SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2024
ORGANIZATION AND MANAGEMENT SERIES NO. 201

PREDICTION OF VOLUNTARY EMPLOYEE TURNOVER USING MACHINE LEARNING

Marcin NOWAK

Faculty of Engineering Management, Poznan University of Technology; marcin.nowak@put.poznan.pl,
ORCID: 0000-0001-5005-7820

Purpose: The aim of the article was to develop a method for predicting the occurrence of voluntary employee turnover intentions.
Design/methodology/approach: The objectives are achieved through the employment of machine learning algorithms, specifically decision tree algorithms, support vector machines, k-nearest neighbors, and naive Bayes classifiers. The article includes a literature review on voluntary employee turnover and the fundamentals of machine learning. It then presents the developed method for predicting employee turnover, which is evaluated under real-world conditions.
Findings: The research demonstrates that the proposed machine learning methods can effectively predict voluntary employee turnover intentions. The analysis and results indicate that these predictive models can identify early signs of turnover with significant accuracy, providing valuable insights into employee retention dynamics.
Research limitations/implications: (The study's limitations include the potential for overfitting in machine learning models and the need for large, high-quality datasets to train the models. Future research should focus on testing the proposed methods in various organizational settings and exploring additional variables that may influence employee turnover intentions.
Practical implications: The practical outcome of this research is the creation of a tool for more effective human resource management, particularly in the context of talent management. Organizations can use this tool to identify employees at risk of leaving and implement targeted retention strategies, ultimately reducing turnover rates and associated costs.
Social implications: By reducing voluntary employee turnover, organizations can foster more stable and supportive work environments, contributing to overall employee well-being and job satisfaction. This can enhance public perception of corporate social responsibility and positively influence industry standards.
Originality/value: This paper introduces a novel application of machine learning techniques to predict voluntary employee turnover intentions. The findings are valuable to human resource professionals, organizational managers, and scholars in the fields of management and quality sciences, offering a data-driven approach to improving employee retention strategies.
Keywords: voluntary employee turnover, employee turnover, talent management, machine learning.
Category of the paper: Research paper.

1. Introduction

Talent management has become one of the key research areas in human resource management (Wójcik, 2018). One element of retaining the most talented employees within an organization is monitoring the risk of their voluntary departure. The aim of this article was to develop a method for predicting the occurrence of voluntary employee turnover intentions. Turnover intention is the strongest predictor of actual turnover (Steel, Ovalle, 1984). On the other hand, forecasting the occurrence of turnover intentions among competent employees creates the possibility to take steps aimed at retaining talent within the organization. To achieve the goal set forth in the article, selected machine learning models were utilized, including the decision tree algorithm, support vector machine algorithm, k-nearest neighbors algorithm, and naive Bayes algorithm. The empirical part of the article presents the results of the author's research, which involved evaluating the method under real-world conditions in a large IT company based in Poznań. The research was conducted from October to December 2023.

The achievement of the article's goals was conditioned by the following structure. After the introduction (first part), the literature review section describes the theoretical foundations of the issue of voluntary employee turnover and the essence of machine learning. In the third part of the study, the structure of the proposed method for forecasting voluntary employee turnover intentions is presented. The fourth part of the study comprises the results of the author's research, which detail the method of predicting the occurrence of voluntary employee turnover intentions under real-world conditions.

2. Literature review

2.1. Voluntary Employee Turnover

Voluntary employee turnover refers to any job departure initiated by the employee (Bolt et al., 2022). Voluntary employee turnover is a significant area of research in contemporary management sciences (Madigan, Kim, 2021; Aswale, Mukul, 2020). This research is typically part of the human resource management field, particularly talent management (Holtom et al., 2005). Talent management is understood as "the organization's effort to attract, select, develop, and retain talented key employees" (Stuss, 2021, p. 52). The most important areas of interest in talent management include identifying talents, their development and training, and talent retention processes (Miś, 2009; Ingram, 2011). It is noted that organizations that effectively manage talents not only achieve better financial results but also enjoy a better reputation in the labor market, which attracts new talents (Knap-Stefaniuk, Karna, 2017). The loss of talent due to voluntary departures entails various costs for companies, both financial and non-financial (Moczydłowska, 2014; Dolot, 2019).

Research on voluntary employee turnover has been conducted for over 100 years. As early as 1915-1920, the causes of voluntary employee departures were analyzed (Diemer, 1917; Eberle, 1919). In 1925, studies were conducted to determine whether appropriate questions during the recruitment stage could predict voluntary employee turnover (Bills, 1925). It is indicated that the first mature theory of voluntary employee turnover was presented by March and Simon in 1958 (1993). Another significant concept for the development of this theory was Mobley's model of voluntary employee turnover (1979). It explains the employee departure process as a sequence from dissatisfaction, through considering resignation and evaluating alternatives, to deciding to change jobs and leaving (Lee et al., 2017). Among other important models related to voluntary employee turnover are:

- Price's model (1977) – which emphasized the significant impact of social factors on the decision to leave.
- Mobley et al. (1979) – which introduced, among other things, the concept of the subjective expected utility of the current job.
- Price and Mueller's model (1986) – which highlighted the significant impact of family relationships on the decision to voluntarily leave.

In the 21st century, further work on the issue of voluntary employee turnover continues. The most current theories include those developed by Mitchell et al. (2001), Nyberg and Ployhart (2013), and Hom et al. (2017, 2020). These contemporary theories focus on the causes of turnover being a result of both individual employee characteristics and organizational context.

A key area of research since the 1950s has been the identification of the causes of voluntary employee turnover. The most important factors that may influence turnover intention typically include: age, education, job satisfaction, tenure, relationships with supervisors, relationships with colleagues, job performance, assessment of working conditions, job level, job satisfaction, promotion opportunities, perception of workplace fairness, assessment of management style, perceived stress at work, job monotony, workload assessment, work-life balance, and ease of commuting to the workplace (Hom, 2017). Further analyses confirmed that intention is the main and direct factor influencing voluntary employee turnover (Steel, Ovalle, 1984). The concept of turnover intention is defined as "the conscious and deliberate willfulness of an employee to leave the organization" (Tett, Meyer, 1993).

2.2. Machine learning

In a general sense, machine learning can be understood as an area of artificial intelligence focused on algorithms that automatically and autonomously improve through experience derived from exposure to data (Cichosz, 2007). The development of machine learning is closely related to advancements in computer science and statistics.

Several types of machine learning are distinguished (Rebala et al., 2019):

- Supervised learning – involves training a machine learning model using data that contains both inputs and corresponding output labels, allowing the model to learn to map inputs to outputs.
- Semi-supervised learning – uses both labeled and unlabeled data to train machine learning models, enabling more efficient learning and generalization.
- Unsupervised learning – involves analyzing data without predefined labels, where the machine learning model identifies hidden patterns and structures in the data on its own.
- Reinforcement learning – is based on the trial-and-error method, where the machine learning model learns to make decisions by interacting with the environment and receives feedback in the form of rewards or penalties for actions taken, aiming to maximize the sum of rewards.

A variety of algorithms are used within machine learning. The most popular include artificial neural networks, support vector machines (linear and nonlinear), linear regression, logistic regression, nearest neighbor algorithms, and naive Bayes classifiers (Alsariera et al., 2022). Machine learning helps, for example, optimize HR, including recruitment and talent management. Algorithms enable the analysis of resumes and employee performance, supporting candidate identification and career development planning.

3. Methods

The proposed method for predicting voluntary employee turnover intentions can be presented as a procedure consisting of five steps:

Step 1. Preparing the database for machine learning

At the beginning of the learning process, a catalog of variables influencing turnover intentions, such as salary or job satisfaction, is established and represented as a vector $X = \{X_1, X_2, \ldots, X_n\}$. The variables can be of different types – from categorical to quantitative, binary, or multistate, which is an advantage of machine learning. Next, a survey questionnaire is created to collect data on the intensity of these characteristics from at least 50 employees, which is a requirement for most machine learning algorithms.

Step 2. Selecting significant variables through dimensionality reduction

First, all variables in the database are normalized. Then, to avoid performing machine learning procedures on an excessively large catalog of variables, dimensionality reduction is performed. It is proposed to use correlation statistics between the values of individual variables and the label for dimensionality reduction in the proposed method. Since the label in predicting the occurrence of voluntary turnover intentions is binary and most variables are quasi-continuous multistate, it is proposed to use the point-biserial correlation coefficient (1).

$$r_{pb} = \frac{M_1 - M_0}{\sigma} \cdot \sqrt{\frac{n_1 \cdot n_2}{n^2}} \quad (1)$$

where:
r_{pb} is the point-biserial correlation coefficient,
M_1 is the mean value of the variable for those exhibiting turnover intention,
M_0 is the mean value of the variable for those not exhibiting turnover intention,
σ is the standard deviation of the variable,
n_1 is the number of individuals exhibiting turnover intention,
n_2 is the number of individuals not exhibiting turnover intention,
n is the total number of individuals in the dataset.

For variables that are not quasi-continuous multistate, Cramér's V correlation coefficient will be used.

Step 3. Selecting machine learning models and their key hyperparameters

In this step, the machine learning models to be tested are selected. Possible models include linear regression, logistic regression, linear and nonlinear support vector machines, nearest neighbor models, naive Bayes classifiers, and artificial neural networks (Alsariera et al., 2022). For each of the analyzed models, their hyperparameters are defined, representing the necessary assumptions for training the algorithms.

Step 4. Splitting the database into training and test sets

In machine learning, the dataset is divided into two subsets – the training set and the test set. Both sets contain records with assigned feature values and labels. The training set usually comprises 70-90% of the data (Nguyen et al., 2021). The division of records in the database into training and test sets must be random. The purpose of splitting the dataset into two subsets is that the algorithm learns on the training set and is verified on the test set. The essence of a well-trained model is to maximize classification accuracy on the test set.

Step 5. Training prediction models for voluntary turnover intentions and selecting the most effective one

In this step, the training process is performed on the training set for each of the analyzed machine learning algorithms. The most effective machine learning model is selected using the accuracy metric (2).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where:
Accuracy is the effectiveness of the machine learning model,
TP (True Positives) is the value correctly classified as True,
FP (False Positives) is the value incorrectly classified as True,
TN (True Negatives) is the value correctly classified as False,
FN (False Negatives) is the value incorrectly classified as False.

The effectiveness of the machine learning model – considering a single training cycle – depends on the specific randomly selected test sample. To avoid the risk of a "lucky draw", it is recommended to train all machine learning models one hundred times, each time on a different training data set randomly selected from the same database (cross-validation mechanism is recommended). The average accuracy values for each algorithm are then calculated, and the one that proved to be the most effective is selected. This model can be used to predict voluntary employee turnover in the organization.

4. Results

Step 1. Developing the database on voluntary employee turnover

Utilizing the analysis of the literature, a catalog of twenty variables potentially significant in the emergence of turnover intentions was identified. The developed set is presented in Table 1.

Table 1.
Catalog of Variables Potentially Influencing Turnover Intentions

<table>
<tr><th>X_i</th><th>Variable</th><th>Notes</th></tr>
<tr><td>X_1</td><td>Age</td><td>Expressed as an integer</td></tr>
<tr><td>X_2</td><td>Education</td><td>{elementary, vocational, high school, higher education}</td></tr>
<tr><td>X_3</td><td></td><td>{female, male, other}</td></tr>
<tr><td>X_4</td><td>Job Satisfaction</td><td rowspan="15">Scale of 1-10 (higher value indicates stronger characteristic)</td></tr>
<tr><td>X_5</td><td>Salary Assessment</td></tr>
<tr><td>X_6</td><td>Perceived Fairness</td></tr>
<tr><td>X_7</td><td>Promotion and Personal Development</td></tr>
<tr><td>X_8</td><td>Work Performance</td></tr>
<tr><td>X_9</td><td>Working Conditions</td></tr>
<tr><td>X_{10}</td><td>Team Atmosphere</td></tr>
<tr><td>X_{11}</td><td>Recognition and Rewards</td></tr>
<tr><td>X_{12}</td><td>Quality of Relationship with Supervisors</td></tr>
<tr><td>X_{13}</td><td>Job Security</td></tr>
<tr><td>X_{14}</td><td>Communication in the Company</td></tr>
<tr><td>X_{15}</td><td>Work-Life Balance</td></tr>
<tr><td>X_{16}</td><td>Autonomy and Independence at Work</td></tr>
<tr><td>X_{17}</td><td>Level of Engagement</td></tr>
<tr><td>X_{18}</td><td>Possibility of Remote Work</td></tr>
<tr><td>X_{19}</td><td>Feeling of Fatigue and Burnout</td><td rowspan="2">Scale of 1-10 (higher value indicates lower characteristic)</td></tr>
<tr><td>X_{20}</td><td>Workload</td></tr>
</table>

Source: Own elaboration based on (Hom, 2020; Bolt et al., 2022).

As part of the conducted research, a survey questionnaire consisting of twenty items corresponding to variables potentially influencing the emergence of turnover intentions was developed. The survey was administered to 100 employees of a large IT company based in Poznań. The study was conducted from July to September 2023. In addition to the questions on the 20 specified criteria, respondents were asked to indicate the presence of turnover intentions (as a binary variable).

Step 2. Selecting significant variables through dimensionality reduction

At this stage, point-biserial correlation coefficients and V-Cramér correlation coefficients (for education and gender) were determined. Table 2 presents the variables most strongly correlated with the label.

Table 2.
Variables most strongly correlated with the label

X_i	X_6	X_{17}	X_5	X_{16}	X_8	X_{14}	X_1	X_{10}	X_{13}	X_7
r_{pb} /r_v	0.419	0.352	0.256	0.207	0.196	0.183	0.167	0.164	0.160	0.130

Source: Own elaboration.

In the target database for machine learning, only the variables from Table 3 will be included. It was found that the highest correlation with the label (indicating the intention to leave) is demonstrated by: sense of fairness (0.419), level of engagement (0.352), and salary evaluation (0.256). Therefore, in the case of the analyzed empirical database, these three variables are the strongest predictors of the intention to leave.

Step 3. Selection of machine learning models and their key hyperparameters

For predicting the intention of voluntary employee departures, the following machine learning algorithms were used in this article: decision tree algorithm, support vector machine algorithm (linear), k-nearest neighbors algorithm, and naive Bayes algorithm. For clarity in the implemented algorithms, the default hyperparameters contained in the Python scikit-learn library dedicated to machine learning were chosen each time.

Step 4. Splitting the database into training and test sets

The dataset was divided into two subsets, such that the training set contained 75 records (75% of the data), and the test set contained 25 records (25% of the data). To perform the split, the ready-made "train_test_split" function from the Python scikit-learn library dedicated to machine learning was used. The split was random.

Step 5. Training prediction models of voluntary departure intentions and selecting the most effective one

According to the procedure presented in the previous chapter, the data set was split into training and test sets one hundred times (step 4 of the method), and for each of the one hundred random splits, the training process was conducted for all analyzed algorithms (according to the cross-validation mechanism). Then, the average accuracy metrics for each algorithm were calculated. The training results of the individual algorithms are presented in Table 3.

Table 3.
Effectiveness results of selected machine learning mechanisms in predicting the intention to leave

Algorithm	**Decision Trees**	**Linear Support Vector Machine**	**k-Nearest Neighbors**	**Naive Bayes**
Average accuracy from one hundred training processes	72%	79%	82%	78%

Source: Own elaboration.

Thus, it turned out that the most effective machine learning algorithm in predicting the intention of voluntary employee departures in the studied organization is the k-nearest neighbors algorithm. Its average accuracy metric reached 82%, which should be assessed very positively. Histogram 1 presents the accuracy metric for the naive classifier (the most effective of the studied algorithms in the empirical database).

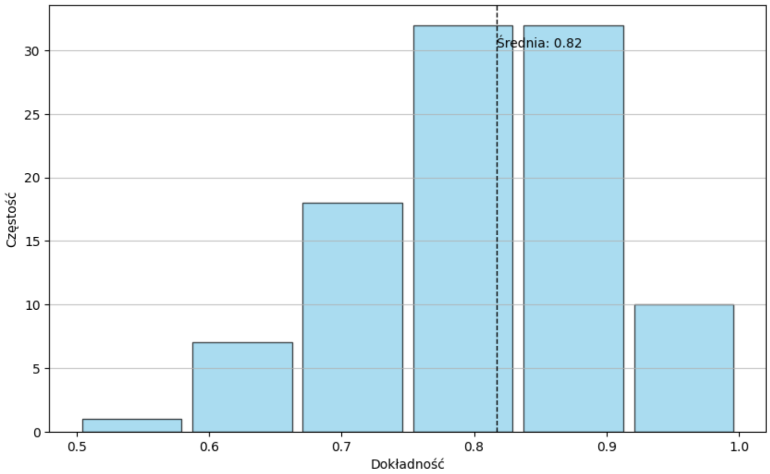

Figure 1. Accuracy metric for the naive classifier.

Source: Own elaboration.

In the empirical part of the article, research was conducted to evaluate the proposed method for predicting the intention of voluntary employee departures. The study was carried out in a large IT company based in Poznań, analyzing data collected from 100 employees using a questionnaire survey. Ten variables most strongly correlated with the intention to leave, such as sense of fairness, level of engagement, and salary evaluation, were identified. Based on these variables, four machine learning algorithms were tested: decision trees, linear support vector machine, k-nearest neighbors, and naive Bayes classifier. The best results were achieved by the k-nearest neighbors algorithm, with an average classification accuracy of 82%. The result of the conducted research is a predictive tool that enables the organization to identify talents exhibiting the intention of voluntary departure. This knowledge forms the basis for effective talent management in the company. Such an approach allows for a more personalized response to the needs and expectations of employees, and can also help improve overall job satisfaction and engagement within the organization.

5. Summary

The article presents a method for predicting the occurrence of voluntary employee departure intentions, utilizing machine learning algorithms. The realization of the article's objectives contributes to a better understanding and prediction of employee turnover dynamics initiated by the workforce. This is one of the fundamental areas of talent management. Practically, the presented method serves as a decision-support tool for management, creating an early warning system for the risk of voluntary employee departures. This allows for actions to be taken to retain talent within the organization. The result of the conducted research is a predictive tool enabling the organization to identify talents exhibiting the intention of voluntary departure. This knowledge forms the basis for effective talent management in the company. Such an approach allows for a more personalized response to the needs and expectations of employees, and can also help improve overall job satisfaction and engagement within the organization.

The developed method is universal and can be utilized in a range of enterprises. However, the limitation of the developed method is that it can only predict whether a given employee has the intention to leave. This method does not allow for determining whether the departure of a particular employee is beneficial for the organization or not. Another problem is that the developed method does not enable managers to select individualized tools to increase the probability of retaining talents within the organization. These limitations simultaneously outline future research areas, which may focus on integrating predictive tools with more individualized talent management solutions and analyzing the long-term benefits associated with retaining key employees.

Acknowledgements

The publication was financed from the funds for the statutory activity of the Faculty of Engineering Management at Poznan University of Technology, under the grant: “Application of Selected Machine Learning Methods in Management 4.0”, No. 0811/SBAD/1074, Project leader – Marcin Nowak, PhD.

References

1. Alsariera, Y.A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A.A., Ali, N.A. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*.
2. Aswale, N., Mukul, K. (2020). Role of data analytics in human resource management for prediction of attrition using job satisfaction. In: *Data Management, Analytics and Innovation* (pp. 57-67). Singapore: Springer.
3. Bills, M.A. (1925). Social status of the clerical work and his permanence on the job. *Journal of Applied Psychology, 9*, 424-427. http://dx.doi.org/10.1037/h0065881.
4. Bolt, E.E.T., Winterton, J., Cafferkey, K. (2022). A century of labour turnover research: A systematic literature review. *International Journal of Management Reviews, 24(4),* 555-576.
5. Cichosz, P. (2007). *Systemy uczące się*. WNT.
6. Diemer, H. (1917). Causes of "turnover" among college faculties. *The Annals of the American Academy of Political and Social Science*, *71*, 216-224.
7. Dolot, A. (2019). Przyczyny odchodzenia pracowników z organizacji. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie [Cracow Review of Economics and Management], 5(977),* 129-142.
8. Eberle, G.J. (1919). Labor turnover. *The American Economic Review, 9,* 79-82.
9. Holtom, B.C., Mitchell, T.R., Lee, T.W., Inderrieden, E.J. (2005). Shocks as causes of turnover: What they are and how organizations can manage them. *Human Resource Management, 44(3),* 337-352.
10. Hom, P.W., Allen, D.G., Griffeth, R.W. (2020). *Employee Retention and Turnover: Why Employees Stay or Leave.* New York, NY: Routledge.
11. Hom, P.W., Lee, T.W., Shaw, J.D., Hausknecht, J.P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102(3),* 530.
12. Ingram, T. (2011). *Zarządzanie talentami. Teoria dla praktyki zarządzania zasobami ludzkimi*. Warszawa: PWE.
13. Knap-Stefaniuk, A., Karna, W.J. (2017). Zarządzanie talentami jako wyzwanie w międzynarodowym zarządzaniu zasobami ludzkimi. *Perspektywy Kultury, 16(1),* 101-120.
14. Lee, T.W., Hom, P.W., Eberly, M.B., Junchao (Jason) Li, Mitchell, T.R. (2017). On the next decade of research in voluntary employee turnover. *Academy of Management Perspectives, 31(3),* 201-221.
15. Madigan, D.J., Kim, L.E. (2021). Towards an understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers' intentions to quit. *Teaching and Teacher Education, 105*, 103425.

16. March, J.G., Simon, H.A. (1993). *Organizations*. John Wiley & Sons.
17. Miś, A. (2009). Zarządzanie talentami w organizacji. *Zeszyty Naukowe, 810*. Uniwersytet Ekonomiczny w Krakowie, 39-54.
18. Mitchell, T.R., Holtom, B.C., Lee, T.W., Sablynski, C.J., Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal, 44*, 1102-1121.
19. Mobley, W.H., Griffeth, R.W., Hand, H.H., Meglino, B.M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin, 86*, 493-522. http://dx.doi.org/10.1037/0033-2909.86.3.493.
20. Moczydłowska, J., Kowalewski, K. (2014). *Nowe koncepcje zarządzania ludźmi*. Difin.
21. Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Le, H.V., Tran, V.Q., Prakash, I., Pham, B.T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 1-15.
22. Nyberg, A.J., Ployhart, R.E. (2013). Context-emergent turnover (CET) theory: A theory of collective turnover. *The Academy of Management Review, 38*, 109-131.
23. Price, J.L. (1977). *The study of turnover*. Ames, IA: Iowa State Press.
24. Price, J.L., Mueller, C.W. (1986). *Absenteeism and turnover of hospital employees*. Greenwich, CT: JAI Press.
25. Rebala, G., Ravi, A., Churiwala, S. (2019). *An introduction to machine learning*. Springer.
26. Steel, R.P., Ovalle, N.K. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. *Journal of Applied Psychology, 69*, 673-686.
27. Stuss, M.M. (2021). *Zarządzanie talentami: koncepcje, modele i praktyki*. Wydawnictwo UJ.
28. Tett, R.P., Meyer, J.P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analysis based on meta-analytic findings. *Personnel Psychology, 46*, 259-293.
29. Wójcik, P. (2017). Shortage of talents–a challenge for modern organizations. *International Journal of Synergy and Research, 6*.