Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 12 NO: 2 PAGE:344-353 YEAR: 2023
DOI:10.17798/bitlisfen.1196174

Predicting the Work-Life Balance of Employees Based on the Ensemble Learning Method

Vedat TÜMEN[1*], Ayşe Saliha SUNAR[1]

[1]*Bitlis Eren University, Faculty of Engineering and Architecture, Department of Computer Engineering (ORCID: 0000-0003-0271-216X) (ORCID: 0000-0002-0836-5616)*

Keyword: Work Life Balance, Ensemble learning, Extreme Gradient Boosting, Random forest.

Abstract
Working life has a great impact on other areas of people's lives. Efforts made at work lead to attrition, exhaustion, and health problems. Employers need to take the necessary steps to keep employees motivated by helping them balance work and personal lives. Employers can use many different techniques to measure and analyse the work-life balance of their employees, such as questionnaires and machine learning techniques. This research was conducted to group workers based on turnover levels using effort and work-life balance parameters. Machine learning, including ensemble learning techniques, is used to achieve this. One ensemble learning algorithm, Random Forest, performed almost as well as Support Vector Machine with the highest score of 95%. Almost all algorithms, regardless of whether they are part of ensemble learning or not, achieved an f-score of 86%. However, one of the ensemble learning models, xGBoost, performed poorly with the lowest f-score of 69%. All algorithms predicted the lowest and highest work-life balance scores but were confused when predicting the middle scores (class 2 and class 3).

1. Introduction

While it is a challenge in the labour market to find a suitable job or a competent employee, one of the biggest problems after recruitment is the attrition rate. This problem affects both employees and employers, and it is not a sector-specific problem. Whether it is government or private institutions in various sectors, from medicine [1] to education [2, 3] to engineering [4], attrition rate is considered a problem. Researchers therefore examined the problem from different angles, e.g. private government institutions [5], employee age [6, 7], maternity leave [8], and so on.

Evidence from the literature shows that many factors can lead to a high attrition rate. For example, researchers from India found that for engineers, salary is the most important factor affecting attrition rates, while for non-engineers, the most important factor is the boss, which is made up of boss, stress and salary factors [9]. In [10], the causes of turnover were analysed in depth using both contextual factors (gender, generation, tenure) and 12 human resource factors, including better career opportunities, enriching work content, reward and recognition, and work-life balance. The authors also found that gender and work-life balance are positively correlated. This is also noted by [4] in their study. The respondents identified the problem of work-life balance and categorised it into lack of personal life, long working hours and work pressure, shift hours and night shifts, and insufficient holidays and lack of national holidays. Nevertheless, work-life balance is becoming more and more important in our lives, as one of the consequences of the COVID -19 pandemic is remote working with flexible working hours, which could make it more difficult to balance private and professional life [11-13].

These identified characteristics, including work-life balance, help researchers predict attrition before it occurs so employers can be aware of the problem to make preventative decisions. Machine learning techniques are often used to predict attrition [14]. Some research uses regression or tree-based models to determine which features best predict

*Corresponding Author: vtumen@beu.edu.tr
Received: 31.10.2022, Accepted: 17.04.2023

turnover. For example, [15] uses XGBoost and [16] uses a regression model. Apart from that, [17] compares a number of machine learning algorithms and finds that the decision tree model has the best predictive model.

In this study, ensemble learning algorithms were used to evaluate, compare and optimise performance metrics. This method, in which predictions are made by using many different model algorithms, has been widely used in the field of artificial intelligence for a variety of applications in recent years [18]. Ensemble learning algorithms produce successful results in various machine learning systems, such as feature extraction, confidence estimation, error correction, unbalanced data, and learning the variance in non-stationary distributions. Ensemble learning is a method of combining multiple models and usually improves the predictive performance of the model and increases confidence in the decision to be made [19].

The ensemble learning method has led to successful results in many different fields. For example, in health science, machine learning techniques were used to diagnose Alzheimer's disease [20], and in sports science, machine learning techniques were used to determine a position based on the performance of football players [21], and high success was achieved. To predict the presence of salmonella in agricultural surface waters [22], a study [23] was conducted to estimate how many sensor nodes are needed to install sensors in the field of wireless sensor networks and how large the distance between nodes should be examined [24].

How machine learning can predict effective work schedules and patterns that increase worker productivity. Estimation results were evaluated using the Random Forest Classifier, SVM, and Naïve Bayes algorithms. The algorithms estimated WLB with a best accuracy of 71.5%. The selectKbest algorithm was used to examine the factors that influence the subjective feeling of work-life balance of 800 workers in a country. In the tests, it predicted WLB with 81% accuracy based on the prepared data set and the selected characteristics [25].

While these studies predict turnover based on parameters that include work-life balance, in this study we predict work-life balance values based on some indicative characteristics from the data collected from the workers at IBM. The aim of this study is to use ensemble learning to predict the work-life balance of employees in a company.

2. Dataset

The dataset used is a freely available Kaggle dataset, namely the IBM HR Analytics Employee Attrition & Performance Dataset (https://www.kaggle.com/ datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset IBM HR Analytics Employee Attrition & Performance Dataset). As it is a freely available, anonymous dataset provided by IBM and Kaggle, it is also used as a trusted resource by other researchers for the implementation of the developed models [16].
There are 35 non-null features derived from the responses of the 1470 employees of IBM. The information collected consists of data on demographics, position, salary, and promotion status in the company, as well as personal opinions on job satisfaction and work-life balance. The parameters that impact work-life balance in this 35-attribute database include the attributes of education, satisfaction with work environment, job engagement, job satisfaction, performance evaluation, satisfaction with relationship, and work-life balance, as shown in Table 1.

Table 1. Possible flood flow rates of E26A010 AGI.

Education	Environment Satisfaction	Job Involvement	Job Satisfaction	Performance Rating	Relationship Satisfaction	Work Life Balance
Below College College Bachelor Master Doctor	Low / Medium / High / Very High	Low / Medium / High / Very High	Low / Medium / High / Very High	Low / Medium / High / Very High	Low / Medium / High / Very High	bad / good / better / the best

3. Methodology

This study takes a quantitative research approach to quantitatively analyse the results of the IBM employee survey by applying machine learning techniques to predict work-life balance. To perform the basic machine learning and ensemble learning algorithms, the Python language and its libraries were used.

The questionnaire asks employees to rank their work-life balance in one of four categories: bad, good, better, and best. Apart from these questions, the

IBM survey asks employees about demographic information, income, distance from home, job satisfaction, environment satisfaction, business travel, working hours, etc.

The algorithms used are divided into two sets: 80% for training and 20% for testing. They are trained on the parameters contained in the answers to the questions in order to categorize workers based on the answers to the work-life balance question.

Figure 1. Methodological approach taken.

3.1. Ensemble Learning

Researchers and developers are searching for the most appropriate algorithm to achieve more accurate prediction results for classification and regression problems. One of the approaches in this search is ensemble learning, which combines the predictions of multiple machine learning models to achieve better predictive performance. Ensemble learning aims to create multiple classifiers with similar bias (a systematic error that occurs due to incorrect assumptions in the ML process) and reduce the variance by combining their results [18]. Although an unlimited number of ensemble learning models can be developed, there are three main families of ensemble learning methods: Bagging, Stacking and Boosting [19].

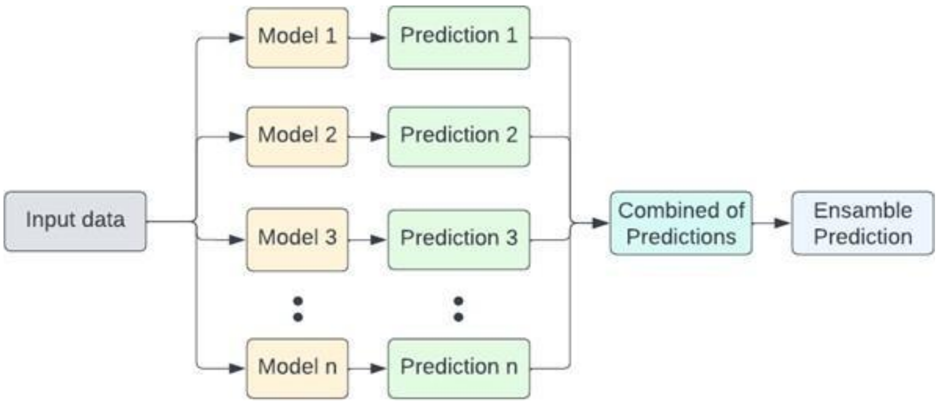

Figure 2. Proposed Ensemble Learning Method.

Bagging fits many decision trees to different samples of the same data set and averages the predictions, i.e., packed decision trees, Random Forest [26]. Stacking applies many different models to the same data and uses a different model for learning to find the best combination of predictions, i.e. voting, weighted average. Boosting uses the prediction errors of the previous models and adds models to the ensemble one by one so that the last model corrects the predictions of the previous models and outputs a weighted average of the predictions [26], e.g., AdaBoost, XGBoost.

In this study, we used the ensemble learning techniques k-nearest neighbours (kNN), random forest (RF), gradient boosting (GB), and extreme gradient boosting (XGBoost) in comparison to the decision tree (DT) and support vector machine (SVM) models as non-ensemble learning techniques.

4. Analysis and Result

Since the aim of the study is to predict work-life balance using ensemble learning models, Figure 4 shows the distribution of workers according to the degree of work-life balance. The classification was divided into four categories. The classification parameters obtained are shown in Table 2.

Table 2. Classification results.

Classes	Precision	Recall	F1-score	Accuracy	Data Size
Bad	0.96	1.00	0.98	0.87	173
Good	0.79	0.89	0.84		184
Better	0.84	0.56	0.67		168
Best	0.88	1.00	0.94		190

Table 2 shows that the average accuracy value is 0.87. It can be seen that the highest success parameters are in the "Poor" category (Precision:0.96, Recall:1.00, f-score:0.98). The lowest results were obtained in the "Better" class with 168 data.
The results of the ensemble model are shown in the rest of the section, and then they are compared at the end.

4.1. Imbalanced Data Problem

The distribution of employees according to work-life balance classes is uneven. Class 3 has the highest proportion at over 60.75%, while Class 1 has the lowest proportion at almost 5.44%. This may result in the minority class being left out. To solve this problem, we need to balance the data. We used the random oversampling method [27] as one of the methods to manipulate the data. In this method, the minority class data is duplicated until the distribution of the classes is balanced. After redistribution of the data, the data is split into a test and a training data set at a rate of 20% and 80%, respectively.

4.2. k-Nearest Neighbor Method

The kNN algorithm classifies similar things in close proximity. It is an ensemble learning algorithm on a single basis. k stands for the distance between neighbors. The calculation of the distance can vary, but Euclidean distance is the most commonly used method. To determine the number k, we plotted the test error rate. The kNN algorithm is one of the sample-based classification methods based on the different distance measures (Euclidean distance, Manhattan distance, and Minkowski distance) of the samples in the known class dataset. It is usually calculated using the Euclidean distance given in Equation (1):

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Figure 3 shows the k in the range 1 to 30 and the error rate. The k for this model is determined as 1.

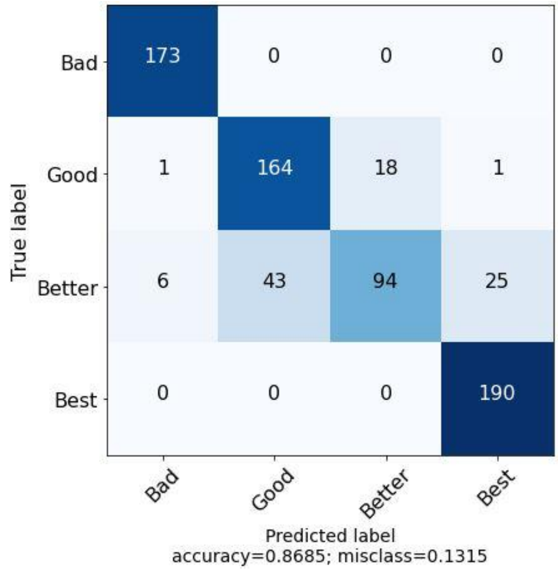

Figure 3. The result confusion matrix of kNN.

Figure 3 shows the heat map of the predicted and true labels. The figure shows that the algorithm correctly classifies between "bad' and 'best" in most cases, but confuses some between "good' and "better". The accuracy rate of the kNN algorithm was found to be 86.85%.

4.3. Support Vector Machine Method

Founded in 1963 by Vladimir Vapnik and Alexey Chervonenkis, the Support Vector Machine (SVM) is one of the supervised learning methods generally used in classification problems. The SVM algorithm draws a line to separate points lying on a plane by linear or non-linear methods. It is an ensemble learning algorithm on a single basis. These separations are to achieve the maximum distance between the points of the two classes [28]. The goal of the SVM algorithm is to maximize the distance between the data to be separated. Figure 3 shows the operating principle of the SVM algorithm.

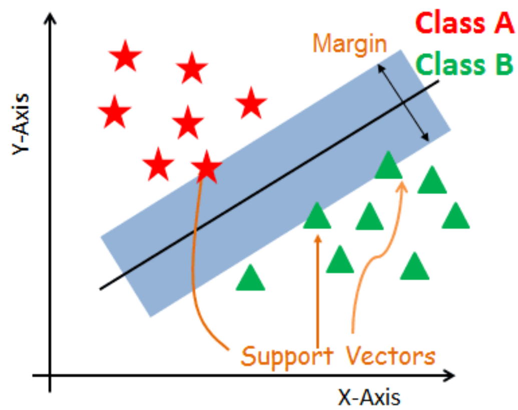

Figure 4. The goal of the SVM algorithm [29].

Looking at Figure 4, the lines are drawn using linear or non-linear methods based on the furthest data in the clustered data with different values. By optimizing these lines, the most distant lines and the lines closest to the data are selected and a new line is drawn over which the average is taken and the separation process is performed. Parameters of the SVM algorithm: The gamma value was set at 0.001 and the class number at 4. The result of the classification process with the SVM algorithm is shown in Figure 5 as a confusion matrix.

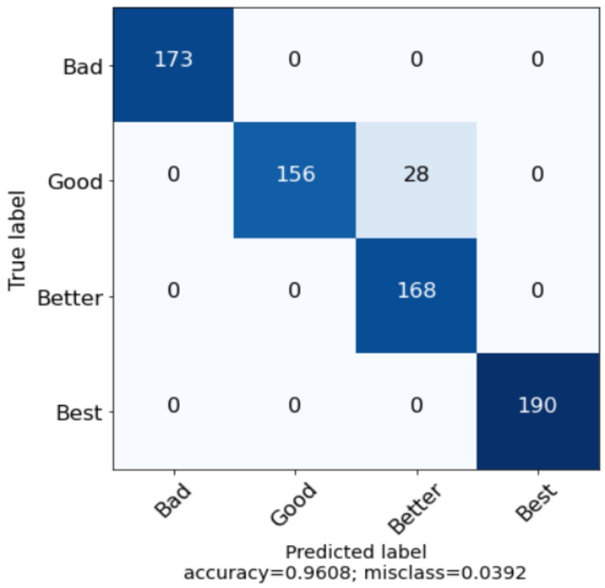

Figure 5. The confusion matrix result of SVM algorithm.

Figure 5 shows that all the data in the 'Bad', 'Better', and 'Best' classes were correctly classified, while 28 of the 'Good' data were incorrectly classified as 'Better'. Figure 5 shows the heat map of the predicted and true labels. The accuracy rate of the SVM algorithm was found to be 96.08%.

4.4. Random Forest

The RF model is one of the most widely used ensemble learning methods that use a set of decision trees and find the most optimal combination from the results of the decision trees. The RF method consists of compilations of the classification tree or the RF tree, depending on the purpose, as many as the number of trees to be created. Therefore, one of the most commonly used algorithms among ensemble methods is RF [30]. In Random Forest, the estimator is 100 and the random value is 101. In addition, the value for the maximum features is set to auto. Since the algorithm randomly selects estimators, the accuracy of the model is higher because less correlation is achieved between the relationships formed in a data set [31].

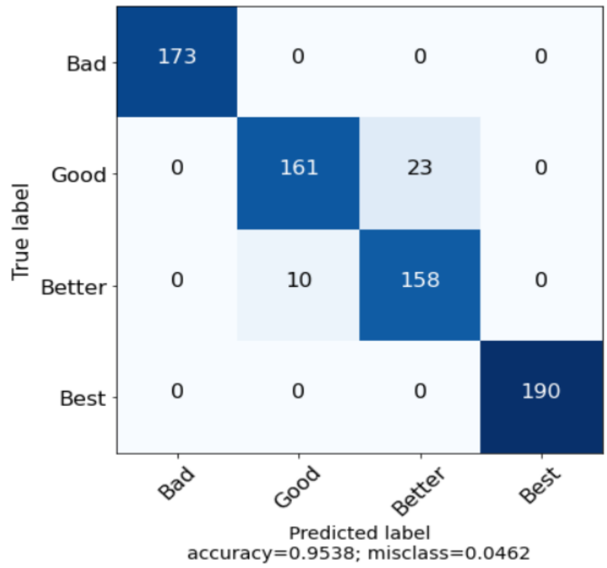

Figure 6. The confusion matrix result of Random Forest method.

Figure 6 shows that the model predicts 'bad' and 'best" perfectly but confuses "good" and "better". The accuracy rate of the random forest method was found to be 95.38%.

4.5. Decision Tree Algorithm

Results were obtained using the Decision tree method, and the comparison matrix is shown in Figure 7.

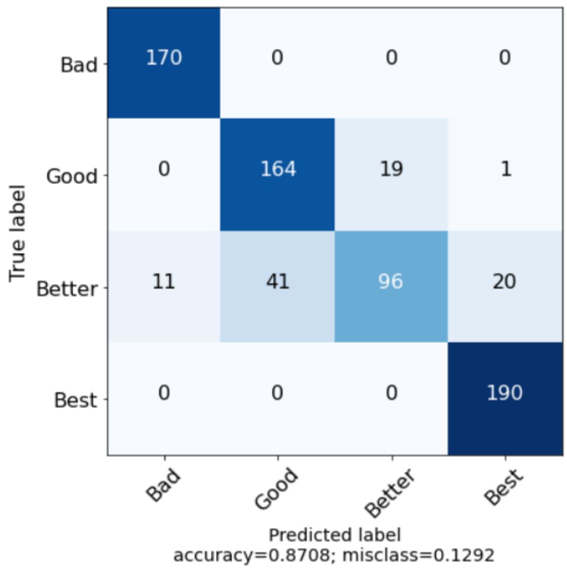

Figure 7. The confusion matrix result of Decision Tree method.

Figure 7 shows that the model predicts 'bad' and 'best" perfectly but confuses "good" and "better". For the "better" data, the model DT showed high mixing (error). The accuracy rate of the decision tree method was found to be 87.08%.

4.6. Gradient Boosting

The gradient boosting method (GBM) is a technique used to solve classification and regression problems and it is a very successful model used to combine several models whose individual performance can be considered poor [32]. The GBM is one of the very popular boosting examples for ensemble learning models. It uses Loos functions, weak learners (decision trees), and an additive model to improve performance over basic algorithms. The parameters of Gradient Boosting are listed in Table 3.

Table 3. The parameters of Gradient Boosting.

Parameters	Value
Number of estimators	100
Max features	0.9
Learning rate	0.3
Depth	4
Minimum samples leaf	2
Sub sample	1

Figure 8 shows the predictions made by the Gradient Boosting method.

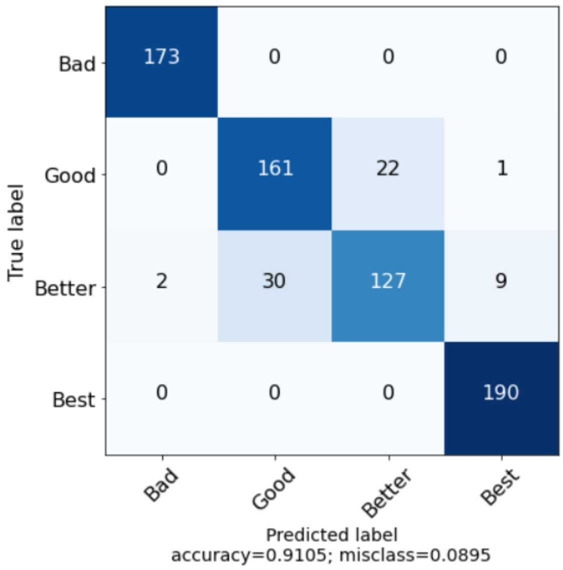

Figure 8. The confusion matrix result of Gradient Boosting method.

4.7. Light Gradient Boosting

Light Gradient Boosting (LGBoost) is a based algorithm that offers leaf-shaped growth with depth constraints to speed up the training process and reduce memory consumption. The structure produces a histogram of width k by separating k bins of continuous floating point eigenvalues. The parameters of LGBoost are set to the values 0.9 for the bagging portion, 0.9 for the feature portion, and 0.2 for the reg lambda [33]. Figure 9 shows the predictions made by the Light Gradient Boosting method.

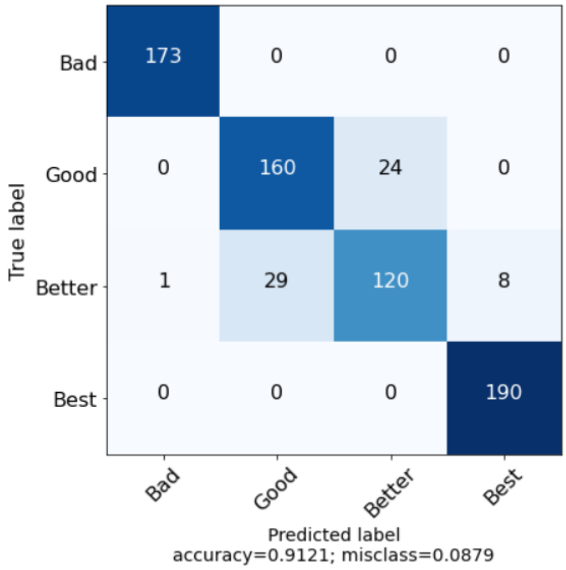

Figure 9. The confusion matrix result of Light Gradient Boosting.

4.8. Comparison of the Results

Table 4 shows the precision, recall, f-score, and accuracy of each algorithm, including the decision tree algorithm and the support vector machine model, for comparison with the models without ensemble learning. DT models propose a tree-like model of decisions with conditional control statements. The SVM algorithm finds a hyperplane in a dimensional space that provides a unique classification of the data. The dimension is the number of features. The algorithm is quite successful in terms of accuracy, misclassifying only some class 2 members as class 3.

Table 4. Classification performance results of algorithms.

Models	Precision	Recall	f-Score	Accuracy
Decision Tree	0.87	0.87	0.86	0.87
SVM	0.97	0.96	0.96	0.96
kNN	0.87	0.86	0.86	0.87
RF	0.96	0.95	0.95	0.95
GBM	0.91	0.91	0.91	0.91
LightGBM	0.91	0.91	0.91	0.91
Average	**0.92**	**0.91**	**0.91**	**0.91**

To evaluate the performance of a classification algorithm, there are a number of metrics. Among them, accuracy and f-score are the most commonly used. Accuracy simply calculates the proportion of correctly categorised examples out of the total number of examples [34]. However, it does not take into account the unbalanced categories or the false negative and false positive examples. On the other hand, the f-score is calculated as the harmonic mean of the precision and recall of the model, which takes into account the proportion of false-positive and false-negative examples. All performance values of the algorithms used in this study are listed in Table 3. Shown are the average performance values of the algorithms used in the study. The average accuracy is 91%.

5. Discussion

In this study, the SVM algorithm achieved the best result, while the k-NN and DT algorithms achieved the worst result. While the RF method achieved high success with an accuracy of 0.95, the ensemble models studied showed lower success. Although the performance values of the models differed fundamentally, their results were close to each other and showed similarity.

5.1. Benefits of Proposed Study

In this proposed study, information about the work-life balance of employees at a workplace was analyzed using machine learning techniques, and conclusions were drawn. This result can be used to evaluate employee satisfaction in any workplace. In addition, other benefits are mentioned below.

- With the help of artificial intelligence techniques, it has been possible to determine work-life balance with great success.
- The proposed method can be used to determine employee satisfaction in a workplace with high efficiency without manager intervention.
- The managers of workplaces can increase the efficiency of their employees by considering the results of this method.
- Considering this study, the performance of monthly or yearly employees can be measured again with this study and the changes can be evaluated.
- With an account to be opened at Google Colab, the results of this study can be accessed free of charge by providing the data.

6. Conclusion

The random forest model as an ensemble of decision trees outperforms the decision tree model by 9%. RF also provides the best results among the ensemble models. Furthermore, the boosting ensemble models (GBM, LightGBM) show relatively poor performance. The reasons for these results are the small number of data, the inconsistency of some relationships between values, and the large imbalance between classes. Table 5 shows the average result of this study and the results of similar studies in the literature.

Table 5. Comparison of own method with the literature.

Models	Accuracy
Radha et all.[24]	*71.50*
Pawlicka et all. [25]	*81.00*
Toğaçar [36]	*87.76*
Mansor [37]	*84.59*
Our Method	*91.05%*

In the [24] study, a total of 12,756 individuals' work-life balance data were used with different machine learning techniques, and the highest success rate (SVM algorithm) was 71.50%. In the [25] study, the artificial neural network method was used to measure the work-life balance of 800 people (Male: 389, Female: 411) in a workplace, and a success rate of 81.00% was achieved. The average success rate of this study was found to be 91.05%. There are many similar studies in the literature. A large amount of data was used in these studies. The study examined in [36] found 87.66% success for work-life balance, while [37] found 84.59% success. Although we used a small number of data in our study, better results were obtained. Although the amount of data was insufficient and had low quality features, the method we proposed was quite successful.

Contributions of the Authors

The authors confirm that the contribution is equally for this paper.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] S. M. Crow, S. A. Smith, and S. J. Hartman, "Attrition in nursing: Perspectives from the national survey of college graduates," *Health Care Manag. (Frederick)*, vol. 24, no. 4, pp. 336–346, 2005.

[2] W. P. Durow and B. Brock, "The retention and attrition of catholic school principals," *Journal of Catholic Education*, vol. 8, no. 2, 2004.

[3] P. Weldon, "Early career teacher attrition in Australia: evidence, definition, classification and measurement," *Aust. J. Educ.*, vol. 62, no. 1, pp. 61–78, 2018.

[4] N. Pandey and G. Kaur, "Factors influencing employee attrition in Indian ITeS call centres," *Int. J. Indian Cult. Bus. Manage.*, vol. 4, no. 4, p. 419, 2011.

[5] L. Kanteh and A. Gibba, "A study on employees' attrition in public and private institutions in the Gambia, 2007-2017," *Arabian J Bus Manager Review*, 2007.

[6] D. Kennedy, "Attrition rates of mature engineers," *Eng. Manag. J.*, vol. 18, no. 3, pp. 36–40, 2006.

[7] C. Marcelo, *Does Employee Training and Development Programs Mitigate Attrition in the Millennial Workforce?* 2022.

[8] N. Aariya, "A study on the impact of maternity benefit on teacher's career and increase in post maternity attrition rate in India," *Journal for Educators, Teachers and Trainers*, vol. 13, no. 2, pp. 207–214, 2022.

[9] S. Bhardwaj and A. Singh, "Factors affecting employee attrition among engineers and non-engineers in manufacturing industry," *Bus. IT*, vol. VII, no. 2, pp. 26–34, 2017.

[10] S. Barpanda and S. Athira, "Cause of Attrition in an Information Technology-Enabled Services Company: A Triangulation Approach," *International Journal of Human Capital and Information Technology Professionals (IJHCITP)*, vol. 13, no. 1, pp. 1–22, 2022.

[11] D. Ayar, M. A. Karaman, and R. Karaman, "Work-life balance and mental health needs of health professionals during COVID-19 pandemic in turkey," *Int. J. Ment. Health Addict.*, vol. 20, no. 1, pp. 639–655, 2022.

[12] C. Calderwood, R. Breaux, L. L. Ten Brummelhuis, T. Mitropoulos, and C. S. Swanson, "When daily challenges become too much during COVID-19: Implications of family and work demands for work-life balance among parents of children with special needs," *J. Occup. Health Psychol.*, vol. 27, no. 5, pp. 516–527, 2022.

[13] L. Vyas, "New normal" at work in a post-COVID world: work-life balance and labor markets," *Policy and Society*, vol. 41, no. 1, pp. 155–167, 2022.

[14] K. K. Mohbey, "Employee's attrition prediction using machine learning approaches," in *Machine Learning and Deep Learning in Real-Time Applications*, IGI Global, 2020, pp. 121–128.

[15] R. Jain and A. Nayyar, "Predicting employee attrition using XGBoost machine learning approach," in *2018 International Conference on System Modeling & Advancement in Research Trends (SMART)*, 2018.

[16] S. Najafi-Zangeneh, N. Shams-Gharneh, A. Arjomandi-Nezhad, and S. Hashemkhani Zolfani, "An improved machine learning-based employees attrition prediction framework with emphasis on feature selection," *Mathematics*, vol. 9, no. 11, p. 1226, 2021.

[17] N. El-Rayes, M. Fang, M. Smith, and S. M. Taylor, "Predicting employee attrition using tree-based models," *Int. J. Organ. Anal.*, vol. ahead-of-print, no. ahead-of-print, 2020.

[18] C. Zhang and Y. Ma, Eds., *Ensemble machine learning: methods and applications*. Springer Science & Business Media, 2012.

[19] R. Polikar, "Ensemble Learning," in *Ensemble Machine Learning*, Boston, MA: Springer US, 2012, pp. 1–34.

[20] S. Buyrukoglu, "Improvement of Machine Learning Models' Performances based on Ensemble Learning for the detection of Alzheimer Disease," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 2021.

[21] S. Buyrukoğlu and S. Savaş, "Stacked-based ensemble machine learning model for positioning footballer," *Arab. J. Sci. Eng.*, 2022.

[22] S. Buyrukoğlu, "New hybrid data mining model for prediction of Salmonella presence in agricultural waters based on ensemble feature selection and machine learning algorithms," *J. Food Saf.*, vol. 41, no. 4, 2021.

[23] A. Akbas and S. Buyrukoglu, "Stacking Ensemble Learning-Based Wireless Sensor Network Deployment Parameter Estimation," *Arabian Journal for Science and Engineering*, pp. 1–10, 2022.

[24] K. Radha and M. Rohith, "An experimental analysis of work-life balance among the employees using machine learning classifiers," *Int. J. Comput. Trends Technol.*, vol. 69, no. 4, pp. 39–48, 2021.

[25] A. Pawlicka, M. Pawlicki, R. Tomaszewska, M. Choraś, and R. Gerlach, "Innovative machine learning approach and evaluation campaign for predicting the subjective feeling of work-life balance among employees," *PLoS One*, vol. 15, no. 5, p. e0232771, 2020.

[26] J. Brownlee, "Ensemble Learning Algorithms With Python: Make Better Prediction with Bagging, Boosting, and Stacking," *Machine Learning Mastery*, 2021.

[27] J. Brownlee, "A Gentle Introduction to Ensemble Learning Algorithms," *Machine Learning Mastery*, 2021.

[28] X. Wang, B. Yu, A. Ma, C. Chen, B. Liu, and Q. Ma, "Protein-protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique," *Bioinformatics*, vol. 35, no. 14, pp. 2395–2402, 2019.

[29] A. Navlani, "Support vector machine classification in scikit-learn," *Medium*, 16-Aug-2020. [Online]. Available: https://avinashnavlani.medium.com/support-vector-machine-classification-in-scikit-learn-3800bc4979ce. [Accessed: 05-Mar-2023].

[30] S. Suthaharan, "Big Data Essentials," in *Machine Learning Models and Algorithms for Big Data Classification*, Boston, MA: Springer US, 2016, pp. 17–29.

[31] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.

[32] B. Suchetana, B. Rajagopalan, and J. Silverstein, "Assessment of wastewater treatment facility compliance with decreasing ammonia discharge limits using a regression tree model," *Sci. Total Environ.*, vol. 598, pp. 249–257, 2017.

[33] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, 2013.

[34] D. Chakraborty, H. Elhegazy, H. Elzarka, and L. Gutierrez, "A novel construction cost prediction model using hybrid natural and light gradient boosting," *Adv. Eng. Inform.*, vol. 46, no. 101201, p. 101201, 2020.

[35] M. Toğaçar, B. Ergen, and V. Tümen, "Use of dominant activations obtained by processing OCT images with the CNNs and slime mold method in retinal disease detection," *Biocybern. Biomed. Eng.*, vol. 42, no. 2, pp. 646–666, 2022.

[36] N. Mansor, N. S. Sani, and M. Aliff, "Machine learning for predicting employee attrition," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, 2021.

[37] N. Darapaneni *et al.*, "A detailed analysis of AI models for predicting employee attrition risk," in *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, 2022.