

Grey clustering machine learning model for predicting voluntary employee turnover

Grey Systems:
Theory and
Application

771

Marcin Nowak
Faculty of Engineering Management, Poznań University of Technology,
Poznań, Poland

Received 27 February 2025
Revised 11 May 2025
7 July 2025
Accepted 9 July 2025

Abstract

Purpose – The aim of this study is to develop a method for predicting voluntary employee turnover using grey machine learning and to demonstrate its effectiveness as a decision-support tool in HR management.

Design/methodology/approach – This study presents an innovative approach that integrates machine learning with grey system theory to predict voluntary employee turnover. The developed method follows a six-step procedure, encompassing variable selection, data preparation, and model training. A grey cluster-based decision model is applied as an activation function within a novel machine learning architecture, enabling the incorporation of imprecise, vague, and error-prone data. A quantitative study was conducted using survey data from 1,030 employees, collected through a CAWI method. The dataset was analyzed using machine learning techniques, and model validation was performed through cross-validation. This methodology enhances HR decision-making by addressing uncertainty in turnover prediction.

Findings – The study demonstrates that integrating grey system theory with machine learning enhances the prediction of voluntary employee turnover by effectively handling imprecise and uncertain data. The developed grey cluster-based decision model improves the representation of social science complexities, capturing key turnover factors such as salary, job satisfaction, and organizational culture. The model's performance was validated through a quantitative study, achieving an average accuracy of 80.94%. These findings confirm its applicability in HR decision-making, suggesting that grey machine learning provides a robust framework for predicting employee departures and enabling organizations to implement proactive retention strategies while minimizing turnover-related costs.

Originality/value – The originality of this study lies in the unique integration of grey system theory with machine learning architecture through the application of a grey cluster-based decision model as an activation function. This approach has not been widely applied to predicting voluntary employee turnover, and its novelty stems from the ability to operate directly on imprecise and incomplete data without the need for full “whitening”. The developed model captures complex, multidimensional, and often subjective factors influencing employee decisions, making it more resilient to informational noise than traditional algorithms. An additional value is the method's potential transferability to other areas of social science research where uncertainty and incomplete data pose significant challenges, opening opportunities for further development in both theoretical and practical analytical contexts.

Keywords Voluntary employee turnover, Talent management, Machine learning, Grey system theory

Paper type Research article

1. Introduction

Employee turnover, especially voluntary departures, is one of the key challenges in modern human resource management. In a dynamic market environment where acquiring and retaining talent is of particular importance, accurately predicting employees' intentions to leave is essential for effective strategic planning. Traditional analytical methods based on classical statistical models often fail to account for the complexity and uncertainty inherent in social data, limiting their effectiveness.

JEL Classification — M51, J63, C45, J24

© Marcin Nowak. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at [Link to the terms of the CC BY 4.0 licence](#)

Funding: The research was funded by a grant from the National Science Centre, Poland, under project title “A method for predicting voluntary employee turnover using grey machine learning” (2024/08/X/HS4/00155).



Grey Systems: Theory and Application
Vol. 15 No. 4, 2025
pp. 771-791
Emerald Publishing Limited
e-ISSN: 2043-9385
p-ISSN: 2043-9377
DOI 10.1108/GS-02-2025-0020

This article proposes an innovative approach that integrates machine learning with grey system theory—an original grey cluster-based decision model. The proposed model introduces a novel integration of grey system theory into a machine learning framework, addressing the limitations of conventional predictive methods in handling ambiguous and incomplete social data. Grey system theory allows the use of so-called grey numbers, which effectively represent informational uncertainty in social research, thereby capturing the nuances influencing employees' decisions more accurately. In the presented approach, grey system theory is integrated with machine learning by applying a grey cluster-based decision model as an activation function in a novel machine learning architecture. This model allows operations on grey numbers, enabling effective representation of imprecise and incomplete information typical for social research. The innovation of this method lies in combining the adaptive capabilities of learning algorithms with the uncertainty-handling features of grey models. As a result, a predictive tool is created that not only learns from empirical data but also maintains strong resilience to informational noise, making it particularly valuable in the context of human resource management.

The aim of this study is not only to introduce a novel model for predicting voluntary employee turnover but also to demonstrate how integrating grey machine learning can serve as a valuable decision-support tool in HR management. The research findings suggest that applying this methodology enhances the ability to account for uncertainty in social data, leading to more accurate predictions of employee behavior.

The article fits into a broader trend of developing new models related to grey systems theory, as exemplified by (Zhang *et al.*, 2025; Hu and Xie, 2025; Li and Li, 2024; Nowak *et al.*, 2021). These models typically address predictive problems (Jia and Pang, 2024; Rathnayaka and Seneviratna, 2024; Zhang and Li, 2024) as well as decision-making problems (Guo *et al.*, 2024; Qin *et al.*, 2025). At the same time, the integration of tools from grey systems theory with machine learning constitutes an important research aspect (Wang *et al.*, 2024; Yao and Liu, 2024; Qin and Zhao, 2024; Hakan *et al.*, 2024).

Section 2 of the article presents a literature review on the fundamentals of machine learning and decision models within grey system theory. Section 3 introduces the formal model of the proposed method for predicting voluntary employee turnover intentions using grey machine learning. Section 4 presents the results of empirical research in which the developed grey machine learning method was applied.

2. Literature review

2.1 Voluntary employee turnover

Voluntary employee turnover is one of the most critical topics in organizational behavior research, impacting both business efficiency and the costs associated with recruiting and training new employees. Over the past century, research on employee turnover has evolved, manifesting in a wide range of theoretical and methodological approaches. Classical theories, such as those proposed by March and Simon (1993), Price (1977), and Mobley *et al.* (1978), laid the foundation for subsequent studies. In recent years, with the advancement of information technologies and analytical methods, there has been an intensification of research utilizing advanced statistical techniques and machine learning models. Hom *et al.* (2017), in their work “*One Hundred Years of Employee Turnover Theory and Research*,” conducted an extensive literature review, highlighting the complexity of the processes underlying employee departures. The authors emphasize that turnover is not merely a result of individual employee decisions but rather an outcome of the interaction of numerous factors—ranging from compensation and job satisfaction to perceptions of organizational justice (Hom *et al.*, 2017). Lee *et al.* (2017), in “*On the Next Decade of Research in Voluntary Employee Turnover*,” propose advancing research by integrating traditional theoretical approaches with modern analytical techniques, such as machine learning. They note that while earlier models often relied on statistical regression analyses (Griffeth *et al.*, 2000; Tett and Meyer, 1993), contemporary research is shifting toward

more dynamic methods that better account for the instability and variability of factors influencing employee decisions (Holtom *et al.*, 2008; Maertz and Campion, 2004).

The latest trends in employee turnover research are summarized by Bolt *et al.* (2022) in “A Century of Labour Turnover Research: A Systematic Literature Review.” This work provides a systematic analysis of hundreds of studies, emphasizing the growing importance of interdisciplinary approaches and methodologies that combine traditional statistical analyses with modern computational tools. The authors highlight that integrating quantitative and qualitative data, along with the application of machine learning—including models based on grey system theory—enhances predictive accuracy and deepens the understanding of the mechanisms driving voluntary employee turnover (Bolt *et al.*, 2022).

The literature also includes numerous studies that expand upon classical models and adapt them to contemporary market conditions. For example, Griffeth *et al.* (2000) conducted a meta-analysis that identified key determinants of turnover, emphasizing the importance of motivational and demographic factors. Similar conclusions were drawn by Maertz and Campion (2004), who critically evaluated previous approaches and proposed new research directions that account for organizational and cultural contexts. In their view, it is crucial not only to identify influencing factors but also to understand their interactions (Mobley *et al.*, 1978; Price, 1977).

Holtom *et al.* (2008) highlight the role of the work environment and organizational support in shaping employees’ intentions to leave. Their research indicates that high levels of employee engagement and effective internal communication can significantly reduce turnover rates, a finding also confirmed by Hausknecht *et al.* (2009). Similarly, Tett and Meyer (1993) suggest that job satisfaction and organizational commitment are key predictors of employee retention, though these relationships can be moderated by industry- and culture-specific conditions. Another important research stream examines the impact of organizational factors on turnover. Eisenberger *et al.* (2005) emphasize that supervisor support and fair treatment of employees contribute to a lower willingness to leave. Allen *et al.* (2010) expand on these findings, suggesting that implementing evidence-based retention strategies can yield significant benefits in reducing turnover.

Thus, voluntary employee turnover is a multidimensional phenomenon, shaped by the interaction of various factors—individual, organizational, and external (Nowak, 2024). Given the dynamic changes in the labor market and increasing competition in attracting and retaining talent, the development of analytical tools is a key component of human resource management strategies. Therefore, future research on voluntary turnover should focus not only on identifying its determinants but also on developing tools that effectively integrate these factors into comprehensive predictive models (Allen *et al.*, 2010; Holtom *et al.*, 2008).

In recent years, researchers have begun to combine grey systems theory with human resource analytics to better capture uncertainty, incomplete data, and nonlinear decision-making processes. This integration allows for enhanced modeling of employee behavior and strategic HR forecasting, particularly in complex or information-poor environments. For instance, Mallick and Mukhopadhyay (2024) developed a grey-based multicriteria decision-making model for “boomerang recruitment,” providing an intelligent rehiring framework that accounts for multiple uncertain inputs. Jin and Yang applied a grey backpropagation neural network to forecast HR demand in port enterprises, demonstrating how grey modeling enhances prediction in sectors with limited historical data. Liu and Zhang (2024) used grey correlation analysis to study the relationship between strategic human resource management and corporate performance, showing that grey models can effectively handle vague causalities in organizational data. Tao (2024) explored the impact of HRM strategies on organizational structure and brand image using grey correlation analysis, bridging soft managerial aspects with measurable outputs. Esangbedo and Xue (2025) proposed a more general application of grey system theory to business decision-making, suggesting its relevance in HR contexts through multi-criteria frameworks. These studies collectively demonstrate the growing potential of grey system methods in HR analytics, particularly in contexts where data is partial, subjective, or hard to quantify. Their relevance to voluntary turnover research lies in the ability of grey models to incorporate uncertain perceptions, such as job satisfaction or organizational fairness, into predictive architectures.

2.2 Machine learning

Machine learning, which is now at the core of artificial intelligence (AI) and data science development, has its origins in the mid-20th century (Nowak and Pawłowska-Nowak, 2024). The first significant studies on the concept of machine “learning” date back to the 1950s when Alan Turing proposed the idea of a “learning machine” in 1950 (Naeini and Prindle, 2018).

In an analytical perspective, machine learning can be described as the process in which a computer program develops the ability to solve tasks T based on experience E , in accordance with a specific evaluation criterion p , with the quality Q of solving these tasks improving as experience E accumulates (Mitchell, 1997). Fields closely related to machine learning include primarily probability theory, statistics, formal logic, information theory, control theory, cybernetics, as well as psychology and neurophysiology. It is often emphasized that machine learning represents an area of knowledge at the intersection of computer science and cybernetics (Sharma and Kumar, 2017).

In 1957, F. Rosenblatt developed an early machine learning model known as the perceptron, which was designed to mimic aspects of human cognitive processes (Rosenblatt, 1957, 1958). This initial approach, later referred to as Rosenblatt’s perceptron, was capable of identifying handwritten characters and laid the foundation for the creation of artificial neural networks (ANNs). Later, in 1959, A. Samuel helped bring the concept of “machine learning” to the forefront by defining it as a discipline that enables computers to learn autonomously without being directly programmed (Samuel, 1959).

The 1980s saw key breakthroughs, including the development of genetic algorithms, which drew inspiration from evolutionary theory and natural adaptation processes. John Holland’s work on genetic algorithms was crucial for the advancement of this field (Othman *et al.*, 2022; Jones *et al.*, 2018). At the turn of the 21st century, machine learning witnessed explosive growth in both its theoretical underpinnings and real-world applications. This rapid expansion was driven by several major developments (Kaplan, 2016):

- (1) A significant boost in the computational speeds of modern computer networks.
- (2) A notable increase in the memory capacity available in IT systems.
- (3) A sharp decline in the cost per unit of both data processing and memory.
- (4) A shift from physical media to digital data storage.
- (5) Improved access to data, especially facilitated by the Internet.
- (6) The emergence of affordable, high-resolution digital sensors.
- (7) The fast-paced advancement of big data technologies.

3. Methods

The developed method for predicting voluntary employee turnover, which integrates grey system theory (grey cluster-based decision model) with machine learning, can be presented as a six-step procedure:

Step 1. Defining the set of variables influencing the intention of voluntary employee turnover

In the developed method, the following set of variables is applied: gender (x_1), age (x_2), education (x_3), job position characteristics (x_4), salary (x_5), job satisfaction (x_6), sense of fairness (x_7), promotion opportunities (x_8), professional development opportunities (x_9), work performance (x_{10}), working conditions (x_{11}), team atmosphere (x_{12}), recognition and rewards (x_{13}), relationships with supervisors (x_{14}), job stability (x_{15}), communication within the company (x_{16}), work-life balance (x_{17}), independence at work (x_{18}), level of autonomy at work (x_{19}), job responsibility (x_{20}), work engagement (x_{21}), remote work availability (x_{22}), flexible working hours (x_{23}), sense of burnout (x_{24}), workload (x_{25}), commuting time (x_{26}), recognition

at work (x_{27}), organizational management (x_{28}), job monotony (x_{29}), employer reputation (x_{30}), and organizational culture (x_{31}). Each variable can be measured using different value scales. All variables are expressed as grey numbers. The selection of 31 variables applied in the first step of the method was based on both theoretical premises and empirical findings in the literature on voluntary employee turnover. As highlighted by numerous studies (Hom *et al.*, 2017; Bolt *et al.*, 2022; Lee *et al.*, 2017; Holtom *et al.*, 2008), decisions to leave a job result from complex interactions among individual, organizational, and contextual factors. The included variables reflect key areas of influence such as salary, job satisfaction, organizational culture, sense of fairness, relationships with supervisors, opportunities for professional development, and work-life balance. The initial inclusion of a broad set of variables is also in line with standard practices in machine learning, where feature selection is performed in later stages—here, through correlation analysis (Step 3). This approach allows for theoretical comprehensiveness while reducing model dimensionality to the most predictive variables, thereby enhancing both methodological validity and predictive efficiency of the developed tool.

Step 2. Preparing the dataset for machine learning

In the second step of the method, employee surveys must be conducted. Given that machine learning is used in this approach, the sample does not need to be representative—conclusions should be limited to the studied group. A sample research group may consist of employees from a large enterprise. The survey questionnaire must include questions related to the variables identified in Step 1.

Step 3. Data Preparation

As part of this stage, the model involves the following actions:

- (1) rescaling variables to uniform scales (or standardizing variables) – a crucial step for facilitating the machine learning process, necessary for many algorithms.
- (2) transforming multi-value nominal data into a set of binary variables – converting categorical variables into numerical representations.
- (3) conducting correlation analysis and selecting the 10 most strongly correlated variables with the target label – different correlation coefficients are used depending on the nature of the variables.

The actions outlined in this step are performed on white numbers (following the *kernel of the grey number* concept) to define a catalog of the 10 most strongly correlated features with the target variable. In the subsequent steps, the model learns using grey numbers.

Step 4. Splitting the Dataset into Training and Testing Data

In machine learning, it is standard practice to randomly split the entire dataset into two distinct groups: one used for training and the other reserved for testing (Nguyen *et al.*, 2021). Both groups include records with specific feature values and their corresponding labels. Typically, the training group comprises between 70 and 90% of the overall data (Nguyen *et al.*, 2021). This separation allows the algorithm to develop its predictive capabilities using the training data, while its effectiveness is independently evaluated on the test data. A robustly trained model is one that achieves high classification accuracy when applied to this test set.

Step 5. Implementing the Machine Learning Process Using the Grey Cluster-Based Model as an Activation Function

The author of this study has frequently applied classification, relational, and predictive methods in previous research on management problems, all of which are part of grey system theory. This theory models informational uncertainty using so-called grey operators,

particularly grey numbers, weighted whitening functions, grey matrices, and grey numerical sequences (or series).

In 2019, the author developed a decision-classification model based on grey system theory—the grey cluster-based decision model. This model utilizes grey numbers and the concept of reference points, making it particularly suitable for modeling the informational uncertainty characteristic of social sciences. However, like many similar models, it is not inherently capable of learning in the machine learning sense, meaning its full potential has not been realized.

The integration of this model with machine learning allows for the incorporation of imprecise, vague, and error-prone data. This approach ensures an accurate representation of the uncertainty inherent in social sciences. The grey machine learning process for each object can be represented by the following procedure:

- (1) Determining the Attribute Vector for Each Object in the Training Set

$$D_k = [\otimes g_{1k}, \otimes g_{2k}, \dots, \otimes g_{nk}], \tag{1}$$

Where $k = \{1, 2, \dots, m\}$ – represents the number of decision objects.

- (2) Initializing Weights in the Grey Machine Learning Model

The weights are selected as small random numbers, following standard machine learning methodology. The weight vector is represented as follows:

$$W = [\otimes w_1, \otimes w_2, \dots, \otimes w_n] \tag{2}$$

- (3) Determining the Number of Epochs – it is assumed that the learning process will be completed in no more than 50 epochs

- (4) Determining the Weighted Attribute Vectors According to the Following Formula:

$$D_k^* = [\otimes v_{1k}, \otimes v_{2k}, \dots, \otimes v_{nk}], \tag{3}$$

where $\otimes v_{ik} = \otimes g_{ik} \times \otimes w_{ik}$.

Which can also be expressed by [formula \(4\)](#):

$$D_k^* = W^T \cdot D_k \tag{4}$$

- (5) Determining the Total Weighted Score for All Objects in the Training Set

In step five, the so-called weighted overall evaluation $\otimes R$ is calculated for each decision-making object. This overall evaluation, represented as a grey number, has a lower bound equal to the sum of the lower bounds of the gray numbers $\otimes v_i$ from the weighted gray decision vector D^* , and an upper bound equal to the sum of the upper bounds of these numbers, as specified by [formula \(5\)](#).

$$\otimes R = \left[\sum_{i=1}^n \underline{v}_i, \sum_{i=1}^n \bar{v}_i \right] \tag{5}$$

The [appendix](#) to this article presents the pseudocode for the developed grey machine learning algorithm for predicting voluntary employee turnover.

- (6) Assigning Labels to the Training Set

It is assumed that the cluster-based model will assign objects to two categories: the presence or absence of turnover intention. In this case, a single reference vector is sufficient. If the total

weighted grey score for a given object is greater than the total weighted score of the reference vector $\otimes r_{ref}$, the object is assigned a value of 0 (absence of turnover intention). Conversely, if the total weighted grey score is lower, the object is assigned a value of 1 (presence of turnover intention). When comparing the total weighted score $\otimes R$ with the total weighted score of the reference vector $\otimes r_{ref}$, the probability that one grey number is greater than another is determined according to [formula \(6\)](#) ([Eshtaiwi et al., 2017](#)).

$$P\{\otimes g_1 \leq \otimes g_2\} = \frac{\max\left[0, L^* - \max\left(0, \bar{g}_1 - \underline{g}_2\right)\right]}{L^*}, \quad (6)$$

where $L^* = L(\otimes g_1) + L(\otimes g_2)$.

(7) Updating Weights in Each Epoch

The next step of the method involves comparing the labels for the training set with the known reference labels. If the computed label for a given decision object does not match the known label, the weights are updated according to [formula \(7\)](#).

$$\Delta w_j = \eta(y - \hat{y})x_j, \quad (7)$$

where:

Δw_j – change in the weight of the j -th criterion,

η – learning rate coefficient,

y – known label values,

\hat{y} – predicted label values,

x_j – attribute value in the j -th criterion.

To adapt [formula \(8\)](#) for the entire training dataset, the following notations are introduced:

$$\Delta w_j = \eta(y^i - \hat{y}^i)x_j^i, \quad (8)$$

where:

i – sample number.

After the update, the new weight for the j -th criterion is given by:

$$w_j := w_j + \Delta w_j \quad (9)$$

If the algorithm’s predicted label matches the expected one, the weight for that particular criterion remains unaltered. The learning rate, denoted by η , sets the magnitude of the weight adjustments—high values of η result in substantial updates, whereas lower values lead to more gradual changes. This learning rate is one of the most pivotal hyperparameters in any machine learning architecture. The model’s weights continue to be updated until either convergence is reached (i.e. the predicted labels in the training set align closely with the actual labels) or the predetermined number of training epochs has been exceeded.

Step 6. Validating the Effectiveness of the Grey Machine Learning Model in Predicting Voluntary Employee Turnover Intentions

The effectiveness of a model within the training set can be evaluated using various metrics ([Wei et al., 2022](#)). One of the most commonly used measures is accuracy, which is

recommended for this purpose. The method for calculating accuracy is presented in [Equation \(10\)](#).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (10)$$

where:

- (1) accuracy – machine learning model accuracy,
- (2) TP (true positives) – value correctly classified as true,
- (3) FP (false positives) – value incorrectly classified as true,
- (4) TN (true negatives) – value correctly classified as false,
- (5) FN (false negatives) – value incorrectly classified as false.

4. Research results

A prediction method for voluntary employee turnover intentions has been developed using the proposed formal model of a new machine learning architecture based on grey system theory. The structure of the implemented actions within the empirical research followed a procedure consisting of six steps.

Step 1. Defining the set of variables influencing the intention of voluntary employee turnover

In the developed method, based on the catalog of criteria and utilizing the grey machine learning architecture, the prediction of voluntary employee turnover intentions is conducted. Based on a review of the relevant literature, a catalog of variables has been formulated and is presented along with the applied scales in [Table 1](#) ([Bolt et al., 2022](#); [Holtom et al., 2008](#); [Hom et al., 2017](#); [Lee et al., 2017](#)).

The label indicating the presence of turnover intention is marked as 1 ([1;1]), while the absence of this intention is marked as 0 ([0;0]).

Step 2. Preparing the dataset for machine learning

The data was collected through a CAWI survey conducted on a purposefully selected sample of 1,030 individuals, consisting of Polish employees regardless of the type of employment contract. The survey was carried out between August 1 and September 30, 2024, allowing for data collection under various organizational conditions. A fragment of the dataset is presented in [Table 2](#).

To ensure a uniform representation of all input variables in the grey machine learning model, both quantitative and qualitative variables were converted into grey numbers. In the case of nominal variables, such as gender or education level, each category was assigned a single-point grey interval—for example, [1;1] for female and [0;0] for male. This conversion was purely technical and did not imply any ordinal or continuous structure. Despite the use of interval notation, these variables were interpreted by the model as categorical data. Their role was limited to class differentiation, without modeling any quantitative relationships between categories. This approach is consistent with the logic of grey system theory, which allows for the use of singleton intervals to represent certain, unambiguous information. As a result, it was possible to maintain a consistent structure of the input data without affecting the classification performance or violating the assumptions of predictive modeling.

A total of 515 women and 515 men participated in the study. The distributions of age, education level, and place of residence are presented in [Figures 1 and 3](#).

Table 1. Set of variables used in the method for predicting voluntary employee turnover

Variable	Name	Scale for the variable				
x ₁	gender	female	male	other		
x ₂	age	expressed as a natural number				
x ₃	education	primary	middle school	vocational	secondary	higher
x ₄	job position characteristics	regular employee	low-level manager	mid-level manager	high-level manager	
		<i>Attribute Evaluation</i>				
		very poor	poor	average	good	very good
x ₅	salary	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₆	job satisfaction	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₇	sense of fairness	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₈	promotion opportunities	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₉	professional development opportunities	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₀	work performance	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₁	working conditions	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₂	team atmosphere	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₃	recognition and rewards	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₄	relationships with supervisors	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₅	job stability	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₆	communication within the company	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₇	work-life balance	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₈	independence at work	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₁₉	level of autonomy at work	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₀	job responsibility	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₁	work engagement	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₂	remote work availability	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₃	flexible working hours	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₄	sense of burnout	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₅	workload	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₆	commuting time	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]

(continued)

Table 1. Continued

Variable	Name	Scale for the variable				
x ₂₇	recognition at work	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₈	organizational management	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₂₉	job monotony	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₃₀	employer reputation	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]
x ₃₁	organizational culture	[0.0; 0.2]	[0.2; 0.4]	[0.4; 0.6]	[0.6; 0.8]	[0.8; 1.0]

Source(s): Author's own elaboration

Table 2. Fragment of the dataset on voluntary employee turnover intentions

	Employee 1 – Linguistic evaluations	Employee 1 – Evaluations expressed as grey numbers
x ₁	female	[1;1]
x ₂	31	[31; 31]
x ₃	higher education	[1;1]
x ₄	mid-level manager	[2;2]
...	...	
x ₂₇	4	[0.6;0.8]
x ₂₈	1	[0.0; 0.2]
x ₂₉	3	[0.4; 0.6]
x ₃₀	2	[0.2; 0.4]
x ₃₁	4	[0.6;0.8]
label	0	[0; 0]

Source(s): Author's own elaboration

The sample reflects considerable diversity in demographic characteristics. As shown in [Figure 1](#), the age structure is skewed towards younger respondents, with the largest group aged 26–30 (253 individuals), followed by the 18–25 and 31–35 age groups (159 each). This concentration of younger participants aligns with the labor market's increased turnover in early career stages. In terms of education ([Figure 2](#)), most respondents held a higher education degree (558), while a substantial group reported upper secondary education (376). This reflects a relatively well-educated sample, which corresponds with recent trends in workforce qualifications in Poland. As illustrated in [Figure 3](#), the place of residence is varied. Respondents come from all types of settlements, with a slight predominance of medium towns (243) and rural areas (237), followed by large cities and metropolitan areas. This distribution ensures a broad contextual basis for modeling employee turnover intentions and enhances the generalizability of the findings within the national labor context.

Step 3. Data Preparation

In the first step of this stage, the variables (expressed as white numbers) were rescaled to a 0–1 scale. Next, multi-state nominal data was transformed into a set of binary variables.

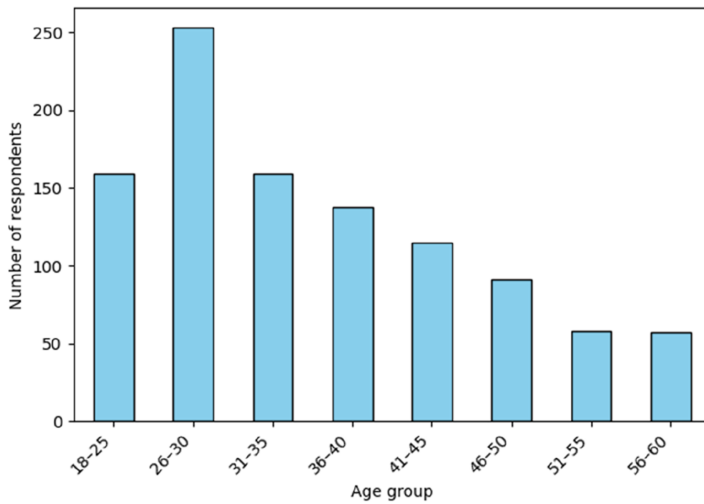


Figure 1. Age group distribution of respondents. Source: Author's own elaboration

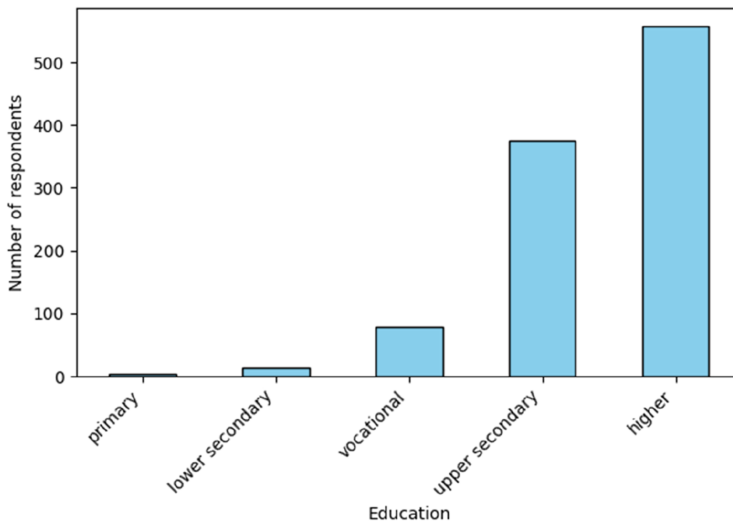


Figure 2. Education level distribution of respondents. Source: Author's own elaboration

Subsequently, a correlation analysis was conducted. Given that the label is binary, a point-biserial correlation coefficient (formula 11) was applied to quasi-continuous variables, while Cramér's V (formula 12) was used for the remaining variables.

$$r_{pb} = \frac{M_1 - M_0}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n^2}}, \quad (11)$$

where:

- (1) r_{pb} -point-biserial correlation coefficient,

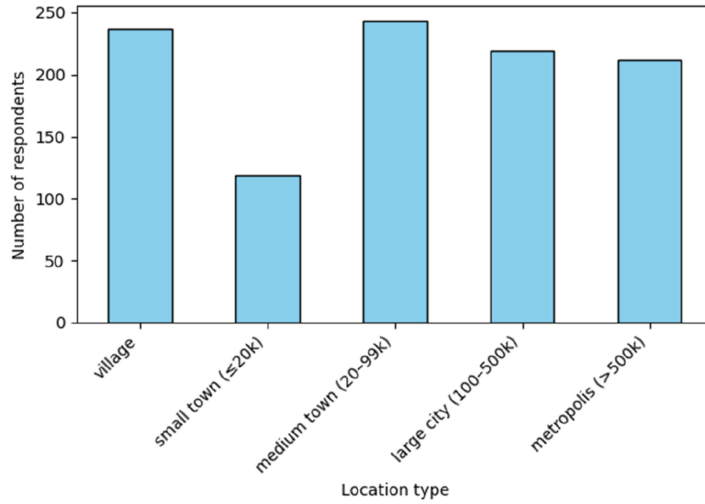


Figure 3. Place of residence distribution of respondents. Source: Author’s own elaboration

- (2) M_1 is the mean value of the variable for individuals exhibiting turnover intention,
- (3) M_0 is the mean value of the variable for individuals not exhibiting turnover intention,
- (4) σ is the standard deviation of the variable,
- (5) n_1 is the number of individuals exhibiting turnover intention,
- (6) n_2 is the number of individuals not exhibiting turnover intention,
- (7) n is the number of individuals in the dataset.

$$V = \sqrt{\frac{\chi^2}{n \cdot \min(k - 1, r - 1)}} \tag{12}$$

where:

- χ^2 – Chi-square goodness-of-fit test result,
- n – total number of observations,
- $\min(k - 1, r - 1)$ – the minimum of $(k-1)$ and $(r-1)$,
- k – number of columns in the contingency table,
- r – number of rows in the contingency table.

In the following section of the study, the strength of relationships between variables was analyzed based on the calculated values of both the point-biserial correlation coefficient and Cramér’s V coefficient. The adopted scale for this analysis is as follows:

- (1) 0 to 0.2 – weak relationship between variables,
- (2) 0.2 to 0.4 – moderate relationship between variables,
- (3) 0.4 to 0.6 – medium relationship between variables,
- (4) 0.6 to 0.8 – strong relationship between variables,

(5) 0.8 to 1.0 – very strong relationship between variables.

The list of the ten variables most strongly correlated with the label is as follows:

- (1) Salary – 0.601 ($p < 0.001$)
- (2) Job satisfaction – 0.482 ($p < 0.001$)
- (3) Sense of fairness – 0.379 ($p < 0.001$)
- (4) Recognition at work – 0.296 ($p < 0.001$)
- (5) Professional development opportunities – 0.282 ($p < 0.001$)
- (6) Relationships with supervisors – 0.262 ($p < 0.001$)
- (7) Team atmosphere – 0.259 ($p < 0.001$)
- (8) Work engagement – 0.252 ($p < 0.001$)
- (9) Organizational culture – 0.251 ($p < 0.001$)
- (10) Working conditions – 0.250 ($p < 0.001$)

For further research, the dataset will include only these ten variables, as they have the most significant impact on the label. The extremely low p -values (all $p < 0.001$) indicate a high level of statistical significance, reinforcing the reliability of these correlations for the analysis.

Step 4. Splitting the Dataset into Training and Testing Data

To prepare the data for the machine learning process, the dataset was divided into two subsets: training and test sets. Both sets contain records described by feature (variable) values along with their corresponding labels. The division was carried out so that the training set includes 750 records, accounting for 75% of the total data, while the test set consists of 250 records (25% of the data). This process was randomized and executed using the `train_test_split` function from the scikit-learn library, which is designed for machine learning applications in Python.

Step 5. Implementing the Machine Learning Process Using the Grey Cluster-Based Model as an Activation Function

(1) Determining a Sample Attribute Vector for Each Object in the Training Set

A sample attribute vector for a selected employee is as follows:

$$D_1 = \{[0.6; 0.8], [0.6; 0.8], [0.8; 1.0], [0.4; 0.6], [0.6; 0.8], [0.8; 1.0], [0.6; 0.8], [0.6; 0.8], [0.8; 1.0], [0.6; 0.8]\}$$

(2) Initializing Weights for the Grey Machine Learning Model

The weight vector, which was randomly initialized in the study on predicting voluntary employee turnover intentions, took the following form in the training set:

$$W = \{[-0.0031; -0.0031], [-0.0023; -0.0023], [0.0014; 0.0014], [-0.0023; -0.0023], [0.0003; 0.0003], [0.0033; 0.0033], [0.0008; 0.0008], [-0.0023; -0.0023], [0.0022; 0.0022], [0.0009; 0.0009]\}$$

(3) It was established that the learning process would take place over 50 epochs.

The selection of 50 epochs in the training process of the grey machine learning model was based on preliminary experimental tests aimed at identifying the point at which the error curve

stabilizes and minimizing the risk of overfitting. The conducted analysis showed that after approximately the 40th epoch, the decrease in error began to slow down, and further increases in the number of epochs did not result in significant improvements in classification performance. At the same time, extending the training process beyond 50 epochs led to greater fluctuations in test set errors, which could indicate an increasing risk of model overfitting. For this reason, a compromise value of 50 epochs was adopted as an optimal balance between sufficient training duration and maintaining proper generalization.

(4) Determining the Weighted Attribute Vectors for Each Object in the Training Set

A sample weighted attribute vector for a selected employee is as follows:

$$D_1^* = \{[-0.00248; -0.00186], [-0.00184; -0.00138], [0.00112; 0.0014], \\ [-0.00138; -0.00092], [0.00018; 0.00024], [0.00264; 0.0033], [0.00048; 0.00064], \\ [-0.00184; -0.00138], [0.00176; 0.0022], [0.00054; 0.00072]\}$$

(5) Assigning Labels to the Training Set

After the first epoch of machine learning using the grey machine learning architecture, the following labels were obtained in the training set:

$$[0, 0, 0, 1, 1, \dots, 0, 1, 1, 0, 1] \text{ (There were 750 elements.)}$$

However, after the first epoch, it became evident that most labels were assigned incorrectly, necessitating weight adjustments and rerunning the algorithm in subsequent epochs.

(6) Updating Weights in Each Epoch

After running the algorithm for 50 epochs on the training set, the resulting weight values are as follows:

$$W = \{[-0.002128; -0.002142; 0.000059; -0.004417; -0.000833; 0.003681; \\ -0.000942; -0.003060; 0.001398; -0.000287]\}$$

After executing the machine learning algorithm, the following set of labels was obtained in the training set:

$$[0, 0, 1, 1, 0, \dots, 1, 1, 1, 0, 0] \text{ (There were 750 elements.)}$$

Step 6. Validating the Effectiveness of the Grey Clustering Machine Learning Model in Predicting Voluntary Employee Turnover Intentions

Figure (1) presents the number of incorrect predictions in the training set, while Figure (2) shows the number of incorrect predictions in the test set as a function of the number of epochs. Both figures refer to a single example of the machine learning model training process (with a specific random seed).

Figure 4 presents the number of misclassifications in each epoch of training for a sample learning process. At the beginning (early epochs), the highest number of errors is observed—reaching approximately 250. Then, as training progresses, this number systematically decreases and stabilizes around 200–220 errors. Some fluctuations occur during the learning process, which is typical for iterative optimization algorithms. However, the overall downward trend indicates an improvement in model quality over successive epochs (see Figure 5).

The presented figure shows the number of misclassifications in the test set across successive epochs for a sample learning process. The number of errors often remains within the range of approximately 40–60, which may indicate a generally acceptable, though not

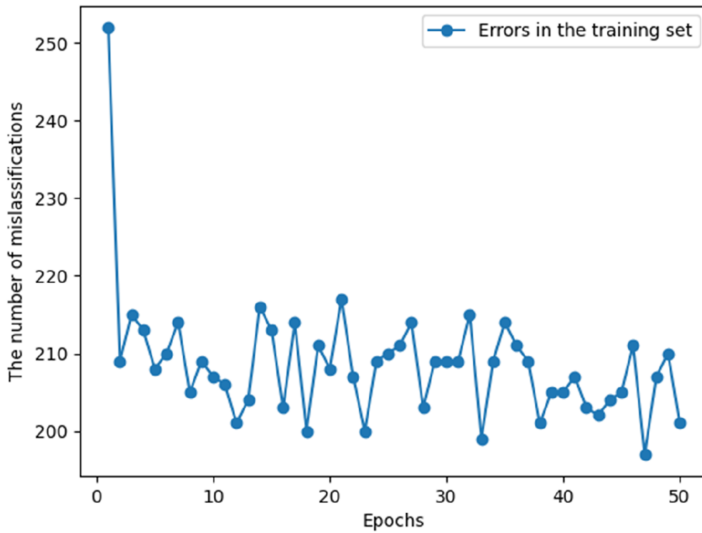


Figure 4. Number of misclassifications in each epoch (Training Data). Source: Author's own elaboration

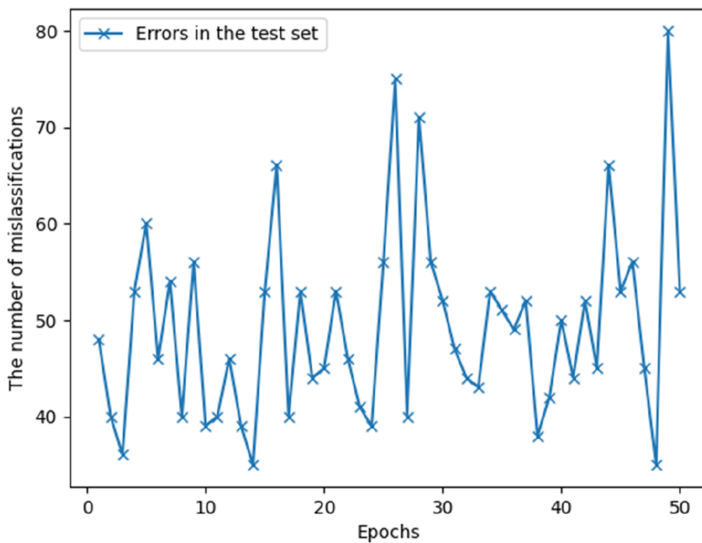


Figure 5. Number of misclassifications in each epoch (Test Data). Source: Author's own elaboration

perfect, model performance in predicting new data. It should be emphasized that the observed fluctuations in the number of errors in both the training and test sets are typical for iterative learning processes, especially when working with imprecise data and noise-resistant models. These variations do not indicate model instability but rather reflect the adaptive nature of the algorithm as it adjusts to the heterogeneous structure of the input data. Importantly, the overall downward trend and the comparable number of misclassifications in the training and test sets suggest that the model is not overfitted and demonstrates good generalization capability.

For the sample learning process, a confusion matrix was prepared to better illustrate the model’s quality (for both the training and test sets) (Figure 6).

The data presented in the confusion matrices indicate that the accuracy for the training set was 72.7%, while for the test set, it was 79.5%. However, these values depend, among other factors, on the randomly selected training and test sets.

To avoid the effect of a “lucky draw,” a cross-validation procedure was conducted for the developed model. A total of 100 training processes were performed, and the accuracy metric was calculated for each. The results of this procedure are presented in Figure 7.

It turned out that the average accuracy metric for the grey clustering machine learning model in predicting voluntary employee turnover intentions was 0.8094 with a standard deviation of 0.0368. It is worth emphasizing that the empirical data revealed the highest correlations with the label as 0.601 for salary, 0.482 for work satisfaction, and 0.379 for a sense of fairness. These correlations are relatively modest, underscoring the intricate and unpredictable nature of voluntary employee departure intentions. Nonetheless, the machine learning architecture devised using grey system theory manages to predict turnover intentions with a success rate above 80%—an excellent performance. This suggests that integrating machine learning methods with grey system theory models can provide highly effective instruments for social science research, especially in the fields of management.

The article pointed out that the key advantage of the developed method over traditional machine learning algorithms is its ability to model uncertainty using grey numbers. In such cases, the application of other machine learning algorithms is not feasible. However, for the purpose of comparative analysis, the respondents’ answers were “whitened” and an assessment was conducted to determine the predictive performance of machine learning algorithms such as: the naïve Bayes classifier, linear and nonlinear support vector machines, decision trees, and logistic regression. In this scenario, the level of uncertainty in the whitened responses is inherently lower than at the outset of the grey clustering machine learning model. It was therefore assumed that if the developed model performs no worse than the average accuracy metric value of the five comparative models (with cross-validation applied to all models), it would be positively validated. Table 3 presents the average accuracy metric value for classical machine learning models (after data whitening) in predicting voluntary employee turnover intentions.

The average accuracy metric across all the traditional machine learning algorithms analyzed was 0.7706, with a standard deviation of 0.0513. This indicates that, although the data in these algorithms do not exhibit the same level of uncertainty as in the proposed algorithm based on grey system theory, their average accuracy is lower than that of the proposed method. The only model that slightly outperformed the proposed approach was the linear support vector machine (0.8105 compared to 0.8094 in the proposed method). The nonlinear support vector machine and logistic regression algorithms produced results very

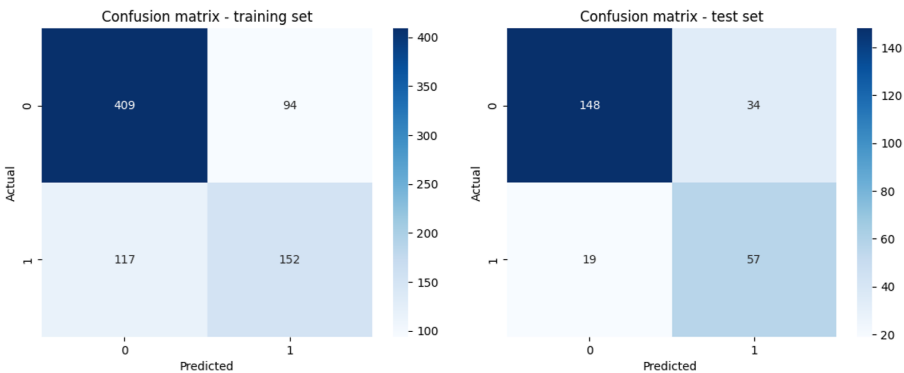


Figure 6. Confusion matrix for the training and test set. Source: Author’s own elaboration

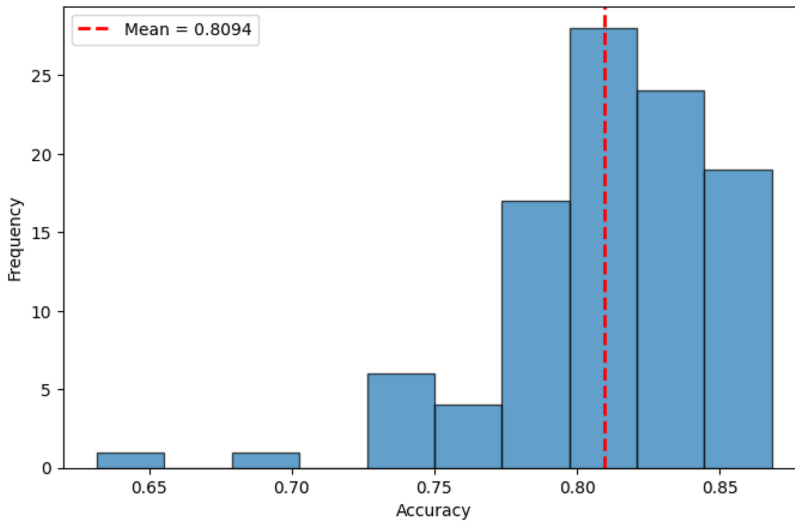


Figure 7. Histogram of the accuracy metric for the grey clustering machine learning model in predicting voluntary employee turnover intentions. Source: Author's own elaboration

Table 3. Average accuracy for classical models in predicting voluntary employee turnover intentions (after data whitening)

Model	Mean accuracy	Standard deviation
Naïve Bayes classifier	0.7368	0.0000
Linear support vector machine	0.8105	0.0000
Nonlinear support vector machine	0.8027	0.0000
Decision trees	0.6966	0.0108
Logistic regression	0.8066	0.0000

Source(s): Author's own elaboration

similar to the developed method. Considering the obtained results, it can be concluded that the machine learning method developed based on grey system theory has also been positively validated in comparison to traditional machine learning algorithms.

5. Summary

This article makes a significant theoretical contribution to the field of employee turnover forecasting by integrating machine learning with grey system theory. The proposed grey clustering machine learning model, based on the grey cluster-based decision mechanism, introduces the concept of operating with grey numbers, which enables the representation of imprecise and uncertain data—characteristic of social science research. This approach allows for the consideration of the complexity and multidimensionality of the determinants of voluntary employee turnover, offering an innovative extension of traditional predictive models.

The final performance of the model was influenced by several key training parameters, particularly the number of learning epochs and the choice of learning rate. These were fine-tuned through preliminary experiments to achieve a stable training process and a satisfactory generalization performance. While not the primary focus of this study, their selection significantly impacted predictive accuracy and should be carefully calibrated in future implementations.

The practical application of the developed model is highly relevant for human resource management. With an accuracy exceeding 80%, it can serve as a decision-support tool in HR, enabling the early detection of employee turnover intentions. As a result, organizations can implement preventive measures, minimize costs associated with recruitment and training, and improve retention strategies. The integration of imprecise data into machine learning models allows for a better understanding of the nuances influencing employee decisions.

In real-world HR applications, several implementation challenges should be considered. First, although the model effectively handles data uncertainty, its interpretability may be limited for individuals without analytical training, which could hinder acceptance among decision-makers. Second, the implementation requires access to standardized employee data, which is often not systematically collected. Additionally, the preparation and deployment of such systems may entail technical and financial resources, especially when scaled.

Despite promising results, this study has certain limitations. The size and specificity of the dataset may limit the generalizability of the findings to other organizational contexts. Reliance on survey data, based on respondents' self-assessment, may introduce subjective biases. These may arise from social desirability, recall inaccuracies, or varying interpretations of questions. Respondents may, consciously or not, give answers aligned with expected norms rather than objective self-evaluation.

Additionally, while the transformation of qualitative variables into grey numbers is methodologically innovative, it may not fully reflect the depth of human perception and behavior. The model is also dependent on appropriate feature selection, which can affect its performance. Although the empirical part of this study was based on data collected in Poland using a structured employee survey, the developed method is not limited to this dataset. The model is architecture-independent from the content of specific variables and can be applied to other datasets, provided that predictors of voluntary turnover are available and can be expressed in the form of grey numbers.

Future research should validate the model using larger and more diverse datasets across sectors and cultures. Particular attention should be paid to cultural differences that may influence employee perceptions, communication norms, and turnover motivations. Since the grey machine learning model operates on subjective and imprecise input data, its effectiveness may vary depending on how social and organizational constructs are understood and expressed in different cultural contexts.

Further refinement of the transformation process for qualitative variables and exploration of alternative grey operators may increase accuracy, especially when dealing with culturally nuanced concepts such as job satisfaction, fairness, or loyalty. Additionally, integration with advanced machine learning techniques and the use of longitudinal datasets would allow for testing the model's predictive stability over time and across regions. Such research could contribute to a deeper, context-sensitive understanding of employee turnover dynamics and support the development of adaptable, culturally aware HR analytics tools.

Appendix

Algorithm Pseudocode

Input:

- Dataset D with N employee records and 31 features (converted to grey numbers)
- Learning rate η
- Maximum number of epochs E (e.g. 50)
- Reference vector R (grey values defining turnover intention threshold)

Output:

- Trained weight vector W

- Predicted labels Y_{pred}

Procedure:

1. Initialize:

- Convert all variables in D to grey numbers
- Standardize and encode nominal variables
- Perform correlation analysis → select 10 most relevant features
- Split D into training set D_{train} and test set D_{test}
- Initialize weight vector W with small random grey values

2. Training loop (epochs):

For epoch = 1 to E do:

For each object i in D_{train} :

- Compute weighted grey sum $G_i = \sum(W_j * X_{ij})$ for $j = 1$ to 10
- Compare G_i with grey reference R using grey relational ordering
 - If $G_i < R \rightarrow Y_{pred_i} = 1$ (turnover intention)
 - Else $\rightarrow Y_{pred_i} = 0$ (no turnover intention)
- If prediction incorrect:

For each feature j :

Update W_j :

$$W_j = W_j + \eta * (Y_{true_i} - Y_{pred_i}) * X_{ij}$$

3. After training:

- Predict labels for D_{test} using final weight vector W
- Evaluate performance (accuracy, confusion matrix, etc.)

Return W, Y_{pred}

References

- Allen, D.G., Bryant, P.C. and Vardaman, J.M. (2010), "Retaining talent: replacing misconceptions with evidence-based strategies", *Academy of Management Perspectives*, Vol. 24 No. 2, pp. 48-64, doi: [10.5465/amp.2010.51827775](https://doi.org/10.5465/amp.2010.51827775).
- Bolt, E.E.T., Winterton, J. and Cafferkey, K. (2022), "A century of labour turnover research: a systematic literature review", *International Journal of Management Reviews*, Vol. 24 No. 4, pp. 555-576, doi: [10.1111/ijmr.12294](https://doi.org/10.1111/ijmr.12294).
- Esangbedo, O.M. and Xue, J. (2025), *Grey Systems Theory in Business Management: Applying Multiple Criteria Decision-Making Methods*, MDPI.
- Eisenberger, R., Jones, J.R., Stinglhamber, F., Shanock, L. and Randall, A.T. (2005), "Flow experiences at work: for high need achievers alone?", *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, Vol. 26 No. 7, pp. 755-775, doi: [10.1002/job.337](https://doi.org/10.1002/job.337).
- Eshtaiwi, M.I., Badi, I.A., Abdulshahed, A.M. and Erkan, T.E. (2017), "Assessment of airport performance using the grey theory method: a case study in Libya", *Grey Systems: Theory and Application*, Vol. 7 No. 3, pp. 426-436, doi: [10.1108/gS-07-2017-0023](https://doi.org/10.1108/gS-07-2017-0023).
- Griffeth, R.W., Hom, P.W. and Gaertner, S. (2000), "A meta-analysis of antecedents and correlates of employee turnover: update, moderator tests, and research implications for the next millennium", *Journal of Management*, Vol. 26 No. 3, pp. 463-488, doi: [10.1177/014920630002600305](https://doi.org/10.1177/014920630002600305).

- Guo, S., Guan, L., Li, Q. and Jia, J. (2024), "Grey multi-criteria group consensus decision-making based on cobweb model", *Grey Systems: Theory and Application*, Vol. 14 No. 3, pp. 451-472, doi: [10.1108/gS-08-2023-0079](https://doi.org/10.1108/gS-08-2023-0079).
- Hakan, H., Canlier, E.C., Özari, Ç. and Can, E.N. (2024), "Modeling and predicting the socio-economic performance of countries using grey relational analysis and K-NN algorithm", *Journal of Grey System*, Vol. 36 No. 1, p. 4.
- Hausknecht, J.P., Rodda, J. and Howard, M.J. (2009), "Targeted employee retention: performance-based and job-related differences in reported reasons for staying", *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, Vol. 48 No. 2, pp. 269-288, doi: [10.1002/hrm.20279](https://doi.org/10.1002/hrm.20279).
- Holtom, B.C., Mitchell, T.R., Lee, T.W. and Eberly, M.B. (2008), "Turnover and retention research", *The Academy of Management Annals*, Vol. 2 No. 1, pp. 231-274, doi: [10.1080/19416520802211552](https://doi.org/10.1080/19416520802211552).
- Hom, P.W., Lee, T.W., Shaw, J.D. and Hausknecht, J.P. (2017), "One hundred years of employee turnover theory and research", *Journal of Applied Psychology*, Vol. 102 No. 3, pp. 530-545, doi: [10.1037/apl0000103](https://doi.org/10.1037/apl0000103).
- Hu, A. and Xie, N. (2025), "Construction and application of a novel grey relational analysis model considering factor coupling relationship", *Grey Systems: Theory and Application*, Vol. 15 No. 1, pp. 1-20, doi: [10.1108/gS-04-2024-0046](https://doi.org/10.1108/gS-04-2024-0046).
- Jia, L. and Pang, M. (2024), "A novel grey forecasting model with generalised fractal derivative and its optimisation", *Grey Systems: Theory and Application*, Vol. 13 No. 3, pp. 543-560, doi: [10.1108/gS-11-2023-0109](https://doi.org/10.1108/gS-11-2023-0109).
- Jones, L.D., Golan, D., Hanna, S.A. and Ramachandran, M. (2018), "Artificial intelligence, machine learning and the evolution of healthcare: a bright future or cause for concern?", *Bone and Joint Research*, Vol. 7 No. 3, pp. 223-225, doi: [10.1302/2046-3758.73.bjr-2017-0147.r1](https://doi.org/10.1302/2046-3758.73.bjr-2017-0147.r1).
- Kaplan, J. (2016), *Artificial Intelligence: What Everyone Needs to know*, Oxford University Press.
- Lee, T.W., Hom, P.W., Eberly, M.B., Junchao (Jason), Li and Mitchell, T.R. (2017), "On the next decade of research in voluntary employee turnover", *Academy of Management Perspectives*, Vol. 31 No. 3, pp. 201-221, doi: [10.5465/amp.2016.0123](https://doi.org/10.5465/amp.2016.0123).
- Li, L. and Li, X. (2024), "The grey decision model and its application based on generalized greyness of interval grey number", *Grey Systems: Theory and Application*, Vol. 14 No. 4, pp. 641-670, doi: [10.1108/gS-01-2024-0003](https://doi.org/10.1108/gS-01-2024-0003).
- Liu, Y. and Zhang, H. (2024), "Gray correlation analysis of strategic human resource management and corporate performance", *Applied Mathematics and Nonlinear Sciences*, Vol. 9 No. 1, 20240980, doi: [10.2478/amns-2024-0980](https://doi.org/10.2478/amns-2024-0980).
- Maertz, C.P., Jr. and Campion, M.A. (2004), "Profiles in quitting: integrating process and content turnover theory", *Academy of Management Journal*, Vol. 47 No. 4, pp. 566-582.
- Mallick, M.A. and Mukhopadhyay, S. (2024), "Boomerang recruitment: an intelligent model for rehiring using a grey-based multicriteria decision-making methodology", *Journal of Global Operations and Strategic Sourcing*, Vol. 17 No. 3, pp. 574-595, doi: [10.1108/jgoss-08-2022-0093](https://doi.org/10.1108/jgoss-08-2022-0093).
- March, J.G. and Simon, H.A. (1993), *Organizations*, John Wiley & Sons.
- Mitchell, T.M. (1997), *Machine Learning*, McGraw-Hill Science.
- Mobley, W.H., Horner, S.O. and Hollingsworth, A.T. (1978), "An evaluation of precursors of hospital employee turnover", *Journal of Applied Psychology*, Vol. 63 No. 4, pp. 408-414, doi: [10.1037/0021-9010.63.4.408](https://doi.org/10.1037/0021-9010.63.4.408).
- Naeini, E.Z. and Prindle, K. (2018), "Machine learning and learning from machines", *The Leading Edge*, Vol. 37 No. 12, pp. 886-893, doi: [10.1190/tle37120886.1](https://doi.org/10.1190/tle37120886.1).
- Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Le, H.V., Tran, V.Q., Prakash, I. and Pham, B.T. (2021), "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil", *Mathematical Problems in Engineering*, Vol. 2021 No. 1, pp. 4832864-15, doi: [10.1155/2021/4832864](https://doi.org/10.1155/2021/4832864).

- Nowak, M. (2024), *Prediction of voluntary employee turnover using machine learning*. Scientific Papers of Silesian University of Technology. Organization and Management/Zeszyty Naukowe Politechniki Slaskiej. Seria Organizacji i Zarzadzanie, No. 201, pp. 263-273.
- Nowak, M., Rabczun, A. and Łopatka, P. (2021), "Impact of electrification on african development-analysis with using grey systems theory", *Energies*, Vol. 14 No. 16, p. 5181, doi: [10.3390/en14165181](https://doi.org/10.3390/en14165181).
- Nowak, M. and Pawłowska-Nowak, M. (2024), "Dynamic pricing method in the E-commerce industry using machine learning", *Applied Sciences (2076-3417)*, Vol. 14 No. 24, p. 11668, doi: [10.3390/app142411668](https://doi.org/10.3390/app142411668).
- Othman, P.S., Ihsan, R.R. and Abdulhakeem, R.M. (2022), "The genetic algorithm (GA) in relation to natural evolution", *Academic Journal of Nawroz University*, Vol. 11 No. 3, pp. 243-250, doi: [10.25007/ajnu.v11n3a1414](https://doi.org/10.25007/ajnu.v11n3a1414).
- Price, J.L. (1977), *The Study of Turnover*, Iowa State Press, Ames, IA.
- Qin, D. and Zhao, B. (2024), "Hydrogen load demand prediction in unified energy system based on grey ridgelet neural network", *Journal of Grey System*, Vol. 36 No. 4, pp. 26-32.
- Qin, Y., Xie, N. and Wang, Y. (2025), "Capacity assessment for single batch machine production with interval grey processing time", *Grey Systems: Theory and Application*, Vol. 15 No. 3, pp. 403-424, doi: [10.1108/gS-11-2024-0139](https://doi.org/10.1108/gS-11-2024-0139).
- Rathnayaka, R.K.T. and Seneviratna, D.M.K.N. (2024), "Predicting of aging population density by a hybrid grey exponential smoothing model (HGESM): a case study from Sri Lanka", *Grey Systems: Theory and Application*, Vol. 14 No. 3, pp. 601-617, doi: [10.1108/gS-01-2024-0002](https://doi.org/10.1108/gS-01-2024-0002).
- Rosenblatt, F. (1957), The perceptron, A perceivingand recognizing automaton, project para, Report No. 85-460-1, Cornell Aeronautical Laboratory (CAL).
- Rosenblatt, F. (1958), "The perceptron: a probabilistic model for information storage and organization in the brain", *Psychological Review*, Vol. 65 No. 6, pp. 386-408, doi: [10.1037/h0042519](https://doi.org/10.1037/h0042519).
- Samuel, A.L. (1959), "Some studies in machine learning using the game of checkers", *IBM Journal of Research and Development*, Vol. 3 No. 3, pp. 210-229, doi: [10.1147/rd.33.0210](https://doi.org/10.1147/rd.33.0210).
- Sharma, D. and Kumar, N. (2017), "A review on machine learning algorithms, tasks and applications", *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, Vol. 6 No. 10, pp. 2278-1323.
- Tao, X. (2024), "The impact of corporate human resource management on corporate structure and image power based on grey correlation analysis", *Applied Mathematics and Nonlinear Sciences*, Vol. 9 No. 1, 20231217, doi: [10.2478/amns.2023.2.01217](https://doi.org/10.2478/amns.2023.2.01217).
- Tett, R.P. and Meyer, J.P. (1993), "Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta-analytic findings", *Personnel Psychology*, Vol. 46 No. 2, pp. 259-293, doi: [10.1111/j.1744-6570.1993.tb00874.x](https://doi.org/10.1111/j.1744-6570.1993.tb00874.x).
- Wang, H., Li, K., Wang, L. and Cheng, X. (2024), "Research on the optimization of flatness grey prediction control based on CNN-BiLSTM", *Journal of Grey System*, Vol. 36 No. 6, pp. 54-68.
- Wei, X., Li, H., Yue, W., Gao, S., Chen, Z., Li, Y. and Shen, G. (2022), "A high-accuracy, real-time, intelligent material perception system with a machine-learning-motivated pressure-sensitive electronic skin", *Matter*, Vol. 5 No. 5, pp. 1481-1501, doi: [10.1016/j.matt.2022.02.016](https://doi.org/10.1016/j.matt.2022.02.016).
- Yao, T. and Liu, X. (2024), "Research on China's GDP growth forecast based on grey machine learning model", *Journal of Grey System*, Vol. 36 No. 4, pp. 1-13.
- Zhang, L., Ma, X., He, Q. and Li, T. (2025), "A novel unbiased multi-variable grey model", *Grey Systems: Theory and Application*, Vol. 15 No. 2, pp. 209-238, doi: [10.1108/gS-07-2024-0090](https://doi.org/10.1108/gS-07-2024-0090).
- Zhang, L. and Li, X. (2024), "Entropy-weighted TOPSIS multi-attribute decision-making model and its applications based on generalized greyness", *Journal of Grey System*, Vol. 36 No. 5, pp. 15-26.

Corresponding author

Marcin Nowak can be contacted at: marcin.nowak@put.poznan.pl

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com