South-Eastern Europe Journal of Economics 2 (2024) 129-146 SEEJE

TREE-BASED APPROACH TO PREDICT EMPLOYEE TURNOVER

ALEXANDRA-DANIELA DAMIR*

Alexandru Ioan Cuza University of Iasi, Romania

Abstract

Employee turnover is a challenge faced by organisations due to negative impact on employee productivity, morale and general performance. Understanding the factors that affect employee turnover contributes to developing effective employee retention strategies. The objective of this paper is identifying the best classification model for predicting employee turnover and detecting the key factors that affect employee churn. To achieve the research objective, we conducted a comparative analysis between decision trees and random forest algorithms. The results of the analysis show that workplace satisfaction, workload, review score, average number of working hours per month, and tenure are the main factors affecting employee turnover. Of the two algorithms employed, random forest exhibited superior performance across all evaluation metrics utilised. This study contributes to existing literature by providing empirical proof regarding factors that affect employee turnover and by comparing different machine learning algorithms. Findings highlight the value of machine learning techniques in understanding complexities of the workforce, in general, and in providing empirical evidence for building human resource strategies.

Keywords: machine learning; predictive analysis; employee turnover; classification; random forest; decision trees

JEL Classification: D20, C31, C52

The paper was among the five nominated papers for the ***Tsekouras Prize for Young Economists***, awarded by the *Association of Economic Universities of South and Eastern Europe and the Black Sea Region (ASECU)* which will be announced at the ASECU International Conference, Montenegro, July 2025

---

* ***Corresponding author:*** Alexandra-Daniela DAMIR, "Alexandru Ioan Cuza" University of Iași, B-dul Carol I, Nr. 20A, RO-700505, Iaşi Romania. E-mail: alexandra.damir@staff.uaic.ro

1. Introduction

Employee turnover is a significant challenge faced by organisations across various industries. Productive employees' decision to leave, driven by factors such as work pressure, unsuitable environments or unsatisfactory compensation, can have profound implications on productivity, morale, and overall organisational success. The departure of valuable employees not only reduces an organisation's productivity but also places additional strain on the human resource department to recruit, train, and integrate new hires. This transition period can lower the morale of remaining employees, who are often required to take on extra work, potentially having a cascading effect if more employees follow their departing colleagues in search of better opportunities.

Understanding the factors contributing to employee turnover and developing effective strategies to mitigate it have become important tasks for HR professionals and organisational leaders. Existing literature on employee turnover offers valuable insights into the various factors influencing employees' decisions to leave a company. However, despite extensive research, there is no consensus on specific factors driving employee turnover.

This study aims to deepen the understanding of factors influencing employee turnover and to develop a robust classification model to assist organisations in making informed decisions and implementing effective strategies to manage and reduce turnover. To achieve our aim, we established the following objectives:

1. Apply tree-based machine learning algorithms to classify employee turnover.
2. Identify the factors influencing employees' decisions to leave the company.
3. Determine which algorithm provides the most accurate and robust classification of employee turnover.

Predicting employee turnover before it occurs can help a company's management prevent or, at least, mitigate the impact of such departures. Recent advances in machine learning and data analysis have provided new opportunities to help stakeholders delve into the complexities of employee turnover. By leveraging the power of machine learning algorithms, organisations can analyse large datasets to unmask hidden patterns, identify factors, and predict employee turnover with greater accuracy.

To achieve the proposed goals and objectives, this study undertook several steps. It reviewed existing literature on employee turnover, examining various theoretical frameworks and empirical findings to establish a solid foundation for understanding the complexity and dynamics of employee turnover. An empirical analysis was conducted using tree-based machine learning algorithms known, from previous research, for their performance in predicting employee turnover. Decision trees and random forests were employed to analyse information on current and former employees using a dataset of a U.S. company. An initial exploratory data analysis and preprocessing step ensured data quality and consistency. Subsequently, various performance metrics were used to evaluate the classification capabilities of the models developed.

This research aims to contribute to the existing literature by conducting a comprehensive analysis of employee turnover and considering a wide range of factors that may influence employees' decisions to leave an organisation. By creating multiple models using different machine learning algorithms, the study seeks to capture a nuanced understanding of factors affecting turnover decisions.

In conclusion, this study highlights the factors impacting employee turnover and their relative importance, as identified through decision trees and random forests. Addressing these factors can help organisations foster a positive work environment, enhance employee engagement, and improve employee retention rates, ultimately contributing to an organisation's long-term success.

2. State of the Art

Employee turnover is a highly studied phenomenon due to its significant importance for organisations. Employees play an important role in the success of companies, and their replacement can be difficult and time-consuming (Kaur & Vijay, 2016). Turnover affects the stability and performance of organisations. Costs associated with employee turnover are estimated to range from the equivalent of an annual salary for each departing employee (Boroş & Curşeu, 2013) to millions of dollars in recruitment, training, and lost productivity (Perryer et al., 2010). The departure of employees also impacts the morale of those who remain, as they are often required to take on additional work until replacements are found, which can lead to feelings of being left behind in an organisation that has driven others to resign. The resignation of one employee can bring further losses as others may follow their former colleagues to seek new opportunities (Felps et al., 2009).

Despite extensive research, there is no single universal reason why employees choose to leave their organisations. Turnover represents the movement of employees within the labour market, between companies, jobs, and occupations, and between employment and unemployment (Abassi & Hollman, 2000). The employee turnover rate is defined as the ratio of the number of members who left the organisation during a given period to the average number of individuals in that organisation over the same period (Price, 1977). Managers often refer to employee turnover as the entire process associated with filling a vacant position: whenever a position is vacated, either voluntarily or involuntarily, a new employee must be hired and trained. This cycle of replacement is known as employee turnover (Kramer, 1995) and the term is frequently used to measure the rate at which employees leave an organisation, regardless of the reason.

2.1 Machine learning algorithms used

Alduayj and Rajpoot (2018) developed multiple machine learning models to predict employee turnover, including random forests, k-nearest neighbours, and support vector machines. They utilised different versions of the IBM HR dataset: the original unbalanced dataset and two synthetically balanced datasets (one over-sampled and one under-sampled). While these authors achieved high accuracy with the synthetically balanced datasets, the accuracy for the original dataset was low.

Najafi-Zangeneh et al. (2021) introduced a three-step framework for predicting employee turnover. The first step involved data cleaning using the "max-out" variable selection method. In the second step, they trained a logistic regression model for prediction. The third step involved performing confidence analysis to assess the prediction model usefulness. Despite these efforts, the model suffered from poor accuracy and high complexity due to extensive pre- and post-processing.

Pratt et al. (2021) employed classification trees and random forests for predicting turnover. Prior to classification, they pre-processed the data by removing unwanted variables using Pearson correlation. However, their model showed only slight improvement in accuracy compared to other machine learning algorithms.

Taylor et al. (2020) used decision tree-based models, including random forests and gradient-boosted trees, to predict employee turnover. These models demonstrated the highest performance. They used their own dataset containing 5,550 records.

2.2 Factors contributing to employee turnover

Understanding the factors influencing employee turnover is important for organisations aiming to enhance retention and maintain a stable, productive workforce. These factors can be broadly categorised into individual factors and work-related factors. These factors interact with each other in complex ways to influence resigning intentions. By comprehensively examining both individual and work-related factors, organisations can develop strategies targeted to address the root causes of employee turnover and to foster a more engaging and supportive work environment.

2.2.1 Individual factors

Individual factors refer to the set of employees' characteristics relevant to employee turnover and can be intrinsic, such as an individual's personality, or acquired, such as technical aptitude.

- Gender - in a meta-analysis (Park and Shaw, 2013) it was concluded that there is a similar turnover rate between men and women. Existing literature (Humpert and Pfeifer, 2013) contains empirical evidence that older women have a lower turnover rate when compared to men of the same age. Additionally, recent papers conclude that women have a higher turnover rate than men (Ono, 2023).

- Age - some researchers demonstrated that the intention of leaving an organisation is higher in younger individuals (Pitts et al., 2011). However, in most papers, the intent of leaving is negatively correlated to age (Carmeli and Weisberg, 2006; Ng and Feldman, 2009).
- Tenure - some research papers claim that age is studied in its interaction with tenure (Griffeth et al., 2000). Additionally, these authors claim that a shorter tenure leads to a higher turnover intention. It has been shown that there is job instability in the first years of employment (Singh and Schwab, 2000); Marital status - the number of children and responsibilities that come with being part of a family also have an impact on an employee's intention to leave the organisation in search of stability (Krau, 1981). This is also confirmed by a meta-analysis by Griffeth et al. (2000).

2.2.2 Work-related factors

There are several extrinsic factors affecting employees and impacting turnover rates.

- Stress - undoubtedly, stress is one of the most critical factors when it comes to employee turnover. Factors that in turn lead to stress at the workplace are ambiguity of one's role, employees' task overload and the conflict between worktime and time spent with the family (Trevor, 2001; Guimaraes, 1997).
- Satisfaction - individuals not satisfied with their current post will look for opportunities in other organisations (Carsten and Spector, 1987; Silla et al., 2009; Rode et al., 2007).
- Number of work hours - Higgins et al. (2000) showed that a shorter work schedule prevents the work-family conflict for women and leads to greater workplace satisfaction. D'Addio et al. (2007) concluded that for men, the opposite is true; in other words, men with full-time jobs have greater job satisfaction than those with part-time jobs.
- Workload - multiple research papers claim that there is a positive correlation between the workload, stress and turnover intention (Brannon et al., 2007).
- Promotion - there is very high correlation between promotion and job satisfaction, which in turn impacts employee retention rates (Pergamit and Veum, 1999; House et al., 1996).
- Salary - the salary variable has a modest impact on the turnover decision, according to Griffeth et al. (2000). These authors concluded that when high-performing employees are not rewarded appropriately, they leave the organisation.

2.3 Research Hypotheses

To confirm or reject the importance of the factors identified following the literature review, we have formulated two research hypotheses.

Although this factor has led to contradictory results, most studies (De Cuyper et al., 2009; Rode et al., 2007) support the notion that low job satisfaction is an important determinant in one's decision to resign and individuals with low job satisfaction are more likely to resign.

High Performers are often sought after by other companies and may receive more attractive and better-paid offers. Such employees may view these opportunities as career advancements and may decide to resign to take advantage of benefits offered elsewhere. Highly performing employees are constantly seeking opportunities for development and professional growth (Prince, 2005). If they feel their current job does not provide enough learning and advancement opportunities, they may seek alternative work environments offering more opportunities for development and skill enhancement and, therefore, are more likely to resign.

Lack of promotion opportunities for employees can lead to loss of valuable talent for the organisation (Eyster et al., 2008). When the organisation fails to provide career growth and development prospects, talented employees may be tempted to seek other opportunities in companies that offer such benefits. Employees who have not been promoted are more likely to leave a company.

Employees with a heavy workload may be overburdened and at risk of exhaustion and stress (Brannon et al., 2007). If they do not receive adequate support, such as additional resources or delegated responsibilities, they may reach a point when they feel they cannot cope with their workload and may decide to resign. In other words, employees with a heavy workload are more likely to resign.

Employees who are relatively new and have a short tenure in the company are not as connected and involved in the organisation's culture and values (Singh & Schwab, 2000). Without a strong connection, they may be less motivated to remain with a company in the long term and may be more open to exploring other job opportunities. This is why employees with a short tenure in the company are more likely to resign.

Although the salary level is no longer the primary decision-making factor when it comes to a job, it is still considered an important factor. There is research suggesting that one's salary level may play a role in an employee's decision to leave a job (Silbert, 2005). Employees with lower salaries may be more motivated to seek other opportunities that offer a competitive salary and are, therefore, more likely to resign.
Based on the above, we formulated the following hypothesis:

Hypothesis 1: Employees that are more likely to leave the company are those with low satisfaction, high performance, a heavy workload, no prospect of promotion, a short tenure, and a low salary level.

After analysing several studies (Usha & Balaji, 2019; Fallucchi et al., 2020) applying machine learning algorithms to predict employee turnover, we observed that decision trees were included in all these studies and showed good classification performance. Other researchers (Pratt et al., 2021; Taylor et al., 2020) have shown that random forest performs even better considering it is composed of multiple decision trees. Thus, we formulated the following hypothesis:

Hypothesis 2: A random forest algorithm provides the best performance in classifying resignation decisions.

These hypotheses represent preliminary statements based on the literature analysed, providing a framework for investigating the relationship between different factors and employee turnover, as well as determining the tree-based machine learning algorithm that will provide the best performance in classifying one's decision to leave a company.

3 Methodology

For achieving the purpose and objectives of this paper, two tree-based machine learning algorithms were used: two decision trees and a random forest one.

For determining the performance of the algorithms, multiple evaluation metrics were used. First, classification *accuracy* was used for measuring the overall correctness of the predictions made by the models. Additionally, *recall*, *specificity* and *precision* were also used as evaluation metrics to assess the algorithms' capability to distinguish between positive and negative instances. To account for the trade-off between sensitivity and specificity, the *ROC curve* and its associated index, *AUC*, were also used.

3.1 Decision trees

The decision tree algorithm is particularly useful for classification tasks because it can handle both categorical and numerical data and it is relatively easy to understand and interpret (Hastie et al., 2009). In addition, decision trees can be visualised, which helps understand the decision-making process.

3.2 Random forest

In the case of a random forest, when training classifiers some data may be used multiple times while others may never be used. Thus, greater stability of the classifier is achieved, as it becomes more robust to slight variations in input data, and, at the same time, it increases classification accuracy (Breiman, 2001). Several studies have shown that bagging-based methods such as RF, unlike other boosting-based methods, are not sensitive to noise or overfitting (Briem et al., 2002; Chan and Paelinckx, 2008; Pal and Mather, 2003).

4 Empirical analysis

In this chapter, we present the empirical analysis conducted using a dataset that includes detailed information on both current and former employees, to investigate the effectiveness of decision tree and random forest algorithms in predicting employee turnover.

4.1 Database structure

The database (Pujar, G., 2017) on which we will perform the analysis contains details about approximately 15,000 employees of a company in the United States of America. The information in this database was collected by the company's human resources department in order to find out the reasons why employees left the company.

The database contains 10 employee characteristics described as follows:

- satisfaction level: employee satisfaction score derived from surveys and ranging from 0 to 1;
- last evaluation: score the employee received in the last evaluation, which varies between 0 and 1;
- number of projects: the number of projects the employee is involved in;
- average monthly hours: the average number of hours the employee worked in a month;
- tenure: number of years the employee has been with the organisation;
- work accident: whether the employee had an accident at work or not;
- promoted: whether the employee was promoted or not in the last 5 years;
- salary: for privacy reasons, the salary is divided to three level groups: low, medium and high;
- departure (left): whether the employee left the company or not.

4.2 Exploratory data analysis

We began the analysis by doing a descriptive analysis of the database, as shown in Figure 1.

```
 satisfaction         review             projects        avg_hrs_month
Min.   :0.090000   Min.   :0.360000   Min.   :2.00000   Min.   : 96.00
1st Qu.:0.440000   1st Qu.:0.560000   1st Qu.:3.00000   1st Qu.:156.00
Median :0.640000   Median :0.720000   Median :4.00000   Median :200.00
Mean   :0.612834   Mean   :0.716102   Mean   :3.80305   Mean   :201.05
3rd Qu.:0.820000   3rd Qu.:0.870000   3rd Qu.:5.00000   3rd Qu.:245.00
Max.   :1.000000   Max.   :1.000000   Max.   :7.00000   Max.   :310.00

     tenure         accident   left      promoted        department       salary
Min.   : 2.00000   0:12830   0:11428   0:14680    sales       :4140   high  :1237
1st Qu.: 3.00000   1: 2169   1: 3571   1:   319   technical   :2720   low   :7316
Median : 3.00000                                  support     :2229   medium:6446
Mean   : 3.49823                                  IT          :1227
3rd Qu.: 4.00000                                  product_mng: 902
Max.   :10.00000                                  marketing  : 858
                                                  (Other)     :2923
```

Figure 1: Database Description

It is apparent that variables “accident”, “departure (left)”, “promotion (promoted)”, “department” and “salary” are categorical variables while the rest of them are numerical. There were no missing values in any of the variables in the dataset.

The descriptive analysis (Figure 1) indicates means and measures of dispersion for numerical variables, and frequency for categorical ones. For example, it can be observed that the average employee satisfaction score is 0.61, and the average evaluation score is 0.71. Furthermore, of the total number of employees only 319 were promoted. Although the satisfaction score ranges from 0 to 1, there are no employees with a satisfaction score of 0 and the lowest score recorded for this variable is 0.09. Similarly, the score obtained in the evaluation ranges between 0 and 1, but the lowest score obtained by employees is 0.36. On average, employees stay with the company for 3 and a half years but there are employees who had been with the company for 10 years.

4.3 Results

This section will show the outcomes achieved when implementing the two tree-based algorithms: two decisions trees and a random forest one.

Figure 2 summarizes the values for each of the performance metrics for all created models.

Algorithm	Accuracy	Sensitivity	Specificity	AUC
DT	0.972	0.9132	0.9904	0.972
DT pre-pruned	0.95	0.9197	0.9595	0.966
DT post-pruned	0.9569	0.9160	0.9697	0.964
RF	0.9904	0.9636	0.9988	0.994
Simplified RF	0.9896	0.9608	0.9985	0.994
NR	0.9727	0.9031	0.9942	0.962

Figure 2: Model performance

Three decision trees were built using different techniques: the first tree did not have any constrains, the second one was pre-pruned, and the third tree was post-pruned. The first decision tree, the most complex one, presented the best performance. The pre-pruned tree, designed for simplicity, showed the worst performance. Finally, the post-pruned tree achieved a good level of accuracy with low complexity.

The post-pruned decision tree, albeit of the lowest accuracy in comparison to Random Forest models, offers the advantage of being less complex and easier to interpret. Its simplicity and transparency make it a viable option for organisations that want a better understanding of how factors influence an employee's decision to resign.

The random forest algorithm was initially trained without specifying any parameters and 500 decision trees were obtained. Analysing the connection between the number of random forest trees and the error, it was concluded that the error did not significantly decrease above 100 trees. Consequently, we built another random forest with 100 trees, which achieved a slightly lower performance but close to the one achieved by the more complex random forest. Given its reduced complexity and good performance, the second random forest was chosen as the more appropriate model of the two.

In all evaluation metrics, the random forest algorithm had the best classification performance.

In the forthcoming subchapters, we will detail the implementation of the model selected for each algorithm.

4.3.1 Decision trees

Out of all the DTs built, the post-pruned version performed the best. To decide which complexity parameter to use, we plotted the relative error and the complexity parameter, as in Figure 3.

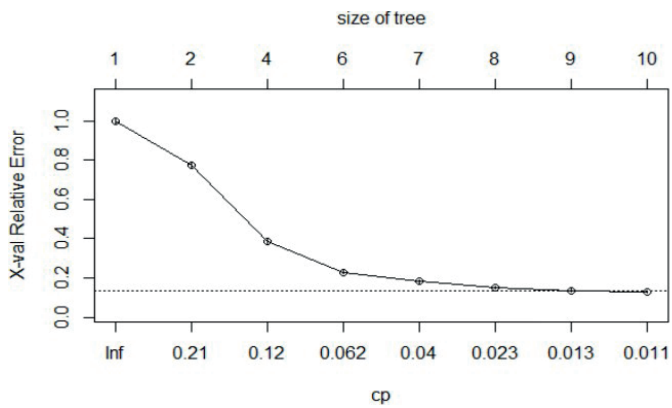

Figure 3: Relative error based on complexity parameter

We can see in this figure that the relative error does not significantly decrease once the 0.04 threshold is crossed. Consequently, the 0.04 complexity parameter is chosen. The resulting DT is represented in Figure 4.

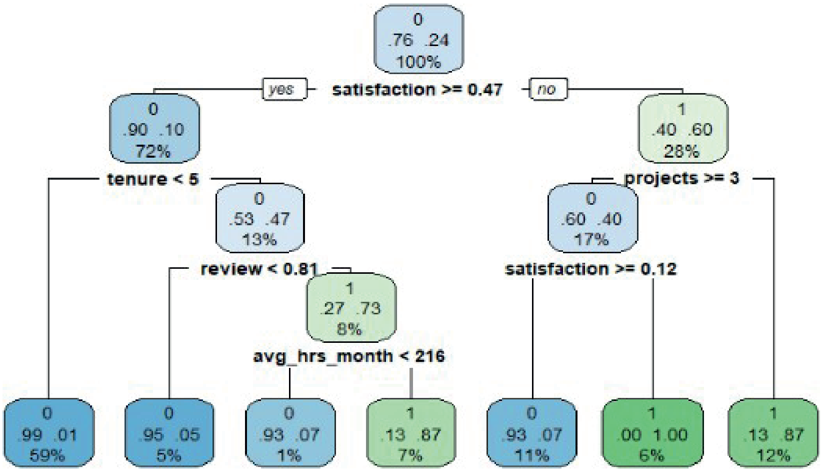

Figure 4: Decision tree with optimal complexity parameter

The decision tree obtained has 13 nodes, of which 7 are leaves, and 4 levels.

A detailed view of the classification matrix and evaluation metrics of this tree are presented in Figure 5.

```
          Reference
Prediction    0    1
         0 3324   90
         1  104  981

               Accuracy : 0.9569
            Sensitivity : 0.9160
            Specificity : 0.9697
         Pos Pred Value : 0.9041
         Neg Pred Value : 0.9736
      Balanced Accuracy : 0.9428
```

Figure 5: Classification matrix and evaluation metrics for the optimal DT

To show how well the decision tree differentiates the two classes, the AUC index and the ROC curve are analysed as follows:

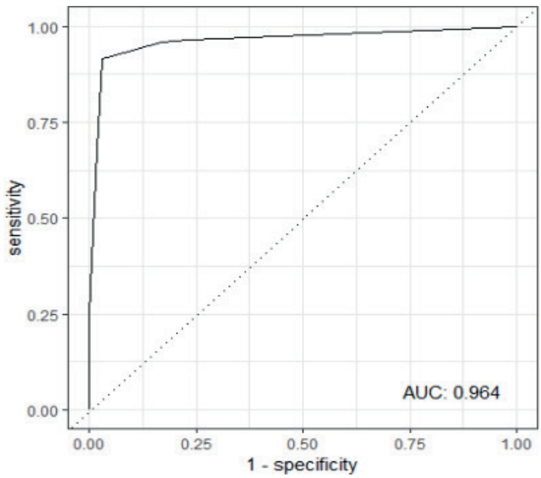

Figure 6: ROC curve and AUC index for the optimal DT

As per Figure 6, the value of the AUC index is lower compared to the pre-pruned DT. This is most likely due to the lower sensitivity of the pre-pruned DT. The value of the AUC index of 0.964 is quite close to the value of 1, which means that the decision tree manages to differentiate well between the two categories. The fact that the ROC curve is close to the upper left corner indicates the same thing.

4.3.2 Random forest

Initially we created a random forest without specifying the parameters that resulted in a random forest of 500 trees. To see whether it was possible to reduce the complexity of the model, we created a graph (Figure 7) with the error and the number of trees from the random forest.

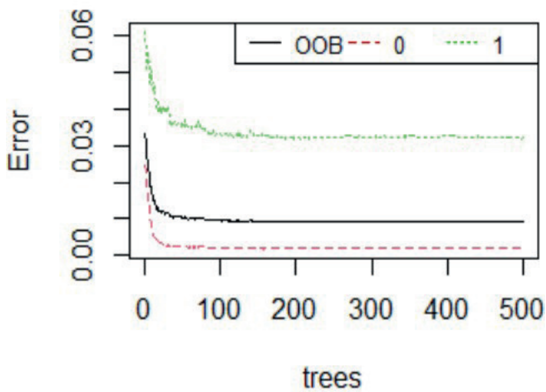

Figure 7: Random Forest error level relative to the number of its trees

In Figure 7 we present the incorrect classifications recorded on the data that were not used when creating trees. In addition, these incorrect classifications were divided by classes, which indicates that the overall classification error, as well as the classification error by classes, do not significantly decrease once approximately 100 decision trees have been created. In other words, it is redundant to create 400 more trees if they don't help improve performance. Thus, a new random forest with 100 decision trees was created to verify whether a similar performance could be obtained with less complexity.

```
Call:
 randomForest(formula = left ~ ., data = train.data, importance = T,
              ntree = 100)
              Type of random forest: classification
                    Number of trees: 100
No. of variables tried at each split: 3

        OOB estimate of  error rate: 0.95%
Confusion matrix:
     0    1 class.error
0 7983   17    0.002125
1   83 2417    0.033200
```

Figure 8: Training results for a random forest with 100 trees

Figure 8 shows that the misclassification rate on OOB data is 0.95%. The rate of the class with employees who did not resign is 0.21%, and the rate of the class of employees who resigned is 3.3%. Overall, decreased complexity did not have a significant effect on performance, at least in the case of OOB data. Performance on the testing dataset is presented below.

```
          Reference
Prediction    0    1
         0 3423   42
         1    5 1029

            Accuracy : 0.9896
         Sensitivity : 0.9608
         Specificity : 0.9985
      Pos Pred Value : 0.9952
      Neg Pred Value : 0.9879
   Balanced Accuracy : 0.9797
```

Figure 9: Random forest classification matrix and evaluation metrics

The 99.9% accuracy obtained from the training set is very close to that obtained from the testing set, which means that the model maintains its performance on new data (Figure 9). Accuracy on the testing dataset dropped very slightly from 99% to 98.96% as compared to the first random forest model built. The same applies to sensitivity and specificity, which are very close to those of the previous model. Reducing the number of trees from 500 to 100 did not have much of an effect on performance.

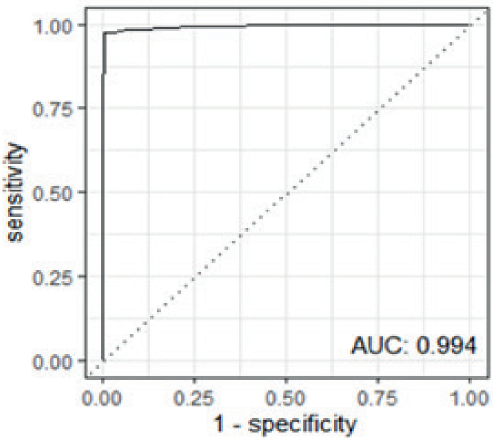

Figure 10: ROC curve and AUC index for the simplified Random Forest

The fact that the ROC curve is so close to the top left corner and the AUC index of 0.994 is very close to 1 indicate that the random forest model is highly effective at distinguishing between employees who left the company and employees who didn't.

4.4 Validation of Research Hypotheses

In all evaluation metrics, the random forest algorithm demonstrated the best classification performance. Thus, hypothesis 2, which states that the random forest algorithm provides the best performance in classifying employee turnover, is validated. The decision tree, although having weak accuracy compared to Random Forest, offers the advantage of being less complex and easier to interpret. Its simplicity and transparency make it a viable option for organisations seeking better understanding of how factors influence employee turnover.

Based on the results provided by the decision tree, we could identify the relationships between factors and one's decision to resign. There is an inverse relationship between the independent variable 'satisfaction' and the dependent variable. Meanwhile, there is a direct relationship between the independent variables 'average number of hours worked per month', 'score obtained in the last evaluation', 'tenure', and the dependent variable. Categorical variables are excluded from the model indicating that they are not significant in predicting employee turnover. Factors such as promotion and salary do not impact the model's predictions. Thus, hypothesis 1, which suggests that employees predisposed to resign are those with low satisfaction, high performance, a heavy workload, no prospects of promotion, a short tenure, and a low salary level, is partially validated. The hypothesis states that employees who were not promoted and had low salaries were more likely to leave the company. However, in our decision tree model, these factors did not demonstrate significant importance.

Overall, the hypotheses developed at the beginning of the research are partially validated. The random forest model stands out for its superior performance, while decision tree haves presented different strengths and weaknesses. The choice between decision trees and random forests depends on the organisation's preference for model complexity and interpretability, as well as its performance.

5 Conclusion

A high employee turnover rate is an important problem for companies. Losing highly performing employees is considered a major loss. Finding replacements of a similar level of performance is difficult and costs the company both money and time. The main objective of this research was to identify factors that influence staff turnover and develop a robust classification model using tree-based machine learning algorithms to predict employee turnover based on employees' characteristics.

Employees who experience low job satisfaction, high performance, heavy workloads, and short tenures are more likely to leave their current company. Low job satisfaction can lead to disengagement and lack of motivation, making employees more inclined to seek opportunities in places where they will feel more valued and content. Highly performing employees, if not adequately recognised or rewarded, may feel unappreciated and overburdened, which prompts them to look for environments that offer better acknowledgment of their contribution to the firm. Heavy workloads can lead to burnout and stress, pushing employees to search for positions with more manageable demands. Lastly, employees with a short tenure might feel less attached to the company and more willing to leave if they perceive better opportunities elsewhere.

Regarding the empirical analysis performed via two tree-based machine learning algorithms, results showed that a random forest model offers the best performance in the classification of one's resignation decision, due to clearly superior results in different evaluation metrics. The decision tree presented a slightly lower accuracy compared to that of a random forest one; however, the former offers the advantage of simplicity and ease of interpretation. This trade-off between performance and interpretability should be considered when selecting the most suitable model for practical implementation. The simplicity of the model can be particularly beneficial for organisations looking for an easy-to-understand representation of their decision-making process.

Overall, this paper provided insights into the factors influencing one's resignation decision and developed predictive models using decision trees, a random forest and neural networks. Research findings lay the foundation for future deeper research into the dynamics of staff turnover, for exploring new data sources, for developing strategies to mitigate staff turnover and for promoting a positive work environment.

References

Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. Public Personnel Management, 29(3), 333–342.

Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. In 2018 International Conference on Innovations in Information Technology (IIT) (pp. 93–98). IEEE.

Aquino, K., Freeman, D., Reed II, A., Lim, V. K. G., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. Journal of Personality and Social Psychology, 97(1), 123.

Booth, S., & Hamer, K. (2007). Labour turnover in the retail industry: Predicting the role of individual, organisational and environmental factors. International Journal of Retail & Distribution Management, 35(4), 289–307.

Boroș, S., & Curșeu, P. L. (2013). Is it here where I belong? An integrative model of turnover intentions. Journal of Applied Social Psychology, 43(8), 1553–1562.

Brannon, D., Barry, T., Kemper, P., Schreiner, A., & Vasey, J. (2007). Job perceptions and intent to leave among direct care workers: Evidence from the better jobs better care demonstrations. The Gerontologist, 47(6), 820–829.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Briem, G. J., Benediktsson, J. A., & Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 40(10), 2291–2299.

Carmeli, A., & Weisberg, J. (2006). Exploring turnover intentions among three professional groups of employees. Human Resource Development International, 9(2), 191–206.

Carsten, J. M., & Spector, P. E. (1987). Unemployment, job satisfaction, and employee turnover: A meta-analytic test of the Muchinsky model. Journal of Applied Psychology, 72(3), 374.

Chan, J. C.-W., & Paelinckx, D. (2008). Evaluation of random forest and AdaBoost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112(6), 2999–3011.

D'Addio, A. C., Eriksson, T., & Frijters, P. (2007). An analysis of the determinants of job satisfaction when individuals' baseline satisfaction levels may differ. Applied Economics, 39(19), 2413–2423.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. Journal of Management, 26(3), 463–488.

Grossi, E., & Buscema, M. (2007). Introduction to artificial neural networks. European Journal of Gastroenterology & Hepatology, 19(12), 1046–1054.

Guimaraes, T. (1997). Assessing employee turnover intentions before/after TQM. International Journal of Quality & Reliability Management, 14(1), 46–63.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (Vol. 2). Springer.

Heneman, H. G., Judge, T. A., & Kammeyer-Mueller, J. D. (2019). Staffing organizations. McGraw-Hill.

Higgins, C., Duxbury, L., & Johnson, K. L. (2000). Part-time work for women: Does it really help balance work and family? Human Resource Management, 39(1), 17–32.

House, R. J., Shane, S. A., & Herold, D. M. (1996). Rumors of the death of dispositional research are vastly exaggerated. Academy of Management Review, 21(1), 203–224.

Humpert, S., & Pfeifer, C. (2013). Explaining age and gender differences in employment rates: a labor supply-side perspective. Journal for Labour Market Research, 46(1), 1-17.

Krau, E. (1981). Turnover analysis and prediction from a career developmental point of view. Personnel Psychology, 34(4), 771–790.

Najafi-Zangeneh, S., Shams-Gharneh, N., Arjomandi-Nezhad, A., & Zolfani, S. H. (2021). An improved machine learning-based employees attrition prediction framework with emphasis on feature selection. Mathematics, 9(11), 1226.

Ng, T. W. H., & Feldman, D. C. (2009). How broadly does education contribute to job performance? Personnel Psychology, 62(1), 89–134.

Ono, H. (2023). Career urgency and turnover intention among young adult workers: a comparison by gender and employment status. BMC psychology, 11(1), 389.

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment, 86(4), 554–565.

Park, T. Y., & Shaw, J. D. (2013). Turnover rates and organizational performance: a meta-analysis. Journal of applied psychology, 98(2), 268.

Pergamit, M. R., & Veum, J. R. (1999). What is a promotion? ILR Review, 52(4), 581–601.

Perryer, C., Jordan, C., Firns, I., & Travaglione, A. (2010). Predicting turnover intentions: The interactive effects of organizational commitment and perceived organizational support. Management Research Review, 33(9), 911–923.

Pitts, D., Marvel, J., & Fernandez, S. (2011). So hard to say goodbye? Turnover intention among US federal employees. Public Administration Review, 71(5), 751–760.

Pratt, M., Boudhane, M., & Cakula, S. (2021). Employee attrition estimation using random forest algorithm. Baltic Journal of Modern Computing, 9(1), 49–66.

Price, J. L. (1977). The study of turnover (1st ed.). Iowa State University Press.

Pujar, G. (2017). HR Analytics. Kaggle. Retrieved February 2024, from https://www.kaggle.com/datasets/giripujar/hr-analytics

Rode, J. C., Rehg, M. T., Near, J. P., & Underhill, J. R. (2007). The effect of work/family conflict on intention to quit: The mediating roles of job and life satisfaction. Applied Research in Quality of Life, 2, 65–82.

Silla, I., De Cuyper, N., Gracia, F. J., Peiró, J. M., & De Witte, H. (2009). Job insecurity and well-being: Moderation by employability. Journal of Happiness Studies, 10, 739–751.

Singh, D. A., & Schwab, R. C. (2000). Predicting turnover and retention in nursing home administrators: Management and policy implications. The Gerontologist, 40(3), 310–319.

Taylor, S., El-Rayes, N., & Smith, M. (2020). An explicative and predictive study of employee attrition using tree-based models.

Trevor, C. O. (2001). Interactions among actual ease-of-movement determinants and job satisfaction in the prediction of voluntary turnover. Academy of Management Journal, 44(4), 621–638.