.1. Measurement Scales for Employee Voluntary Turnover Intentions

Turnover intention refers to the likelihood or propensity of an employee to exit their current organisational affiliation voluntarily [14]. This construct is typically operationalised through temporal measurement frameworks within empirical research, capturing the individual's deliberative process regarding organisational departure [15]. Prior studies have demonstrated a significant positive association between turnover intentions and actual voluntary turnover behaviour, underscoring the predictive validity of the construct [16].

Voluntary turnover intention is one of organisational behaviour research's most frequently analysed variables. The literature indicates that turnover intentions are a reliable predictor of actual employee departures [17]. A key issue in this area is the selection of appropriate measurement tools, namely, scales for assessing turnover intentions and related psychological and organisational variables. One of the most commonly used instruments is the three-item scale developed by Mobley and colleagues [14], which includes questions about thoughts of leaving, intentions to search for a new job, and the likelihood of leaving in the near future—this scale has demonstrated good validity and reliability [18].

Subsequent research has introduced extended and multidimensional scales for measuring voluntary turnover intentions, for example, the following:

- Maertz and Campion [19] distinguish eight dimensions of turnover (e.g., avoidance, calculative);
- Tett and Meyer [20] propose separating the measurement of intentions from the emotional reasons for leaving;

- Lee et al. [21] develop a “push-pull” scale assessed using 5-point Likert scales;
- Bothma and Roodt [22] confirm the factorial validity as well as the reliability of the TIS-6 scale;
- Ike et al. [23] proposed and evaluated twenty-five items with a five-factor scale of turnover intention.

In these proposed scales, turnover intentions are strongly associated with factors such as job satisfaction [24], organisational commitment [25], and stress and burnout [26]. Schaufeli and colleagues [20] point out that indicators such as voluntary turnover intention are conceptualised as latent changes in SEM models or aggregated into composite scales. Table 1 provides a comparative overview of key psychometric instruments used to measure voluntary turnover intentions, detailing their length, dimensional structure, scale format, and validation evidence.

Table 1. Comparative overview of psychometric scales for voluntary turnover intentions.

Scale (Author, Year)	Number of Items	Dimensionality	Scale Type	Validated Constructs	Notes
Mobley et al. [14]	3	Unidimensional	5-point Likert	Reliability, Predictive Validity	Most widely used baseline scale; concise but limited in scope. Separates cognitive intention from emotional antecedents.
Tett & Meyer [20]	4–6 (varies)	Multidimensional	5-point Likert	Affective and Behavioral Intent	Explores motivational paths to quitting; rich theoretical basis. Context-sensitive scale; validated in longitudinal studies.
Maertz & Campion [19]	~24 (8 × 3)	Eight dimensions	5-point Likert	Avoidance, Calculative, Normative Job Embeddedness, Intent to Leave Construct	Shortened validated scale; good internal consistency ($\alpha > 0.8$). Recent comprehensive scale with psychometric robustness.
Lee et al. [21]	~12–16	Two dimensions (Push vs. Pull)	5-point Likert	Validity, Reliability	
Bothma & Roodt [22]—TIS-6	6	Unidimensional	5-point Likert	Factorial Validity, Model Fit	
Ike et al. [23]	25	Five dimensions	5-point Likert		

As shown in Table 1, shorter unidimensional instruments—such as Mobley et al.’s three-item scale—offer parsimony and ease of administration but may lack the breadth to capture multifaceted turnover drivers. In contrast, extensive multidimensional scales (e.g., Maertz & Campion’s eight-factor model or Ike et al.’s five-factor inventory) deliver richer diagnostic insight at the cost of increased respondent burden. The choice of scale should, therefore, balance theoretical comprehensiveness, empirical robustness (factorial validity, reliability), and practical considerations related to survey length and predictive utility.

However, an increasing number of contemporary studies are linking psychometric scale development with the construction of machine learning models. In measurement scales used as datasets for machine learning, various variables—most commonly rated on a 5-point Likert scale—are collected. Predictive analyses of this type frequently employ algorithms such as logistic regression, support vector machines, and decision trees [27].

Despite the availability of numerous measurement instruments, existing validation approaches often present methodological limitations. Traditional psychometric validation focuses heavily on internal consistency and factorial validity, usually confirmed via con-

firmatory factor analysis or structural equation modelling. However, these approaches frequently neglect the external, predictive utility of the instruments, particularly their performance in real-world classification or decision-making contexts. Moreover, item retention decisions are commonly based solely on model fit indices (e.g., RMSEA, CFI), which may inadvertently compromise the predictive capacity of the scale. Conversely, machine learning-based validations typically prioritise accuracy but disregard the theoretical coherence of the construct, leading to a lack of interpretability or construct-level insight. This bifurcation between theory-driven and data-driven validation creates a gap in psychometric practice, where neither approach alone ensures both conceptual soundness and practical effectiveness. The need to address this dual objective motivates the integrated SEM-ML framework proposed in this study [28,29].

2.2. Structural Equation Modelling

Structural equation modelling (SEM) is an advanced statistical method for analysing relationships between observed and latent variables. SEM combines features of factor analysis and regression modelling, allowing for the testing of complex theoretical models through the use of matrix equations [5]. An SEM model consists of two main components:

1. The measurement model, which describes the relationship between latent variables and observed variables according to Formula (1):

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \delta \mathbf{y} = \Lambda_y \boldsymbol{\eta} + \epsilon \quad (1)$$

where:

- \mathbf{x}, \mathbf{y} —vectors of observed variables;
 - $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ —exogenous and endogenous latent variables;
 - Λ_x, Λ_y —factor loading matrices;
 - δ, ϵ —measurement errors.
2. The structural model, which describes the relationships between latent variables, is as follows (Formula (2)):

$$\boldsymbol{\eta} = B\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\varsigma} \quad (2)$$

where:

- B —matrix of regression coefficients among endogenous variables;
- Γ —matrix of regression coefficients from exogenous to endogenous variables;
- $\boldsymbol{\varsigma}$ —vector of structural errors.

The most commonly used method for parameter estimation in SEM is the Maximum Likelihood (ML) method, which involves minimising function (3).

$$F_{ML} = \ln |\Sigma(\theta)| - \ln |S| + \text{tr}(\Sigma^{-1}) - p \quad (3)$$

where:

- $\Sigma(\theta)$ —model-implied covariance matrix;
- S —observed covariance matrix;
- p —number of observed variables.

Alternative estimation methods include Generalised Least Squares (GLS), Unweighted Least Squares (ULS), and Bayesian SEM [13]. The fit of an SEM model to the data is assessed using multiple indices, such as those presented in Table 2 [30,31].

Table 2. SEM model fit indices.

Index	Formula or Description	Interpretation
Chi-square(χ^2)	$\chi^2 = (N - 1)F_{ML}$	High p value (>0.05) indicates good fit
RMSEA (Root Mean Square Error of Approximation)	$RMSEA = \sqrt{\frac{\max(\chi^2 - df, 0)}{df \cdot (N - 1)}}$	<0.05 —very good fit 0.05 – 0.08 —moderate fit
CFI (Comparative Fit Index)	$CFI = 1 - \frac{\max(\chi^2 - df, 0)}{\max(\chi^2_{baseline} - df_{baseline}, 0)}$	>0.95 —very good fit
TLI (Tucker–Lewis Index)	Takes model complexity into account (penalises complex models)	>0.95 —very good fit
GFI (Goodness-of-Fit Index)	Proportion of explained variance	>0.90 —very good fit
AGFI (Adjusted GFI)	GFI adjusted for degrees of freedom	>0.90 —very good fit
AIC/BIC	Information criteria (relative measures)	Lower AIC/BIC \rightarrow better model (only for model comparisons)

Source: [6].

The main advantages of SEM include the ability to model latent variables while accounting for measurement error, testing complex theoretical hypotheses, and assessing both direct and indirect effects. The most commonly cited limitations of the method are its high sample size requirements (recommended $N > 200$), sensitivity to deviations from data normality, and the possibility of fitting a model with low theoretical validity [32].

2.3. Machine Learning

Machine learning (ML) offers a range of algorithms for classification and regression that allow for modelling relationships in data without the need to specify their functional form strictly. The present article employs several key machine learning algorithms, including naive Bayes, linear and nonlinear support vector machines, decision trees, k-nearest neighbours, and logistic regression.

The first algorithm analysed is the naive Bayes classifier. This model is based on Bayes' theorem and the assumption of conditional independence of features [33] (Formula (4)):

$$P(C_k|x) = \frac{P(C_k)\prod_{i=1}^n P(x_i|C_k)}{P(x)} \quad (4)$$

where:

Gdzie:

$P(C_k|x)$ —probability of belonging to class C_k ;

$P(C_k)$ —prior probability of class;

$P(x_i|C_k)$ —conditional probability of feature x_i given class C_k .

In the Gaussian classifier, a normal distribution of features is assumed (Formula (5)):

$$P(x_i | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad (5)$$

The next algorithms addressed in this study are linear and nonlinear support vector machines (SVMs). In the linear SVM model, for a dataset $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, the objective is to determine a decision function of the form $f(x) = w^T x + b$

that separates the classes while simultaneously solving the optimisation problem defined by the objective function (6) [34]:

$$\min_{w,b} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right) \quad (6)$$

Under the assumption that the following margin constraints are satisfied:

$$y_i (w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0; \quad \forall i$$

In the mathematical context, nonlinear SVM addresses the classification problem in its dual form by maximising the objective function (7):

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (7)$$

subject to the following constraints: $0 \leq \alpha_i \leq C$; $\sum_{i=1}^n \alpha_i y_i = 0$.

where $K(x_i, x_j)$ is a kernel function, e.g., RBF. Once the coefficients α_i are determined, the classification of a new observation x is based on function (8):

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \quad (8)$$

In contrast to the linear variant, which operates directly on the original features, nonlinear SVM uses a kernel function to transform the data space, allowing it to handle more complex patterns more effectively [35].

Another algorithm applied in this study was decision trees. Decision trees are constructed based on data splits that maximise information gain [36]. This occurs for the entropy function as follows (9):

$$H(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (9)$$

The information gain from splitting by an attribute is as follows (10):

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v) \quad (10)$$

where:

- p_i —frequency of class i ;
- S_v —subset of data with value v of attribute A .

The article also applied the k-nearest neighbours (k-NN) method. In this algorithm, for a given point, the closest training points are found (11) (e.g., using the Euclidean metric) [37]:

$$d(x, x_i) = \|x - x_i\|_2 = \sqrt{\sum_{j=1}^d (x_j - x_{ij})^2} \quad (11)$$

The decision is made through majority voting of the classes (12):

$$\hat{y} = \arg \max_y \sum_{i \in N_k(x)} 1(y_i = y) \quad (12)$$

where $1(\cdot)$ is an indicator function that takes the value 1 if the condition is met and 0 otherwise. The final algorithm applied in the article is logistic regression. This algorithm models the probability of belonging to class 1 using the function [38] (13):

$$P(y = 1|x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (13)$$

To fit the model to the data, the log-likelihood function is maximised, expressed as follows (14):

$$\max_{w,b} \sum_{i=1}^n [y_i \log P(y_i|x_i) + (1 - y_i) \log (1 - P(y_i|x_i))] \quad (14)$$

Optimisation is performed, for example, using the gradient descent method.

2.4. Integration of SEM Models and Machine Learning Methods

Recent methodological innovations aim to combine the explanatory power of structural equation modelling (SEM) with the predictive potential of machine learning [39]. One significant line of research focuses on enhancing SEM through regularisation and tree-based algorithms—referred to as regularised SEM and SEM trees—to prevent overfitting and manage high-dimensional sets of indicators [40]. Probabilistic SEM frameworks have begun to incorporate ensemble learners to capture complex, nonlinear interactions among latent variables, as exemplified by Super Learner Equation Modelling (SLEM), which integrates super learner algorithms with path analysis for robust causal inference [41]. Partial least squares SEM (PLS-SEM) is also routinely combined with classifiers—such as support vector machines and random forests—to optimise both measurement validity and predictive accuracy in domains such as marketing and supply chain management [42,43]. Bayesian variants of SEM, enriched with machine learning routines, have demonstrated increased estimation stability and predictive robustness, particularly in the context of small samples or complex models [44]. Early hybrid approaches automated the item-reduction process by concurrently evaluating multiple psychometric criteria and classification performance, paving the way for more efficient scale refinement [45].

In psychometric research, integrating SEM diagnostics with ML-based feature selection has led to scalable procedures for constructing concise, high-performing measurement instruments. Studies combining decision trees, support vector machines, and Naïve Bayes classifiers with confirmatory modelling have systematically evaluated the trade-off between construct validity (e.g., RMSEA, CFI) and predictive utility (accuracy, AUC), facilitating the elimination of redundant items [46]. Conceptual reviews advocate embedding ML feature importance metrics within latent variable frameworks to preserve interpretability while leveraging data-driven selection [47]. Furthermore, practitioners apply machine learning optimisation techniques in test development to enhance item quality and respondent engagement, underscoring the practical significance of SEM–ML integration in psychological assessment. A growing consensus across these approaches highlights the promising potential of integrated SEM–ML methodologies for replicable, parsimonious, and empirically robust scale evaluation [48].

3. Materials and Methods

Before performing the analyses, the dataset underwent a comprehensive data preparation process to ensure its suitability for both structural equation modelling (SEM) and machine learning (ML). All responses from the 27-item questionnaire were screened for missing values and outliers. Cases with incomplete or inconsistent responses were removed, resulting in a final sample size of 854. The data were assessed for normality, and, although Likert-type scales are ordinal, they were treated as continuous for SEM

purposes, as is standard practice with large samples. For SEM, the observed variables were standardised, and assumptions related to multivariate normality were examined to validate the use of maximum likelihood estimation. In the context of machine learning, the dataset was further preprocessed by normalising the features using min–max scaling to ensure comparability across items and improve model convergence. The binary target variable—voluntary turnover intention—was extracted and coded consistently for classification purposes. Stratified sampling techniques were applied during training–test splits to preserve class distribution. These preparation steps ensured the reliability of both theoretical modelling and predictive analytics.

The proposed method for evaluating measurement scales using structural equation modelling and machine learning can be presented as a five-step procedure.

Step 1. Development of a dataset based on the prepared measurement scale.

After the measurement scale is developed, a questionnaire study is conducted on a selected research sample. The respondents' answers are collected into a dataset. Considering the formal and substantive requirements of SEM methodology, the research sample should not be smaller than 200 participants.

Step 2. Construction of a structural model in which the latent variable is the selected psychometric construct.

In this step, an SEM model is developed consisting of two components:

1. Measurement model—this model tests whether all the scale's factors can be reduced to a single component (the examined psychometric construct).
2. Structural model—this model tests the regression relationship between the analysed factors and the label. In this case, the label is the dependent variable, and its predictors are the factors from the psychometric scale.

At this research stage, it is necessary to determine the key SEM model fit indices, especially χ^2 , RMSEA, CFI, and TLI. In the proposed method, it is standardly assumed that an acceptable model fit corresponds to an RMSEA value not exceeding 0.08. If the RMSEA value exceeds 0.08, it indicates that the psychometric scale is not suitable for measuring the selected psychometric construct. Although the developed method is primarily intended to enhance the performance of well-constructed psychometric scales, improving the SEM model to achieve the desired fit level is still possible even when the RMSEA slightly exceeds 0.08.

The application of machine learning tools in human resource management (HRM) has gained significant traction in recent years, particularly for tasks involving employee retention, talent acquisition, and performance prediction [49]. In the context of voluntary turnover intention, ML techniques are increasingly used to identify subtle patterns in employee survey data that may predict attrition risk. Studies have demonstrated the effectiveness of algorithms such as logistic regression, decision trees, and support vector machines in predicting turnover with high accuracy, often outperforming traditional statistical methods. These models offer the added advantage of handling complex, nonlinear relationships and large feature spaces commonly found in psychometric datasets. Consequently, integrating ML algorithms into the scale evaluation process not only enhances predictive accuracy but also aligns the research with modern HR analytics practices [50].

In the structural equation modelling (SEM) component, all variables from the questionnaire were treated as continuous and modelled as indicators of a single latent construct—voluntary turnover intention. Each item was measured on a 5-point Likert scale and treated as approximately continuous in line with common SEM practice [12]. The latent construct was modelled using a reflective approach, with each observed item serving as an indicator influenced by the underlying psychological factor. The dependent variable (label) used in the structural model was binary, coded as 0 (no intention to leave) and 1 (intention to

leave), based on a self-reported item in the survey. Model estimation was conducted using the Maximum Likelihood (ML) method, which assumes multivariate normality and is suitable for continuous observed variables. This method optimises the likelihood function to estimate model parameters that best reproduce the observed covariance matrix. Given the relatively large sample size ($N = 854$), the use of ML estimation was justified despite the presence of ordinal data, as ML is considered robust under such conditions when sample sizes are sufficient. All standard fit indices (e.g., RMSEA, CFI, TLI) were computed based on the ML estimates.

Step 3. Selection of the best machine learning algorithm for predicting the selected psychometric construct.

In this step of the method, a machine learning process is conducted on the dataset using the following algorithms: naive Bayes, linear and nonlinear support vector machines, decision trees, k-nearest neighbours, and logistic regression. To avoid the issue of “lucky sampling”, each algorithm is evaluated using cross-validation and repeated random splits of the data into training and test sets. For each algorithm, the average value of the prediction quality metric accuracy is calculated across all learning processes, along with the standard deviation of this metric. The algorithm with the highest average accuracy is then selected for further analysis.

The selection of machine learning algorithms in this study was based on their established utility in classification tasks within the domain of human resources and behavioural prediction. Naive Bayes, despite its simplifying assumption of feature independence, offers robustness and interpretability, especially when dealing with categorical or Likert-type inputs [51]. Support Vector Machines (SVMs), both linear and nonlinear, are known for their strong generalisation capabilities and effectiveness in handling high-dimensional data, which is often characteristic of psychometric scales [52]. Decision trees provide a transparent decision-making process that is particularly useful in applied HR contexts, albeit sometimes at the cost of overfitting. The k-nearest neighbours algorithm, although sensitive to feature scaling, is valuable in identifying local structure in datasets with ambiguous class boundaries [53]. Finally, logistic regression remains a staple baseline model in predictive HR analytics due to its interpretability and statistical grounding [54]. Together, this ensemble of classifiers enables a robust comparison across a spectrum of model complexities and underlying assumptions.

Step 4. Simulation of the impact of removing factors on the SEM model and the effectiveness of the machine learning model.

In this step, SEM model fit simulations are conducted by iteratively removing items from the scale. If the scale consists of n items, n SEM simulations are performed. The difference between the initial RMSEA (with no items removed) and the RMSEA after item removal is computed for each simulation.

Analogous simulations are carried out for the best-performing machine learning model (as selected in Step 3). That is, items are sequentially removed from the machine learning model, and the average accuracy metric is calculated after each removal. Differences between accuracy before and after elimination are also determined.

In the structural equation modelling literature, it is common practice to eliminate items solely based on improvements in fit indices (e.g., RMSEA, CFI, or TLI) [55,56]. Although such a procedure may lead to a less complex model and a formally better fit, lowering the RMSEA alone does not guarantee the maintenance or improvement of the scale’s predictive capacity. In practice, removing even a single item may reduce the measurement tool’s validity in terms of classification or forecasting, thereby limiting its practical utility.

Therefore, the proposed method balances the SEM fit criterion with an assessment of each variable’s contribution to the effectiveness of the machine learning model. Machine

learning enables the quantification of the impact of removing a particular item on prediction quality (measured, for example, by accuracy), allowing for the selection of variables whose elimination does not deteriorate—and in the best case even improves—both SEM fit and classification quality. As a result, the scale achieves an optimal compromise: it retains theoretical construct coherence (good fit indices) while preserving the tool's real predictive power.

The decision thresholds of $\Delta\text{RMSEA} \geq 0$ and $\Delta\text{Accuracy} \leq 0$ were adopted to ensure a balanced trade-off between theoretical model fit and predictive utility. A non-negative change in RMSEA ($\Delta\text{RMSEA} \geq 0$) indicates that removing an item does not worsen the structural model's approximation error and may even improve overall fit. This aligns with the goal of refining the scale without compromising construct validity. Similarly, a non-positive change in prediction accuracy ($\Delta\text{Accuracy} \leq 0$) ensures that the classification performance of the ML model is not degraded by the removal of an item. These thresholds were intentionally conservative to avoid overfitting and to maintain both psychometric rigour and applied classification capability. Their combined use allows for identifying items whose exclusion simultaneously preserves or improves both aspects of scale quality.

Step 5. Refinement of the Psychometric Scale Based on SEM-ML Simulations.

Following the simulations conducted in Step 4, variables are identified whose removal improves one of the two components—SEM model fit or machine learning prediction quality—without simultaneously worsening the other. According to the proposed method, such variables should be excluded from the psychometric scale. This results in at least a non-deteriorated SEM model fit and no decrease in the predictive quality of the selected psychometric construct, with the additional benefit of a shortened scale.

In the most favourable scenario, beyond reducing the number of items in the measurement scale (which is a significant benefit in itself), both the SEM model fit and the predictive accuracy of the psychometric construct using machine learning are improved.

4. Results

Step 1. Development of a dataset based on the prepared measurement scale.

Table 3 presents a custom-developed measurement scale regarding the occurrence of employee voluntary turnover intentions (after the whitening process of grey numbers).

Additionally, for machine learning purposes in particular, the survey questionnaire included a question asking whether the respondent demonstrates an intention to leave their job voluntarily. The survey was conducted between 1 August and 30 September 2024. The sample included in the present study comprised 854 individuals.

Step 2. Construction of a structural model in which the latent variable is the occurrence of employee voluntary turnover intention

The developed SEM model consisted of two components:

- Measurement model—the latent factor is voluntary turnover intention, onto which all 27 items are loaded. This model tests whether all 27 items can be reduced to a single component (turnover intention);
- Structural model—this model tests the regression relationship between the 27 items and the label, which is the occurrence of turnover intention. The label is the dependent variable in this model, and all 27 items are predictors.

The key parameters of the measurement and structural models are presented in Table 4.

Table 3. Custom scale for measuring employee voluntary turnover intentions.

Variable	Name	Scale for the Variable				
		Attribute Evaluation				
		Very Poor	Poor	Average	Good	Very Good
x ₁	salary	1	2	3	4	5
x ₂	job satisfaction	1	2	3	4	5
x ₃	sense of fairness	1	2	3	4	5
x ₄	promotion opportunities	1	2	3	4	5
x ₅	professional development opportunities	1	2	3	4	5
x ₆	work performance	1	2	3	4	5
x ₇	working conditions	1	2	3	4	5
x ₈	team atmosphere	1	2	3	4	5
x ₉	recognition and rewards	1	2	3	4	5
x ₁₀	relationships with supervisors	1	2	3	4	5
x ₁₁	job stability	1	2	3	4	5
x ₁₂	communication within the company	1	2	3	4	5
x ₁₃	work–life balance	1	2	3	4	5
x ₁₄	independence at work	1	2	3	4	5
x ₁₅	level of autonomy at work	1	2	3	4	5
x ₁₆	job responsibility	1	2	3	4	5
x ₁₇	work engagement	1	2	3	4	5
x ₁₈	remote work availability	1	2	3	4	5
x ₁₉	flexible working hours	1	2	3	4	5
x ₂₀	sense of burnout	1	2	3	4	5
x ₂₁	workload	1	2	3	4	5
x ₂₂	commuting time	1	2	3	4	5
x ₂₃	recognition at work	1	2	3	4	5
x ₂₄	organisational management	1	2	3	4	5
x ₂₅	job monotony	1	2	3	4	5
x ₂₆	employer reputation	1	2	3	4	5
x ₂₇	organisational culture	1	2	3	4	5

Table 4. Key parameters of the measurement and structural models.

Indicator	Estimate	Std. Err	z-Value	p-Value
x ₁ ~all scale items	1.000	—	—	—
x ₂ ~all scale items	1.222	0.0538	22.702	<0.001
x ₃ ~all scale items	1.224	0.0535	22.875	<0.001
x ₄ ~all scale items	1.110	0.0556	19.965	<0.001
x ₅ ~all scale items	1.145	0.0536	21.356	<0.001
...
x ₂₇ ~all scale items	1.042	0.0484	21.498	<0.001
Path/Variance	Estimate	Std. Err	z-Value	p-Value
label~all scale items	−0.332	0.0199	−16.671	<0.001
all scale items~~all scale items (variance)	0.497	0.0429	11.582	<0.001
label~~label (variance)	0.104	0.00514	20.233	<0.001

In the measurement model, all loadings are statistically significant ($p < 0.001$) and generally high (>0.8), which confirms that each indicator effectively reflects the latent construct. In the structural model, we examine the influence of this construct on the label (turnover intention). The negative coefficient (estimate = -0.332 , $p < 0.001$) indicates that a higher level of the latent construct is associated with a lower probability of turnover intention. Both variances are significant, suggesting meaningful variability in both the construct and the intention to leave. The SEM model fit indices are presented in Table 5.

Table 5. SEM model fit indices for the occurrence of employee voluntary turnover intentions.

Indicator	Value
Chi ²	1961.138
df	350
p-value	0
CFI	0.878
TLI	0.868
RMSEA	0.073
GFI	0.856
AGFI	0.844
NFI	0.856
AIC	107.4
BIC	373.4

The overall model fit can be considered good despite the statistically significant Chi² test ($p < 0.001$)—a typical result for large samples. The key RMSEA index of 0.073 falls below the 0.08 threshold, indicating an acceptable approximation error. The CFI = 0.878 and TLI = 0.868, though slightly below the conventional 0.90 cutoff, still suggest satisfactory model fit. Additionally, GFI = 0.856, AGFI = 0.844, and NFI = 0.856 confirm that the model structure adequately reflects the data. The AIC and BIC values can be used for comparison with alternative models, but, in themselves, they raise no concerns.

Step 3. Selection of the Best Machine Learning Algorithm for Predicting the Occurrence of Voluntary Employee Turnover.

Following the methodology outlined in the previous section, a training process was carried out using the following algorithms: naive Bayes, linear and nonlinear support vector machines, decision trees, k-nearest neighbours, and logistic regression. Cross-validation was used in the analysis. Table 6 presents the training process results for all algorithms, along with the standard deviations of the accuracy metric.

Table 6. Performance of applied machine learning algorithms in predicting voluntary employee turnover intentions.

Algorithm	Accuracy (Mean)	Standard Deviation of Accuracy
RBF SVM	0.862	0.017
Logistic Regression	0.857	0.017
Linear SVM	0.854	0.015
Naive Bayes	0.835	0.033
K-Nearest Neighbor	0.834	0.029
Decision Tree	0.810	0.022

Based on the obtained results, it can be concluded that the analysed models perform well in predicting the occurrence of voluntary employee turnover intentions. Each of the analysed models demonstrates over 80% accuracy. For further research, the nonlinear support vector machine algorithm was the most effective of the analysed algorithms.

Step 4. Simulations of the impact of factor removal on the SEM model and on the effectiveness of the machine learning model.

At the beginning of this step, simulations of the SEM model were conducted by successively excluding individual items from the scale. The results of the SEM model fit, measured by changes in the RMSEA index following successive reductions, are presented in Table 7.

Figure 2 illustrates the Δ RMSEA values resulting from the iterative removal of each item, with x_4 yielding the largest improvement and x_9 and x_{18} producing smaller yet still favourable reductions.

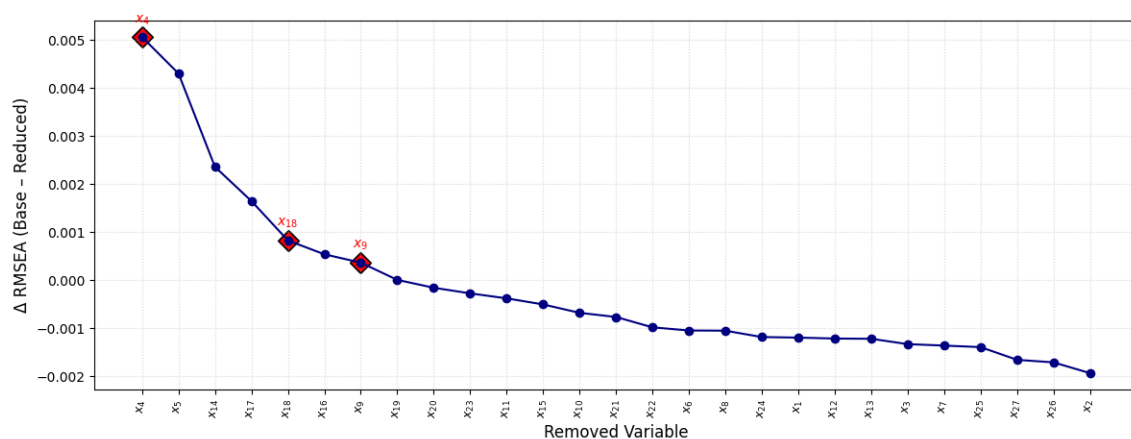


Figure 2. Change in RMSEA after removing each item.

Table 7. Change in RMSEA after removing subsequent variables from the measurement scale.

Removed	Δ RMSEA	Δ AIC	Δ BIC
x ₄	0.00506	−3.19403	−12.69389
x ₅	0.00430	−3.26161	−12.76147
x ₁₄	0.00236	−3.43795	−12.93782
x ₁₇	0.00165	−3.50375	−13.00361
x ₁₈	0.00083	−3.58040	−13.08026
x ₁₆	0.00054	−3.60751	−13.10737
x ₉	0.00037	−3.62397	−13.12383
x ₁₉	0.00001	−3.65796	−13.15783
x ₂₀	−0.00016	−3.67355	−13.17341
x ₂₃	−0.00027	−3.68472	−13.18458
x ₁₁	−0.00038	−3.69464	−13.19450
x ₁₅	−0.00050	−3.70670	−13.20656
x ₁₀	−0.00068	−3.72347	−13.22333
x ₂₁	−0.00076	−3.73177	−13.23163
x ₂₂	−0.00098	−3.75251	−13.25237
x ₆	−0.00104	−3.75881	−13.25867
x ₈	−0.00105	−3.75924	−13.25910
x ₂₄	−0.00118	−3.77198	−13.27184
x ₁	−0.00119	−3.77312	−13.27298
x ₁₂	−0.00121	−3.77500	−13.27487
x ₁₃	−0.00122	−3.77547	−13.27533
x ₃	−0.00133	−3.78626	−13.28612
x ₇	−0.00136	−3.78919	−13.28905
x ₂₅	−0.00139	−3.79226	−13.29212
x ₂₇	−0.00166	−3.81793	−13.31779
x ₂₆	−0.00171	−3.82308	−13.32295
x ₂	−0.00193	−3.84489	−13.34475

This visual confirms that excluding x₄, x₉, and x₁₈ consistently lowers or maintains RMSEA, reinforcing their selection for removal under the joint SEM–ML optimisation criteria. To determine whether the removal of specific items significantly improved model fit, chi-square difference tests (χ^2) were conducted and changes in the Comparative Fit Index (Δ CFI) were calculated. The removal of item x₄ produced the strongest effect: $\Delta\chi^2 = 344.15$ with $\Delta df = 26$ ($p < 0.001$) and $\Delta CFI = 0.017$, meeting both the statistical criterion ($p < 0.05$) and the practical threshold ($\Delta CFI \geq 0.01$). For items x₉ and x₁₈, the χ^2 tests showed minor decreases ($p > 0.05$) and ΔCFI values below 0.01, indicating that the model fit improvements were not statistically significant. Nonetheless, both items reduced RMSEA (Δ RMSEA = 0.00037 and 0.00083, respectively) and lowered the AIC/BIC information criteria, while their removal did not negatively affect ML classification performance. Therefore, considering the parallel criteria of model fit, parsimony, and preservation of predictive power, we recommend eliminating x₄, x₉, and x₁₈ as the optimal approach to scale simplification.

In the next step, the impact of removing successive variables on the accuracy metric of the best-performing model—the nonlinear support vector machine—was verified. The results of these simulations are presented in Table 8.

Table 8. Change in the accuracy metric after removing subsequent variables from the measurement scale.

Removed Variable	CV Accuracy	CV Accuracy Std	CV Accuracy Drop
x ₁₀	0.865	0.022	−0.004
x ₆	0.865	0.016	−0.004
x ₁₁	0.864	0.019	−0.002
x ₂₃	0.864	0.014	−0.002
x ₁₃	0.863	0.020	−0.001
x ₂₅	0.863	0.017	−0.001
x ₉	0.863	0.023	−0.001
x ₈	0.863	0.011	−0.001
x ₂₇	0.863	0.015	−0.001
x ₄	0.862	0.021	−0.000
x ₁₈	0.862	0.018	0.000
x ₂₁	0.862	0.022	0.000
x ₁₄	0.861	0.018	0.001
x ₂₄	0.861	0.017	0.001
x ₅	0.859	0.018	0.002
x ₁₉	0.859	0.015	0.002
x ₂	0.859	0.017	0.002
x ₇	0.858	0.023	0.004
x ₁	0.858	0.016	0.004
x ₁₅	0.858	0.018	0.004
x ₁₂	0.858	0.016	0.004
x ₂₀	0.858	0.015	0.004
x ₂₂	0.857	0.019	0.005
x ₃	0.857	0.020	0.005
x ₁₇	0.856	0.015	0.006
x ₂₆	0.856	0.015	0.006
x ₁₆	0.855	0.018	0.007

Figure 3 plots the difference in mean 5-fold cross-validation accuracy (Δ CV Accuracy) obtained by removing each item in turn, revealing that most values cluster tightly around zero.

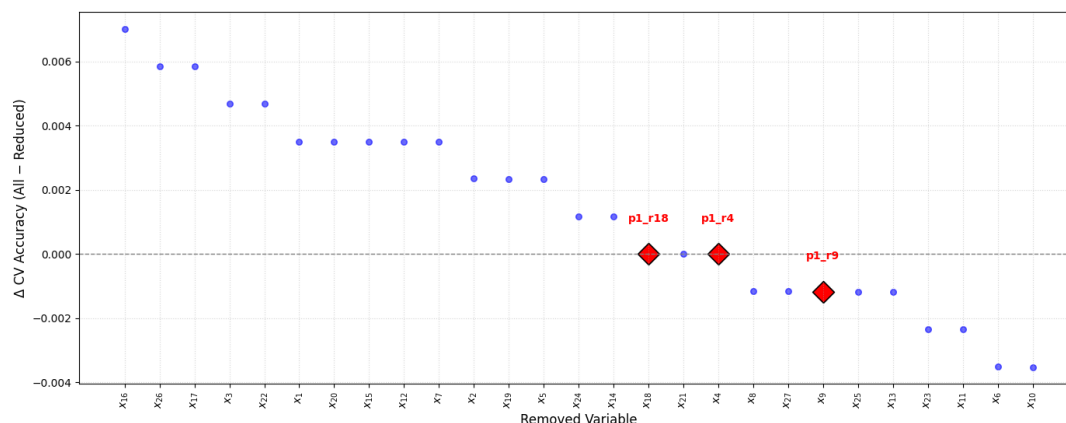


Figure 3. Change in 5-fold CV accuracy following item removal.

As the plot shows, the removal of x_4 , x_9 , and x_{18} produces negligible shifts in accuracy ($\Delta CV \approx 0$), visually confirming the paired t-test and bootstrap findings that the ML model’s predictive performance remains stable despite scale simplification.

To assess the stability of the ML model’s accuracy after item removal, two complementary techniques were applied: the paired t-test for comparing five 5-fold cross-validation results and the estimation of 95% confidence intervals using the bootstrap method (1000 replications). Both methods consistently demonstrated that the mean accuracy differences following the removal of x_4 , x_9 , and x_{18} were near zero, with p -values far from the significance threshold ($p \gg 0.05$). This form of statistical validation is particularly valued in ML research as it does not rely solely on point estimates of accuracy but also accounts for their variability and uncertainty. These findings indicate that the item selection process—while reducing the scale and simplifying the model—does not adversely affect its predictive performance. As a result, a more concise and parsimonious scale is achieved without compromising classification power, which provides strong justification for the practical application of the proposed method.

Step 5. Improvement of the psychometric scale based on the conducted SEM-ML simulations.

In the next step, those factors were identified whose potential removal neither worsens the fit of the SEM model (i.e., leads to a decrease in the RMSEA index or maintains it at the same level) nor reduces the predictive performance of the machine learning model (measured by the average accuracy value in the cross-validation method).

It was found that, out of the 27 analysed items in the scale measuring turnover intention, three indicators met the exclusion criteria. These factors are presented in Table 9.

Table 9. Factors to be removed from the measurement scale based on the SEM-ML method.

Removed Variable	Decrease in Average ML Model Accuracy (Prediction Quality Worsening) Caused by Its Presence in the Measurement Scale	Increase in Average RMSEA (Fit Error Worsening) in the SEM Model Caused by Its Presence in the Measurement Scale
X_9	0.00118	0.00037
X_4	0.00001	0.00506
X_{18}	0.00000	0.00083

Table 9 identifies three items— x_4 (promotion opportunities), x_9 (recognition and rewards), and x_{18} (remote work availability)—whose exclusion from the “voluntary turnover

intentions” scale does not deteriorate either the measurement validity assessed by SEM ($\Delta\text{RMSEA} \geq 0$) or the predictive power of the best ML model ($\Delta\text{accuracy} \leq 0$). Notably, the removal of x_9 and x_{18} even leads to a slight reduction in RMSEA without any loss of classification performance, while the removal of x_4 results in the most considerable improvement in model fit ($\Delta\text{RMSEA} = -0.00506$) with a negligible change in prediction accuracy (-0.00001).

These findings offer meaningful insight into the structure of the turnover intention construct. Item x_4 reflects perceptions of career advancement, which, in this sample, appear to play a minor role in influencing voluntary departure. Item x_9 captures employee recognition, which may overlap conceptually with other variables like job satisfaction. Item x_{18} represents access to remote work—a timely but possibly redundant factor in this organisational context, potentially subsumed by broader constructs like work-life balance. Their exclusion leads to a more concise and psychometrically sound scale without compromising theoretical coherence or practical utility.

In the final step, a SEM model was constructed, and the accuracy metric was calculated for the measurement scale after removing factors X_9 , X_4 , and X_{18} . Table 10 presents the fit indices of the new, simplified SEM model for voluntary employee turnover intention.

Table 10. Fit indices of the SEM model for voluntary employee turnover intention after removing the three variables X_9 , X_4 , and X_{18} .

Indicator	Value
Chi ²	1278.480
df	275
<i>p</i> -value	0
CFI	0.911
TLI	0.903
RMSEA	0.065
GFI	0.890
AGFI	0.880
NFI	0.890
AIC	97.0
BIC	334.5

The results presented in the table for the simplified SEM model (after removing X_9 , X_4 , and X_{18}) show a clear improvement in all key fit indices compared to the initial model. RMSEA decreased from 0.073 to 0.065, indicating a significant enhancement in model quality. At the same time, CFI increased from 0.878 to 0.911, and TLI from 0.868 to 0.903—both now exceed the commonly accepted threshold of 0.90, signalling a strong representation of the theoretical structure. GFI (0.856 \rightarrow 0.890), AGFI (0.844 \rightarrow 0.880), and NFI (0.856 \rightarrow 0.890) also improved by more than 0.03 points, confirming the overall better quality of the model. Lower values of the information criteria AIC (from 107.4 to 97.0) and BIC (from 373.4 to 334.5) indicate that a more economical model was obtained with fewer parameters, offering a better balance between parsimony and accuracy.

The new machine learning model (nonlinear support vector machine), without variables X_9 , X_4 , and X_{18} , achieved an average accuracy metric (calculated via cross-validation) of 0.8630 with a standard deviation of 0.017565.

The predictive performance of the selected ML classifier (nonlinear SVM) for the shortened scale also appears promising—the mean accuracy increased from 0.862 to 0.863,

and the standard deviation remained at a similar level. Although the accuracy gain is modest, it demonstrates that eliminating the three variables improved SEM fit without any loss, and even with a slight enhancement of the ML model's predictive power. These results confirm that the applied method for item selection achieves its intended trade-off: it yields a more concise and theoretically coherent measurement tool while maintaining (and even slightly improving) its practical utility in classifying turnover intentions.

The results of this study demonstrate that integrating structural equation modelling with machine learning allows for a more nuanced and evidence-based refinement of psychometric scales. The removal of three items from the original 27-item turnover intention scale resulted in improved model fit indices (e.g., RMSEA dropped from 0.073 to 0.065) and maintained or slightly enhanced classification accuracy (from 0.862 to 0.863). These findings highlight the dual benefit of the proposed approach: a scale that is both theoretically coherent and empirically predictive.

From a practical standpoint, the method supports researchers and HR professionals in identifying items that contribute marginally to construct measurement or predictive accuracy. This leads to more efficient and interpretable scales, reducing respondent burden and improving deployment in real-world contexts. In organisational settings, particularly in human capital management, the ability to reliably predict voluntary turnover intentions using a leaner and validated instrument has clear operational value, supporting proactive retention strategies and data-driven workforce planning.

Moreover, the approach offers a replicable framework that can be generalised to other constructs beyond turnover intention. By aligning theoretical psychometrics with algorithmic performance assessment, the method promotes a more integrated paradigm for instrument development, bridging the gap between conceptual modelling and applied analytics.

5. Conclusions

This article makes a significant contribution to the field of applied mathematics by extending the methodology of psychometric scale evaluation through an integrated approach that combines covariance-based SEM with machine learning algorithms. By proposing a general algorithm that simultaneously assesses the impact of individual indicators on model fit measures (RMSEA, CFI, TLI) and their role in classification performance (accuracy), this article sheds new light on the trade-offs between construct validity and the practical utility of measurement tools. Unlike traditional studies, where SEM optimisation and predictive validation are treated separately, the proposed procedure integrates SEM parameter estimation (via maximum likelihood) with variable selection processes in the context of classifiers such as nonlinear SVMs, logistic regression, and decision trees. This approach enriches the theoretical foundations of structural equation modelling with an algorithmic learning perspective and demonstrates how optimisation tools and simultaneous data analysis can be used to construct more concise and effective psychometric scale structures.

From a practical standpoint, the method provides researchers and HR professionals with a useful tool for optimising the length and validity of applied scales. It enables the identification of items whose exclusion leads to maintained or improved structural fit without degrading the predictive capacity of ML models, ultimately resulting in a shorter and more easily implementable questionnaire. In the context of human capital management, this allows for faster and more precise diagnosis of employee turnover intentions under limited research resources and reduces respondent burden. The case study on voluntary turnover intentions, conducted with a sample of over 850 individuals, demonstrates that removing three indicators from a 27-item scale is possible without any significant negative impact on construct validation or predictive accuracy. Crucially, the

proposed SEM-ML pipeline is not tied to a specific domain: by abstracting the item-removal and retraining steps into a generalisable algorithmic routine, it can be applied to any psychometric instrument—whether measuring consumer preferences, clinical symptoms, or educational outcomes. This flexibility underscores the method’s potential for broad adoption across different research areas and practical settings. As a result, the tool becomes more economical and adaptable, while organisations benefit from a scale better suited for the rapid identification of turnover risk.

At the same time, practitioners should be mindful of the inherent tension between theory-driven SEM and data-driven ML: the SEM component demands a clearly specified, interpretable latent structure, whereas the ML component optimises purely for predictive accuracy. Balancing these objectives requires careful judgment—avoiding overfitting in ML while preserving theoretical coherence in SEM—which remains a non-trivial challenge for future refinements of the framework.

Despite its clear advantages, the developed method has certain limitations. First, the application of covariance-based SEM assumes compliance with sample size and distribution normality requirements—conditions not always met in field research. Second, the ML analysis was limited to selected classifiers and the accuracy metric; alternative performance measures were not considered, which may affect optimal variable selection. Additionally, the procedure is based on cross-sectional data and cross-validation, which does not eliminate the potential for overfitting to a specific sample. Finally, this study pertains to one specific turnover intention scale—generalising the results to other psychometric tools or organisational cultures requires further verification.

The methodological development should progress in several directions. First, it would be valuable to explore the adaptation of the procedure in the context of PLS-SEM or Bayesian SEM, allowing application in complex models and smaller samples. Second, expanding the range of ML algorithms and evaluation metrics (including multiclass scenarios or continuous data) would enable a more comprehensive assessment of the predictive utility of scales. Moreover, longitudinal analysis using panel data could reveal how stable the selection recommendations are over time and what factors influence the variability of turnover intention intensity. Finally, applying the method to areas beyond human resource management—such as social psychology or consumer research—would allow verification of the universality and scalability of the proposed approach. Such a broadening of research horizons would contribute to the fuller integration of applied mathematics and machine learning techniques in the process of creating and validating measurement tools.

Author Contributions: Conceptualisation, M.N. and R.Z.; methodology, M.N. and R.Z.; software, M.N.; validation, R.Z. and M.N.; formal analysis, M.N.; investigation, M.N. and R.Z.; resources, M.N. and R.Z.; data curation, M.N.; writing—original draft preparation, M.N. and R.Z.; writing—review and editing, R.Z. and M.N.; visualisation, M.N. and R.Z.; supervision, M.N. and R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre, Poland 2024/08/X/HS4/00155.

Data Availability Statement: Marcin Nowak has unlimited and free-of-charge access to the dataset.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Harris, K.J.; James, M.; Boonthanom, R. Perceptions of Organizational Politics and Cooperation as Moderators of the Relationship between Job Strains and Intent to Turnover. *J. Manag. Issues* **2005**, *17*, 26–42.
2. Van Breukelen, W.; der Vlist, R.; Steensma, H. Voluntary Employee Turnover: Combining Variables from the ‘Traditional’ Turnover Literature with the Theory of Planned Behavior. *J. Organ. Behav. Int. J. Ind. Occup. Organ. Psychol. Behav.* **2004**, *25*, 893–914. [[CrossRef](#)]

3. Will, M.G. Voluntary Turnover: What We Measure and What It (Really) Means. Available online: <https://ssrn.com/abstract=2909718> (accessed on 10 June 2025).
4. Nowak, M. Prediction of Voluntary Employee Turnover Using Machine Learning. In *Scientific Papers of Silesian University of Technology; Seria Organizacja i Zarządzanie; Organization & Management/Zeszyty Naukowe Politechniki Śląskiej*: Zabrze, Poland, 2024; pp. 263–273, ISSN 1641-3466. [\[CrossRef\]](#)
5. Kline, R.B. *Principles and Practice of Structural Equation Modeling*; Guilford Publications: New York, NY, USA, 2023.
6. Schermelleh-Engel, K.; Moosbrugger, H.; Müller, H. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods Psychol. Res. Online* **2003**, *8*, 23–74.
7. Yarkoni, T.; Westfall, J. Choosing Prediction over Explanation in Psychology: Lessons from Machine Learning. *Perspect. Psychol. Sci.* **2017**, *12*, 1100–1122. [\[CrossRef\]](#)
8. Trognon, A.; Cherifi, Y.I.; Habibi, I.; Demange, L.; Prudent, C. Using Machine-Learning Strategies to Solve Psychometric Problems. *Sci. Rep.* **2022**, *12*, 18922. [\[CrossRef\]](#)
9. Orsoni, M.; Benassi, M.; Scutari, M. Information Theory, Machine Learning, and Bayesian Networks in the Analysis of Dichotomous and Likert Responses for Questionnaire Psychometric Validation. *Psychol. Methods* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Yu, B. Veridical Data Science. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 3–7 February 2020; pp. 4–5.
11. Hebart, M.N.; Baker, C.I. Deconstructing Multivariate Decoding for the Study of Brain Function. *Neuroimage* **2018**, *180*, 4–18. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Cano, F.; Pichardo, M.C.; Berbén, A.B.G.; Fernández-Cabezas, M. An Integrated Test of Multidimensionality, Convergent, Discriminant and Criterion Validity of the Course Experience Questionnaire: An Exploratory Structural Equation Modelling. *Assess. Eval. High. Educ.* **2021**, *46*, 256–268. [\[CrossRef\]](#)
13. Bollen, K.A. *Structural Equations with Latent Variables*; John Wiley & Sons: Hoboken, NJ, USA, 1989.
14. Mobley, W.H.; Horner, S.O.; Hollingsworth, A.T. An Evaluation of Precursors of Hospital Employee Turnover. *J. Appl. Psychol.* **1978**, *63*, 408. [\[CrossRef\]](#)
15. Wong, Y.; Wong, Y.-W.; Wong, C. An Integrative Model of Turnover Intention: Antecedents and Their Effects on Employee Performance in Chinese Joint Ventures. *J. Chin. Hum. Resour. Manag.* **2015**, *6*, 71–90. [\[CrossRef\]](#)
16. Hancock, J.I.; Allen, D.G.; Bosco, F.A.; McDaniel, K.R.; Pierce, C.A. Meta-Analytic Review of Employee Turnover as a Predictor of Firm Performance. *J. Manag.* **2013**, *39*, 573–603. [\[CrossRef\]](#)
17. Griffeth, R.W.; Hom, P.W.; Gaertner, S. A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the next Millennium. *J. Manag.* **2000**, *26*, 463–488. [\[CrossRef\]](#)
18. Hom, P.W.; Griffeth, R.W.; Sellaro, C.L. The Validity of Mobley’s (1977) Model of Employee Turnover. *Organ. Behav. Hum. Perform.* **1984**, *34*, 141–174. [\[CrossRef\]](#)
19. Maertz Jr, C.P.; Campion, M.A. Profiles in Quitting: Integrating Process and Content Turnover Theory. *Acad. Manag. J.* **2004**, *47*, 566–582. [\[CrossRef\]](#)
20. Tett, R.P.; Meyer, J.P. Job Satisfaction, Organizational Commitment, Turnover Intention, and Turnover: Path Analyses Based on Meta-Analytic Findings. *Pers. Psychol.* **1993**, *46*, 259–293. [\[CrossRef\]](#)
21. Lee, T.H.; Gerhart, B.; Weller, I.; Trevor, C.O. Understanding Voluntary Turnover: Path-Specific Job Satisfaction Effects and the Importance of Unsolicited Job Offers. *Acad. Manag. J.* **2008**, *51*, 651–671. [\[CrossRef\]](#)
22. Bothma, C.F.C.; Roodt, G. The Validation of the Turnover Intention Scale. *SA J. Hum. Resour. Manag.* **2013**, *11*, 1–12. [\[CrossRef\]](#)
23. Ike, O.O.; Ugwu, L.E.; Enwereuzor, I.K.; Eze, I.C.; Omeje, O.; Okonkwo, E. Expanded-Multidimensional Turnover Intentions: Scale Development and Validation. *BMC Psychol.* **2023**, *11*, 271. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Judge, T.A.; Thoresen, C.J.; Bono, J.E.; Patton, G.K. The Job Satisfaction–Job Performance Relationship: A Qualitative and Quantitative Review. *Psychol. Bull.* **2001**, *127*, 376. [\[CrossRef\]](#)
25. Meyer, J.P.; Allen, N.J. A Three-Component Conceptualization of Organizational Commitment. *Hum. Resour. Manag. Rev.* **1991**, *1*, 61–89. [\[CrossRef\]](#)
26. Maslach, C.; Jackson, S.E. The Measurement of Experienced Burnout. *J. Organ. Behav.* **1981**, *2*, 99–113. [\[CrossRef\]](#)
27. Zhao, Y.; Hryniewicki, M.K.; Cheng, F.; Fu, B.; Zhu, X. Employee Turnover Prediction with Machine Learning: A Reliable Approach. In Proceedings of the Intelligent Systems and Applications: 2018 Intelligent Systems Conference (IntelliSys), London, UK, 6–7 September 2018; Volume 2, pp. 737–758.
28. Geisinger, K.F. The Metamorphosis to Test Validation. *Educ. Psychol.* **1992**, *27*, 197–222. [\[CrossRef\]](#)
29. Esposito, G.; Marôco, J.; Passeggia, R.; Pepicelli, G.; Freda, M.F. The Italian Validation of the University Student Engagement Inventory. *Eur. J. High. Educ.* **2022**, *12*, 35–55. [\[CrossRef\]](#)
30. MacCallum, R.C.; Browne, M.W.; Sugawara, H.M. Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychol. Methods* **1996**, *1*, 130. [\[CrossRef\]](#)

31. Hu, L.; Bentler, P.M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Struct. Equ. Model.* **1999**, *6*, 1–55. [\[CrossRef\]](#)
32. Lei, P.-W.; Wu, Q. Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educ. Meas. Issues Pract.* **2007**, *26*, 33–43. [\[CrossRef\]](#)
33. Guo, L.; Hao, R.; Yu, J.; Yang, M. Privacy-Preserving Naive Bayesian Classification for Health Monitoring Systems. *IEEE Trans. Ind. Inform.* **2024**, *20*, 11622–11634. [\[CrossRef\]](#)
34. Guido, R.; Ferrisi, S.; Lofaro, D.; Conforti, D. An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information* **2024**, *15*, 235. [\[CrossRef\]](#)
35. Khan, M.; Hooda, B.K.; Gaur, A.; Singh, V.; Jindal, Y.; Tanwar, H.; Sharma, S.; Sheoran, S.; Vishwakarma, D.K.; Khalid, M.; et al. Ensemble and Optimization Algorithm in Support Vector Machines for Classification of Wheat Genotypes. *Sci. Rep.* **2024**, *14*, 22728. [\[CrossRef\]](#)
36. Wang, Z.; Gai, K. Decision Tree-Based Federated Learning: A Survey. *Blockchains* **2024**, *2*, 40–60. [\[CrossRef\]](#)
37. Razavi-Termeh, S.V.; Sadeghi-Niaraki, A.; Razavi, S.; Choi, S.-M. Enhancing Flood-Prone Area Mapping: Fine-Tuning the K-Nearest Neighbors (KNN) Algorithm for Spatial Modelling. *Int. J. Digit. Earth* **2024**, *17*, 2311325. [\[CrossRef\]](#)
38. Jain, M.; Srihari, A. Comparison of Machine Learning Algorithm in Intrusion Detection Systems: A Review Using Binary Logistic Regression. *Authorea* **2025**, *13*, 45–53. [\[CrossRef\]](#)
39. Nowak, M.; Pawłowska-Nowak, M. Dynamic Pricing Method in the E-Commerce Industry Using Machine Learning. *Appl. Sci.* **2024**, *14*, 24. [\[CrossRef\]](#)
40. Brandmaier, A.M.; Jacobucci, R.C. Machine Learning Approaches to Structural Equation Modeling. In *Handbook of Structural Equation Modeling*; Guilford Press: New York, NY, USA, 2023; pp. 722–739.
41. Vowels, M.J. SLEM: Machine Learning for Path Modeling and Causal Inference with Super Learner Equation Modeling. *arXiv* **2023**, arXiv:2308.04365.
42. Richter, N.F.; Tudoran, A.A. Elevating Theoretical Insight and Predictive Accuracy in Business Research: Combining PLS-SEM and Selected Machine Learning Algorithms. *J. Bus. Res.* **2024**, *173*, 114453. [\[CrossRef\]](#)
43. Sarkale, S.L.; Bhide, H.N.; Tatia, A.; Mahajan, Y.; Sharma, V. Integration of Structural Equation Modeling and Machine Learning in Supply Chain Management. In *Digital Transformation for Improved Industry and Supply Chain Performance*; IGI Global: Hershey, PA, USA, 2024; pp. 93–107.
44. Magasi, C. Evaluating Machine Learning Approaches in Structural Equation Modelling to Improve Predictive Accuracy in Marketing Research. *Indones. J. Bus. Entrep.* **2025**, *11*, 93. [\[CrossRef\]](#)
45. Browne, M.; Rockloff, M.; Rawat, V. An SEM Algorithm for Scale Reduction Incorporating Evaluation of Multiple Psychometric Criteria. *Sociol. Methods Res.* **2018**, *47*, 812–836. [\[CrossRef\]](#)
46. Gonzalez, O. Psychometric and Machine Learning Approaches to Reduce the Length of Scales. *Multivar. Behav. Res.* **2021**, *56*, 903–919. [\[CrossRef\]](#)
47. Orrù, G.; Monaro, M.; Conversano, C.; Gemignani, A.; Sartori, G. Machine Learning in Psychometrics and Psychological Research. *Front. Psychol.* **2020**, *10*, 2970. [\[CrossRef\]](#)
48. Goretzko, D.; Bühner, M. Note: Machine Learning Modeling and Optimization Techniques in Psychological Assessment. *Psychol. Test. Assess. Model.* **2022**, *64*, 3–21.
49. Nowak, M.; Pawłowska-Nowak, M. Prediction of the Type of Organizational Culture Using Machine Learning Approach. *Prz. Organ.* **2023**, *3*, 264–272. [\[CrossRef\]](#)
50. Veglio, V.; Romanello, R.; Pedersen, T. Employee Turnover in Multinational Corporations: A Supervised Machine Learning Approach. *Rev. Manag. Sci.* **2025**, *19*, 687–728. [\[CrossRef\]](#)
51. Chakraborty, R.; Mridha, K.; Shaw, R.N.; Ghosh, A. Study and Prediction Analysis of the Employee Turnover Using Machine Learning Approaches. In Proceedings of the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Kuala Lumpur, Malaysia, 24–26 September 2021; pp. 1–6.
52. Kittikunsiri, M.; Vateekul, P.; Nupairoj, N. Predicting Newcomer's Turnover Using Machine Learning Algorithms: A Case Study of Thai Financial Firm in Bangkok, Thailand. In Proceedings of the 2023 International Conference on Data, Information and Computing Science (CDICS), Singapore, 8–10 December 2023; pp. 39–45.
53. Adeusi, K.B.; Amajuoyi, P.; Benjami, L.B. Utilizing Machine Learning to Predict Employee Turnover in High-Stress Sectors. *Int. J. Manag. Entrep. Res.* **2024**, *6*, 1702–1732. [\[CrossRef\]](#)
54. Ajit, P. Prediction of Employee Turnover in Organizations Using Machine Learning Algorithms. *Algorithms* **2016**, *4*, C5.

-
55. MacCallum, R.C.; Roznowski, M.; Necowitz, L.B. Model Modifications in Covariance Structure Analysis: The Problem of Capitalization on Chance. *Psychol. Bull.* **1992**, *111*, 490. [[CrossRef](#)] [[PubMed](#)]
 56. Brown, T.A. *Confirmatory Factor Analysis for Applied Research*; Guilford Publications: New York, NY, USA, 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.