# Employee Turnover in Multinational Corporations
## A Supervised Machine Learning Approach

Veglio, Valerio ; Romanello, Rubina; Pedersen, Torben

ORIGINAL PAPER

# Employee turnover in multinational corporations: a supervised machine learning approach

Valerio Veglio[1,4] · Rubina Romanello[2] · Torben Pedersen[3]

## Abstract
This research explores the potential of supervised machine learning techniques in transforming raw data into strategic knowledge in the context of human resource management. By analyzing a database with over 205 variables and 2,932 observations related to a telco multinational corporation, this study tests the predictive and analytical power of classification decision trees in detecting the determinants of voluntary employee turnover. The results show the determinants of groups of employees who may voluntarily leave the company, highlighting the level of analytical depth of the classification tree. This study contributes to the field of human resource management by highlighting the strategic value of the classification decision tree in identifying the characteristics of groups of employees with a high propensity to voluntarily leave the firm. As practical implication, our study provides an approach that any organization can use to self-assess its own turnover risk and develop tailored retention practices.

✉ Rubina Romanello
   Rubina.romanello@units.it

   Valerio Veglio
   Valerio.veglio@unipv.it

   Torben Pedersen
   Tp.si@cbs.dk

1   Department of Economics and Management, University of Pavia, Pavia, Italy

2   Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy

3   Department of Strategy and Innovation, Copenhagen Business School, Copenhagen, Denmark

4   Institute for Transformative Innovation Research, University of Pavia, Pavia, Italy

Published online: 21 May 2024

🙂 Springer

# 1 Introduction

Data and analytics have captured the attention of human resource management (HRM) scholars, as multinational corporations (MNCs) increasingly have at their disposal large volumes of data and techniques for analyzing large amounts of data that could be used to support decision making related to complex problems, such as task organization, employee turnover, career development, and training design. Extrapolating strategic knowledge from large datasets through supervised Machine Learning (ML) techniques is becoming one of the main challenges for decision-makers in MNCs. In the digital era, techniques based on machine learning algorithms play an increasingly important role in extrapolating strategic knowledge from raw data (Canhoto and Clear 2020). Since 1956, when John McCarthy coined the term "Artificial Intelligence" (AI), the interest in this topic has grown exponentially, in line with the increasing number of applications, permeating different disciplines and research areas. In this context, management studies have started to analyze the potential applications of ML, which is a subset of AI and represents a way to achieve AI through the development of algorithms capable of improving themselves with experience (Garg et al. 2022). Supervised ML techniques can analyze large volume of data from different sources to discover hidden patterns with high strategic value for organizations (Pereira et al. 2018), providing insights for prediction, classification, and decision-making purposes (Cui et al. 2006; Naeem et al. 2024). These techniques have some specific features, such as scalability, because they can handle and process large amounts of data; interactivity, because they can learn new variables from new data; and dynamism, because they can periodically reassess and reevaluate hypotheses by taking into account incoming data, even without human interaction (Garg et al. 2022). With these capabilities, managers could make rapid and contextualized decisions based on data-driven evidence (Gupta et al. 2018; Wirges and Neyer 2023). For instance, supervised machine learning techniques are useful in marketing to estimate the probability of customer churn (Archaux et al. 2004; Gordini and Veglio 2017; Hung et al. 2006; Rosset et al. 2003; Wei and Chiu 2002), in social and economic analyses (Blazquez and Domenech 2018), in smart city applications (Iqbal et al. 2020), in finance to predict customer credit risk (Kruppa et al. 2013), and in HRM in the field of recruitment, performance management, and team dynamics (Garg et al. 2022; Koechling et al. 2023). As has happened in marketing and finance over the past decade, the application of ML techniques in HRM is growing rapidly (Garg et al. 2022; Yang et al. 2023), although only a limited number of studies have aimed at predicting employee turnover (Rombaut and Guerry 2018; Saradhi and Palshikar 2011). Voluntary employee turnover refers to why people leave an organization (Lee et al. 2017) and is considered a serious issue for MNCs (Saradhi and Palshikar 2011). High levels of employee turnover generate unexpected costs in terms of hiring, training, search, selection, and replacement (Mobley 1982; Price 1977; Staw 1980). In fact, the cost of hiring new employees is substantially higher than retaining the existing employees, negatively affecting firm performance and competitiveness (Holtom et al. 2005; Mitchell et al. 2001).

Despite the relevance of this topic, extant research has barely explored the potential, but not yet investigated, the real added value of the application of supervised

ML techniques to identify the determinants of voluntary employee turnover (Yang et al. 2023). Particularly in the context of employee turnover, the few existing works employing supervised ML techniques have so far been published in conference proceedings (Garg et al. 2022).

Responding to the call for new methodological approaches (Hom et al. 2017), this study tests the analytical and predictive power of a classification decision tree based on the CHAID (Chi-square Automatic Interaction Detector) algorithm to identify the characteristics of employees who voluntarily leave the company. By analyzing a large dataset of employees in the context of a telco MNC, we apply a CHAID classification decision tree to a sample of 2,932 employees working in the firm's headquarters in Norway and in a subsidiary located in Denmark. This approach makes it possible to exploit the potential of this technique to identify in advance those groups of employees who have a higher likelihood of leaving the firm and, ultimately, to implement retention practices targeted at these groups.

This study contributes to the turnover literature from a methodological perspective by encouraging researchers to measure and/or analyze employee turnover in nontraditional ways using supervised ML models to test the validity of well-known historical explanatory constructs in this area of research (Hom et al. 2017). In addition, we contribute to the HRM literature by demonstrating the potential of the classification decision tree as a method for solving complex problems in HRM (Garg et al. 2022), which can be used to complement previous studies and further advance the literature on this topic. Our results suggest the complementary, rather than substitutive, role of supervised ML in assessing the risk of employee turnover. From a practical perspective, our study demonstrates the ability of these techniques to reveal hidden relationships between data, allowing decisionmakers and scholars to identify new, previously unknown relationships and evidence. Using this technique, companies could self-assess their own employee turnover risk levels and identify high-risk employees, allowing them to develop timely and effective retention strategies tailored to their needs. In this sense, supervised ML techniques become an additional toolkit that supports, but does not replace, human resource management decisions.

## 2 Theoretical background

### 2.1 A brief introduction to employee turnover research

Voluntary employee turnover – employees' unilateral, unwanted, and often surprising termination of their employment contract – is a phenomenon of practical relevance for all organizations for a variety of reasons (e.g. Lee et al. 2008), as a negative economic impact is generally expected, mainly due to the additional recruitment costs or tacit knowledge drains (Glebbeck and Bax 2004; Holtom et al. 2005; Mitchell et al. 2001; Reiche 2008).

Reflecting its importance, research on employee turnover can be traced back to 1920 (Hom et al. 2017; Lee et al. 2017). Since its inception, the literature has continually grown, resulting in an impressive number of theories and models over the subsequent one hundred years mainly aimed at explaining the motivations, antecedents,

processes, and consequences of this organizational phenomenon (Hom et al. 2017; Lee et al. 2017; Rubenstein et al. 2018). Employee turnover is related to a wide range of factors. Researchers agree on the importance of job satisfaction (or dissatisfaction) and individual perceptions of the perceived desirability and ease of moving to another job based on the assumption that employees who are satisfied with their job and do not have other job options are more likely to stay in the organization (Griffeth et al. 2000; Mobley 1977). The seminal work of March and Simon (1958) triggered a stream of studies attempting to explain voluntary employee turnover by focusing on why people leave (Barrick and Zimmerman 2005). This initial approach was followed by attempts to identify the primary antecedents of employee turnover (e.g. Lee and Mitchell 1994; O' Reilly III et al. 1991). With the aim of retaining employees, academics then attempted building an employee turnover model that was as accurate as possible to minimize voluntary employee turnover. Over the years, several models ensued, such Mobley's (1977) process model explaining how dissatisfaction leads to turnover in the attempt to explain why employees leave their jobs (Lee et al. 2017). Other scholars focused more on the content than the process, identifying a variety of determinants of turnover, including factors related to the workplace, labor market causes, community, and occupational aspects (Hom et al. 2017; Price 1977, 2001), emphasizing the importance of both individual and environmental attributes. As scholars have criticized extant turnover models for their lack of explanatory and predictive power, turnover research has included variables that are not necessarily related to the employee's affective state and the decision to quit (Morrell et al. 2001). Subsequently, researchers studied the role of shocks and jarring events that drive employees to choose alternative career paths (Hom and Griffeth 1991), demonstrating that voluntary leave is not necessarily related to job dissatisfaction alone. The resulting models and analyses of voluntary employee turnover therefore included "shocks" (Lee and Mitchell 1994), while others focused on motives for leaving (Maertz and Campion 2004).

Another research stream has adopted the opposite perspective (Porter and Steers 1973), focusing on why people stay, proposing the "job embeddedness" construct that considers contextual factors both related to the workplaces and off-the-job aspects (Coetzer et al. 2019; Mitchell et al. 2001). Following this radical paradigm shift, scholars developed further conceptualizations, explanations, and empirical approaches (Lee et al. 2017). Along these lines, some researchers more recently proposed integrative frameworks that attempt to explain both why and how employees quit (Maertz and Campion 2004).

In relation to the IT context, which is particularly affected by turnover issues (Rode et al. 2007), Ghapanchi and Aurum (2011) developed a systematic literature review of the antecedents of IT employee turnover, showing that the determinants can generally be grouped into five main categories: individual, organizational, job-related, psychological, and environmental. This summarizing perspective is consistent with the more general perspective that emerges from findings of the broader turnover literature (Lee et al. 2017; Rubenstein et al. 2018). The first category includes individual attributes, motivational, and professional behavior constructs. For example, motivational factors may be related to mindset types such as self-positiveness or low core self-evaluation (Hom et al.

2012). Organizational factors relate to individual perceptions of the organization, such as remuneration, benefits, human resource practices, organizational culture (O' Reilly III et al. 1991; Rubenstein et al. 2018) and the centrality of the functional department in the intraorganizational network of the MNE (Castellacci et al. 2018). For instance, in terms of organizational culture, the person-organization fit predicts employee job satisfaction, which also affects turnover (O' Reilly III et al. 1991). Also, knowledge sharing practices within the department and with colleagues outside the function (e.g., Cabrera and Cabrera 2005; Dasí et al. 2017; Garg et al. 2022) and knowledge flows across firm boundaries (Gupta and Govindarajan 2000). Job-related antecedents instead concern the characteristics, support, difficulties, and attractiveness of jobs, whereas psychological factors include individuals' satisfaction in terms of jobs, career prospects, and organizational aspects (e.g. commitment). For instance, job design can influence job characteristics such as competence, autonomy, or task identity, which can motivate knowledge sharing among employees (e.g., Foss et al. 2009; Ryan and Deci 2000), aspects that can also influence employee well-being and turnover. Finally, environmental factors include aspects that are external to the workplace related to job alternatives, family support, work-family balance, etc. However, Ghapanchi and Aurum (2011) underline a prevalence of antecedents at the job-related, organizational, and psychological levels, whereas fewer antecedents pertain to the remaining categories. Similar to other widely and long studied management and organizational phenomena, the study of employee turnover has developed into distinct streams of research. This research system-immanent development has led to ever more complexity and an abundance of highly specialized empirical findings rather than convergence and consolidation. This can also be concluded from the number and increasing specializations of literature reviews and meta-analyses (which are briefly described in Table 4 in the appendix).

A closer look at these reviews reveals that research on voluntary employee turnover has not led to maturity and consolidation, in the sense that the (i) results converge, (ii) the most important factors and cause-effect relationships are clearly known, and (iii) only nuances are debated. Quite the opposite, although research on voluntary employee turnover has developed and propagated advanced models and theories, the findings remain inconsistent and at times even conflicting (Hom et al. 2017). Partly fueled by the paradigm prevailing in the social sciences and peer-reviewed journals documenting the novelty of research by virtue of previously unstudied or understudied determinants, the number of variables studied has increased significantly in recent decades. It seems that research on voluntary employee turnover – like many other organizational phenomena – is in search of the famous needle in the haystack, or as renowned scholars in the field put it, in "Search of the Holy Grail" (Holtom et al. 2008). The search for impact factors that allow forecasting turnover has predominantly adopted regression models (e.g. OLS, logit, and logistic), structural equation modelling (SEM), and other traditional statistical techniques, such as cluster analysis and dimension reduction (Garg et al. 2022; Lee et al. 2017).

As the numerous and increasingly specialized reviews and especially meta-analyses demonstrate, the voluntary employee turnover phenomenon is a mature and

increasingly differentiated field of management research with high scientific and practical relevance. Based on these reviews, different approaches can be adopted for the theoretical and empirical development of the field. On the one hand, replication studies could be conducted for known but possibly inconsistent correlations, or testing new independent variables, moderators, or mediators. We refer to this as the 'more of the same' approach. On the other hand, the extensive datasets obtained from the increasing digitalization of HRM and web-based surveys (Holtom et al. 2008, p. 259) could be used to identify previously unrecognized influencing factors, effect relationships, and patterns. More recent calls in the field of employee turnover also emphasize the need to apply new and innovative methodological approaches, such as the implementation of machine learning techniques, to predict employee turnover (Choudhury et al. 2021; Garg et al. 2022; Lee et al. 2017; Rombaut and Guerry 2018; Yang et al. 2023).

## 2.2 Applications of ML techniques in the HRM context

Although HRM is a somewhat unexplored area with regard to big data analytics (BDA) and supervised ML applications (e.g. Ekawati 2019; Sheng et al. 2017), interest has significantly grown as a consequence of the ongoing digitalization of firms (Raguseo and Vitari 2018; Rombaut and Guerry 2018; Saradhi and Palshikar 2011; Sexton et al. 2005; Shah et al. 2017). The comparatively few studies that apply data mining techniques in the HRM field focus on employee selection (Aiolli et al. 2009), employee competences (Zhu et al. 2005), career planning (Lockamy and Service 2011), predicting employee performance and evaluation (Zhao 2008), candidates' preliminary evaluation and training success (Aviad and Roy 2011), and employee turnover (Quinn et al. 2002; Saradhi and Palshikar 2011; Sexton et al. 2005). Besides the studies adopting other advanced statistical techniques, such as regressions, SEM models and Bayesan Model Averaging (BMA) (e.g., Coetzer et al. 2019; Nandialath et al. 2018; Sandhya and Sulphey 2021), the handful of studies that have applied supervised ML to the issue of employee turnover evaluate or compare neural network solutions (Quinn et al. 2002; Sexton et al. 2005), random forests, support vector machines and naïve Bayes (Saradhi and Palshikar 2011), and classification decision trees (Choudhury et al. 2021; Rombaut and Guerry 2018; Saradhi and Palshikar 2011). Table 1 summarizes the main characteristics and contributions of these studies.

We next provide a summary of the rather diverse and contradictory assessments of these techniques with a particular focus on their predictive power. While Sexton and colleagues (2005) emphasize the accuracy of neural network techniques in predicting and solving classification business problems, Quinn et al. (2002) find that they perform worse than logistic regression. Saradhi and Palshikar (2011) compare naïve Bayes, support vector machines, decision tree and random forests, highlighting the superiority of support vector machines. Rombaut and Guerry (2018) point out the superiority of decision tree techniques compared to logistic regression. Choudhury and colleagues (2021) partially confirm this finding in their recent comparison of decision trees, random forests, and neural networks, documenting that

**Table 1** Previous applications of ML techniques in voluntary employee turnover research

| Article | Purpose | Sample | ML technique(s) | Results |
|---|---|---|---|---|
| Quinn et al. (2002), *Journal of Technology in Human Services* | Development of a model to predict employee turnover in a regional child protective services agency | 536 employees (caseworkers and supervisors) of a regional child welfare agency; over 100 independent variables | Comparison between logistic regression and neural networks | Neural networks more accurately predicted employee turnover than logistic regression by only 1% |
| Sexton et al. (2005), *Computers & Operation Research* | Predicting turnover and identifying at-risk employees | 447 employees from a small (fewer than 100 employees) family-owned mid-western stainless-steel manufacturer; 29 independent variables | Neural networks and modified genetic algorithm to train neural networks | The neural network simultaneous optimization algorithm predicts the turnover rate with a high degree of accuracy |
| Saradhi and Palshikar (2011), *Journal of Business Research* | Building and comparing predictive employee turnover models, proposing a simple model to identify churn of "valuable" employees | 1,575 employees of a specific client unit within a large organization; 25 independent variables | Naïve Bayes, support vector machines, decision tree, random forests, and logistic regression | Machine learning techniques (e.g. support vector machine) can be used to build reliable and accurate predictive models for employee turnover |
| Rombaut and Guerry (2018), *Management Research Review* | Development of a methodological approach to investigate and predict turnover (not turnover intention) exclusively through the analysis of data storage in the internal human resource database | 4,041 employees of a Belgian branch of a private company offering HR solutions; 13 independent variables | Logistic regression and decision tree | Data collected from the internal human resource database are able to accurately predict employee turnover |
| Ekawati (2019), *Journal of Management Information and Decision Science* | Identifying the more accurate predictive techniques of employee turnover | Systematic literature review of scholarly articles related to both employees and customers published from 2000 to 2018 in major databases, such as IEEE, ACM, Science Direct, Emerald, and Willey Online | Decision tree, naïve Bayes, neural network, support vector machine, and random forest | The most used technique to predict employee churn is the neural network even if decision tree and naïve Bayes are also very accurate tools. Instead, for customer churn, the support vector machine performs the best compared to the other methods |

**Table 1** (continued)

| Article | Purpose | Sample | ML technique(s) | Results |
|---|---|---|---|---|
| Choudhury et al. ([2021](#)), *Strategic Management Journal* | Demonstrating the power of machine learning techniques for pattern discovery as well as the development of a guide to evaluate model performance, highlighting human decisions in the process, and warning of common misinterpretation pitfalls | 1,191 employees of a large Indian technology firm; 9 independent variables | Decision tree | Machine learning methods can be powerful for pattern discovery. However, they need to be seen as a complement to, not a substitute of, traditional econometric methods |

classification decision trees have high predictive and analytical power in identifying employee turnover probability. However, their study does not address the characteristics of employees highly inclined to voluntarily leave, instead limited to evaluating the statistical performance of different ML techniques. Although these first attempts provide initial evidence of the potential applications of these techniques in the field of HRM, there is still a lack of research that applies the ML technique to identify the characteristics of employees who are more likely to leave. Our study contributes to filling this gap by identifying the determinants of voluntary employee turnover using a classification decision tree.

## 3 Research method

Through an illustrative example on data from a telco MNC, this study investigates the root causes of voluntarily employee turnover through the CHAID classification decision tree. IBM SPSS Statistics (v.27) has been applied to run the CHAID classification decision tree.

### 3.1 Data collection, sample, and measures

We used a database derived from an online survey submitted to the employees of a leading telco MNC in Northern European countries and Asia. The company has a strong position in mobile, broadband, and TV services with 180 million global customers worldwide and annual revenues of approximately USD 12 billion. It is headquartered in Norway with more than 12 subsidiaries in Europe and Asia.

The company's HRM department provided the dataset. The data was collected in 2016 through an email-based survey sent via the firm's internal system to all 7,786 employees working in the headquarters and Nordic subsidiaries. Prior to the survey, an invitation letter signed by the CEO was sent out, emphasizing the importance and reasons for the survey, as well as the fact that there was no obligation to respond. Employees were clearly informed of the mechanism in place to protect their privacy. Employee email addresses were retrieved from the central HRM system. Then, when employees returned the questionnaire, the research department temporarily used their email addresses to retrieve some general (e.g. demographic) information from the HRM system and to link responses to previous surveys. The research department ensured that an encrypted employee email address was developed prior to any use of the data to ensure the anonymity of responses.

The survey was sent out via the head office and each subsidiary's local intranet. After three weeks, the average response rate was around 56%, of which 66% from Norway and 48% from Denmark, which is considered an acceptable response rate for this type of analysis. The decision to focus on both the headquarter and a subsidiary was prompted by a discussion with the CEO, who claimed that voluntary employee turnover was a serious problem in Norway and Denmark, thus confirming the relevance of the setting for our research. Table 2 shows the percentage of the company's voluntary employee turnover for Norway and Denmark in 2016.

The final database was constructed from a few different datasets. The HR department merged the survey data with internal data on voluntary employee turnover after two years, resulting in a dataset based on 2,932 usable responses, of which 834 were from Denmark and 2,098 from Norway, and 209 variables. We removed 95 variables from the database because unrelated to the issues of voluntary employee turnover and only 9 responses due to missing values. We chose to exclude these answers instead of replacing missing values in order to minimize the risk of bias in our results.

The dependent variable is dichotomous and takes the value 1 if voluntary employee turnover occurred and 0 otherwise. Instead, the 113 independent variables are nominal (dichotomous), categorical, and single-item 7-point Likert scales drawn from previous literature. Table 5 in the Appendix A shows the variables included in the analysis and their coding. Independent variables pertain to three main categories of determinants: individual attributes, job-related determinants, and organizational determinants, in line with Ghapanchi and Aurum (2011)'s classification.

We employed several procedural remedies to reduce common method bias, including guaranteeing anonymity to respondents, emphasizing the importance and reasons for the survey, collecting data from different sources and at different points of time, using different datasets to build the final database, and included questions with different shapes to reduce the risk of response set (Podsakoff et al. 2012). Moreover, data on the dependent variable is collected through an internal database after two years from the survey collecting data on independent variables; therefore, the issue related to common method bias in not relevant for the analysis (Kock et al. 2021).

### 3.2 Research methodology

We applied a supervised ML technique, known as the classification decision tree technique, based on CHAID algorithms, to identify the determinants that characterize the employees with a high probability of voluntarily leave the firm. This technique is, particularly suitable for discovering patterns of meaningful relationships (both linear and nonlinear) and rules from large databases (Jain et al. 2016). It is particularly efficient with dichotomous, nominal, and scale-ordinal variables (Ture et al. 2009).

This technique has numerous advantages, such as: 1) it is simple to understand and interpret, 2) it requires little data preparation, 3) it can handle both numerical and categorical data, 4) it uses a white box model, 5) it has a high explanatory power, 6) it performs well with large data in a short time (Alao and Adeyemo 2013; Choudhury et al. 2021; Perner et al. 2001; Rombaut and Guerry 2018), 7) it is more "fair" as it shows an improved ability to make unbiased decisions (Garg et al. 2022), 8) it provides clear information about the importance of significant factors for prediction or

| **Table 2** Voluntary employee churn | Location | Voluntary Employee Turnover |
|---|---|---|
| | Norway (Headquarters) | 5.39% |
| | Denmark (Subsidiary) | 16.55% |

classification (Tso and Yau 2007), 9) multicollinearity is not a problem; attempts to eliminate it resulted in poor classification performance (Piramuthu 2008), thus eliminating the need to apply dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) (Reddy et al. 2020), and 10) it can be applied to a large sample for data reduction purposes. This technique can also handle missing values, provides stopping rules that account for statistical significance at the 1%, 5%, or 10% level, does not assume a priori the type of distribution of independent variables, and its probabilistic estimation is based on the chi-square test (Díaz-Pérez and Bethencourt-Cejas 2016; Kass 1980; Nisbet et al. 2018). However, despite the advantages of this technique, the classification decision tree is prone to overfitting (Giudici 2010), leading to biased results, even with large databases. However, various remedies, such as the use of cross-validation, can be employed to successfully address overfitting concerns. This method divides the sample into several subsamples (typically 10 samples). Tree models are then generated, excluding the data from each subsample generated. For each tree, the risk of misclassification is estimated by applying the tree to the subsample excluded in its generation. This approach produces a single, final tree model for which the cross-validated risk estimate is calculated as the average of the risks for all trees (Blockeel and Struyf 2002).

The CHAID classification decision tree systematically breaks down data to classify patterns found in the data set and make rule-based predictions (Berry and Linoff 2000). It can be seen as a recursive procedure in which a set of $n$ observations is progressively partitioned into groups according to a division rule – based on the $p$ value derived from multiple chi-square tests – aimed at maximizing a measure of homogeneity or purity of the dependent variable in each of the obtained groups (Giudici 2010). Then, in further step stage, the identification of the best variable as the first variable of the dependent variable is considered. Once again, a chi-square test is applied and calculated for each contingency table derived from the intersection of the dependent variable with each individual predictor (Cerchiello and Giudici 2012). The first split falls on the predictor with the highest chi-square value and the lowest $p$ value. The tree continues to branch into child nodes until it reaches the terminal node for each branch (Jain et al. 2016). Each terminal node identifies subgroups defined by different sets of predictors (Tan et al. 2006). The procedure stops when the chosen stopping rule is satisfied (Cerchiello and Giudici 2016). To obtain a final partition of the observation, it is necessary to specify stopping criteria for the division process. The criteria developed for selecting the best partition are often based on the degree of impurity of the child nodes (Tan et al. 2005). The concept of impurity refers to a measure of the variability of the response values of the observation (Giudici 2010). The lower the degree of impurity, the more skewed the class distribution. Specifically, in a regression tree, a node is pure if it has zero variance (all observations are equal) and impure if the variance of the observations is high, while for classification decision trees, alternative measures such as misclassification impurity, Gini impurity, and entropy impurity should be considered (Tan et al. 2005). In this case, the misclassification impurity (the distance between the observed and expected frequencies) was applied to obtain the final partition of the tree. The expected frequencies are calculated using the hypotheses of homogeneity for the

observation in the node considered or using the Chi-square index (Giudici 2010). Assuming that a final partition consisting of $g$ groups ($g < n$) has been reached, then for any given response variable observation $y_i$, a CHAID classification decision tree will produce the fitted value $\widehat{y}_i$ or the fitted probabilities of belonging to a single group, assuming only two classes (binary classification). Then, the fitted probability of success is given by the equation:

$$\widehat{y}_i = \frac{\sum_{l=1}^{m_n} y_{lm}}{n_m},$$

where the observation $y_{lm}$ can take the value 0 or 1, and the fitted probability corresponds to the observation proportion of success in group $m$ with $\widehat{y}_i$ constant for all observations in the group (Giudici and Figini 2009; Linoff and Berry 2011; Tan et al. 2005). Figure 1 shows the basic structure of classification decision trees.

The statistical performance of the classification decision tree was assessed through the area under the ROC curve (AUC) (Deng et al. 2016; Hanley and McNeil 1982; Pendharkar 2009; Tan et al. 2006) and the cross-validation test (Choudhury et al. 2021). AUC values equal to 0.5 indicate that the test is not informative, values between 0.5 and 0.7 indicate an inaccurate test, values between 0.7 and 0.9 indicate a moderate test. Instead, values between 0.9 and 1 or equal to 1 indicate that the test is highly accurate or perfect (Swets 1988). However, values of AUC in the range between 0.9 and 1 may indicate overfitting and analysis bias (Foucher and Danger 2012).

## 4 Findings

Figure 2 shows the predictive and analytical power of the CHAID classification decision tree. The tree starts with the root node (node 0), which indicates that 8.6% of permanent employees have recently left their job voluntarily. The classification
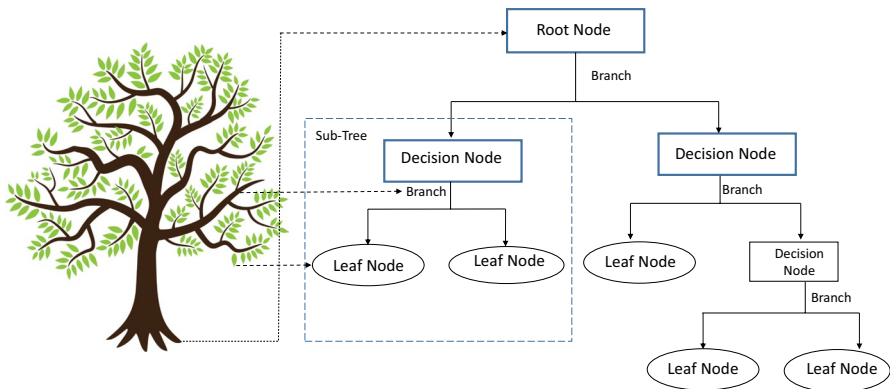


**Fig. 1** Structure of the classification decision tree model. Source: Adapted from Li et al. (2022)

tree then identifies three layers of predictors with different predictive power (from highest to lowest), which can profile groups of employees with a high propensity to voluntarily quit the firm.

The first layer (node 1 and node 2) pinpoints the most powerful predictor that influences voluntary employee turnover. The country location of employees is the predictor with the greatest discriminatory power. Two different scenarios—in terms of employee turnover determinants—emerge from the analysis.

In Norway, the classification decision tree identifies three different groups of employees at risk of voluntary turnover. The first group includes employees with very low variety in their jobs (e.g. in terms of tasks) who also consider social media (e.g. Facebook) as important tools to share work-related knowledge in the firm (20.7%). The second group includes employees with low-medium job variety who are highly inclined to share work-related knowledge with colleagues to get promoted (16.9%). The third group of employees includes females with high job variety (12.1%). Instead, in Denmark, the classification decision tree identified six groups of employees with a high risk to voluntarily leave. The first three groups, respectively, include the employees with very low, low, and medium job freedom in terms of deciding how to manage their work (33.3%, 13.7%, 34.1%). The fourth group identifies the employees with high freedom in managing their jobs who also work in organizational contexts that consider job rotation an important tool to share work-related knowledge within the organization (28.3%). The fifth group includes employees with high job freedom working in organizational contexts where there is a medium propensity to consider job rotation as an important activity to share work-related knowledge (16.4%). Finally, another important group includes the employees with high job freedom working in contexts where there is a lower propensity to share work-related knowledge through job rotation (8.4%).

Table 3 summarizes the terminal nodes (or leaves) of the CHAID classification decision tree highlighting the split value used in the development of the predictive classification and the $p$-value for each predictor. Predictors such as country,
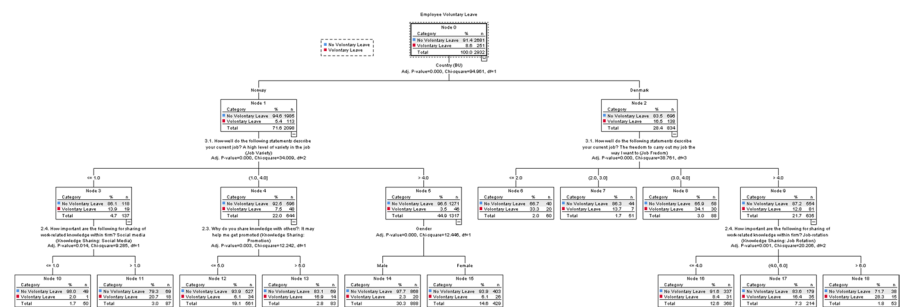


**Fig. 2** CHAID classification decision tree

**Table 3** Classification decision tree terminal nodes

| Node | Country | No Voluntary Leave | Voluntary Leave | Primary Independent Variable | Sig.[a] | Split Values |
|---|---|---|---|---|---|---|
| | | Percent | Percent | Variable | | |
| 0 | Norway and Denmark | 91.4% | 8.6% | | | |
| **Norway** | | | | | | |
| First Layer | | | | | | |
| 1 | Norway | 94.6% | 5.4% | Country (organizational) | 0.000 | Norway |
| Second Layer | | | | | | |
| 3 | Norway | 86.1% | 13.9% | Job variety (job related) | 0.000 | <=1.0 |
| 4 | Norway | 92.5% | 7.5% | Job variety (job related) | 0.000 | (1.0,4.0] |
| 5 | Norway | 96.5% | 3.5% | Job variety (job related) | 0.000 | >4.0 |
| Third Layer | | | | | | |
| 10 | Norway | 98.0% | 2.0% | Knowledge-sharing: social media (organizational) | 0.014 | <=1.0 |
| 11 | Norway | 79.3% | 20.7% | Knowledge-sharing: social media (organizational) | 0.014 | >1.0 |
| 12 | Norway | 93.9% | 6.1% | Know-sharing: promotion (job related) | 0.003 | <=5.0 |
| 13 | Norway | 83.1% | 16.9% | Knowledge-sharing: promotion (job related) | 0.003 | >5.0 |
| 14 | Norway | 97.7% | 2.3% | Gender (individual attribute) | 0.000 | Male |
| 15 | Norway | 93.9% | 6.1% | Gender (individual attribute) | 0.000 | Female |
| **Denmark** | | | | | | |
| First Layer of Prediction | | | | | | |
| 2 | Denmark | 83.5% | 16.5% | Country (organizational) | 0.000 | Denmark |
| Second Layer of Prediction | | | | | | |
| 6 | Denmark | 66.7% | 33.3% | Job freedom (job related) | 0.000 | <=2.0 |
| 7 | Denmark | 86.3% | 13.7% | Job freedom (job related) | 0.000 | (2.0, 3.0] |
| 8 | Denmark | 65.9% | 34.1% | Job freedom (job related) | 0.000 | (3.0, 4.0] |

**Table 3** (continued)

| Node | Country | No Voluntary Leave | Voluntary Leave | Primary Independent Variable | | |
|---|---|---|---|---|---|---|
| | | Percent | Percent | Variable | Sig. [a] | Split Values |
| **Third Layer of Prediction** | | | | | | |
| 16 | Denmark | 91.6% | 8.4% | Knowledge-sharing: job rotation (organizational) | 0.001 | < =4.0 |
| 17 | Denmark | 83.6% | 16.4% | Knowledge-sharing: job rotation (organizational) | 0.001 | (4.0, 6.0] |
| 18 | Denmark | 71.7% | 28.3% | Knowledge-sharing: job rotation (organizational) | 0.001 | > 6.0 |
| Robustness checks | | | | | | |
| AUC | 74.5 | | | | | |
| AUC (Cross-validation test) | 74.5 | | | | | |

[a] Bonferroni-adjusted ($p < 0.000$ ***; $p < 0.01$ **; $p < 0.005$*), N = 2,932 (N = 834 Denmark, N = 2,098 Norway)

job variety, gender, and job freedom have high statistical relevance ($p<0.000$) in predicting the determinants of the voluntary employee turnover, while "knowledge-sharing with colleagues in order to get promoted", "importance of knowledge-sharing through social media in the company" and "importance of job rotation as an activity to share work-related knowledge within the organization" have a medium statistical relevance ($p<0.001$). The predictor related to knowledge-sharing through social media has low predictive value ($p<0.05$).

The goodness of fit of the CHAID classification decision tree is moderate, as indicated by the AUC equal to 74.5%. The cross-validation test confirms the moderate accuracy of the model, excluding overfitting issues. Figure 3 provides a representation of the ROC curve.

## 5 Discussion

The CHAID classification decision tree identified seven predictors of voluntary employee turnover. The country location of employees has the highest predictive power, suggesting that the organizational context plays a crucial role in understanding why permanent employees voluntarily leave the firm. This is consistent with previous research showing the importance of organizational variables in this decision (Rubenstein et al. 2018). Then, the model identified six predictors of which

**Fig. 3** ROC (receiver operating characteristic) curve
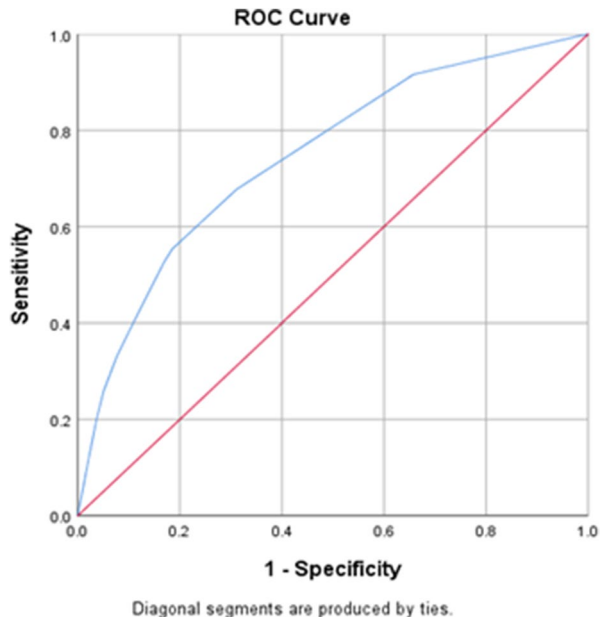
four refer to the Norwegian context (job variety, importance of sharing work-related knowledge through social media in the firm, propensity to share work-related knowledge to get promoted, and gender) and two to the Danish one (job freedom, importance of sharing work-related knowledge within the firm through job rotation). The analysis identifies predictors which are organizational, job-related, and related to individual attributes in line with previous studies in high-tech contexts (Ghapanchi and Aurum 2011).

The model identifies groups of employees, who are internally homogeneous and heterogeneous with respect to each other and who have a high propensity to leave voluntarily. This approach allows for the profiling of the employees who are more likely to leave voluntarily, highlighting the concurrent effect of predictors for specific groups of employees. For example, our analysis has led to the identification of groups of employees who leave voluntarily the company, which we labelled for the discussion. In the Norwegian context, job variety is the most important predictor, and the model identifies three groups of employees. The first group, which we labelled "bored workers", includes employees who have relatively repetitive job tasks (Foss et al. 2009) and who want to use the social media to share knowledge within the organization (Cabrera and Cabrera 2005). The second one is "ambitious workers" who are employees with a medium job variety (Foss et al. 2009), but who consider important to share knowledge in order to get promoted (Ryan and Deci 2000). The third group identifies "female workers" who have a high job variety (Foss et al. 2009). An imposed high level of tasks job variety may imply more effort and resources for employees who need to change job frequently. Further studies could explore these relationships and their motivations in more detail. Instead, in Denmark, job freedom is the most significant predictor (Foss et al. 2009), which identifies six groups of employees, which we have labeled as follows. "Rebels I", "Rebels II" and Rebels III" are the employees who are at risk of voluntary turnover because they have limited job freedom (Foss et al. 2009). "Impatients I", "Impatients II" and "Impatients III" are the groups of employees who have a high degree of job freedom but work in an organizational context that views job rotation as a tool for sharing knowledge within the organization (Cabrera and Cabrera 2005). Employees who are relatively free in their work organization tend to be reluctant to engage in job rotation. Job rotation may involve the reorganization of tasks and active confrontation with other employees and departments, thus threatening to restrict freedom in the way and timing of individual work organization. This suggests that job rotation may have a constraining effect on the employees who are free to organize their work. Future studies could explore the interactions between these variables.

In addition, the CHAID classification decision tree identifies the set or bundle of determinants that influence the decision to leave, enabling a predictive profiling of employees. Thus, unlike traditional statistical techniques (Hom et al. 2017; Garg et al. 2022), this method reveals which combinations of variables influence the decision to voluntarily leave the company by identifying specific groups of employees

operating in the organizational context. This method highlights a multiplicity of linear and non-linear effects among different groups of employees, paving the way for new lines of research to explore these aspects.

Our study makes two important theoretical contributions to the field of HRM. First, from a theoretical perspective, our study shows both the application and the benefits of the CHAID classification decision trees in selecting the determinants that characterize voluntary employee turnover and in profiling groups of employees at risk of voluntary turnover. As suggested by Hom et al. (2017), the application of supervised ML techniques could be used to advance the turnover literature. In this sense, the CHAID classification decision tree can be used to uncover linear and non-linear relationships in large databases. Supervised ML techniques could be used to analyze in depth relationships that have emerged in the past to complement past evidence and, eventually, lead to the identification of new relationships, especially in the presence with large databases. Second, in line with the considerations of Garg et al. (2022) considerations, the classification decision tree emerges as an effective technique that can be used to solve complex management problems, such as the employee turnover. However, this approach could also be combined with other ML tools to support decision making on complex HRM issues that require an understanding of socio-cultural phenomena (Garg et al. 2022; Yang et al. 2023). This suggestion could pave the way for the birth and growth of a research stream lying at the intersection of the HRM and ML literature, which could also properly assess the benefits and risks arising from this interaction.

This article also provides relevant managerial implications. By using ML techniques, in particular the classification decision trees, MNCs could develop effective ad hoc retention plans that precisely and accurately target the groups of employees who are more likely to leave voluntarily. For instance, in this case, our results suggest that a retention strategy that increases job variety and targets female employees in Norway would not be as effective for Danish employees whose propensity to leave is more related to job freedom (Foss et al. 2009). This shows that MNCs could use this method to effectively self-assess their employee turnover risk, identify employees at risk, and create tailored retention strategies that address the needs of different employee groups. In doing so, they can improve their chances of reducing voluntary employee turnover rates (e.g., Reiche 2008). In fact, developing effective retention strategies allows for the retention of current employees, thereby avoiding additional costs due to turnover (Mobley 1982; Price 1977; Saradhi and Palshikar 2011; Staw 1980) and negative effects on the overall organizational effectiveness and business success (Holtom et al. 2005; Mitchell et al. 2001).

## 6 Conclusion and limitations

This study shows the predictive and analytical power of ML techniques through the application of the CHAID classification decision tree to predict the determinants of voluntary employee turnover to profile groups of employees at risk of voluntary turnover. This research goes beyond merely identifying the probability

of employees at risk of voluntary turnover, as previously done in past studies (Choudhury et al. 2021). Indeed, this research shows the predictive and analytical power of the CHAID classification decision tree, and more generally of the supervised ML techniques, in analyzing large databases. In particular, through this research we highlight the advantages of this technique in the context of HRM by seeking to open new avenues for future applications related to classification problems in other management contexts where supervised ML techniques could be successfully used to support and improve the quality of the decision-making process (Iqbal et al. 2020; Janssen et al. 2017). ML techniques, especially the CHAID classification decision tree, appear to be a realistic way for decision-makers to obtain strategic knowledge from raw data considered relevant to steering the firm and could be used not only to solve simple management problems related to recruitment and performance management, but also complex ones such as forecasting employees' turnover, also eventually being used in combination with other ML techniques (Garg et al. 2022). However, while the mere implementation of supervised ML techniques to support the decision-making process is a good starting point, when not integrated with strategic thinking risks to produce biased results with a negative impact on the quality of business decisions (Choudhury et al. 2021). In fact, we would like to emphasize that supervised ML techniques it should not be seen as a replacement for human resource reasoning.

Our research also has some limitations that highlight opportunities for future research. First, the choice of classification algorithm influences the predictive power of the resulting model. However, to date, no complete theory or conceptual guidelines are available to assist researchers in choosing or developing appropriate classification decision tree algorithms. Thus, more research is needed to assess the selection of these algorithms according to a specific type of classification problem. In addition, we recall the importance of acknowledging that classification decision trees are sensitive to noisy data and could also not perform as good as neural network on non-linear data (Curram and Mingers 1994). Future research should test this technique on different contexts, databases, and, also, in comparison with other techniques. In addition, future studies could test the application of supervised ML techniques (e.g., decision trees, support vector machine, neural networks) on panel data, which could provide a more robust estimation of employee turnover predictors and represent an effective tool to accurately develop customized retention strategies able to considerably reduce the employee turnover risk over the time.

To conclude, we hope our study will generate interest and new stimuli in the application of supervised ML techniques, particularly, among management scholars, opening new lines of research not only limited to HRM but also in fields where the use of these techniques is still in its embryonic phase.

## Appendix

**Table 4** Overview of the review and meta-analysis articles on voluntary employee turnover

| Author(s) Publication Year), Journal | Research objective | Kind of review/meta-analysis | 'Sample' | Main results | Selective recommendations for future research | GS citations* |
|---|---|---|---|---|---|---|
| Griffeth et al. (2000), *Journal of Management* | Analysis of the predictive strength of turnover antecedents extending Hom and Griffeth's (1995) study | Meta-analysis | 500 correlations from 42 studies in the 1990s | Highest predictive strength: quit intentions, job search, organizational commitment, overall job and work satisfaction Modest to low predictive strength: most demographic predictors, and perceived alternatives | Call for new comprehensive turnover meta-analyses by extending tests of a priori specified moderators to earlier studies | 6,892 |
| Morrell et al. (2001), *International Journal of Management Reviews* | Review of the literature on labor turnover in organizations | Narrative Review | Not specified | Criticism of poor explanatory and predictive power of most turnover models | Call for turnover research with high predictive power and studies focused on variables that exceed the relationship between an affective state and turnover | 330 |
| Holtom et al. (2008), *Academy of Management Annals* | Critical review of prior research and extensive summary of the field coupled with recommendations for future research | Narrative Review | Building on previous reviews up to 1995 | Expansion from immediate causes and consequences to more distal ones, and from a focus on individual attitudes and turnover to other levels of interest, such as group or organizational variables | Call for international turnover research with more consideration of time in the turnover process and more precision in measuring and modeling both independent and dependent variables | 1,608 |

**Table 4** (continued)

| Author(s) Publication Year), Journal | Research objective | Kind of review/meta-analysis | 'Sample' | Main results | Selective recommendations for future research | GS citations* |
|---|---|---|---|---|---|---|
| Hom et al. (2012), *Psychological Bulletin* | Review of the literature aimed at identifying shortcomings with prevailing turnover dimensions | Narrative Review | Not specified | Development of a framework that links both the configurational approach with motivational states and the antecedents for staying and turnover criteria | Call for research investigating the influence of different mindset types (e.g. self-positiveness vs low core self-evaluations) | 800 |
| Jiang et al. (2012), *Journal of Applied Psychology* | Introduction of an overall model of the relationships between job embeddedness and turnover outcomes | Meta-analysis | 52 studies and 65 independent samples ($N = 42,907$) | The negative relationships between the job embeddedness and turnover criteria are stronger in female-dominated samples and public organizations. Turnover intentions, job search behavior, and job performance fully (partially) mediate the effect of on-the-job embeddedness on actual turnover | Call for research investigating the unique predictive power of job embeddedness, beyond other constructs, as a major predictor of employee turnover, as well as the boundary conditions and the mechanism underlying its relationship with turnover | 608 |

**Table 4** (continued)

| Author(s) Publication Year), Journal | Research objective | Kind of review/meta-analysis | 'Sample' | Main results | Selective recommendations for future research | GS citations* |
|---|---|---|---|---|---|---|
| Lee et al. (2017), *Academy of Management Perspectives* | Conceptual article starting with a brief synopsis of major theories to set the stage for our suggestions on future research directions | Essay | Not specified | Description of the most relevant theories used in turnover research and constructs explaining employee decisions to participate in organizations. Data on predictors collected at time 1, data on employee leave collected at time 2 (e.g. after 6–24 months). Employee turnover is mainly estimated through regression or structural equation models, mostly explaining only 25% of turnover variance | Call for the application of novel data collection (e.g. data collected in real time and as continuous ratings) and analytical (e.g. social network analysis, natural language processing, and machine learning techniques) tools to understand the reasons why employees leave from firms | 246 |

**Table 4** (continued)

| Author(s) Publication Year, Journal | Research objective | Kind of review/meta-analysis | 'Sample' | Main results | Selective recommendations for future research | GS citations* |
|---|---|---|---|---|---|---|
| Hom et al. (2017), *Journal of Applied Psychology* | Literature review aimed at highlighting key aspects that have shaped turnover research or management practice | Narrative Review | 100 years of employee turnover publications | Key theoretical and methodological contributions to the turnover literature published in the Journal of Applied Psychology and elsewhere | Theorizing and studying change in turnover antecedents and consequences, considering the role of "time", in the analyses. Need to investigate post-turnover implications for individuals and organizations. Capturing the role of context to underline under which conditions some factors are more relevant than others | 1070 |
| Rubenstein et al. (2018), *Personnel Psychology* | Identifying predictors of voluntary employee turnover to propose future research directions | Meta-analysis | 316 articles | Identifying voluntary employee turnover predictors at different levels, such as individual attribute predictors, predictors reflecting aspects of the job, traditional job attitude predictors, newer personal condition predictors, organizational context predictors, person-context interface predictors, etc | Developing complex multilevel path models, employing methods to capture turnover processes in real time to establish the temporal order and capture interactions | 501 |

**Table 4** (continued)

| Author(s) Publication Year), Journal | Research objective | Kind of review/meta-analysis | 'Sample' | Main results | Selective recommenda-tions for future research | GS citations* |
|---|---|---|---|---|---|---|
| Field et al. (2020), *Journal of Business and Psychology* | Assessing the robust-ness of turnover research | Meta-analysis | 205 distributions from seven recently pub-lished meta-analyses on turnover between 2005 and 2009 | Identifying the reasons why turnover research might not be robust. Turnover research largely influenced by outliers and publica-tion bias | Need for further rigor-ous investigations of the "role conflict and turnover" relation | 25 |

* Google Scholar citations as of 4 December 2023

**Table 5** Description and measure of the predictors

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| Individual attributes | | | |
| Gender | Gender | Nominal | 1 = Male 2 = Female |
| Organizational determinants | | | |
| BU | Country of the subsidiary | Nominal | 1 = Norway 2 = Denmark |
| *To what degree do you think the following statements are characteristic of the department in which you work? (O' Reilly III et al. 1991)* | | | |
| curcult1 | Risk taking | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult2 | Working in collaboration with others | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult4 | Having integrity | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult5 | Paying attention to detail | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult6 | Being results-oriented | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult7 | Being customer-oriented | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |
| curcult8 | Sharing information freely | Scale | 7-points Likert scale 1 = not characteristic at all 7 = highly characteristic |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| curcult10 | Being team-oriented | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult11 | High ethical standards | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult12 | Empowering employees | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult13 | Being precise | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult14 | Having high expectations for performance | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult15 | Listening to customers | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult16 | Individual goals are transparent | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult17 | Predictability | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult18 | Being aggressive | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| curcult19 | Being respectful | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult20 | Being honest | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult21 | Emphasizing quality | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult23 | Being market-driven | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN1 | Being adaptable | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN2 | Employees can influence important decisions | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |

*How important are the following activities for sharing work-related knowledge in the company?* (Cabrera and Cabrera 2005)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| act1 | Face-to-face meetings with colleagues | Scale | 7-points Likert scale<br>1 = Not Important<br>7 = Very Important |
| act2 | Telephone, Lync live meeting conferences | Scale | 7-points Likert scale<br>1 = Not Important<br>7 = Very Important |
| act3 | Video conferences with colleagues | Scale | 7-points Likert scale<br>1 = Not Important<br>7 = Very Important |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| act4 | Job-rotation | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| act8 | Projects of task groups | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| act9 | Formal best-practice sharing organized by HQ/Group Units | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| act10 | Information sharing in Sharepoint/WOW | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| actN1 | Attend specialized-skills training in my field of expertise | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| actN2 | Attend general-skills training programs | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| actN3 | Participation in seminars and workshops | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| actN4 | Management meeting or other social interaction | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |
| actN5 | Performance appraisals (TDP) | Scale | 7-points Likert scale 1 = Not Important 7 = Very Important |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| actN6 | Social media (Facebook, Facebook at work, other) | Scale | 7-points Likert scale<br>1 = Not Important<br>7 = Very Important |
| *In your opinion, to what degree do the following effects of knowledge-sharing actually occur in the company?(Castellacci et al. 2018)* | | | |
| knout1 | Creation of new products/services | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knout2 | Creation of new ways of work that make us more efficient | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knout3 | Improvement of existing products / services | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knout4 | Improvements of existing ways of work | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knout5 | Adoption products/services from one BU to another | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knout6 | Adoption of ways of work from one BU to another | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| *To what degree do you think the following statements are characteristic of the department in which you work? (O' Reilly III et al. 1991)* | | | |
| curcult25 | Tolerating minor compromises on ethics to achieve important business goals | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| curcult26 | Being willing to experiment | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult28 | High levels of conflict | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult29 | Being analytical | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult30 | Achievement-oriented | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult31 | Being innovative | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult32 | Cooperative | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult33 | Keeping promises | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult35 | Fast moving | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult36 | Learning from mistakes | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| curcult37 | Action oriented | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult38 | Being quick to take advantage of opportunities | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult39 | Being supportive | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult41 | Taking initiative | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcult46 | Following ethical guidelines | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN3 | Not being constrained by many rules | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN4 | Being fair | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN5 | Confronting conflict directly | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN6 | Being praised for individual performance | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| curcultN7 | Being flexible | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN8 | Management promotes a culture in which corruption and bribery is unacceptable | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN9 | Top-down decision-making | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| curcultN10 | Fear of failure | Scale | 7-points Likert scale<br>1 = not characteristic at all<br>7 = highly characteristic |
| *Sharing knowledge can sometimes be challenging. To what degree do you think the following statements are characteristic to your department?* (Dasí et al. 2017) | | | |
| knch1 | Employees find it difficult to find experts outside of their own department | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knch2 | Employees don't share knowledge for fear of becoming less valuable | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knch3 | Employees find it difficult to share knowledge with colleagues they never worked with before | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knch4 | There is a prevailing attitude that employees ought to fix their own problems within the department | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| knch5 | Management actively discourages employees from sharing knowledge | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| knch6 | I have the required time to help others | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |

*Successful innovation depends on many conditions. To what degree do you think the following statements are characteristic of your department? (Castellacci et al. 2018)*

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| innch1 | There is adequate funding for executing on new ideas | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| innch2 | People have sufficient time to develop new ideas | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| innch3 | Our KPIs incentivize innovation | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| innch4 | Managers support improvement ideas from employees | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| innch5 | When employees have ideas, they know where to go to get sponsorship | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |
| innch6 | People have the necessary tools to execute on new ideas | Scale | 7-points Likert scale<br>1 = To a Limited Degree<br>7 = To a High Degree |

Job-related Determinants

*How often do you share the following types of knowledge with colleagues who work in other departments within your business unit? (Gupta and Govindarajan 2000)*

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| knbatl1 | Knowledge about customer groups and markets (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| knbat12 | Knowledge about new service development (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat13 | Knowledge about new ways to serve the customers (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat14 | Knowledge about technology and telecom infrastructure (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat15 | Best practices about, for instance, marketing, technology or sourcing (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat17 | Knowledge that originates from external collaborators (across functions within BUs)? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| *How often, approximately, do you share the following types of knowledge with colleagues who work in other BUs?* (Gupta and Govindarajan 2000) | | | |
| knbat21 | Knowledge about customer groups and markets? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat22 | Knowledge about new service development? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat23 | Knowledge about new ways to serve the customers? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat24 | Knowledge about technology and telecom infrastructure? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| knbat25 | Best practices about for instance, marketing, technology or sourcing? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| knbat27 | Knowledge that originates from external collaborators? | Scale | 7-points Likert scale<br>1 = Never<br>7 = Very Often |
| *Why do you share knowledge with others?* (Ryan and Deci 2000) | | | |
| knbat31 | I want my supervisor(s) to appreciate me | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat32 | I want my colleague(s) to appreciate me | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat33 | I think it is an important part of my job | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat34 | I find it personally satisfying? | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat35 | I might get a reward | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat36 | It may help me get promoted | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |
| knbat37 | I like sharing knowledge | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| knbat38 | I feel I have knowledge that can be useful for others | Scale | 7-points Likert scale<br>1 = Strongly Disagree<br>7 = Strongly Agree |

*In the firm, there is a number of opportunities for training and career development. How often have you taken part in these in the last 24 months?* (Foss et al. 2009)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| jobdes22 | Job rotation across departments within my own BU | Scale | 7-points Likert scale<br>1 = Never<br>4 = Occasionally<br>7 = Often |
| jobdes23 | Job rotation across different BUs | Scale | 7-points Likert scale<br>1 = Never<br>4 = Occasionally<br>7 = Often |
| jobdes25 | Specialized training in my field of expertise within my own BU | Scale | 7-points Likert scale<br>1 = Never<br>4 = Occasionally<br>7 = Often |
| jobdes26 | Specialized training in my field of expertise across different BUs | Scale | 7-points Likert scale<br>1 = Never<br>4 = Occasionally<br>7 = Often |
| jobdes27 | Seminars and workshops within my own BU | Scale | 7-points Likert scale<br>1 = Never<br>4 = Occasionally<br>7 = Often |

*How well do the following statements describe your current job?* (Foss et al. 2009)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| jobdes11 | The freedom to carry out my job the way I want to | Scale | 7-points Likert scale<br>1 = Not at All<br>7 = Very Well |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| jobdes12 | A high level of variety in the job | Scale | 7-points Likert scale<br>1 = Not at All<br>7 = Very Well |
| jobdes13 | The opportunity for independent initiative | Scale | 7-points Likert scale<br>1 = Not at All<br>7 = Very Well |
| jobdes14 | The opportunity to complete work that I started | Scale | 7-points Likert scale<br>1 = Not at All<br>7 = Very Well |
| jobdes16 | I can conduct the tasks independent of others | Scale | 7-points Likert scale<br>1 = Not at All<br>7 = Very Well |
| innfilt | How much of your time approximately do you spend on working directly with innovation? | Categorical | 0 = 0%<br>1 = Less than 25%<br>2 = Between 25 and 50%<br>3 = More than 50% Less than 75%<br>4 = Between 75 and 100% |

**Table 5** (continued)

| Variable | Description | Measure | Encoding |
|---|---|---|---|
| funcfilt | In which functional area do you work? Please choose the most appropriate alternative | Categorical | 1 = Marketing/sales<br>2 = Technology<br>3 = Network<br>4 = IS/IT<br>5 = Sourcing<br>6 = Operations<br>7 = Research<br>8 = Product development<br>9 = Strategy<br>10 = Finance<br>11 = HR<br>12 = Legal<br>13 = Top leadership (CXO level) |
| collfilt | Do you collaborate with colleagues who work outside of your own functional area? | Nominal | 1 = Yes<br>2 = No |
| collfilt2 | Do you collaborate with colleagues who work outside of your own business unit? | Nominal | 1 = Yes<br>2 = No |

## Declarations

**Competing interests and funding** The authors do not have competing interests and funding that are directly or indirectly related to the work to disclose.

## References

Aiolli F, De Filippo M, Sperduti A (2009) Application of the preference learning model to a human resources selection task. IEEE Symposium on Computational Intelligence and Data Mining, Nashville, pp 203–210. https://doi.org/10.1109/CIDM.2009.4938650

Alao D, Adeyemo AB (2013) Analyzing employee attrition using decision tree algorithms. Comput Inform Syst Dev Inform Allied Res J 4(1):17–28

Archaux C, Martin A, Khenchaf A (2004) An SVM based churn detector in prepaid mobile telephony. International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, pp 459–460. https://doi.org/10.1109/ICTTA.2004.1307830

Aviad B, Roy G (2011) Classification by clustering decision tree-like classifier based on adjusted clusters. Expert Syst Appl 38(7):8220–8228

Barrick MR, Zimmerman RD (2005) Reducing voluntary, avoidable turnover through selection. J Appl Psychol 90(1):159–166

Berry MA, Linoff GS (2000) Mastering data mining: the art and science of customer relationship management. Ind Manag Data Syst 100(5):245–246

Blazquez D, Domenech J (2018) Big data sources and methods for social and economic analyses. Technol Forecast Soc Chang 130:99–113

Blockeel H, Struyf J (2002) Efficient algorithms for decision tree cross-validation. J Mach Learn Res 3(12):621–650

Cabrera EF, Cabrera A (2005) Fostering knowledge sharing through people management practices. Int J Human Resour Manag 16(5):720–735

Canhoto AI, Clear F (2020) Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential. Bus Horiz 63(2):183–193

Castellacci F, Gulbrandsen M, Hildrum J, Martinkenaite I, Simensen E (2018) Functional centrality and innovation intensity: employee-level analysis of the Telenor group. Res Policy 47(9):1674–1687

Cerchiello P, Giudici P (2012) Non parametric statistical models for on-line text classification. Adv Data Anal Classif 6(4):277–288

Cerchiello P, Giudici P (2016) Big data analysis for financial risk management. Journal of Big Data 3(1):18–30

Choudhury P, Allen RT, Endres MG (2021) Machine learning for pattern discovery in management research. Strateg Manag J 42(1):30–57

Coetzer A, Inma C, Poisat P, Redmond J, Standing C (2019) Does job embeddedness predict turnover intentions in SMEs? Int J Product Perform Manag 68(2):340–361

Cui G, Wong ML, Lui HK (2006) Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. Manage Sci 52(4):597–612

Curram SP, Mingers J (1994) Neural networks, decision tree induction and discriminant analysis: an empirical comparison. J Oper Res Soc 45(4):440–450

Dasí À, Pedersen T, Gooderham PN, Elter F, Hildrum J (2017) The effect of organizational separation on individuals' knowledge sharing in MNCs. J World Bus 52(3):431–446

Deng X, Liu Q, Deng Y, Mahadevan S (2016) An improved method to construct basic probability assignment based on the confusion matrix for classification problem. Inf Sci 340:250–261

Díaz-Pérez FM, Bethencourt-Cejas M (2016) CHAID algorithm as an appropriate analytical method for tourism market segmentation. J Destin Mark Manag 5(3):275–282

Ekawati AD (2019) Predictive analytics in employee churn: a systematic literature review. J Manag Inform Decis Sci 22(4):387–397

Field JG, Bosco FA, Kepes S (2020) How robust is our cumulative knowledge on turnover? J Bus Psychol 36(3):349–365

Foss NJ, Minbaeva DB, Pedersen T, Reinholt M (2009) Encouraging knowledge sharing among employees: how job design matters. Hum Resour Manage 48(6):871–893

Foucher Y, Danger R (2012) Time dependent ROC curves for the estimation of true prognostic capacity of microarray data. Stat Appl Genet Mol Biol 11(6):871

Garg S, Sinha S, Kar AK, Mani M (2022) A review of machine learning applications in human resource management. Int J Product Perform Manag 71(5):1590–1610

Ghapanchi AH, Aurum A (2011) Antecedents to IT personnel's intentions to leave: a systematic literature review. J Syst Softw 84:238–249

Giudici P (2010) Scoring models for operational risk. In: Kenett RS, Raanan Y (eds) Operational risk management: a practical approach to intelligent data analysis. Wiley, pp 125–135

Giudici P, Figini S (2009) Applied data mining for business and industry. Wiley, Chichester

Glebbeck AC, Bax EH (2004) Is high employee turnover harmful? An empirical test using company records. Acad Manag J 47(2):277–286

Gordini N, Veglio V (2017) Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. Ind Mark Manage 62:100–107

Griffeth RW, Hom PW, Gaertner S (2000) A meta-analysis of antecedents and correlates of employee turnover: update, moderator tests, and research implications for the millennium. J Manag 26(3):463–488

Gupta AK, Govindarajan V (2000) Knowledge flows within multinational corporations. Strateg Manag J 21(4):473–496

Gupta S, Kar AK, Baabdullah A, Al-Khowaiter WA (2018) Big data with cognitive computing: a review for the future. Int J Inf Manage 42:78–89

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36

Holtom BC, Mitchell TR, Lee TW, Eberly MB (2008) Turnover and retention research. Acad Manag Ann 2(1):231–274

Holtom BC, Mitchell TR, Lee TW, Inderrieden EJ (2005) Shocks as causes of turnover: what they are and how organizations can manage them. Hum Resour Manage 44(3):337–352

Hom PW, Griffeth RW (1991) Structural equations modeling test of a turnover theory: cross-sectional and longitudinal analyses. J Appl Psychol 76(3):350–366

Hom PW, Griffeth RW (1995) Employee turnover. South-Western College Publishing. Cincinnati, OH

Hom PW, Lee TW, Shaw JD, Hausknecht JP (2017) One hundred years of employee turnover theory and research. J Appl Psychol 102(3):530–545

Hom PW, Mitchell TR, Lee TW, Griffeth RW (2012) Reviewing employee turnover: focusing on proximal withdrawal states and an expanded criterion. Psychol Bull 138(5):831–858

Hung SY, Yen DC, Wang HY (2006) Applying data mining to telecom churn management. Expert Syst Appl 31(3):515–524

Iqbal R, Doctor F, More B, Mahmud S, Yousuf U (2020) Big data analytics: computational intelligence techniques and application areas. Technol Forecast Soc Chang 153:119–253

Jain RK, Natarajan R, Ghosh A (2016) Decision tree analysis for selection of factors in DEA: an application to banks in India. Glob Bus Rev 17(5):1162–1178

Jiang K, Liu D, McKay PF, Lee TW, Mitchell TR (2012) When and how is job embeddedness predictive of turnover? A meta-analytic investigation. J Appl Psychol 97(5):1077–1096

Janssen M, van der Voort H, Wahyudi A (2017) Factor influencing big data decision-making quality. J Bus Res 70:338–345

Kass GV (1980) An exploratory technique for investigating large quantities of categorical data. J Roy Stat Soc: Ser C (Appl Stat) 29(2):119–127

Kock F, Berbekova A, Assaf AG (2021) Understanding and managing the threat of common method bias: detection, prevention and control. Tour Manage 86:104–330

Koechling A, Wehner MC, Warkocz J (2023) Can I show my skills? Affective responses to artificial intelligence in the recruitment process. RMS 17(6):2109–2138

Kruppa J, Schwarz A, Arminger G, Ziegler A (2013) Consumer credit risk: Individual probability estimates using machine learning. Expert Syst Appl 40(13):5125–5131

Lee TH, Gerhart B, Weller I, Trevor CO (2008) Understanding voluntary turnover: path-specific job satisfaction effects and the importance of unsolicited job offers. Acad Manag J 51(4):651–671

Lee TW, Hom PW, Eberly MB, Li J, Mitchell TR (2017) On the next decade of research in voluntary employee turnover. Acad Manag Perspect 31(3):201–221

Lee TW, Mitchell TR (1994) An alternative approach: the unfolding model of voluntary employee turnover. Acad Manag Rev 19(1):51–89

Li X, Yi S, Cundy AB, Chen W (2022) Sustainable decision-making for contaminated site risk management: a decision tree model using machine learning algorithms. J Clean Prod 371:133612

Linoff GS, Berry MJ (2011) Data mining techniques: for marketing, sales, and customer relationship management. Wiley

Lockamy A, Service RW (2011) Modeling managerial promotion decisions using Bayesian networks: an exploratory study. J Manag Dev 30(4):381–401

Maertz CP, Campion MA (2004) Profiles in quitting: Integrating content and process turnover theory. Acad Manag J 47(4):566–582

March J, Simon H (1958) Organizations. Wiley, New York

Mitchell TR, Holtom BC, Lee TW, Sablynski CJ, Erez M (2001) Why people stay: Using job embeddedness to predict voluntary turnover. Acad Manag J 44(6):1102–1121

Mobley WH (1977) Intermediate linkages in the relationship between job satisfaction and employee turnover. J Appl Psychol 62(2):237–240

Mobley WH (1982) Some unanswered questions in turnover and withdrawal research. Acad Manag Rev 7(1):111–116

Morrell K, Loan-Clarke J, Wilkinson A (2001) Unweaving leaving: the use of models in the management of employee turnover. Int J Manag Rev 3(3):219–244

Naeem R, Kohtamäki M, Parida V (2024) Artificial intelligence enabled product–service innovation: past achievements and future directions. Rev Manag Sci 44–1. https://doi.org/10.1007/s11846-024-00757-x

Nandialath AM, David E, Das D, Mohan R (2018) Modeling the determinants of turnover intentions: a Bayesian approach. Evid-based HRM: Glob Forum Empir Scholarsh 6(1):2–24 (**Emerald Publishing Limited**)

Nisbet R, Miner G, Yale K (2018) Handbook of statistical analysis and data mining applications, 2nd edn. Academic Press, Boston

O'Reilly CA III, Chatman J, Caldwell DF (1991) People and organizational culture: a profile comparison approach to assessing person-organization fit. Acad Manag J 34(3):487–516

Pendharkar PC (2009) Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. Expert Syst Appl 36(3):6714–6720

Pereira RB, Plastino A, Zadrozny B, Merschmann LH (2018) Categorizing feature selection methods for multi-label classification. Artif Intell Rev 49(1):57–78

Perner P, Zscherpel U, Jacobsen C (2001) A comparison between neural networks and decision trees based on data from industrial radiographic testing. Pattern Recogn Lett 22(1):47–54

Piramuthu S (2008) Input data for decision trees. Expert Syst Appl 34(2):1220–1226

Podsakoff PM, MacKenzie SB, Podsakoff NP (2012) Sources of method bias in social science research and recommendations on how to control it. Annu Rev Psychol 63:539–569

Porter LW, Steers RM (1973) Organizational, work, and personal factors in employee turnover and absenteeism. Psychol Bull 80(2):151–176

Price JL (1977) The study of turnover. Iowa State Press

Price JL (2001) Reflections on the determinants of voluntary turnover. Int J Manpow 22(7):600–624

Quinn A, Rycraft JR, Schoech D (2002) Building a model to predict caseworker and supervisor turnover using a neural network and logistic regression. J Technol Hum Serv 19(4):65–85

Raguseo E, Vitari C (2018) Investments in big data analytics and firm performance: an empirical investigation of direct and mediating effects. Int J Prod Res 56(15):5206–5221

Reddy GT, Reddy MPK, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G, Baker T (2020) Analysis of dimensionality reduction techniques on big data. IEEE Access 8:54776–54788

Reiche BS (2008) The configuration of employee retention practices in multinational corporation's foreign subsidiaries. Int Bus Rev 17(6):676–687

Rode JC, Rehg MT, Near JP, Underhill JR (2007) The effect of work/family conflict on intention to quit: The mediating roles of job and life satisfaction. Appl Res Qual Life 2(2):65–82

Rombaut E, Guerry MA (2018) Predicting voluntary turnover through human resources database analysis. Manag Res Rev 41(1):96–112

Rosset S, Neumann E, Eick U, Vatnik N (2003) Customer lifetime value models for decision support. Data Min Knowl Disc 7(3):321–339

Rubenstein AL, Eberly MB, Lee TW, Mitchell TR (2018) Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. Pers Psychol 71(1):23–65

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68

Sandhya S, Sulphey MM (2021) Influence of empowerment, psychological contract and employee engagement on voluntary turnover intentions. Int J Product Perform Manag 70(2):325–349

Saradhi VV, Palshikar GK (2011) Employee churn prediction. Exp Syst Appl 38(3):1999–2006

Sexton RS, McMurtrey S, Michalopoulos JO, Smith AM (2005) Employee turnover: a neural network solution. Comput Oper Res 32(10):2635–2651

Shah N, Irani Z, Sharif AM (2017) Big data in an HR context: exploring organizational change readiness, employee attitudes and behaviors. J Bus Res 70:366–378

Sheng J, Amankwah-Amoah J, Wang X (2017) A multidisciplinary perspective of big data in management research. Int J Prod Econ 191:97–112

Staw BM (1980) The consequences of turnover. J Occup Behav 1(4):253–273

Swets JA (1988) Measuring the accuracy of diagnostic system. Science 240(4857):1285–1293

Tan PN, Steinbach M, Kumar V (2005) Association analysis: basic concepts and algorithms. Introduction to data mining. Addison-Wesley, Boston, pp 71–94

Tan PN, Steinbach M, Kumar V (2006) Classification: basic concepts, decision trees, and model evaluation. Introduction to data mining. Pearson Addison-Wesley, pp 25-44

Tso GK, Yau KK (2007) Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy 32(9):1761–1768

Ture M, Tokatli F, Kurt I (2009) Using Kaplan-Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients. Exp Syst Appl 36(2):2017–2026

Yang Y, Shamim S, Herath DB, Secchi D, Homberg F (2023) The evolution of HRM practices: big data, data analytics, and new forms of work. RMS 17(6):1937–1942

Wei CP, Chiu IT (2002) Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl 23(2):103–112

Wirges F, Neyer AK (2023) Towards a process-oriented understanding of HR analytics: implementation and application. RMS 17(6):2077–2108

Zhao X (2008) An empirical study of data mining in performance evaluation of HRM. IntSymp Intell Inform Technol Appl Workshops 82–85https://doi.org/10.1109/IITA.Workshops.2008.235

Zhu J, Gonçalves AL, Uren, VS, Motta E, Pacheco R (2005) Mining web data for competency management. The 2005 IEEE/WIC/ACM international conference on web intelligence.(WI'05), Compiegne, pp 94–100. https://doi.org/10.1109/WI.2005.99

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.