Chapter 10

Examination of Supervised Machine Learning Algorithms in Employee Turnover Prediction

Melisa Dikici[1]

Gökçe Sabriye Hörük[2]

Deniz Efendioğlu[3]

Abstract

This study classifies employee layoffs into two types: voluntary and involuntary. The financial effects of voluntary turnover are highlighted in organizations. The ability to predict redundancies is a very important aspect of employee retention strategies for any organization. Supervised machine learning algorithms, which are more accurate, were analyzed in the context of employee turnover prediction. The literature review identified some of the key parameters that influence turnover, which include working conditions, job satisfaction, management support, pay fairness, and career opportunities. The algorithms to be studied will include Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Naive Bayes. Random Forest showed the best accuracy; hence, it is recommended for complicated datasets. Logistic Regression, though less accurate, is simple and interpretable, hence useful for strategic decision-making. This study henceforth highlights that the choice of algorithms should be fully aligned with data structure and organizational priorities for better human resource management and reduction in turnover ratio.

1 Ankara Yıldırım Beyazıt University, Faculty of Engineering & Natural Science, Industrial Engineering Department, 06010, Ankara,Turkey

2 Ankara Yıldırım Beyazıt University, Faculty of Engineering & Natural Science, Industrial Engineering Department, 06010, Ankara,Turkey

3 Ankara Yıldırım Beyazıt University, Faculty of Engineering & Natural Science, Industrial Engineering Department, 06010, Ankara,Turkey

https://doi.org/10.58830/ozgur.pub865.c3508

INTRODUCTION

The turnover of employees happens in two ways: through voluntary and involuntary turnover. Voluntary turnover takes place when the decision to leave the job comes from the employee himself, that is, resignation. Involuntary turnover takes place when the employer decides on termination of the employment (Hong et al., 2007).

The factors that constitute voluntary turnover are employees looking for better or more lucrative positions for their own career advancement. Employees dissatisfied with their job or unable to get along with managers or colleagues, and/or in stressful working conditions, lead the decisions to leave voluntarily. Furthermore, personal reasons such as relocating, health issues, and responsibilities towards family may drive a person to leave employment. Long working hours, or at times inability or difficulty in maintaining the so-called 'work-life' balance, may also be some reasons for voluntary quits (Russell et al., 2013).

On the other hand, the causes of involuntary turnover are mainly associated with the decisions of employers. Factors such as failure to meet expectations of performance or other disciplinary actions commonly lead to an employer's decision for employee termination. Furthermore, organizational changes like downsizing, mergers, and cost-cutting policies are some of the major reasons for involuntary turnover. Economic fluctuations or the development of technologies that make certain jobs redundant also fall into this category of turnover (Leana et al., 1987).

It might lead to unforeseen consequences for companies, especially when skilled employees leave the company. It could result in high costs because of workforce disruptions, recruitment efforts, and training processes. Since involuntary turnover is initiated by employers themselves, predictions are not required. However, in the case of voluntary turnover, which occurs at the discretion of employees, companies try to estimate which of their employees are likely to quit. These estimates are important for workforce planning and in devising strategies for better employee engagement and retention (Setiawan et al., 2020).

Some of the machine learning algorithms are used for predicting employee turnover. Whether unsupervised or supervised, machine learning techniques are increasingly being used to predict turnover and help inform retention strategies. Which of these algorithms is used will depend upon the size and nature of the data set.

Supervised learning methods have been widely used for the prediction of employee turnover, either because they can handle structured data or due to their better performance in terms of predictive accuracy. Decision trees, random forests, gradient boosting, logistic regression, support vector machines, neural networks, and many other techniques have been greatly tested for this purpose (Zhao et al., 2018) (Punnoose et al., 2016) (Jhaver et al., 2019). Some of the benefits are that several supervised methods have demonstrated high accuracy in predicting turnover even with noisy HR data, hence finding them quite effective for practical applications (Punnoose et al., 2016) (Jhaver et al., 2019). These methods have also been tested across different organizational sizes and complexities, hence providing guidelines for their effective use under diverse HR scenarios (Zhao et al., 2018).

Unsupervised methods, like clustering, are not applied very often to turnover predictions but may provide specific knowledge regarding patterns and groupings of data on employees. An example could be K-Means clustering combined with PCA: this was also tried without giving results that were notably good compared to supervised methods (Birzniece et al., 2022). The unsupervised nature of K-Means has shown limited success regarding the accuracy of the forecast, and it usually needs complementary techniques to enhance their utility (Birzniece et al., 2022).

The following part is about the supervised machine learning algorithms and the role of each of the supervised machine learning algorithms in predicting employee turnover. The article delves into the factors influencing employees' decisions to leave their jobs, offering detailed explanations of these factors. A comprehensive literature review was conducted, analyzing the machine learning methods employed in prior studies. The article provides in-depth insights and analyses of these techniques. The findings aim to guide organizations by highlighting the advantages of supervised machine learning techniques applicable to predicting employee turnover and their impact on the accuracy of such predictions.

In this prediction, some employee turnover parameters and reasons were considered. Many reasons can be present to cause employees to leave their jobs. Some of them are as follows:

Working conditions significantly impact employees' well-being, including their physical and psychological environment. Factors like office ergonomics, safety, and hygiene deficiencies can lead to dissatisfaction and eventual turnover (Salleh, Nair, & Harun, 2012). Job satisfaction, or the lack thereof, is another critical factor; dissatisfaction in one's role is often a precursor to an employee leaving the organization (Arshad & Puteh, 2015). Inadequate training, which hinders employees from developing necessary skills for their roles, further exacerbates this issue (Nketsiah & Nkansah, 2024). A lack of management support negatively impacts employee motivation and engagement, reducing their commitment to the organization (Ribes, Touahri, & Perthame, 2017). Pay fairness, or the perception of receiving lower-than-market-standard salaries, also plays a crucial role in turnover decisions (Effendi, 2024). Work-life imbalance, characterized by excessive workload or inflexible hours, often leads to burnout and dissatisfaction among employees (Ongori, 2007). Similarly, inequitable workload distribution can exacerbate burnout and stress, pushing employees toward resignation (Zhang, Cao, Qu, & Wang, 2024). Organizational commitment is another significant determinant; when employees lose alignment with the organization's values and goals, turnover becomes more likely (Nketsiah & Nkansah, 2024). A lack of career advancement opportunities within the organization further fuels the intention to leave, especially when external opportunities offer better prospects (Hong, Wei, & Chen, 2007). Leadership style mismatches and workplace conflicts, including poor team dynamics, further contribute to dissatisfaction and disengagement (Brown et al., 2017). The infectious effect, wherein employees are influenced by others leaving, can also create a ripple effect, especially in closely-knit teams (Teng et al., 2019). Technological changes that overwhelm employees or the lack of personal development opportunities can make them feel unsupported, further affecting their retention (Yin, Hu, & Chen, 2024). Job security concerns and feelings of social isolation at the workplace are also critical contributors to employee dissatisfaction and eventual turnover (Recilla et al., 2024). Anomalies in project management and organizational processes, along with mismatched business culture, can cause employees to feel disconnected from the organization (Ajit, 2016). Finally, perceptions of injustice regarding organizational decisions and practices can undermine employee trust and commitment, prompting them to seek opportunities elsewhere (Nketsiah & Nkansah, 2024).

MATERIAL AND METHODS

Machine Learning-Supervised Learning

1. Logistic Regression

Since logistic regression can provide firms with an understanding of the factors that affect employee retention, it is a very crucial research area in predicting employee turnover. Logistic regression finds its fit in modeling binary outcomes-whether an employee will stay or leave. Some key elements of this predictive strategy:

Salary: A business should differentiate itself from the competitors with regards to pay. According to studies, pay plays a significant role in influencing the dismissal rate, with those workers who earn better income showing a lower propensity to leave their jobs. (Effendi, 2024). Promotion: career development is made visible to employees of various business units who have opportunities for promotion; this displays different tendencies from the point of dismissal. (Effendi, 2024)(Chen, 2023).

Demographic Factors: These factors have a bearing on commitment and job satisfaction; therefore, age, marital status, and length of service are the key predictors of turnover. (Chen, 2023). Logistic regression was proven to be very powerful for predicting turnover and reaching F1-scores and accuracies that were very high in several works. (Taner et al., 2024).Contrary, it often falls back behind the top models on accuracy, namely Random Forest and Support Vector Machines. (Zhang et al., 2024) (Liao, 2023). Despite the advantages of logistic regression, there are some of its disadvantages that should be considered, such as possible class imbalance in turnover data, which may affect the performance of the model. Future studies should investigate hybrid models incorporating logistic regression with other machine learning methods to improve predicted accuracy. Logistic regression, though simple to predict turnover, usually performs worse in comparison to other more sophisticated machine learning models. Accordingly, the studies indicate that such models as Random Forests and Support Vector Machines result in a higher degree of accuracy in predictions of turnover. Nevertheless, logistic regression is still applicable because it is easy to work with and can clearly show which factors are driving certain phenomena. (Effendi, 2024) (Chen, 2023) (Krishna et al., 2023) In the end, logistic regression is helpful in predicting employee turnover, whereby one gets to understand those factors that influence an employee to leave. While it may not always match the predictive power of more complex models, the simplicity and interpretability make it a very valuable tool for HR analytics.

2. **Decision Trees**

Decision trees, combined with a variety of algorithms and datasets, have produced some promising results in predicting employee turnover. Certain studies were able to provide evidence that decision tree models, like the RandomTree algorithm, have reached high accuracy in the classification of turnover intention, which includes job satisfaction and organizational commitment as important predictors. (Živković et al., 2024) Decision trees are also useful in classifying employees into different categories of attrition, using demographic data and employment history, hence enabling predictions about the future cases of turnover. (Ahmed et al., 2023)

Various studies have replicated the identification of critical factors contributing to employee turnover using the Decision Tree analysis. Some commonly identified predictors are job satisfaction, opportunities for personal growth, affective organizational commitment, salary satisfaction, and interpersonal relationships. . (Živković et al., 2024) Other important factors were monthly income, overtime, age, distance from home, and years in the company. (Gao et al., 2019) Such information has helped the organization understand the root cause for turnover and come up with effective retention policies. It has been contrasted with other well-known machine learning ensembles, including Support Vector Machines, Random Forests, and Gradient Boosting Trees. Despite being straightforward and easy to understand, the model that a decision tree produces performs somewhat worse than ensemble approaches that use gradient-based decision trees, which are more effective with structured and unbalanced data and offer scalable solutions for strategic business analytics (Zhang et al., 2024). Nevertheless, because of their ease of use and ability to highlight the primary causes of turnover, decision trees continue to be very valuable. (Kanuto, 2024)

Decision Trees have practical applications in predicting employee turnover. In this regard, identification of at-risk employee groups will enable organizations to take necessary steps to reduce the rate of turnover. Decision Trees also allow companies to make immediate assessments of turnover risk using existing HR data and provide valuable insights for decision-making. Also, the integration of Decision Trees with other techniques, such as K-means clustering, enhances the prediction and analysis of turnover, hence bringing comprehensive solutions for workforce management. Decision Trees are basically one of the powerhouse tools for employee turnover predictions that provide very valuable insight into the vital factors while allowing organizations to formulate an effective retention policy. The application of the same in human resource management results in informed decisions and ensures a stable workforce. (Yunmeng et al., 2019)

3. Random Forests

The findings demonstrate that Random Forests have so far outperformed other machine learning algorithms in forecasting employee turnover. Several studies have demonstrated that Random Forest models achieve high accuracy rates, typically outperforming alternative techniques like logistic regression, decision trees, and support vector machines (Gao et al.,

2019) (Liao, 2023) (Zhang et al., 2023) (Nayak et al.,2023). According to one such study, Random Forests have a prediction accuracy of 98.8%, making them more effective than logistic regression and KNN (Zhang et al., 2023). A number of these research determined the primary factors that have a significant impact on employee turnover. These often include years of employment, age, distance from home, monthly salary, and overtime (Gao et al., 2019) (Atef et al.,2022). Also, work satisfaction has been identified as a significant predictor of turnover, and Random Forest models support this finding(Chang et al., 2022). These factors are essential for creating an accurate prediction model and could offer guidance on HR tactics to lower attrition.

The literature has proposed innovations in algorithms for the enhancement of the predictive capabilities of Random Forests. An example includes an improved weighted quadratic Random Forest algorithm to handle high-dimensional and unbalanced datasets, which proves to remarkably improve recall and F-measure compared to traditional Random Forest and other algorithms (Gao et al., 2019). Such efforts are important in refining prediction models and improving their applicability in real-world scenarios. Therefore, due to the intricacies and complexities of human behaviors along with dynamic work-place ecologies, employee churn is usually a hard task despite all the success of Random Forests. It is expected that further integration of data sources from more diversified platforms along with more hybrid models that would use ensembles of Random Forests will be explored for making accurate and reliable predictions in further research. (Zhao et al., 2018) (Islam et al., 2018) Moreover, continuous feature engineering for feature selection and embedding/dimensionality reduction contributes a lot to fine-tuning these models (Islam et al., 2018). Random Forests are a strong competitor in employee turnover prediction and achieved high accuracy with complexities of datasets. Innovation and research on this part of the area will promote the effectiveness of these works more and provide many valuable insights into making decisions for organizations targeting reducing turnover and improving employees' retention.

4. Support Vector Machine (SVM)

It is one of the preferred machine learning methods in regression analysis as well as classification problems. Its working principle is basically to divide a certain data set into groups or classes. Thus, when a new data is to be classified, it is easily determined which data set or group it belongs to. If the boundary line is drawn non-linearly rather than straight when determining the groups of the data set, this method is expressed as a non-linear support vector machine (SVM). SVM contributes to the prediction of employee turnover in turnover estimation. In addition, it helps determine strategies. It has been observed that the results obtained from balanced data sets are highly consistent in turnover estimation (Kumar et al., 2023). It provides great support to the decision maker in the decision process thanks to its special functions to model complex problems in linear and non-linear data sets. Since its grouping feature is strong, it has been observed that it is successful in dividing the turnover estimation of employees into groups (Kumar et al., 2023, Teng et al., 2019).

5. **K-Nearest Neighbors (KNN)**

A machine learning algorithm used for classification and regression analysis. KNN makes predictions based on input data and the associated data to reach a result. It is preferred in turnover prediction to model the complex decisions regarding employee resignation. The steps of the KNN algorithm are as follows: In the first step, the value of K (the number of neighbors) is determined. Then, the Euclidean distance is calculated. (Euclidean distance is a method for measuring the distance between two data points.) (Kabak et al., 2016). In the third step, the K nearest neighbors are identified. For this, the positions of the elements in the dataset are checked and the nearest ones are identified. Then, the number of points in each group is determined, and the new point is assigned to the group with the higher count, completing the algorithm's process (Kumar et al., 2023).

6. **Naive Bayes**

One of the machine learning methods, Naive Bayes is used in problems with large datasets and for solving classification problems. Although it is frequently used in many fields, it is observed that there is relatively less research in the literature on predicting employee turnover (Valle and Ruz, 2015). Naive Bayes helps in calculating class probabilities using Bayes' theorem. Bayes' theorem is stated as follows.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

LITERATURE REVIEW

Teng et al. (2019) proposed the Contagious Effect Heterogeneous Neural Network (CEHNN), which is based on the idea that an employee's departure could influence their colleagues to exhibit similar turnover tendencies. They created a random training and test set to evaluate the model's performance.

Their approach highlights the importance of considering social contagion effects in turnover prediction models.

Hong et al. (2007) discussed two types of turnover: voluntary and involuntary. They emphasized that predicting turnover in advance could benefit organizations by enabling preventive measures to retain employees. High turnover rates pose financial disadvantages for businesses. The study attempted to predict turnover using logit and probit models, which aimed to determine the likelihood of employees leaving an organization based on historical data. The success rate for both models reached up to 95.2%. Hong's work compared these two models for predicting employee turnover, examining their accuracy, effectiveness, and practical applications. Accurate turnover predictions can support organizations in improving strategic planning and human resource management.

Yin et al. (2024) emphasized the importance of minimizing employee turnover in the finance sector, as customer relationships are heavily dependent on employees. The study utilized features such as employee satisfaction, performance, tenure, last evaluation scores, the number of projects participated in, average monthly working hours, years spent at the company, work accident history, whether the employee left the company, promotion history within the last five years, department, and salary level (low, medium, high) to analyze turnover.

CatBoost was found to be more effective for categorical data, while XGBoost was faster and more efficient. The research focused on the differences between these two algorithms. Regression algorithms proved effective for imbalanced datasets. When using CatBoost, long processing times for categorical data were unnecessary. It provided quick predictions and avoided overfitting. XGBoost, on the other hand, utilized regularization to prevent overfitting and worked well with datasets containing missing values, offering fast predictions.

Recilla et al. (2024) emphasize the importance of maintaining low employee turnover rates regardless of the industry. Frequent job changes among employees pose an ongoing challenge for companies, making it crucial to control turnover to maintain a competitive edge. Accurate predictions, therefore, become essential in addressing this issue.

The study analyzed 30 variables using data from 2,035 respondents and employed the Genetic Algorithm approach. Four operators—FMSAX, CAX, IBAX, and AX—were applied to compare optimization processes. The results revealed that the FMSAX operator consistently achieved the highest and most reliable performance. For instance, the "Co-workers" variable was optimized from an initial fitness value of 33,489 to 24,360.97. In contrast, IBAX and AX showed less consistent outcomes, while CAX demonstrated moderate performance. Overall, FMSAX emerged as the most effective operator.

Zhang et al. (2024) conducted statistical analyses to explore the relationship between employee turnover and various indicators. Pearson correlation analysis was used to assess the relationships among these indicators, and machine learning models such as Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Logistic Regression were applied to predict employee turnover.

The Random Forest model achieved the highest accuracy at 98.8%, followed by the SVM and KNN models, which also demonstrated strong performance. However, the Naive Bayes and Logistic Regression models showed lower accuracy.The findings indicate that the data used are effective in predicting employee turnover, and machine learning models significantly contribute to improving prediction accuracy. As the accuracy of these predictions increases, businesses can benefit by taking appropriate measures to address potential turnover risks.

Ajit et al. (2016) addressed voluntary turnover, a costly and undesirable issue for companies. Voluntary employee departures often involve significant expenses, including onboarding and training for replacements, particularly when highly skilled employees leave. The study focused on analyzing voluntary turnover and considered key parameters such as age, tenure, pay, overall job satisfaction, and employees' perceptions of fairness.

For prediction, the study utilized the Extreme Gradient Boosting (XGBoost) technique and tested seven different models: Logistic Regression, Naïve Bayes, Random Forest, K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and XGBoost. The importance of XGBoost was highlighted, with the authors emphasizing that future research should shift focus from identifying who might leave to exploring what can be done to retain employees.

This study focuses on measuring employee commitment and determining how satisfied workers are with their current conditions in a retail company in Malaysia. It explores the relationship between job satisfaction and turnover intention. To investigate this relationship, a questionnaire was developed. The developed questionnaire was based on the Job Descriptive Index, Organizational Commitment Questionnaire, and Lee and Mowday's turnover intention items, and was answered by 62 participants. The results of the survey revealed that employees in the retail sector in Malaysia are not satisfied with their salary. Additionally, it was found that employees have a moderate level of commitment to their jobs. There is a negative and significant relationship between job satisfaction factors and turnover intention. In response to the findings, the article suggests that increasing salaries and managers supporting employees could have a positive effect in reducing turnover rates.

One of the most important factors for the success of firms is the employees' commitment to the company. A decrease in employee engagement and situations such as employees leaving the company create obstacles to the success of firms. In their study, Arshad and Puteh (2015) aim to determine the turnover tendencies of employees. A total of 106 employees from four different branches of the company were surveyed, as mentioned in the article. The factors affecting employee turnover in firms are listed as Perceived Organizational Support, Job Stress, Work-Life Balance, and Available Job Alternatives/Opportunities. SPSS was used to analyze the results. The findings show that work-life balance and other job alternatives are key factors leading employees to leave their organizations.

In his literature-based study, Ongori (2007) examined the reasons for employee turnover, its effects on organizations, and the measures companies can take to address employee turnover. The study categorized the reasons for employee turnover into organizational factors, environmental factors, and personal factors. Criteria such as dissatisfaction with management, career goals, insufficient salary, and unfavorable working conditions were discussed. The study also highlighted both the negative and positive impacts of employee turnover on organizations and proposed strategies to address this issue.

In their study, Alla and Rajaa (2019) classified employee turnover into Functional/Dysfunctional turnover and Voluntary/Involuntary turnover. The study listed and analyzed the factors associated with each group, arguing that organizations should not only aim to reduce turnover rates but also analyze them according to these classifications to take appropriate actions. The article also included strategies to address employee turnover.

They have worked on predicting employee turnover and developing strategies to retain employees within the organization. In their research, Ribes et al. (2017) utilized mathematical and statistical models. By applying machine learning methods, they were able to more effectively identify the reasons behind employee turnover. Additionally, they compared various machine learning techniques, such as Random Forest, Linear Discriminant Analysis, and Support Vector Machines (SVMs). They found that more complex models were more successful in prediction compared to simpler models. Through the developed mathematical and statistical models, they were able to predict employees with a high likelihood of turnover in advance, allowing for the implementation of strategies to retain these employees.

In their study, Brown et al. (2024) proposed the use of deep learning techniques to predict employee turnover. They compared deep learning methods with other machine learning techniques, such as Random Forest, Decision Trees, and Support Vector Machines. They argued that artificial intelligence-based deep learning algorithms are more effective than traditional machine learning methods. They suggested that better results are obtained when working with very large datasets. The authors highlighted that deep learning methods provide a significantly higher accuracy in predicting employee turnover. They stated that this could enable organizations to develop more accurate strategies.

In Effendi's (2024) study, the factors influencing employee turnover were examined, and Logistic Regression analysis was used as the method. The study identified factors affecting employee commitment, such as salary, career advancement, teamwork, and working hours. The method revealed how these factors influenced employees' decisions to leave their jobs. According to the model's results, salary was found to be a strong determinant among the factors influencing employee turnover. Furthermore, the article reported that the model could predict employee turnover with an accuracy of 82%.

In their study, Nketsiah and Nkansah (2024) examined the turnover intentions of employees in the banking sector in Ghana based on certain criteria. The research explored how factors such as trust and commitment influence employees' decisions to stay in their current jobs. Data was collected from the sample, and it was found that there is an inverse relationship between employees' level of organizational commitment and their turnover intentions. The authors noted that the turnover rates of employees in local and foreign banking firms may differ from each other. They suggested that Human Resources activities should be restructured to support and enhance employees' commitment to the organization.

The explanation of the numbers corresponding to the turnover prediction parameters in the table is stated below.

Table 1. Parameters of Employee Turnover.

Parameters of Employee Turnover			
1	**Working Conditions**	13	**Team Dynamics**
2	**Job Satisfaction**	14	**Infectious Effect**
3	**Lack of Training**	15	**Technological Changes**
4	**Management Support**	16	**Personal Development**
5	**Pay Fairness**	17	**Unsuitability for Work**
6	**Work-Life Balance**	18	**Lack of Recognition**
7	**Workload Balance**	19	**Job Security**
8	**Organizational**	20	**Lack of Social Ties**
9	**Career Opportunities**	21	**Project Dynamics**
10	**External Opportunities**	22	**Business Culture**
11	**Leadership Style**	23	**Perception of Justice**
12	**Workplace Conflicts**		

Table 2. Employee Turnover Parameters in the Studies

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Salleh, R., et al. (2012)		✓						✓	✓														
Arshad, H., & Puteh, F.	✓			✓			✓																
Ongori, H. (2007)			✓	✓		✓																	
Alla, A. A., & Rajaa, O.		✓								✓		✓											
Teng, M., et al. (2019)				✓									✓	✓									
Hong, W. C., et al. (2007)					✓	✓					✓												
Yin, Z., et al. (2024)	✓	✓													✓								
Recilla, V. J., et al. (2024)				✓												✓							
Zhang, J., et al. (2024)			✓														✓						
Ajit, P. (2016)		✓				✓												✓					
Ribes, E., et al. (2017)		✓																	✓	✓			
Brown, A., et al.										✓													
Effendi, S. R. F. (2024)					✓		✓														✓		
Nketsiah, T. A., &								✓											✓			✓	
Chang, Y., et al. (2024)	✓							✓															✓

According to the data in the table, "job satisfaction" and "organizational commitment" are among the most frequently mentioned factors. This illustrates that the employees' level of job satisfaction and organizational commitment is directly related to their decision to leave the job. Further, factors like "management support" and "work-life balance" are also high on mentions, showing how important it is for workers to receive expectations from leaders in the organization and their personal life balance.

Table 3. Machine Learning Methods Used in the Studies.

Study	Random Forest	Support Vector Machine	CatBoost	XGBoost	Artificial Intelligent	Logistic Regression	Statistical Analysis Methods	KNN	Naive Bayes
Salleh, R., Nair,							✓		
Arshad, H., &							✓		
Teng, M., Zhu, H.,	✓	✓							
Yin, Z., Hu, B., &			✓	✓					
Recilla, V. J.,					✓				
Zhang, J., Cao, Z.,		✓							
Ajit, P. (2016)								✓	✓
Ribes, E., Touahri,	✓	✓							
Brown, A., Davis,					✓				
Effendi, S. R. F.						✓			
Nketsiah , T. A., &							✓		

Studies in the literature have been examined and a table of machine learning methods used in the studies has been stated. In the table, it is concluded that Statistical Analysis Methods and Support Vector Machine methods are used more frequently than other methods. It has been observed that the combination of Support Vector Machine algorithm with Random Forest method increases the accuracy in the estimation of leaving the job and is useful in evaluating the effects of the previously mentioned effective factors in leaving the job.

The figure presents some variations of machine learning and data analytics techniques adopted for the predictions of employee turnover. The most frequent occurrence is the 'supervised' technique, with 14 papers in total. All the well-known algorithms, such as Support Vector Machine, Random Forest, XGBoost, Catboost, Logistic Regression, and Deep Learning, form a part of this category. Techniques like Deep Learning and Logistic Regression occur in two papers. More precisely, the RF and XGBoost algorithms were compared in two papers published in the year 2024. K-Means Clustering and Hierarchical Clustering were the unsupervised techniques applied in only 2 papers. Besides, 2 papers have applied the Hybrid methods. These techniques combined supervised and unsupervised techniques. In one of these, they used GA for prediction, while others used the techniques that combined Big Data Analytics with Machine Learning techniques. In view of this distribution, in most cases, employee churn could be predicted with the application of the supervised machine learning techniques, while only some concrete cases could have unsupervised and hybrid approaches applied.

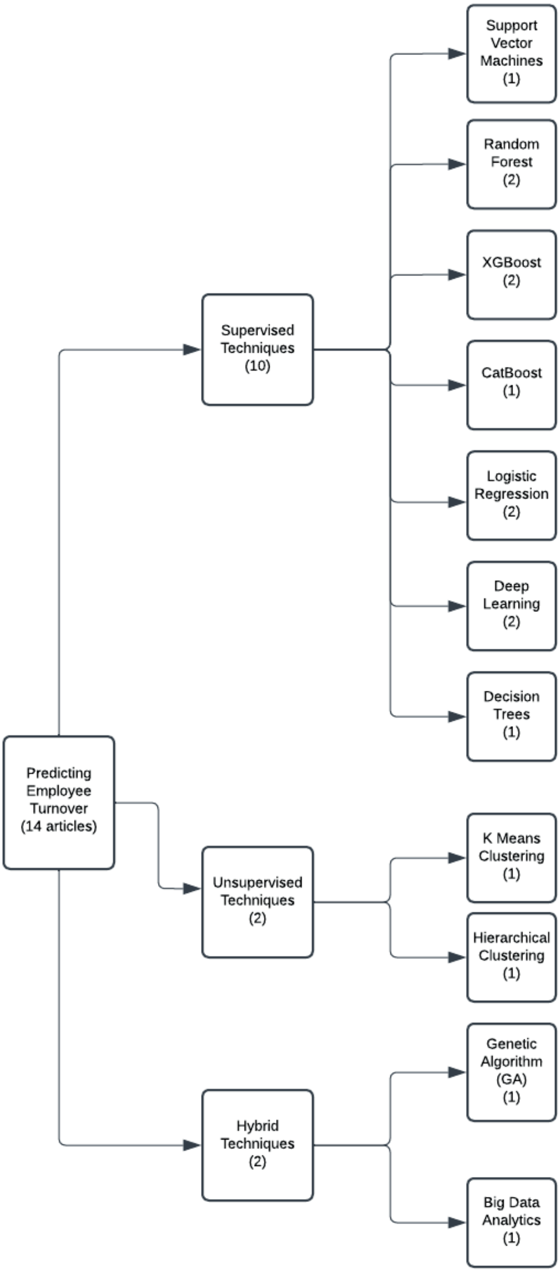

Figure. 1. Machine Learning Techniques for Employee Turnover Prediction in the Literature

CONCLUSION

In this study, employee layoffs were categorized into two types: voluntary and involuntary. voluntary layoffs could result in a financial loss to the company. Predicting the redundancies earlier is a situation that most of the companies want to have so that measures can be taken depending on the situation and retention strategies can be developed for employees who are at risk of quitting their jobs. In order to predict the exit from work, machine learning algorithms are generally used. Supervised machine learning algorithms generally provide more accurate results than machine learning algorithms. For this reason, the study of supervised machine learning algorithms was carried out in the study. While the data were being examined within the literature research in order to be able to use machine learning algorithms, some parameters came to the fore in employee data. The most used ones among these parameters are: working conditions, job satisfaction, lack of training, management support, pay fairness, work-life balance, workload balance, organizational commitment, career opportunities, external opportunities, technological changes, personal development. It has, in fact studied the role of algorithms operating according to these parameters in forecasting. Logistic Regression, Decision Trees, Random Forests, Support Vectos Machines(SVM), K-Nearest Neighbors (KNN), Naive Bayes from Supervised machine learning algorithms have been examined in the literature and their role in forecasting.

Each of the machine learning algorithms used for predicting employee turnover rate has various advantages and disadvantages. Logistic Regression is very powerful, especially in terms of simplicity and interpretability. The transparency provided by employees in understanding the reasons for leaving their jobs makes it easier for human resources teams to make strategic decisions. However, it usually stands behind more complex algorithms, such as Random Forest or Support Vector Machine, in terms of accuracy. On the other hand, Decision Trees, due to their simplicity and visualization, are useful in showing major reasons for the employee turnover rate. However, this algorithm is not as effective as Random Forest or Gradient Boosting Trees, especially on unstable and complex datasets. The Random Forest attracts attention as one of the algorithms showing the highest accuracy rates in employee turnover rate forecasting. Strong performance allows, in particular for large and complex data sets, the elaboration of effective strategies in human resource management. However, these high accuracy rates have to be offset against increased computational costs. SVM is an excellent choice because of its accuracy and success in classifying complex datasets. Other algorithms, like KNN and Naive Bayes, work well in certain situations but are less preferred for large data sets. Therefore, Random Forest yielded very good performance with high accuracy rates, while its application areas are really wide in predicting the rate of employee turnover and recommending that the organization make a decision by taking into consideration the structure of datasets, their balance, and business needs. Simpler models could offer effective alternatives, at least when Logistic Regression, speed of implementation, and interpretability are a priority. The same approach may offer organizations a strategic benefit in the engagement of workers and in limiting turnover.

References

Ahmed, S., Ahmed, A., & Jwmaa, S. (2023). Analyzing The Employee Turnover by Using Decision Tree Algorithm. *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1-4. https://doi.org/10.1109/HORA58378.2023.10156709.

Alaskar, L., Crane, M., & Alduailij, M. (2019). Employee Turnover Prediction Using Machine Learning. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-030-36365-9_25.

Atef, M., Elzanfaly, D., & Ouf, S. (2022). Early Prediction of Employee Turnover Using Machine Learning Algorithms. *International Journal of Electrical and Computer Engineering Systems*. https://doi.org/10.32985/ijeces.13.2.6.

Birzniece, I., Andersone, I., Ņikitenko, A., & Zvirbule, L. (2022). Predictive Modeling of HR Dynamics Using Machine Learning. *Proceedings of the 2022 7th International Conference on Machine Learning Technologies*. https://doi.org/10.1145/3529399.3529403.

Chang, V., Mou, Y., Xu, Q., & Xu, Y. (2022). Job satisfaction and turnover decision of employees in the Internet sector in the US. *Enterprise Information Systems, 17*. https://doi.org/10.1080/17517575.2022.2130013.

Chen, B. (2023). Factors of Employee Attrition: A Logistic Regression Approach. *Advances in Economics, Management and Political Sciences, 20,* 214-225.

Effendi, S. R. F. (2024). The Effect of Job Division, Promotion, Project, Salary, Length of Work With Employee Turnover Using Logistic Regression Analysis. *Jurnal Indonesia Sosial Teknologi, 5(3),* 1204-1210.

Gao, X., Wen, J., & Zhang, C. (2019). An Improved Random Forest Algorithm for Predicting Employee Turnover. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2019/4140707.

Hong, W. C., Wei, S. Y., & Chen, Y. F. (2007). A comparative test of two employee turnover prediction models. *International Journal of Management, 24(4),* 808.

Islam, M., Alam, M., Islam, M., Mohiuddin, K., Das, A., & Kaonain, M. (2018). An Adaptive Feature Dimensionality Reduction Technique Based on Random Forest on Employee Turnover Prediction Model. *Advances in Intelligent Systems and Computing, 269-278*. https://doi.org/10.1007/978-981-13-1813-9_27.

Jhaver, M., Gupta, Y., & Mishra, A. (2019). Employee Turnover Prediction System. *2019 4th International Conference on Information Systems and Computer Networks (ISCON),* 391-394. https://doi.org/10.1109/ISCON47742.2019.9036180.

Kanuto, A. (2024). Identifying Patterns and Predicting Employee Turnover Using Machine Learning Approaches. *International Journal of Science and Business*. https://doi.org/10.58970/ijsb.2373.

Kabak, M., Saglam, F., & Aktaş, A. (2017). Usability analysis of different distance measures on TOPSIS. *Journal of the Faculty of Engineering and Architecture of Gazi University, 32(1)*.

Krishna, S., S., & Borah, D. (2023). Machine Learning for Ensuring Sustainable Development: Predicting Employee Attrition in the Workplace. *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA),* 1-7.

Kumar, P., Gaikwad, S. B., Ramya, S. T., Tiwari, T., Tiwari, M., & Kumar, B. (2023). Predicting Employee Turnover: A Systematic Machine Learning Approach for Resource Conservation and Workforce Stability. *Engineering Proceedings, 59,* 117.

Leana, C. R., & Ivancevich, J. M. (1987). Involuntary job loss: Institutional interventions and a research agenda. *Academy of Management Review, 12(2),* 301-312.

Liao, C. (2023). Employee turnover prediction using machine learning models. *Proceedings of SPIE, 12596,* 125960X-5. https://doi.org/10.1117/12.2672733.

Nayak, S., & Palai, P. (2023). Employee Attrition System Prediction using Random Forest Classifier. *International Journal of Computer and Communication Technology*. https://doi.org/10.47893/ijcct.2023.1445.

Punnoose, R., & Ajit, P. (2016). Prediction of Employee Turnover in Organizations using Machine Learning Algorithms: A case for Extreme Gradient Boosting. *International Journal of Advanced Research in Artificial Intelligence*. https://doi.org/10.14569/IJARAI.2016.050904.

Russell, C. J. (2013). Is it time to voluntarily turn over theories of voluntary turnover? *Industrial and Organizational Psychology, 6(2),* 156-173.

Setiawan, I., Suprihanto, S., Nugraha, A. C., & Hutahaean, J. (2020). HR analytics: Employee attrition analysis using logistic regression. *IOP Conference Series: Materials Science and Engineering, 830(3)*. https://doi.org/10.1088/1757-899X/830/3/032001.

Taner, Z., Hızıroğlu, O. A., & Hızıroğlu, K. (2024). Leveraging Machine Learning Methods for Predicting Employee Turnover Within the Framework of Human Resources Analytics. *Journal of Intelligent Systems: Theory and Applications, 7(2),* 145-158.

Valle, M. A., & Ruz, G. A. (2015). Turnover prediction in a call center: behavioral evidence of loss aversion using random forest and naïve bayes algorithms. *Applied Artificial Intelligence, 29(9),* 923-942.

Zhang, C., & Han, W. (2024). Ensembles of decision trees and gradient-based learning for employee turnover rate prediction. *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.2387.

Zhang, J., Cao, Z., Qu, H., & Wang, M. (2024). Management and prediction of employee turnover in enterprises based on big data analytics and machine learning. *Applied and Computational Engineering, 76,* 103-108.

Živković, A., Šebalj, D., & Franjkovic, J. (2024). Prediction of the Employee Turnover Intention Using Decision Trees. *Proceedings of the 2024 International Conference,* 325-336. https://doi.org/10.5220/0012538400003690.

Zhao, Y., Hryniewicki, M., Cheng, F., Fu, B., & Zhu, X. (2018). Employee Turnover Prediction with Machine Learning: A Reliable Approach. *Lecture Notes in Computer Science, 737-758.* https://doi.org/10.1007/978-3-030-01057-7_56.

Yunmeng, Z., & Chengyi, Z. (2019). The Application of The Decision Tree Algorithm Based on K-means in Employee Turnover Prediction. *Journal of Physics: Conference Series, 1325.* https://doi.org/10.1088/1742-6596/1325/1/012123

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

Melisa Dikici and Gökçe Sabriye Hörük write the Article. Deniz Efendioğlu supervised the article.