Building prediction models with grouped data: A case study on the prediction of turnover intention

Yuan, S.; Kroon, B.; Kramer, A.

DOI
10.1111/1748-8583.12396

Publication date
2024

Document Version
Final published version

Published in
Human Resource Management Journal

Received: 4 January 2020 | Revised: 17 April 2021 | Accepted: 14 June 2021

DOI: 10.1111/1748-8583.12396

SPECIAL ISSUE

Human Resource Management Journal WILEY

Building prediction models with grouped data: A case study on the prediction of turnover intention

Shuai Yuan[1] | **Brigitte Kroon**[2] | **Astrid Kramer**[3]

[1]Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

[2]Department of Human Resource Studies, Tilburg University, Tilburg, The Netherlands

[3]Department of Management, Tilburg University, Tilburg, The Netherlands

Correspondence
Brigitte Kroon, Department of Human Resource Studies, Tilburg University, Warandelaan 2, PO Box 90153, 5000 LE Tilburg, The Netherlands.
Email: b.kroon@tilburguniversity.edu

Abstract

The availability of big data spurred the application of modern prediction analytics (e.g., machine learning methods) in human resource management (HRM) research and practice. Due to the novel and technical nature of prediction analytics, HR professionals and researchers may struggle to collaborate with data experts. We offer a comprehensive introduction to the logic and value of prediction methods. Moreover, we highlight the concern of treating grouped data—commonly seen in HRM research yet rarely discussed in building prediction models. We introduce different strategies to deal with grouped data in applying prediction models. The performance of different modelling approaches and prediction models are compared in an empirical data set consisting of 1454 employees from 199 small and medium sized enterprise's. Following a workflow to compare the relative performance of the prediction models, the model with the best prediction accuracy was the random-effects bagged tree that allows for complex relationships and incorporates random effects. Following the estimates of this model, we identified the five

Abbreviations: CRE, correlated random effects; CV, cross-validation; FE, fixed effects; FE_O, fixed effects, dummy coding; FE_T, fixed effects, target coding; HR, human resources; HRM, human resource management; LMM, linear mixed model; LMX, leader-member exchange; PLMM, penalized linear mixed models; RE, random effects; RE-EM tree, random-effects expectation-maximization tree; RMSE, root mean square error; SL, single level; SME, small and medium sized enterprise.

most influential predictors of turnover intention: perceived fairness, leader-member exchange, career opportunities, pay satisfaction and age. The inductive nature of prediction models is expected to advance theory development and HR analytics for developing effective HRM policies.

KEYWORDS
attitude survey, hierarchical linear modelling, HR analytics, metrics, modern prediction models, multi-level modelling, turnover

1 | INTRODUCTION

In the era of 'Big Data' in human resource management (HRM), machine-learning methods increasingly find their way to human resource (HR) analytics (e.g., Oswald et al., 2020). The objective of machine learning methods is to make accurate predictions about what will happen to an outcome in the future, by integrating and processing a large variety of information and without formulating theory-based hypotheses beforehand (Yarkoni & Westfall, 2017). The machine learning approach, an inductive approach to data analysis, allows HR data analysts to effectively translate large volumes of information into straightforward and concise messages for managerial decisions. Besides, prediction models provide unprecedented opportunities for research, in particular for theory development (Choudhury et al., 2020), and for overcoming the problem of limited generalizability of findings to other than the investigated sample (Yarkoni, 2020).

Despite the analytical potential of prediction models, most HR practitioners are relatively unfamiliar with them (McCartney et al., 2020). This strengthens the concern that HR analytics will be controlled by non-HR data analysts with less knowledge of typical HR issues and HR data structures (Marler & Boudreau, 2017), which can lead to misspecified analyses as input for strategic HRM decisions. Therefore, prudence is required, since the flexibility embedded in prediction models does not mean they could be applied blindly and solve all modelling issues one might face. In this study, we explain the value of data-driven methods for predicting HR outcomes in non-technical terms. We take the prediction of voluntary turnover intentions of employees as an example to highlight modelling issues that also confront modern prediction models. In particular, we investigate how the typically grouped nature of HRM data—gathered in teams, departments and organizations—can be accounted for in prediction models.

Voluntary employee turnover is one of the most studied outcomes in HRM (Holtom et al., 2018), as high levels of turnover induce operational costs for rehiring, decreased productivity, and the destruction of human and social capital (Park & Shaw, 2013). In the era of scarcity of talent, understanding factors predicting employee retention have large practical relevance. Over the past century, models explaining voluntary turnover have become increasingly complex, encompassing a broad range of predictor variables including employee demographics, personal resources, work attitudes, and affective reactions, as well as factors resting in various levels of personal, workplace, organizational and labour market characteristics (Hom et al., 2017). Recent publications call for acknowledging this complexity among relations and levels of turnover predictors (Rubenstein et al., 2018) and exploring the additional value of 'big data' in understanding turnover and retention (Hausknecht et al., 2016). A modelling issue in turnover research is that predictors exist at different levels (Hom et al., 2017). Data of grouped structure (also called multilevel structure, hierarchical structure, or nested structure) prevail in HRM research: Measurements are embedded in individuals, individuals in teams, teams in organizations and organizations in countries (Peccei & van de Voorde, 2019). How to appropriately model grouped data is extensively reported in the *explanation* paradigm (e.g., McNeish et al., 2017; Peccei & van de Voorde, 2019), wherein studies aim at estimating the relationships between variables as accurately as possible. However, in the *prediction* paradigm where studies

Practitioner notes

What is currently known?

- Human resource management analyses often involve employee data that are grouped in teams or organizations. We know that neglecting grouped data structures leads to inaccurate predictions in correlational analyses.
- When data analysts without a background in human resource (HR) or psychology perform HR analytics, there is a risk that they may mis-specify the analyses and (or) misinterpret the results.
- HR practitioners and management, who use HR analytics reports to design policies, often lack skills to evaluate advanced predictive analysis (e.g., machine learning, AI). In case of misspecified analysis or misinterpreted results, this can result in the design of less effective policies.

What this study adds?

- The study demonstrates that neglecting grouped data structures leads to less accurate predictions about employee turnover intentions, also when advanced predictive analyses are used.
- The study explains in non-technical language what prediction models are, illustrated by the example of predicting employee turnover intentions.
- The study proposes a workflow to determine which prediction model best suits the grouped data to achieve the best prediction. The method that performed best in the current analysis is the one that accounted for the grouped structure of the data and for complex relationships between the predictors and employee turnover intentions.

Implications for practitioners:

- For HR practitioners who are familiar with correlational analyses, the study is an easy to read introduction to prediction models for grouped data, which can facilitate their cooperation with data analysts.
- Prediction models are 'theory free'; there is no need for theoretically grounded hypotheses before running the analyses. Identifying the most important predictors of a key outcome can turn HR practitioner's attention to more specific follow-up actions (e.g., focus groups with employees) to develop well-grounded, evidence-based HR practices.

mainly aim at generating the most accurate predictions possible, the problem of modelling grouped data has been largely neglected. To our knowledge, it has not been explicitly addressed in any theoretical or empirical study in the field. As a result, HR data analysts are likely to be confused about and (or) unaware of different modelling approaches and techniques for constructing accurate predictions in grouped data.

In this study, we introduce different modelling approaches for analysing grouped data with prediction models, detail the procedures to use these approaches, and compare the relative prediction accuracy of each model using an empirical data set. Our contribution is threefold. First, we offer an introduction about modelling approaches to readers who are new to modern prediction models. Second, we illustrate how to apply these approaches and how to train and test prediction models in an empirical analysis. Finally, we provide practical recommendations to HRM researchers and practitioners to make informed decisions in building prediction models with grouped data. Although our contribution might be somewhat technical in nature, we believe that it facilitates in opening conversations between HRM professionals and data experts, which are currently taking place in separate publication domains (Garg et al., 2021).

2 | THEORY

2.1 | Introduction to prediction models

Prediction is central in the HRM—performance paradigm, which concentrates on the value of HRM practices and systems as management tools for influencing outcomes of interest to organizations, teams, or individuals. Prediction implies that research is future-oriented, using current information to forecast what will happen to the outcome (Yarkoni & Westfall, 2017). Prediction is distinct from explanation, which primarily concentrates on accurately describing the causal underpinnings of outcomes. Where theory drives the models tested in explanatory research, prediction models depart from the data without making assumptions beforehand (Putka et al., 2018).

Although explanatory methods are dominant in HRM—performance research, prediction models could make at least three important contributions. First, prediction models are able to incorporate large numbers of variables and offer great flexibility in modelling. It allows for example for non-linear relationships between variables, whereas explanatory methods typically select a few variables based on theory and limit to modelling linear relationships between variables. In many practices of HR analytics (e.g., for a recruiter to determine the future job performance of each applicant based on their test results and previous work experience), prediction models are suitable choices, because their primary objective is to make accurate predictions by integrating and processing a large variety of information, rather than to provide the most convincing explanations of causality by systematically ruling out all confounders. Second, critics argue that many behavioural science studies and theories lack generalizability, because their applicability is constrained to a specific sample or a particular subgroup: the findings obtained therein can hardly be extended to other samples and scenarios (Yarkoni & Westfall, 2017; Yarkoni, 2020). In contrast to a common explanatory model that exploits links between the outcome(s) and the predictors *in* the sample, prediction models aim to produce accurate predictions for entries coming *out of* the sample—which is the essence of generalizability. Third, in addition to testing theories, prediction models are also useful for theory development, as they could identify predictors that are neglected in existing theories and thereby provide exciting opportunities for follow-up studies (Choudhury et al., 2020).

2.2 | The essence of prediction models: hyper-parameters and model training

The objective of a prediction analysis is to predict, as accurately as possible, an outcome of interest (e.g., employee turnover intention) with a set of predictors. The goal is to determine a set of rules, often formulated as mathematical equations that can be used to infer the value of the outcome from the set of available predictors. The set of rules is typically adapted from a well-established prediction model. The regression model can be considered as a traditional prediction model, while its modern variants include regression trees (Breiman et al., 1984), lasso regressions (Tibshirani, 1996), support vector machines (Hearst et al., 1998), random forests (Breiman, 2001), and, more recently, convolutional neural networks (Albawi et al., 2017).

To fit the prediction model close to the data, many modern prediction models include one or multiple *hyper-parameters* (also known as tuning parameters; Kuhn & Johnson, 2013) that control the complexities of prediction. Tuning parameters indicate how well a complex model describes a specific sample as well as how generalizable it is to other data sets (Kuhn & Johnson, 2013). The values of these hyper-parameters are not selected a priori but optimized in the analysis. The procedure to employ any prediction model involves a 'training—testing' process, in which the data set under consideration is split into a training set and a test set. The training set serves to determine the prediction model, and the test set serves to test the trained model. The test set consists of data that are not in the training set, which offers an effective way to assess the generalizability of the prediction model. When hyper-parameter(s) are involved, the training process is typically done via the cross-validation (CV) routine—a resampling-based procedure to assess how well a model constructed in one sample could be generalized to another. Depending on the number of subsets created, it is called k-folds CV where k is the number of subsets. Let us take 10-folds CV as an example. In this process, the training sample is randomly divided into 10 equal subsamples (or folds). At each time, nine of the folds are used to fit the model at each possible value (or, in case of multiple tuning parameters, at each possible combination) of the tuning parameter(s), and the resulting model will be tested on the remaining fold. This process will be repeated 10 times. The average prediction accuracy across these 10 tests is compared and the tuning parameter(s) is (are) selected such that the average prediction accuracy is maximized.

When grouped data is under consideration, however, caution should be paid to the sampling procedures embedded in the CV routine. More specifically, statisticians have suggested to design the resampling procedure in such a way that it resembles the data generation procedure (e.g., Fox, 2008). In the presence of a grouped data structure, it is considered that the folding procedure is first applied in the higher level, and then, conditional on the group membership, further used in the lower level (Davison & Hinkley, 1997). To closely follow this idea, the folds in the CV routine should be randomly generated with a special restriction: observations belonging to the same group should always be placed into the same fold (see also Deen & de Rooij, 2020).

2.3 | Five modelling approaches to multi-level data

To increase the understanding of the modelling approaches, we show an example that aims to build prediction models on an illustrative data set consisting of multiple groups (see Table 1). We consider a prediction question where the goal is to predict, as accurately as possible, employee's turnover intention based on a series of predictors at both the firm level and individual level. Table 1 displays the responses of six participants on several variables, including their self-reported turnover intentions (Y; the outcome variable), the firm to which participants belong (G; the group indicator), a set of group-level predictors ($\boldsymbol{X}_{h;}$ e.g., firm), and a set of individual-level predictors ($\boldsymbol{X}_{l;}$ e.g., self-reported fairness perception). In empirical data sets, $\boldsymbol{X}_h$ and $\boldsymbol{X}_l$ can potentially include a large number of predictors.

2.3.1 | The single-level approach

The simplest—and arguably the most straightforward—approach to tackle the prediction question described above is to treat the data set as a single level, that is, neglecting the variable 'Firm'. With this approach, the objective is to determine—based upon the sample—a prediction function **f** that infers Y from $\boldsymbol{X}_h$ and $\boldsymbol{X}_l$ (but not the group indicator G): $Y = \mathbf{f}\ (\boldsymbol{X}_h, \boldsymbol{X}_l)$. However, this approach neglects the dependence of responses collected from the participants of the same firm: for example, participants #2 and #3 work in the same firm, therefore their turnover intentions might be influenced by the same cause of variations (e.g., a recent change in HR policy).

TABLE 1 An illustrative grouped data set

Respondent	Turnover intention (Y)	Firm (G)	Firm Size (X_h)	Fairness perception (X_l)	Other firm-level predictors	Other individual-level predictors
#1	1	A	20	1	...	...
#2	4	B	5	2	...	...
#3	3	B	5	5	...	...
#4	5	C	10	6	...	...
#5	6	C	10	2	...	...
#6	5	C	10	4	...	...

To address this issue, researchers from different academic disciplines have proposed various approaches to account for the grouped structure: the fixed effects (FE) approach in economics and sociology (Bell et al., 2019), the random effects (RE) approach in psychology and management (Bliese et al., 2007; McNeish et al., 2017) and the correlated random effects (CRE) approach (Antonakis et al., 2021).[1] Next, we apply each approach to the context of the illustrative data set.

2.3.2 | The FE approach

The FE approach models the grouped structure by including group affiliation indicator(s) directly into the model as predictor(s). A widely used way to construct these variables is to generate a set of dummy variables. In the machine learning literature, this strategy is also referred to as one-hot encoding (Rodríguez et al., 2018). In our illustrative sample, one-hot encoding the grouping variable G (Firm) results in three dummy variables: D1 ($A = 1, B = 0, C = 0$), D2 ($A = 0, B = 1, C = 0$) and D3 ($A = 0, B = 0, C = 1$). Generally, when one-hot encoding is applied, K dummy variables pertain to K groups. The set of dummy variables explains all variability at the group level and therefore none of the group-level predictors (i.e., $\boldsymbol{X}_h$) should be entered as predictors. The FE approach that incorporates dummy variables (called FE_O approach hereafter) aims to build up a prediction model that calculates Y based upon $\boldsymbol{X}_h$, $\boldsymbol{X}_l$, as well as the set of dummy variables $\boldsymbol{D}$: $Y = \mathbf{f}\ (\boldsymbol{X}_l, D)$.

Despite its popularity, the FE_O approach suffers an important shortcoming. In the presence of an extensive number of groups, unavoidably, the prediction model has to encompass a large number of dummy variables and becomes overwhelmingly complex. An alternative strategy[2] is to apply a target encoding scheme (called FE_T approach in the current study, e.g., Abraham et al., 2014). Here, a single numeric variable (T) is generated to replace the grouping variable (G): for a respondent that belongs to group A, its corresponding value in T is the average value of the outcome variable (Y) across all respondents in group A. Therefore, in our illustrative data set, #1 scores 1 on the numerical group indicator T, #2–#3 score 3.5, while #4–#6 record 5.33. Unlike dummy variables, the single numeric variable does not explain all variations in the group; hence, the prediction function of the FE_T approach can be written as: $Y = \mathbf{f}\ (\boldsymbol{X}_h, \boldsymbol{X}_l, T)$. The implementation of the FE approach is straight-forward: built on the single-level (SL) approach, it includes additional group indicators into the prediction models.

2.3.3 | The RE approach

The RE approach accounts for the grouped structure of the data with RE (Laird & Ware, 1982). RE capture how much the effects pertained to a specific group differ from the average effects across the general population. More technically, the effects of each group form a distribution (often assumed to be normally distributed) where the group-specific RE are unique data points (McNeish & Kelley, 2018). For simplicity, here we only focus on the RE of intercepts (also called random intercepts). We assume that respondents from different firms only differ in the average level of turnover intention, but not in the relationships between predictors and turnover intention. The prediction question can be expressed in $Y = \mathbf{f}\ (\boldsymbol{X}_h, \boldsymbol{X}_l) + U_g$, where U_g is the group-specific intercept that pertains to Group g.

Compared with the SL approach, the addition of RE largely complicates model estimation, as RE have to be estimated with an iterative procedure. To our knowledge, very few RE prediction models have been proposed. Examples include linear mixed model (LMM) based on linear regression, penalized linear mixed models (PLMM; Schelldorfer et al., 2014) based on lasso regression, and random-effects expectation-maximization tree (RE-EM tree) based on regression trees (Sela & Simonoff, 2012).

2.3.4 | The CRE approach

The RE approach has been criticized for not separating individual effects (i.e., the effects at the lower level, e.g., the effect of one's fairness perception on his turnover intention) from context effects (i.e., the effects at the higher level, e.g., the effect of the group's fairness climate on one's turnover intention; e.g., Antonakis et al., 2021; Bell et al., 2019). A simple solution, referred to as the correlated random effects approach or the CRE approach (e.g., Antonakis et al., 2021; Bell et al., 2019; Hamaker & Muthén, 2020), is to slightly modify the RE approach by group-mean centring the lower-level predictors and adding these group means as extra predictors. These group means could be interpreted as aspects of organizational climate. For example, the effects of the group average of fairness perception could be interpreted as the context effects of fairness climate. The CRE approach is expressed in the formulation: $Y = \mathbf{f}\ (\boldsymbol{X}_h, \bar{\boldsymbol{X}}_l, \boldsymbol{X}_l - \bar{\boldsymbol{X}}_l) + U_g$, where $\bar{\boldsymbol{X}}_l$ denotes the group means of the lower-level predictors.

The implementation of the CRE approach is largely identical to the RE approach, with the exception that the lower-level predictors are group-mean centred and that these group means are entered in the prediction model as additional predictors.

2.3.5 | The five approaches compared in a prediction question

Although the five modelling approaches have been discussed and compared extensively in the *explanation* paradigm, their accuracy in making *predictions* is largely unknown. In Table 2, we summarize the formulation, strengths and limitations of these five modelling approaches. Here, we briefly discuss three of the theoretical observations. First, except for the SL approach, the other four approaches all account for group-specific effects. Although this capability is always conceived as an important strength, it may transfer into a deficiency when few employees in a firm took the survey, because this would result in erroneous estimations of firm-specific effects. To illustrate, only 1

TABLE 2 The formulation, strength, and limitations of the five modelling approaches

Modelling Approaches	Inputs	Strengths	Limitations
Single-level (SL)	$\boldsymbol{X}_h$, $\boldsymbol{X}_l$	Easy to implement	Neglects the grouped structure completely
Fixed effects with one-hot encoded group indicators (FE_O)	$\boldsymbol{X}_l$, D	Captures group-specific effects via dummy variables	Exceedingly complex with a large number of groups
Fixed effects with target encoded group indicators (FE_T)	$\boldsymbol{X}_h$, $\boldsymbol{X}_l$, T	Captures group-specific effects via a target variable	The prediction models trained may be too specific to the training set, as the information about the outcome has been used to generate predictors (could be potentially mitigated, see the main text)
Random effects (RE)	$\boldsymbol{X}_h$, $\boldsymbol{X}_l$	Captures group-specific effects with random effects	Mixes the individual effects (i.e., the effects of fairness perception) with the context effects (i.e., the effects of fairness climate)
Correlated random effects (CRE)	$\boldsymbol{X}_h$, $\bar{\boldsymbol{X}}_l$, $\boldsymbol{X}_l - \bar{\boldsymbol{X}}_l$	Captures group-specific effects with random effects and accounts for the context effects	Typically uses a large number of predictors and hence may result in prediction models that are not generalizable

Note: $\boldsymbol{X}_h$ = predictors from the higher level, $\boldsymbol{X}_l$ = predictors from the lower level, D = a set of dummy variables obtained by one-hot encoding the group indicator, T = a numerical grouping variable obtained by target encoding the group indicator, $\bar{\boldsymbol{X}}_l$ = the calculated group means of predictors from the lower level.

employee from our illustrative data set (respondent #1) comes from firm A; therefore, it is not possible to distinguish the effects that are specific to the firm from those specific to the individual (#1). Second, the presence of a large number of groups introduces a large number of dummy variables in the FE_O approach and may thereby reduce its comparative predictability and efficiency. The application of the FE_T approach effectively addresses this limitation, however, in its current format, it uses the information of the outcome directly to generate its predictors, causing the problem known as target leakage. Target leakage may result in prediction models that are too specific to the sample. Although potential remedies have been proposed to mitigate this problem, they are not likely to fully address the problem. Third, although the CRE approach takes organizational climate as additional predictors and thereby offers higher flexibility, the fact that it bears a larger number of predictors, compared to the other approaches, makes it more prone to overfitting (i.e., fit well to the observations in the sample, but poorly to the observations out of the sample), especially, with a small-to-medium sample size. In the next section, we illustrate the application of the five modelling approaches, combined with four prediction functions (i.e., different forms of **f**, including regression, lasso regression, regression trees and bagged trees) in an empirical analysis, and demonstrate the procedures to determine which modelling approach fits the illustrative data set best on employee turnover intentions in small and medium-sized enterprises. The current analysis and the accompanying code serve as an illustration that can be easily adapted in analysing other grouped data. Furthermore, since these prediction functions have only been introduced to HRM researchers on several occasions (e.g., Putka et al., 2018; Oswald et al., 2020; Spisak et al., 2019), they will be briefly explained in the next section.

3 | METHODS

3.1 | Data

A research team collected data of 1510 employees in a convenience sample of 199 small and medium-sized enterprises (10–250 employees) in 2017–2019, according to an ethically approved procedure, which entailed a signed data agreement by the managing director of the firm and informed consent by employees. Two questionnaires were distributed, one for the managing director of the organization and one for employees. To ensure the quality of the prediction analysis we only retained respondents who provided complete answers to the outcome of interest: turnover intention. As a result, a total of 1454 respondents were used in the following analyses.

The variable of interest turnover intentions, was measured with three items developed by Valentine et al. (2006) in the employee questionnaire. According to Valentine et al. (2006), the scale is a valid measure that relates thoughts about the likelihood of leaving with expected job search behaviours within a 3-year timeframe. Next to demographic characteristics like education, age, the current type of contract, and management responsibilities, the employee questionnaires further contained validated measures from previous research on a range of job attitude variables that have been related to employee turnover intentions, such as leader-member exchange (LMX), perceived fairness, pay satisfaction and job proactivity. In order to ensure the generalizability of the sample within organizations, questionnaires were obtained either digitally or hardcopy from at least 30% of all employees in each organization. The questionnaire for the managing director covered contextual variables such as organization size, firm complexity, governance (supervisory board, works council, HR professional) as well as information concerning the organization's level of corporate entrepreneurship, strategic planning and the entrepreneurial orientation of the managing director. A detailed description of the measures and their origins in each questionnaire, as well as the scale statistics, means and standard deviations are available in the Data S1.

The largest group of businesses in the sample operated in the service industry (54.8%), followed by manufacturing and agriculture (20.1%), healthcare, education, and recreation (14.6%), and logistics (10.6%). Firm size as expressed by the average amount of full-time equivalents (FTE's) is 22,92 (SD = 31,850). Given that a substantial proportion of employees work part-time, the average number of real employees per organization is actually higher.

The proportion of family firms in the sample is 60.5%, which corresponds with the estimate by the European Commission (nd). Of the 1510 employee respondents, 57% are male and 38% have acquired at least a bachelor's diploma. The average age is 38,44 (SD = 12,867). Further details on the sample can be read in the Data S1.

3.2 | Design of prediction model comparisons

In our empirical illustration, we examine and compare the prediction accuracy of a total of 20 model configurations that combine the two factors: (1) the modelling approaches, including the SL approach, FE_O approach, FE_T approach, RE approach and CRE approach; and (2) the form of prediction functions, including linear regression, lasso regression, regression trees and bagged regression trees. These four prediction functions were selected because their extensions that incorporate RE have been implemented and well-studied. Below we offer a brief introduction of the three prediction functions—except for the well-known linear regression—and refer interested readers to Putka et al. (2018) for a more detailed discussion.

Lasso regression (Tibshirani, 1996) has been invented to address the shortcoming of linear regression that all predictors—even those irrelevant ones—are retained in the final model, causing inefficiency in predictions as well as complications in interpretations. In essence, lasso regression performs automatic variable selection such that only the few variables with the largest predictive values are retained in the model and the other values are deemed as irrelevant variables, whose regression weights are replaced with zeros. Applied in our illustrative data set, lasso regression may produce an exact zero weight for the variable 'firm size', which implies this variable is not at all relevant to our prediction of one's turnover intention.

Regression trees have been developed to allow for higher-order interactions that are not allowed in ordinary regressions (Breiman et al., 1984). In essence, the regression tree is a recursive partitioning method that, in each step, performs a two-way split to improve the prediction performance criterion, based on the values of a single variable (represented as a node in the visualization of a regression tree). Once fully grown, a regression tree tends to be very large, and, just like the consequence of having too many predictors in a linear regression model, prone to over-fitting. To prevent this issue, during the growth of a decision tree, its size is constrained—or, put it in another way, the tree is pruned—by a number of parameters whose values are determined in training the sample.

However, even with carefully studied pruning procedures, a single tree is still highly prone to overfitting. To further account for this limitation, Breiman (1996) proposed a bootstrap resampling method to develop multiple trees for the prediction question, which is referred to as *bagged trees*. The core idea of this prediction model is to create a number of different trees, each of which is developed from a bootstrap resampling scheme—each bootstrap sample, whose size equals the size of the original sample, is constructed by sampling *with* replacement from the original sample. The final prediction results from the aggregation of all bootstrapped trees: it takes the average of all individual trees. It is interesting to note that, unlike regression trees, estimating bagged trees does not require pruning individual trees, as its final step of averaging has already reduced the adverse effects of overly complex trees.

In combining the five modelling approaches and the four prediction functions, we arrive at a total of 20 model configurations. As discussed above, while the SL approach, the FE_O approach, the FE_T approach utilize prediction functions *without* random effects, the RE approach, and the CRE approach employ prediction functions *with* RE. Therefore, if FE models and RE models are considered separately, a total of eight forms of prediction functions are applied in the current analysis. Table 3 lists the prediction models, the specialized packages that implement these methods, as well as some details on model training (i.e., the hyper-parameter to be tuned, as well as the grid from which it is selected). As noted above, we only included random intercepts (but not random slopes) in all RE models.

Furthermore, we investigated the performance of these 20 models in two types of prediction questions: whether a new respondent to be predicted belongs to (1) an unknown firm—for example, a new employee of an organization that is not yet in the data set, or (2), to a known firm—for example a new employee of an organization that is currently in the data set.

TABLE 3 The package used and the tuning parameters specified to train the 20 prediction models

Modelling approaches	Settings	1. Regressions	2. Lasso regression	3. Regression trees	4. Bagged regression trees
Prediction models without random effects:	*R* package	Base *R*	*GLMMLasso*	*caret* + *rpart*	*caret* + *rpart*
	Tuning parameter	/	λ	*cp*	/
1. SL, 2. FE_O, 3. FE_T	Grid of tuning parameter used	/	1,2,…,14	0.001, 0.003, …, 0.031	/
Prediction models with random effects	*R* package	*lme4*	*GLMMLasso*	*REEMtree*	*REEMtree*
	Tuning parameter	/	λ	*cp*	/
1. RE, 2. CRE	Grid of tuning parameter used	/	1,2,…,14	0.001, 0.003, …, 0.031	/

Note: All of the analyses reported were carried out in R version 3.5. No hyper-parameters were used in training multiple regressions and bagged trees, either with or without random effects.

Abbreviations: CRE, correlated random effects; $FE_{O,}$ fixed effects with one-hot encoded group indicators; $FE_{T,}$ fixed effects with target encoded group indicators; RE, random effects; SL, single-level.

3.3 | Analytical procedure

Figure 1 presents a schematic overview of the analytical procedure. First, we randomly divided the full sample of 1454 respondents into a training set and a test set (step 1). The training set included roughly 9/10 of the respondents and the test set included roughly 1/10. Two types of data splitting were used. The first type restricted all firms presented in the test set to *not* occur in the training set, while the second type demanded that all firms presented in the test set should *also* occur in the training set. Specifically, the first type concerns prediction on respondents from *new* groups while the second type concerns prediction on respondents from *existing* groups. Furthermore, the sample sizes of the **test sets** resulting from the two types of splits were kept constant to partial out the possible confounding effect of different sample sizes.

In step 2, the process of CV, we selected the optimal value of the hyper-parameter from all candidate values—also called a tuning grid—based on prediction accuracy computed upon the training set. Step 2 is only applied to models that included hyper-parameters (i.e., the prediction models based on lasso regression and regression trees). Here, prediction accuracy was quantified by root mean square error between the observed outcome values and the predicted outcome values, where a smaller value corresponds to better prediction accuracy.

In step 3, we trained each model on the training set. If applicable, the fine-tuned value of the hyper-parameter was specified. Then, in step 4, we employed each fitted model, resulting from Step 3, to predict the values of the outcome (turnover intention) of respondents in the test set. The resulting prediction accuracy, quantified by R^2 (i.e., squared correlation between the observed outcome values and the predicted outcome values), indicates the performance of a prediction model, such that a higher R^2 suggests a better performance in prediction.

To yield a stable estimation of (relative) model performance, in step 5, we performed data partition and repeated steps 1 to 4 for 200 times (as suggested by Kuhn & Johnson, 2013). This way, not only the prediction performance of these prediction models could be compared, but we were also able to quantify the certainty about the conclusion drawn.

In the final step, we conducted a supplementary analysis to illustrate potential insights that could be obtained from these prediction models. More specifically, we identify the variables that were of the greatest *predictive*

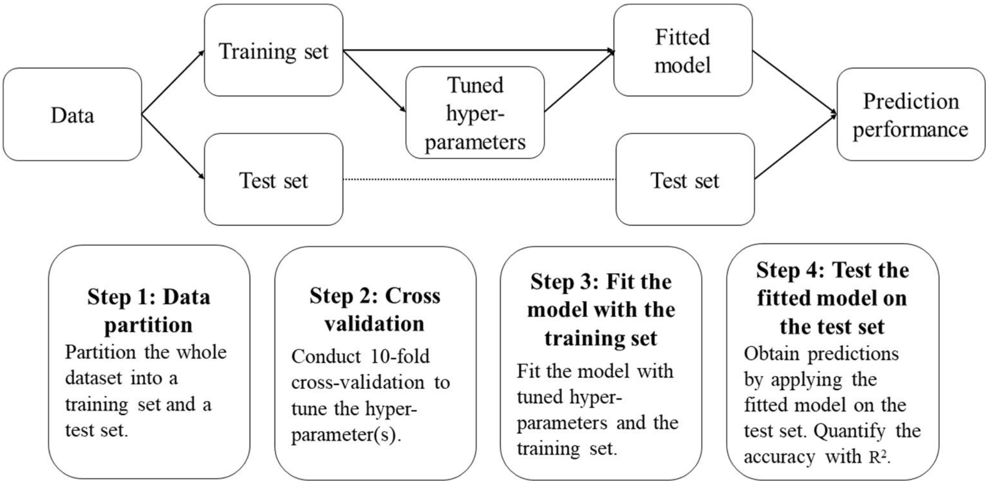

FIGURE 1 The workflow of data analysis in the current study

importance to turnover intentions by examining the relative importance metric—a set of widely used indicators that are generated automatically in model training and that quantify the relative contribution of each predictor—of the models with the best prediction accuracy. The steps to calculate this relative importance metric has been explained in Putka et al. (2018).

4 | FINDINGS

Figure 2 presents the prediction accuracy of each model in predicting out-of-sample respondents across 200 repetitions, quantified by percentages of explained variance (*i.e.*, R^2) of the outcome variable (turnover intention). The results are plotted according to the type of predictions, that is, predicting respondents coming from a new group (the *left* part) versus predicting respondents coming from an existing group (the *right* part).

Several findings emerge from Figure 2. First, when predicting a new respondent from a firm that *not* appeared in the training sample, the two models that predicted most accurately are the bagged trees that adopted the RE approach (average R^2 = 0.32, SD = 0.07) and the CRE approach (average R^2 = 0.31, SD = 0.07). These two models have a clear advantage over all other 18 models. However, this advantage diminished in predicting a new respondent from a firm that appeared in the training sample. In these cases, there are seven models that predict almost equally well: the regressions with the SL (average R^2 = 0.279, SD = 0.07), RE (average R^2 = 0.279, SD = 0.07) and CRE (average R^2 = 0.275, SD = 0.07) approach, the lasso regression with the SL (average R^2 = 0.279, SD = 0.07), RE (average R^2 = 0.279, SD = 0.07) and CRE (average R^2 = 0.275, SD = 0.07) approach; and the bagged trees that adopt the RE approach (the average R^2 = 0.276, SD = 0.07). Taken both types of prediction (i.e., predicting the outcome of new respondents from new groups or existing groups) into account, the current analysis shows that the bagged trees with the RE approach performs best, as it is the only model that performed well on both tasks.

Second, of the four types of prediction functions (i.e., regressions, lasso regressions, regression trees and bagged trees), regression trees performed the worst, as none of the five prediction models that employed regression trees have an average R^2 of more than 0.2 in either type of prediction. In sharp contrast to regression trees, both the simplest function (i.e., regressions; the average R^2 = 0.261) as well as the most complex function examined (i.e., bagged trees; the average R^2 = 0.257) performed adequately, especially, when coupled with the RE approach.

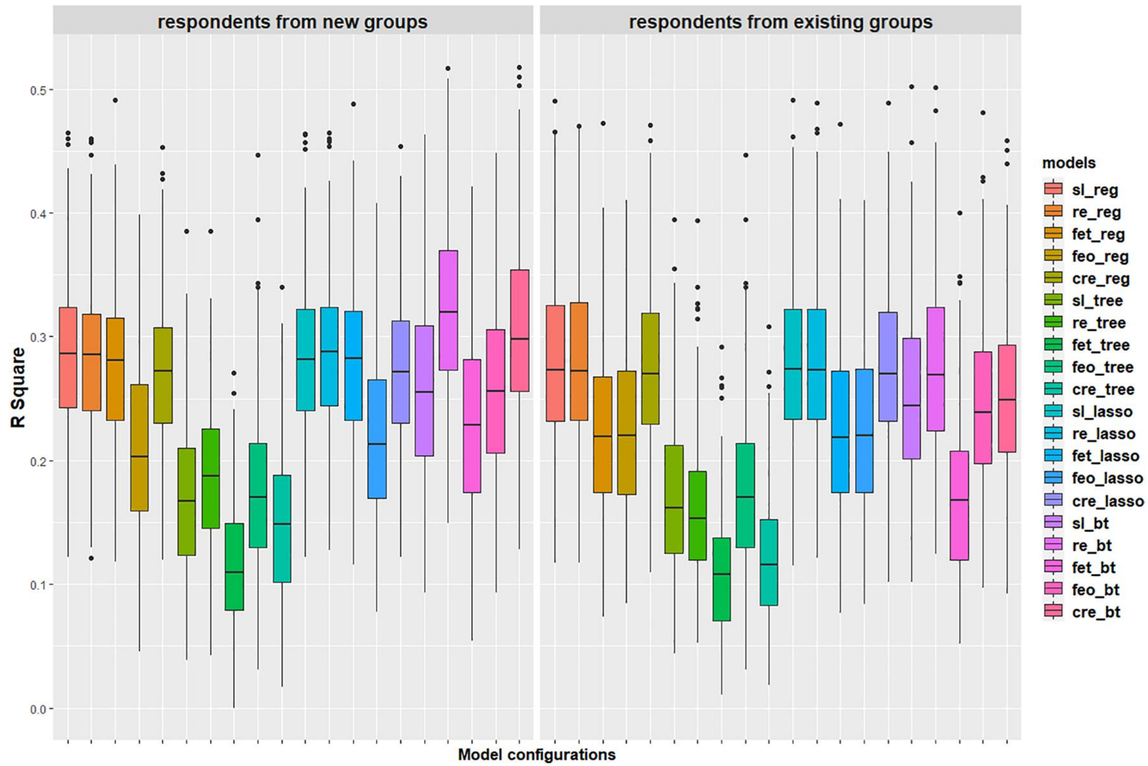

FIGURE 2 The prediction accuracy of each method in predicting out-of-sample respondents. The prefix of the model name indicates the modelling approach: *sl* = the single-level approach, *re* = random effects approach, fe_t = fixed effects with target encoders, fe_o = fixed effects with one-hot encoders, *cre* = correlated random effects. The suffix indicates the type of prediction models: *reg* = regression, *tree* = regression trees, *lasso* = lasso regression, *bt* = bagged random-effects expectation-maximization trees [Colour figure can be viewed at wileyonlinelibrary.com]

Third, of the five modelling approaches (SL, FE_O, FE_T, RE and CRE approach), the RE approach resulted in the best prediction accuracy (average R^2 = 0.26) while the FE_T approach has the worst prediction accuracy (average R^2 = 0.21). That RE approach predicted more accurately than the CRE approach is an interesting discovery, because as compared to RE, CRE is a more generic approach that models group-level predictors. This finding may be due to two reasons. First, the small number of respondents in each group impeded the most important feature of the CRE approach, namely the separation of individual effects (e.g., the effect of one's fairness perception) and the context effects (e.g., the effect of the fairness climate). Second, with the CRE approach, the high complexity of the models results in prediction models that overfit training sets but cannot be generalized well to test sets. Another interesting finding is that the SL approach, which neglects the grouped structure (i.e., the SL approach) has an impressive prediction accuracy with an average R^2 of 0.25. The extent to which this finding can be generalized to other prediction questions remains to be investigated in future research.

Following the analytical workflow from Figure 1, we conclude that the best performing prediction model is the bagged trees with a RE approach: the one that takes the heterogeneity stemming from grouped data into account and that explicitly allowed for higher-order interactions. This model is further examined to determine the relative importance of all variables in predicting turnover intentions. The results are reported in Table 4. The five predictors that are identified as the most important predictors are (ranked in decreasing order) perceived fairness, LMX, career opportunities, pay satisfaction and age.

5 | DISCUSSION

Without some understanding of modern prediction models, HR professionals and–researchers may find it hard to collaborate with data experts. In this contribution, we explained the logic of prediction models in general. Moreover, we introduced, illustrated, and examined different modelling approaches to account for the grouped data structures in HRM data. We first discuss the lessons learned from the empirical comparison of the predictive performance for employee turnover intentions. Second, we discuss possibilities for HRM theory development and implications for HR analytics teams. Finally, we discuss limitations and suggestions for future research.

HRM analytics dealing with prediction questions need a data-driven, model-agnostic mindset: instead of settling at one specific prediction model, researchers are encouraged to select a full set of candidate models. By training and evaluating these models independently, their prediction accuracies can be compared. Only the model with the best prediction accuracy should be picked for making predictions for new data points (see Grimmer et al., 2021 for similar calls). To compare the relative performance of the predictive models in a grouped data set, we provide an analytical workflow (see Figure 1). This strategy echoes the well-known 'no free lunch theorem' (e.g., Ho & Pepyne, 2002) that states there is *no* model that works for every problem. Therefore, the recommended strategy is to try multiple prediction models and find the best-suited model empirically. We demonstrated this is, especially, important in the case of grouped data, as many characteristics of the data structure—for example, the number of variables and groups, and the group sizes—may affect the relative performance of different modelling approaches. For example, although the CRE approach was not particularly advantageous in the current study, we argue that there are situations where the flexibility offered by the CRE approach likely translates to higher prediction accuracy: for instance, in data sets existing of a large number of groups as well as respondents per group.

Methodologically, the current contribution underscores that applying prediction models is not void of accounting for particular characteristics of HRM data sets. In addition to the issue of grouped data, other characteristics that are worth consideration for treatment include the reliability of (e.g., employee survey) measures, the distribution of features and outcomes (e.g., skewedness of absenteeism), correlations among features and patterns of missing data. For example, Jacobucci and Grimm (2020) pointed out that measurement error obstructs the

TABLE 4 The relative importance metrics based on the model with the highest prediction accuracy (i.e., the bagged trees with the random-effects approach)

Predictors	The standardized relative importance	Rank	Predictors	The standardized relative importance	Rank
Gender	−0.64	32	Number of hierarchical levels	−0.33	19
Age	**0.86**	**5**	Number of departments	−0.23	18
Education level	−0.03	11	Number of managers	−0.16	17
Management position	−0.64	34	Family business	−0.61	29
Contract hour	−0.08	14	HRM: turnover	−0.46	20
Contract type	−0.61	30	HRM: Attract well-qualified personnel	−0.51	22
Leader-member exchange	**2.23**	**2**	HRM: Retain key employees	−0.57	27
Information sharing	0.07	9	HRM: Absenteeism level	−0.49	21
Voice	0.35	6	HRM: Amount of labour disputes	−0.58	28
Pay satisfaction	**1.26**	**4**	HRM: Quality of ideas and suggestions	−0.55	24
Perceived fairness	**4.20**	**1**	HRM: Contribution to innovation	−0.52	23
Job proactivity	−0.14	15	HRM: Involvement of employees	−0.57	26
Career opportunities	**1.94**	**3**	HRM: Flexibility of employees	−0.55	25
Number of employees	0.05	10	Availability of a supervisory board	−0.67	35
Number of full-time-equivalent employees	0.13	7	Availability of an advisory board	−0.68	36
Number of FTE employees last year	0.10	8	Corporate entrepreneurship	−0.07	13
Availability of works council	−0.62	31	Strategic planning	−0.16	16
Availability of HR professionals	−0.64	33	Entrepreneurial orientation	−0.07	12

Note: The five most important predictors (i.e., the predictors with the highest score of relative importance) are marked in bold.
Abbreviations: FTE, full-time equivalents; HR, human resource; HRM, human resource management.

discovery of non-linear relationships and weakens the expected advantages of modern prediction models. Likewise, García-Laencina et al. (2010) reviewed approaches to treat missing data when using prediction models, and much like our proposal, suggest an agnostic, data-driven strategy to determine the most adequate method.

An important advantage of prediction models is their power to deal with large and diverse data sets without using any a priori theory and/or hypotheses, which ensures that HR analytics include all kinds of relevant business information. To illustrate, in our analyses all variables ($N = 36$) available in the data on employee turnover intentions in SMEs could be used, as well as the grouped data structure (1454 employees in 199 SMEs). These models could be expected to work well even if the predictors are item scores—meaning that there are hundreds, if not thousands, of variables present. Critically, traditional models may not fare well in dealing with these data sets: researchers would have to be more selective about the number of variables in a study, due to restrictions of sample size and power. Findings like these focus managerial attention on domains that are most predictive for relevant business outcomes. In our analysis for example, the three most predictive factors are perceived fairness, LMX and career opportunities. These are the domains where interventions to improve retention are most likely to be effective. Despite this advantage, findings of prediction analysis typically fall short of interpretation, compared to explanatory models (e.g., ANOVA, structural equation models). This shortcoming prohibits the use of interpretation obtained from prediction models to create high-stake decisions. For HR policy development, the advices following from prediction results are preferably supplemented with theoretical foundations and (or) follow-up confirmatory analyses (Ellmer & Reichel, 2021). Take for example career opportunities as an important predictor of turnover intentions. Before taking actions to boost career opportunities, a detailed investigation can be conducted to confirm a set of practical questions, such as the scale of the effect, or to whom the lack of career opportunities is most detrimental. If used properly, HR analytics can contribute to evidence based HRM in providing inductive, context relevant, quantitative input for informed decision making (Barends & Rousseau, 2018).

The inductive nature of predictive analytics benefits HRM theory development and validation (Choudhury et al., 2020), as can be illustrated looking at the substantial findings of our empirical study. Our findings highlight that five factors, namely perceived fairness, LMX, career opportunities, pay satisfaction and age are the strongest predictors of employee's turnover intentions in SME's. These findings theoretically bridge between the employee turnover literature and what is known about HRM in SMEs (Harney & Alkhalaf, 2021). More specifically, the findings can be interpreted in the light of the centrality of the owner-manager, and the close and personal nature of employment relations in small organizations (LMX), but also illustrate the struggle of small organizations to retain talent (pay satisfaction, career opportunities) and the (lack of) formal HRM procedures that may hamper perceived fairness. Given the assumed disadvantage that SMEs cannot compete with larger organizations on wages to keep employees (Cobb & Linb, 2017), an interesting finding is that pay satisfaction is ranked lower than fairness and LMX. This inductive finding may invoke new hypothesis about employee retention in SME's, which can deductively validated in follow-up research, or stimulate inductive or abductive methods to specify these findings (Choudhury et al., 2020).

While data scientists, predominantly trained in science and technology, are perfectly capable of building prediction models from scratch, HR trained experts who can bridge HRM professionals and data scientists are invaluable to a successful HR analytics team (Marler & Boudreau, 2017). These experts can operationalize HRM business questions to analytical queries and provide analysts with relevant backgrounds and contexts of HRM data. Moreover, these experts can translate the numbers and figures from the analyses into sensible conclusions and decisions, aided by knowledge about HRM theories and the current business. Finally, they should safeguard ethical standards in HR analytics teams to protect individuals. Data obtained from employees and used in prediction models should always be collected with consent from each employee, especially, when their future career or personal development is (in part) affected by the results of these prediction analyses. Moreover, prediction models are known to be prone to different sorts of biases, including but not limited to racial, age and gender bias, resulting from implicit biases underlying existing performance and administrative data (Mehrabi et al., 2019). In agreement with McCartney et al. (2020), we stress that HR analytics teams should include of HR trained experts who can bridge between data experts, the business and HR theory, to reap the full and promising potential of modern prediction models. The current contribution serves as a practical guide to the use of prediction models, thereby helping HRM experts to become the bridge that brings together techniques and domain-specific expertise.

5.1 | Limitations and future research

Like any research, the empirical study is not without limitations. Most importantly, the outcome specified, namely turnover intention, is a self-reported measure that indicates one's *attitude* towards turnover, which, collected at the same time as other measurements, might suffer from common method bias (Podsakoff et al., 2012). We partly addressed this issue by designing a user-friendly questionnaire with many reversed-coding items. Moreover, despite the range of variables included in the dataset, it was not exhaustive of all potential predictors of turnover suggested in previous research (e.g., the presence of job alternatives was not measured). Similarly, data from organizations with more than 250 employees were unavailable in the data set. The findings of any analysis should always be interpreted taking the limitations of the data into account.

As for the prediction analytics, we could only cover a subset of prediction models, while novel models are proposed regularly. For example, Capitaine et al. (2021) suggest a novel prediction model incorporating RE have been proposed based on random forests, which could be another candidate for testing and comparing prediction questions with grouped data. For the sake of brevity, the prediction models that consisted of RE were specified with only random intercepts, but no random slopes (i.e., the effects of the predictors on the outcome is assumed to be constant across all respondents). We invite researchers to further examine the consequences of involving random slopes in cases where random slopes are expected from theories and (or) previous findings. Also, the target encoder as illustrated and implemented in the current contribution—replacing grouping variables with the group average of the outcome—suffers the problem of target leakage: the information about the outcome has been used to create predictors. Several remedies have been proposed to mitigate this problem. For example, the cross-fold target encoding partitions all respondents into *k* folds and, for each fold, replaces the grouping variable with the average of the outcome across respondents of the other $k-1$ folds. Last, the current study computed variable importance metrics following the strategies adapted by Putka et al. (2018) and Spisak et al. (2019). However, we noticed recent studies have proposed several alternative—and arguably more accurate—approaches to facilitate the interpretation of machine learning algorithms (e.g., Azodi et al., 2020; Fisher et al., 2019 for a comprehensive review of interpretable machine learning methods).

Finally, we note that the literature on predictive analysis in HRM has largely developed by technology oriented researchers, outside mainstream HRM journals (Garg et al., 2021). We hope that our contribution in this special issue inspires researchers and HR analytics teams to join us and bridge these separate worlds, by describing, developing and demonstrating applications of modern methods for HR analytics that truly take account of the specifics of HRM data.

ACKNOWLEDGEMENTS

The authors would like to thank Jiayan Xiang MSc, junior lecturer in the department of Human Resource Studies, Tilburg University, for her collaboration in collecting the data and in bringing the author team together. Furthermore, they would like to express their gratitude to Prof. Dr. Martin Edwards and two anonymous reviewers for their useful comments in developing the study throughout the review process. The research project was co-funded by the Tilburg Institute for Family Business and a personal grant from The Netherlands Organization for Scientific Research [NWO-Research Talent 406.17.526] awarded to Shuai Yuan.

CONFLICT OF INTEREST

The authors confirm that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data, supplementary materials and the executive code are available at https://osf.io/yx2us/files/.

ORCID

Shuai Yuan https://orcid.org/0000-0003-1720-4483
Brigitte Kroon https://orcid.org/0000-0001-9968-872X
Astrid Kramer https://orcid.org/0000-0002-7112-7610

ENDNOTES

[1] In other research fields, this is also referred to as the within-between random effects models (Bell et al., 2019), the Mundlak models (Mundlak, 1978) and the contextual models (Hamaker & Muthén, 2020).

[2] We thank an anonymous reviewer to point out this option.

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort A., Thirion B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In International Conference on Engineering and technology (ICET) 1–6 Antalya, Turkey. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *24*(2), 443–483. https://doi.org/10.1177/1094428119877457

Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the black box: Interpretable machine learning for geneticists. *Trends in Genetics*, *36*(6), 443–455. https://doi.org/10.1016/j.tig.2020.03.005

Barends, E., & Rousseau, D. M. (2018). *Evidence-based management: How to use evidence to make better organizational decisions*. Kogan Page Publishers.

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality and Quantity*, *53*(2), 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Bliese, P. D., Chan, D., & Ployhart, R. E. (2007). Multilevel methods: Future directions in measurement, longitudinal analyses, and nonnormal outcomes. *Organizational Research Methods*, *10*(4), 551–563. https://doi.org/10.1177/1094428107301102

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* (1st ed.). CRC Press. https://doi.org/10.1201/9781315139470

Capitaine, L., Genuer, R., & Thiébaut, R. (2021). Random forests for high-dimensional longitudinal data. *Statistical Methods in Medical Research*, *30*(1), 166–184.

Choudhury, P., Allen, R., & Endres, M. G. (2020). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30–57. https://doi.org/10.1002/smj.3215

Cobb, J. A., & Linb, K. H. (2017). Growing apart: The changing firm-size wage premium and its inequality consequences. *Organization Science*, *28*(3), 429–446. https://doi.org/10.1287/orsc.2017.1125

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Cambridge series in statistical and Probabilistic Mathematics). Cambridge University Press. https://doi.org/10.1017/CBO9780511802843

Deen, M., & de Rooij, M. (2020). ClusterBootstrap: An R package for the analysis of hierarchical data using generalized linear models with the cluster bootstrap. *Behavior Research Methods*, *52*, 572–590. https://doi.org/10.3758/s13428-019-01252-y

Ellmer, M., & Reichel, A. (2021). Mind the channel! An affordance perspective on how digital voice channels encourage or discourage employee voice. *Human Resource Management Journal*, *31*(1), 259–276. https://doi.org/10.1111/1748-8583.12297

European Commission. (nd). *Internal market, industry, entrepreneurship and SMEs: Family business*. https://ec.europa.eu/growth/smes/promoting-entrepreneurship/we-work-for/family-business_en

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81.

Fox, J. (2008). *Applied regression analysis and generalized linear models*. SAGE Publications.

García-Laencina, P. J., Sancho-Gómez, J. L., & Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: A review. *Neural Computing & Applications*, *19*, 263–282. https://doi.org/10.1007/s00521-009-0295-6

Garg, S., Sinha, S., Kar, A. K., & Mani, M. (2021). A review of machine learning applications in human resource management. *International Journal of Productivity and Performance Management*, ahead-of-print. https://doi.org/10.1108/IJPPM-08-2020-0427

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine learning for social science: An agnostic approach. *Annual Review of Political Science*, *24*, 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modelling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Harney, B., & Alkhalaf, H. (2021). A quarter-century review of HRM in small and medium-sized enterprises: Capturing what we know, exploring where we need to go. *Human Resource Management*, *60*(1), 5–29. https://doi.org/10.1002/hrm.22010

Hausknecht, M., Li, W. K., Mauk, M., & Stone, P. (2016). Machine learning capabilities of a simulated cerebellum. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(3), 510–522. https://doi.org/10.1109/TNNLS.2015.2512838

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, *13*(4), 18–28. https://doi.org/10.1109/5254.708428

Ho, Y. C., & Pepyne, D. L. (2002). Simple explanation of the no-free-lunch theorem and its implications. *Journal of Optimization Theory and Applications*, *115*(3), 549–570. https://doi.org/10.1023/A:1021251113462

Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2018). Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future. *The Academy of Management Annals*, *2*(1), 231–274. https://doi.org/10.5465/19416520802211552

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, *102*(3), 530–545. https://doi.org/10.1037/apl0000103

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

Kuhn, M., & Johnson, K. (2013). *Applied predictive modelling* (Vol. 26). Springer. https://doi.org/10.1007/978-1-4614-6849-3

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. https://www.jstor.org/stable/2529876

Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR analytics. *International Journal of Human Resource Management*, *28*(1), 3–26. https://doi.org/10.1080/09585192.2016.1244699

McCartney, S., Murphy, C., & McCarthy, J. (2020). 21st century HR: A competency model for the emerging role of HR analysts. *Personnel Review*, ahead-of-print. https://doi.org/10.1108/PR-12-2019-0670

McNeish, D., & Kelley, K. (2018). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. *Psychological Methods*, *24*(1), 20–35. https://doi.org/10.1037/met0000182

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modelling. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85.

Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: Forward progress for organizational research and practice. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(1), 505–533. https://doi.org/10.1146/annurev-orgpsych-032117-104553

Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: A meta-analysis. *Journal of Applied Psychology*, *98*(2), 268–309. https://doi.org/10.1037/a0030723

Peccei, R., & van de Voorde, K. (2019). The application of the multilevel paradigm in human resource management – outcomes research: Taking stock and going forward. *Journal of Management*, *45*(2), 786–818. https://doi.org/10.1177/0149206316673720

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Putka, D. J., Beatty, A. S., & Reeder, M. C. (2018). Modern prediction models. *Organizational Research Methods*, *21*(3), 689–732. https://doi.org/10.1177/1094428117697041

Rodríguez, P., Bautista, M. A., Gonzalez, J., & Escalera, S. (2018). Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing*, *75*, 21–31. https://doi.org/10.1016/j.imavis.2018.04.004

Rubenstein, A. L., Eberly, M. B., Lee, T. W., & Mitchell, T. R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, *71*(1), 23–65. https://doi.org/10.1111/peps.12226

Schelldorfer, J., Meier, L., & Bühlmann, P. (2014). Glmmlasso: An algorithm for high-dimensional generalized linear mixed models using ℓ1-penalization. *Journal of Computational & Graphical Statistics*, *23*(2), 460–477. https://doi.org/10.1080/10618600.2013.773239

Sela, R. J., & Simonoff, J. S. (2012). RE-EM trees: A data mining approach for longitudinal and clustered data. *Machine Learning*, *86*(2), 169–207. https://doi.org/10.1007/s10994-011-5258-3

Spisak, B. R., van der Laken, P. A., & Doornenbal, B. M. (2019). Finding the right fuel for the analytical engine: Expanding the leader trait paradigm through machine learning? *The Leadership Quarterly*, *30*(4), 417–426. https://doi.org/10.1016/j.leaqua.2019.05.005

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Valentine, S., Greller, M. M., & Richtermeyer, S. B. (2006). Employee job response as a function of ethical context and perceived organization support. *Journal of Business Research*, *59*(5), 582–588. https://doi.org/10.1016/j.jbusres.2005.06.004

Yarkoni, T. (2020). The generalizability crisis, *Behavioral and Brain Sciences*, 1-37. https://doi.org/10.1017/S0140525X20001685

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Yuan, S., Kroon, B., & Kramer, A. (2024). Building prediction models with grouped data: A case study on the prediction of turnover intention. *Human Resource Management Journal*, *34*(1), 20–38. https://doi.org/10.1111/1748-8583.12396