DOI https://doi.org/10.5281/zenodo.10049652#104

TWIST, 2024, Vol. 19, Issue 1, pp. 382-387

TWIST

Journal homepage: www.twistjournal.net

Machine Learning Techniques in Employee Churn Prediction

Gillann Earl S. Alcala
College of Development Management, University of Southeastern Philippines – Mintal Campus,
Mintal, Davao City 8000, Philippines

John Vianne B. Murcia*
[1]- College of Development Management, University of Southeastern Philippines – Mintal Campus,
Mintal, Davao City 8000, Philippines
[2]- College of Business Administration Education, University of Mindanao, Bolton Street,
Davao City 8000, Philippines
[**Corresponding author*]

Abstract
An employee has one of the most important roles in a company's continued operation. This study employed different machine-learning techniques to identify the best classifier for employee churn prediction. A total of 14,249 instances were secured of which 14,200 instances were used as a training set and identifying the two best classifiers. The remaining 49 instances with removed class labels were used as a test set. Seven classifiers were employed in this study, these are the trees.J48 (decision tree), trees.RandomForest, trees.RandomTree, the lBk (*k*-NN), Naïve Bayes, Logistic, and the Multilayer Perceptron. Parameter tuning was implemented for the J48 and the *k*-nearest neighbor. The result revealed that the trees. Random Forest (97.76%) and trees.J48 with a confidence factor of 0.25 (96.73%) have the two highest classification accuracy values. In the prediction phase, these two classifiers were used to predict the test set, where 40 instances are predicted to stay and 9 instances are predicted to leave in the actual dataset. On the other hand, the trees. Random Forest predicted 38 employees to stay and 11 to leave, while the trees.J48 with a confidence factor of 0.25 predicted 35 employees to stay and 14 as leaving. Implications are discussed.

Keywords
Employee churn, Prediction, Machine learning, Decision tree, Random forest, Weka

INTRODUCTION

An employee performs a vital role in a company or an organization (Gabčanová, 2011). When an employee leaves the company, a disruption in day-to-day operations may occur. Hence, understanding and even attempting to predict when an employee would leave the company is imperative. The advent of globalization and the internet ameliorated the situation as these conditions have given employees a chance to explore different companies all across the globe. This posed a challenge for employers and companies, seeing that employee churning and company hopping had been a practice for employees who are looking for a greener pasture. Predicting employee turnover is crucial to human resource management because it enables businesses to recognize and address the causes of employee turnover.

Employee churn prediction in several industries, including healthcare, has recently increased (Gentek, 2022). Additionally, Internet of Things (IoT) technology is now being used to improve employee turnover analysis precision and effectiveness (Naz et al., 2022). Organizations may create strategies to keep their valuable staff and preserve a competitive edge in the market by recognizing the elements that lead to employee turnover (Saradhi & Palshikar, 2011). Businesses that have a high staff churn rate sometimes suffer from increased costs and workplace disturbances (Thompson, 2023). Employee turnover has a significant financial impact; it is estimated that on average, it costs six to nine months of a salaried employee's income to replace them (Charaba, 2022). High turnover rates can also result in lost productivity, which costs U.S. businesses an estimated $1.8 trillion yearly (Charaba, 2022). This lost production is in addition to the direct costs of hiring and training new personnel.

Organizations must comprehend the causes of high churn rates to create effective retention strategies. Lack of possibilities for career advancement, inadequate pay, a bad work-life balance, and an undesirable working environment are some typical causes of employee turnover (Bradshaw, 2023). Organizations can enhance employee retention and satisfaction by addressing these problems, which will ultimately lower the expenses incurred by high churn rates (Thompson, 2023). Given these, building prediction models that may predict staff turnover rates requires the application of data mining and statistical approaches (Saradhi & Palshikar, 2011). These models are the result of the application of a wide range of machine-learning strategies, such as logistic regression, support vector machines, and decision trees (Maharjan, 2021; Alamsyah & Salma, 2018). With such models, organizations can identify at-risk individuals and undertake focused actions to keep prized talent by using predictive models for employee attrition (Saradhi & Palshikar, 2011). Organizations can acquire insights into the issues behind employee turnover and design data-driven solutions to increase retention by utilizing advanced data science approaches and IoT-enabled predictive methodologies (Naz et al., 2022).

Several studies have tried to employ machine learning techniques in predicting employee churn. Using the CatBoost method, which is a supervised machine learning algorithm, Jain, Tomar and Jana (2023) proposed a scheme for predicting employee churn and retaining current employees. This method was able to outperform other algorithms that were utilized in the prediction of employee turnover. Moreover, they have employed Support Vector Machine, Random Forest, and Decision Tree in identifying the best classifier to predict employee churn. It was revealed that the random forest and decision tree were the better classifiers compared to SVM. Furthermore, it was also found that between the two, the random forest performs better compared to the decision tree.

Najafi-Zangeneh et. al. (2021) used feature selection in improving the machine-learning prediction model for employee churn. The use of *m*-max-out algorithm was utilized in the pre-processing stage and used logistic regression in prediction. It was found that applying the feature selection improves the performance of logistic regression. A different approach was studied by Cai et. al. (2020), wherein instead of the m-max-out algorithm, the dynamic bipartite graph embedding (DBGE) was utilized in the feature selection. It was also found that this feature selection method significantly improves the performance of the prediction.

Khera and Divya (2018) employed a support vector machine (SVM) to predict employee churn in the IT industry in India. It was found that SVM has a classification accuracy of 85%. It was further revealed that SVM better predicts employees who will leave the company compared to predicting the employees who will stay. Zhang et. al. (2018) employed the Gradient-boosted decision trees algorithm and the logistic regression algorithm in predicting employee churn. However, the use of one or two machine learning algorithms does not capture the capability of other machine learning techniques in predicting employee churn.

Meanwhile, Chaudhary et. al. (2022), found that the CatBoost algorithm has the best prediction performance compared to Support Vector Machine, decision tree, RandomForest, and XGradient Boost machine learning techniques. It was concluded that the CatBoost performed best for both categorized and non-categorized datasets. On the other hand, Srivastava and Eachempati (2021) found that Deep Neural Network (91.2%) has better classification accuracy performance compared to the RandomForest (82.3%) and gradient boosting (85.2%).

There is a lack of research regarding the prediction of employee attrition, particularly regarding the identification of the causes of high churn rates and the development of viable solutions across industries and companies. While some studies have focused on specific industries, such as retail (e.g., Hubbard, 2023) and healthcare (Qualtrics, 2023), a more comprehensive study encompassing a broader range of industries and organizational contexts is necessary. The vast majority of the present research makes use of generalized datasets that are available across several platforms; nevertheless, these datasets may not effectively reflect the specific challenges and dynamics of different enterprises or organizations (Hubbard, 2023). In addition, the prediction of employee turnover through the use of IoT-enabled predictive techniques and machine learning algorithms is still an emerging topic of study with opportunities for additional research and development (Robinson, 2023). Hence, this study aims to employ different machine learning algorithms for employee churn prediction. Evaluating each machine learning algorithm using the classification accuracy and the different statistical techniques. In addition, this study also aims to explore which among the machine learning algorithms being employed is the best classifier for predicting employee churn.

MATERIALS AND METHODS

Dataset

Data on employee churn was obtained from Kaggle dataset employee churn prediction (https://www.kaggle.com/datasets/mzinic/employee-churn-prediction) from Mzinic (2021). This dataset was used to find a machine learning that will accurately predict employee churn. The dataset has 10 variables, nine (9) of which are explanatory variables, while one (1) class variable. In addition, these variables have 14,249 instances of which 14,200 were used as a training set while 49 instances were utilized as the test set which will be used in the prediction. Further, the class attribute had been removed from the test set. Table 1 shows the list of attributes in this study with their corresponding description.

Table 1 List of Attributes of the Employee Churn Dataset

Attribute	Type	Description
AVG_MONTHLY_HRS	Numeric	Employee’s average number of hours worked per month.
DEPARTMENT	Nominal	The department to which the employee is assigned.
FILED_COMPLAINT	Nominal	If the employee filed a complaint for the last three years (Y) or not (N)
LAST_EVALUATION	Numeric	Employee’s most recent evaluation score
N_PROJECTS	Numeric	The number of projects the employee is tasked with.
RECENTLY_PROMOTED	Nominal	If the employee was promoted for the last three years (Y) or not (N)
SALARY	Nominal	Salary level of the employee, coded as low, medium, or high
SATISFACTION	Numeric	Employee’s satisfaction score with the company
TENURE	Numeric	Number of years in the company
Class	Nominal	If the employee stayed (1) or left (0) the company.

Data Preparation

Upon securing the needed data for employee churn prediction, a data cleaning was performed where missing values were replaced and attribute types were converted. In dealing with missing values, an unsupervised machine learning called *Replace Missing Values* was utilized. Upon inspection, there were three attributes with missing values, these are DEPARTMENT (the department where the employee is assigned with 706 missing values), LAST_EVALUATION (the last evaluation score of the employee with 1,528 missing values), and SATISFACTION (the employee’s satisfaction score with 180 missing values). The unsupervised machine learning algorithm that was applied in replacing missing values utilizes the mean and mode in replacing the missing values of numerical and nominal attributes, respectively.

For mislabeled attributes, that is nominal attributes are labeled as numeric, the unsupervised machine learning called *Numeric To Nominal* was utilized. Upon inspection, attributes FILED_COMPLAINT (filed a complaint) and RECENTLY_PROMOTED (recently promoted) were incorrectly labeled as numeric. Since there are only two attributes that need to be converted from numeric to nominal, the *attribute Indices* parameter was set to FILED_COMPLAINT and RECENTLY_PROMOTED. This is to ensure that other attributes which were correctly labeled as numeric will not be affected.

Selection of Attributes

A feature selection procedure was implemented to gain information from the nine attributes considered in this study. The *CorrelationAttributeEval* is a feature selection technique based on the correlation of the individual attributes to the class. Moreover, the *InfoGainAttributeEval* was utilized, which is another feature selection technique. This was implemented to evaluate the worthiness of an attribute by measuring the information gained concerning the class.

After implementing the *Correlation Attribute Eval* feature selection, it was found that SATISFACTION (r=0.3879) has the highest correlation to the class. This implies that the higher the satisfaction of an employee with the company, the higher the chance of staying with the company. Moreover, FILED_COMPLAINT (r=0.1573), TENURE (r=0.1434), SALARY (r=0.1049), and AVG_MONTHLY_HRS (r=0.0733) were the other highly correlated attributes to the class. On the other hand, DEPARTMENT (r=0.0172), N_PROJECTS (r=0.0259), LAST_EVALUATION (r=0.0385), and RECENTLY_PROMOTED (r=0.0615) were the attributes with the least correlation coefficient to the class.

Applying the *Info Gain Attribute Eval* reveal that it was also SATISFACTION (r=0.2559) which has the highest information gain concerning the class. Further, N_PROJECTS (r=0.2547), AVG_MONTHLY_HRS (r=0.1822), TENURE (r=0.1225), and LAST_EVALUATION (r=0.0761) were the other attributes with high information gain. On the contrary, RECENTLY_PROMOTED (r=0.0036), DEPARTMENT (r=0.0044), SALARY (r=0.0202), and FILED_COMPLAINT (r=0.0218) were the least attributes with the lowest gained information concerning the class.

Data Classification and Cross-Validation

This study utilizes seven classifiers to perform classification on the training dataset. These are the trees.J48 (decision tree), trees.RandomForest, lBk (k-NN), NaiveBayes, Logistic, and Multilayer Perceptron. The classifiers trees. Random Tree, Naïve Bayes, Logistic, and Multilayer Perceptron no longer need parameter tuning. Hence, these classifiers were left in default mode. For J48, the confidence factors were tested for 0.25, 0.50, and 0.75, while for lBk, k-NN 1, k-NN 3, k-NN 5, k-NN 7, k-NN 9, and k-NN 11 were utilized. Meanwhile, the feature selection *Wrapper Subset Eval* was used to identify the number of cross-validation folds for accuracy estimation. The result revealed that the number of folds for accuracy estimation is five. Hence, a cross-validation fold of five was utilized in the training dataset.

RESULTS AND DISCUSSION

There are 14 cross-validations conducted using seven classifiers: four using J48 (decision tree), one Random Tree, one Random Forest, six using lBk (k-NN), one Naïve Bayes, one Logistic, and one Multilayer Perceptron. Table 1 shows the classification accuracy for each classifier used in this study. It can be observed that the Random Forest has the highest classification accuracy (97.76%) among all the classifiers. On the other hand, the J48 with a confidence factor of 25% has the highest accuracy (96.73%) among the J48 variants and k-NN 1 has the highest classification accuracy (95.20%) among the lBk variants. It can also be observed that as the hyper parameter of k-NN increases, the classification accuracy decreases. Moreover, the random tree has a classification accuracy of 96.14%, while the Naïve Bayes was able to correctly classify 81.11% of the instances. In addition, Logistic has a 78.94% classification accuracy, while the multilayer perceptron was able to obtain a 96.13% classification accuracy. This result implies that Random Forest and J48 (25% confidence factor) have the highest classification accuracy with 97.76% and 96.73%, respectively.

Examining the number of correctly classified instances revealed that among the J48 variants, the confidence factor of 0.25 correctly classified 13,736 out of 14,200 instances. In terms of the lBk *k*-NN machine learning algorithm, the *k*-NN 1 variant correctly classified 13,519 out of 14,200 which has the greatest number of correctly classified instances. Moreover, the Random Tree correctly classified 13,652 out of 14,200 instances, while the random forest correctly classified 13882 out of 14200. In addition, multilayer perceptron correctly classified 13,651 out of 14200; NaiveBayes correctly classified 11,518 out of 14,200; and Logistic correctly classified 11,209 out of 14,200.

Table 2 Classification accuracy of the classifiers and their variants on the training dataset (*N*=14,200)

Classifier	Variants	Correctly Classified Instances
Decision Tree (trees.J48)	0.25	13736 (96.73%)
	0.50	13731 (96.70%)
	0.75	13678 (96.32%)
Random Tree	-	13652 (96.14%)
Random Forest	-	13882 (97.76%)
lBk (*k*-NN)	1	13519 (95.20%)
	3	13391 (94.30%)
	5	13296 (93.63%)
	7	13263 (93.40%)
	9	13235 (93.20%)
	11	13193 (92.91%)
NaiveBayes	-	11518 (81.11%)
Logistic	-	11209 (78.94%)
Multilayer Perceptron	-	13651 (96.13%)

Overall, the two classifiers with the highest number of correctly classified instances are the trees.RandomForest (13,882 out of 14,200) and the trees.J48 (decision tree) with a confidence factor of 0.25 (13,736 out of 14,200). On the contrary, NaiveBayes (11,518 out of 14,200) and Logistic (11,209) have the lowest number of correctly classified instances. The results implied that trees.RandomForest and trees.J48 (decision tree) with 0.25 as confidence factor are the two candidate machine learning algorithms that can be used in predicting the test set.

Assessing the classification performance of machine learning techniques is another way of deciding to choose a model. Table 3 shows the classification performance of the classifiers using the kappa statistic, true positive (TP) rate, false positive (FP) rate, precision, recall, and F-measure. As a rule, the larger the value of the kappa statistic (Veira, Kaymak & Sousa, 2010), the TP rate, precision, recall, and F-measure means the better the machine learning algorithm (Parikh et al., 2008). On the contrary, the lower the FP rate of the classifier, the better (Parikh et al., 2008).

Table 3 Classification performance of the utilized classifiers and their variants on the training dataset (*N*=14,200)

Classifier	Variants	κ	TP rate	FP rate	Precision	Recall	F-measure
Decision Tree (J48)	0.25	0.9104	0.907	0.013	0.956	0.907	0.931
	0.50	0.9102	0.912	0.015	0.951	0.912	0.931
	0.75	0.9019	0.925	0.023	0.926	0.925	0.925
Random Tree	-	0.8897	0.923	0.029	0.909	0.923	0.916
Random Forest	-	0.9359	0.929	0.008	0.974	0.929	0.951
lBk (k-NN)	1	0.8709	0.933	0.042	0.874	0.933	0.903
	3	0.8476	0.903	0.043	0.867	0.903	0.885
	5	0.826	0.869	0.042	0.866	0.869	0.867
	7	0.8157	0.857	0.043	0.862	0.857	0.859
	9	0.8113	0.852	0.043	0.860	0.852	0.856
	11	0.8029	0.845	0.045	0.854	0.845	0.850
Naïve Bayes	-	0.4029	0.419	0.067	0.662	0.419	0.513
Logistic	-	0.3207	0.347	0.073	0.599	0.347	0.439
Multilayer Perceptron	-	0.8881	0.893	0.019	0.937	0.893	0.914

The result revealed that the random forest (κ=0.9359) and the J48 with 0.25 confidence factor (κ=0.9104) have the highest kappa statistic. This implies that when measure in terms of kappa statistic, these two models performed better compared to others. Furthermore, the result of the TP rate revealed that lBk *k*-NN 1 (TP=0.933) and the RandomForest (TP=0.929) have the highest TP rates. Meanwhile, the result of the precision, recall, and F-measure revealed that the RandomForest (Precision=0.974, Recall=0.929, F-measure=0.951) has the highest value compared to other classifiers and the J48 with 0.25 confidence factor (Precision=0.956), lBk *k*-NN 1 (Recall=0.933), and J48 with 0.25 and 0.50 confidence factors (F-measure=0.931) were seen toi have the highest values in precision, recall, and F-measure, respectively. In terms of FP rate, the RandomForest (FP=0.008) and J48 with 0.25 confidence factor (FP=0.013) have the lowest values. The classification performance revealed that the RandomForest and the J48 (decision tree) with a 0.25 confidence factor are the most performing machine learning algorithms that may best classify and predict class membership of employee churn compared to other classifiers used in this study.

Being able to derive the accuracy of the models, the algorithms were put to test that involves a test dataset containing 49 employees with unlabeled class attribute (whether an employee churn is predicted or otherwise). Table 4 displays the accuracy of the Random Forest and J48 (0.25) classifiers in predicting employee churn. The classifiers were trained using a dataset of 14,200 instances, with the goal of determining whether an employee will stay or leave the company. The current dataset consists of 40 employees classified as 'Staying' and 9 employees classified as 'Leaving'. The evaluation of the classifiers is determined by the number of true positives (TP), true negatives (TN), and their respective rates. The Random Forest method demonstrated significantly greater accuracy in predicting both classes compared to J48 (0.25), indicating that it exceeds the latter in properly identifying employees at risk of leaving. The exceptional performance can be ascribed to the intrinsic attributes of the Random Forest method, which builds numerous decision trees and combines their outcomes, potentially resulting in more robust and precise predictions.

Table 4 Classification performance of the utilized classifiers and their variants on the training dataset (*N*=14,200)

Algorithm	Staying	Leaving	TP	TN	TP Rate (%)	TN Rate (%)
Actual	40	9	-	-	-	-
Predicted (Random Forest)	38	11	7	38	77.78%	95%
Predicted (J48[0.25])	35	14	4	35	44.44%	87.5%

CONCLUSION

This study's exploration of machine learning methods to predict employee churn has yielded valuable insights with significant implications for human resource management and organizational strategy. By applying various classifiers, the research demonstrates the nuanced complexity inherent in understanding employee turnover. The trees.RandomForest classifier emerged as the most accurate, boasting a classification accuracy of 97.76%. This high level of precision suggests that Random Forest is adept at capturing the multifaceted and dynamic nature of factors contributing to employee churn, making it a potentially powerful tool for businesses seeking to predict and mitigate turnover.

Additionally, the study found that the J48 decision tree, particularly with a confidence factor of 0.25, also exhibited high accuracy (96.73%) in predicting employee churn. The J48's strength lies in its ability to model the decision-making process behind employee turnover, simplifying complex relationships into understandable rules. This characteristic makes it not only a practical predictive tool but also a means for businesses to gain qualitative insights into the specific factors influencing employee churn. Businesses can leverage these insights to identify key areas for intervention, such as job satisfaction, work-life balance, and career development opportunities, aligning with findings that employee satisfaction is a significant predictor of retention.

Furthermore, the lower agreement values in NaiveBayes and Logistic classifiers underscore the challenges in capturing the dynamic interplay of elements in predicting employment outcomes. This diversity in classifier performance highlights the necessity for businesses to adopt a multifaceted approach when analyzing employee turnover, considering a range of factors from personal job satisfaction to broader organizational culture and policies.

RECOMMENDATIONS

The study's practical implications for businesses are manifold. Firstly, adopting advanced machine learning models like trees.RandomForest and trees.J48 can significantly enhance the accuracy of employee churn predictions. This, in turn, allows for more targeted and effective human resource strategies, reducing turnover costs and improving overall organizational efficiency. Secondly, the insights gained from these models can guide businesses in identifying at-risk employees and developing tailored retention strategies, thereby fostering a more stable and engaged workforce.

From a strategic standpoint, businesses should focus on creating a work environment that addresses key factors influencing employee turnover. This includes offering competitive compensation, providing clear career progression paths, ensuring a positive work-life balance, and fostering a supportive and inclusive company culture. Furthermore, businesses should consider investing in continuous training and development programs, not only to enhance employee skills but also to boost job satisfaction and loyalty.

Finally, the findings from this study offer critical business insights into predicting and managing employee turnover. By effectively utilizing machine learning models, businesses can gain a deeper understanding of the factors driving employee churn and develop more nuanced strategies to enhance employee retention. This proactive approach to managing turnover not only reduces costs but also contributes to building a more resilient and dynamic organization.

REFERENCES

1. Alamsyah, A., & Salma, N. (2018, August). A comparative study of employee churn prediction model. In *2018 4th International Conference on Science and Technology (ICST)* (pp. 1-4). IEEE.
2. Bradshaw, R. (2023). 19 Employee Retention Statistics That Will Surprise you. Apollo Technical, LLC [Online]. Available at https://www.apollotechnical.com/employee-retention-statistics/
3. Cai, X., Shang, J., Jin, Z., Liu, F., Qiang, B., Xie, W., & Zhao, L. (2020). DBGE: employee turnover prediction based on dynamic bipartite graph embedding. *IEEE Access, 8*, 10390-10402.

4. Charaba, C. (2022). Employee retention: The real cost of losing an employee. People Keep. https://www.peoplekeep.com/blog/employee-retention-the-real-cost-of-losing-an-employee.
5. Chaudhary, M., Gaur, L., Jhanjhi, N. Z., Masud, M., & Aljahdali, S. (2022). Envisaging Employee Churn Using MCDM and Machine Learning. *Intelligent Automation & Soft Computing, 33*(2).
6. Gabčanová, I. V. E. T. A. (2011). The employees–the most important asset in the organizations. *Human Resources Management & Ergonomics, 5*(1), 30-33.
7. Gentek, A. (2022). Employee churn prediction in healthcare industry using supervised machine learning. Dissertation (KTH Royal Institute of Technology). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-321536.
8. Hubbard, C. (2023 May). Employee Turnover: Causes, Effects, and Strategies. HR Profiling Solutions. Available at https://www.hrprofilingsolutions.com.au/blogs/aus-blog/employee-turnover-causes-effects-strategies/.
9. Jain, N., Tomar, A., & Jana, P. K. (2021). A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. *Journal of Intelligent Information Systems, 56*, 279-302.
10. Jain, P. K., Jain, M., & Pamula, R. (2020). Explaining and predicting employees' attrition: a machine learning approach. *SN Applied Sciences, 2*, 1-11.
11. Khera, S. N., & Divya. (2018). Predictive modelling of employee turnover in Indian IT industry using machine learning techniques. *Vision*, *23*(1), 12-21.
12. Maharjan, R. (2021). Employee churn prediction using logistic regression and support vector machine. Master's Projects. 1043. https://doi.org/10.31979/etd.3t5h-excq
13. Mzinic, (2021 January). Employee churn prediction. Kaggle. Available at https://www.kaggle.com/datasets/mzinic/employee-churn-prediction.
14. Najafi-Zangeneh, S., Shams-Gharneh, N., Arjomandi-Nezhad, A., & Hashemkhani Zolfani, S. (2021). An improved machine learning-based employees attrition prediction framework with emphasis on feature selection. *Mathematics, 9*(11), 1226.
15. Naz, K., Siddiqui, I. F., Koo, J., Khan, M. A., & Qureshi, N. M. F. (2022). Predictive modeling of employee churn analysis for IoT-enabled software industry. *Applied Sciences, 12*(20), 10495.
16. Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology, 56*(1), 45.
17. Qualtrics (2023 January). Employee turnover: causes, costs, and combat strategies. Available at https://www.qualtrics.com/experience-management/employee/employee-turnover/
18. Robinson, A. (2023 May). 14 Causes & Reasons for Employee Turnover in 2023. teambuilding.com [Online]. Available at https://teambuilding.com/blog/employee-turnover
19. Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications, 38*(3), 1999-2006.
20. Srivastava, P. R., & Eachempati, P. (2021). Intelligent employee retention system for attrition rate analysis and churn prediction: An ensemble machine learning and multi-criteria decision-making approach. *Journal of Global Information Management (JGIM), 29*(6), 1-29.
21. Thompson, M. (2023 February). Why Churn is High and What You Can Do to Fix It. WeThrive [Online]. Available at https://wethrive.net/employee-retention/why-churn-is-high/.
22. Vieira, S. M., Kaymak, U., & Sousa, J. M. (2010, July). Cohen's kappa coefficient as a performance measure for feature selection. In *International Conference on Fuzzy Systems* (pp. 1-8). IEEE.
23. Zhang, H., Xu, L., Cheng, X., Chao, K., & Zhao, X. (2018, September). Analysis and prediction of employee turnover characteristics based on machine learning. In *2018 18th International Symposium on Communications and Information Technologies (ISCIT)* (pp. 371-376). IEEE.