IJEMD

IJEMD-CSAI, 3 (1) (2024), 1 – 9 https://doi.org/10.54938/ijemdcsai.2024.03.1.352

International Journal of Emerging Multidisciplinaries:
Computer Science and Artificial Intelligence

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2791-0164 ISSN (online): 2957-5036

An Intelligent Analysis and Prediction of Employee Attrition Rate in Healthcare Using Machine Learning Techniques

Egwom .O. Jessica[1]*, Lawrence, Emmanuel[1], Moshood A. Hambali[1], Kefas Rimamnuskeb Galadima[1]

1. Department of Computer Science, Federal University Wukari, PMB 1020 Katsina-Ala Road, Wukari, Taraba State, Nigeria.

ABSTRACT

In the increasingly competitive business environment of the twenty-first century, organizations try to retain their most valuable employees. In the healthcare sector especially, Attrition rates among employees are being affected by a new situation that has arisen as a result of the COVID-19 pandemic and remote working scenarios. Attrition refers to the process by which employees depart from a business for any reason, including their own choice of leave. The application of artificial intelligence (AI) in the past for business expansion and advancement has started influencing human resource management. Improvements to models that predict employee attrition are a constant goal of machine learning (ML) researchers. These models can predict the results of employee turnover, but they have one major drawback: they are mostly used on datasets on attrition in non-healthcare settings. In this study, we offer an exploratory study of using machine learning methodologies to predict employee attrition in healthcare, using the SVM, KNN, XGBoost, RF, and SMOTETOMEK resampling techniques to make healthcare attrition prediction more accurate using the Kaggle dataset of healthcare employees. This gave the healthcare sector a chance to change the narration of cause's employees leaving the system. The effectiveness of the model was assessed using the area under the curve (ROC), accuracy, recall, and precision metrics. The findings indicate that the SMOTETOMEK Random Forest model exhibits a 98.0% accuracy rate, outperforming the other classification models.

Keywords: Employee attrition, Healthcare, Machine learning, Random Forest, K-nearest Neighbor, XGBoost, Support Vector Machine, SMOTETOmek.

Email Addresses: Egwom@fuwukari.edu.ng (Jessica), Lawrence@fuwukari.edu.ng (Lawrence), hambali@fuwukari.edu.ng (Hambali), kefasmanu3@gmail.com (Kefas)

INTRODUCTION

There is a wealth of talent and skill among individuals from all walks of life, and the world is full of opportunities. Regardless of age or demographic differences, this ability can be found in people from all walks of life. Since employees have factors to consider when choosing and working for an organization, and they have the option to resign if they are dissatisfied, it is equally important for organizations to satisfy and retain such talent as it is to recruit it [1]. Employee attrition is a significant concern in HR analytics, as companies invest heavily in training for long-term advantages [2]. When an employee leaves an organization, it results in a loss of potential earnings and opportunity costs for the organization. Attrition can be classified into involuntary, external, and internal attrition. Involuntary attrition occurs when an employer severs ties with a worker, while external attrition occurs when one employee chooses to leave one organization in favor of another [2]. Internal attrition occurs when a person is promoted to a different position within the same organization. The rate of employee attrition is an important indicator of an organization's level of development. High attrition rates are evident in the healthcare sector, making it difficult for governments to achieve universal health coverage and equitable objectives [3]. High attrition rates lead to financial constraints related to recruiting, training, and the loss of valuable knowledge, as well as operational and patient care problems. According to a recent analysis by the World Health Organization (WHO), it is predicted that by 2030 there will be a shortage of about 18 million healthcare professionals in low and lower-income countries such as Nigeria [4]. This shortage of staff is expected to impede the achievement of sustainable development goals [5]. The replacement costs for entry-level positions may be lower, but for higher-level positions in healthcare, they will likely be significantly higher. Accurately predicting attrition in health with the key factors that impact this decision using an employee's customized data available online is necessary to alleviate this problem [1]. Utilizing cutting-edge machine learning techniques will be possible. Machine learning is a branch of AI that allows computers to learn from experience and generate predictions based on that learning [6]. These techniques consider attrition factors when training their model with past employee data. The goal is to identify potential employees who are likely to leave the organization by examining historical employee data, demographic information, and performance metrics, and offering valuable insights into the probability of employee attrition. Several studies [7, 8, 9, and 10] have made predictions about employee turnover, but these studies are deficient in at least one significant domain: healthcare. Furthermore, they do not make adequate use of exploratory data analysis or select appropriate preprocessing techniques to address the issue of data imbalance. This is because they are typically applied to datasets on employee attrition in industries that are not related to healthcare. Therefore, this study aims to use SMOTETomek to fix data imbalances and combine it with machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), K nearest neighbor (KNN), and XGBoost to predict employee attrition in the healthcare sector. Our contributions include: (i) To identify potential reasons for employees in health care, we conducted an exploratory analysis of the dataset. (ii) To predict employee attrition in healthcare using a variety of machine learning models. We evaluate each model based on its performance and interpretability.

The paper is structured into five sections which are as follows: An overview of the research is provided in Section 1. Section 2 presents related research works on employee attrition prediction. The research methodology and materials are covered in Section 3. The results are presented in Section 4, and finally, Section 5 provides a conclusion, summary, and critique of the findings.

REVIEW OF RELATED LITERATURE

The healthcare sector has a high attrition rate, and identifying the factors causing the high employee attrition is an important first step in addressing the issue in the healthcare sector. According to Frye et al. [11,13,12], the main factors why hospital employees leave their jobs are excessive stress, an ever-increasing workload, low job satisfaction, poor working environments, no room for advancement, low pay, and inadequate compensation.

In recent times, the healthcare sector has been trying to determine what makes workers leave their jobs so they can lower employee attrition rates [13]. Human resources (HR) have profited substantially from applying machine learning techniques since these methods can significantly increase employee retention and decrease attrition rates [14, 15, 16, 17, 18]. Over time, different researchers have carried out different research work to help HR managers make accurate predictions on employee attrition in healthcare and other business organizations with most ML techniques having high prediction performance [8, 14, 18, 19]. In their study on employee attrition factors, [9] used several different classifiers and regression models to draw their conclusions. These models included logistic regression, K-NN, decision tree, random forest, support vector machine, and linear support vector machine. The SMOTE method was used to resample the data. Recall was the most important performance indicator, and the study indicated that 51 out of 71 employees' departures were accurately predicted using Gaussian naive Bayes. Improved classification accuracy, precision, and true positive rate with minimized error rates were achieved by [20] through the use of feature selection based on machine learning classifiers. Six classifiers were tested: decision trees, random forests, artificial neural networks, support vector machines, gradient boosting trees, and banning. The results demonstrated that the accuracy of attrition classification was much enhanced when the gradient-boosting tree classifier was used with the chi-square feature selection. [21] used logistic regression to predict employee turnover using three different feature selection methods: information gain, pick k-best and recursive feature elimination. They used 10-fold cross-validation for evaluation. The AUC score for forecasting employee attrition with a logistic regression model was 0.932, with an accuracy of 0.865. [22] utilized various machine learning models to predict employee attrition accurately using IBM's attrition dataset. They trained and evaluated models like Decision Tree, Random Forest Regressor, Logistic Regressor, Adaboost Model, and Gradient Boosting Classifier. The best accuracy, recall, and AUC were achieved using logistic regression and decision tree ensemble techniques. Another study [23] compared three machine learning classifiers: Decision Trees (DT), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) on the IBM Human Resource Analytic Employee Attrition and Performance dataset. They used preprocessing techniques and regularization to combat overfitting. The optimized SVM model achieved the highest accuracy (88.87%) in forecasting employee turnover, followed by ANN and DT. However, the accuracy was and could be improved using different data resampling techniques. [24] Conducts a study to analyze organizational factors causing employee attrition and predict it using machine learning approaches. Four methods were used: Extra Tree Classifier (ETC), Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree Classifier (DTC). The dataset was resampled using the synthetic minority oversampling technique (SMOTE) for a more even distribution. Hyperparameter tuning and K-Fold cross-validation were used for model accuracy. Performance evaluation scores were compared and compared. The improved ETC approach had a 93% predictive accuracy for employee turnover. A research on employee attrition was carried out by [25] utilizing a variety of datasets, ranging from small, medium, and big. Decision trees, random forests, gradient boosting trees, neural networks, linear discriminant analysis, K-nearest neighbor, Naive Bayes, and linear discriminant analysis were among the machine learning techniques used with a recall of 0.778 on the HR IBM Watson Analytics dataset, XGBoost was determined to be the top machine learning method in the research. [26] Conducted a study on employee attrition to predict employee turnover using the XGBoost Machine Learning algorithm. The model achieved an accuracy of 89% when split into training and testing data. However, no data resampling procedures were used, to minimize these effects, a data resampling approach should have been employed. [8] Used a mixed-methods approach to examine employee turnover using machine learning data from an Iranian pharmaceutical business and interviews with HR managers. 89% was achieved as the accuracy rate of predicting high-risk attrition using a gradient boosting machine algorithm model. In a bid to improve employee attrition prediction [27], an ensemble model based on machine learning algorithms was used, specifically by determining the optimal combination of hyperparameters for the model through the application of support vector machines and artificial neural networks.

MATERIALS AND METHOD

Machine Learning (ML) has reached state-of-the-art performances in many research fields nowadays, from computer vision to bioinformatics, from natural language processing to object detection. In our work we use a set of four well known and robust supervised machine learning algorithms to analyze employee attrition in the health sector. The procedure which we followed to design the predictive model for employee attrition in health are data collection, data cleaning, determining the relevance of features, utilising a data sampling approach to solve the problem of class imbalance, model training and evaluation, and determining the degree to which the model is successful. The model's proposed framework is shown in Figure 3.1. Preprocessing the data included a check for missing values and a simplification of the features utilised in the prediction. The features of the dataset were ranked according to their importance in order to ascertain which ones were most useful for predicting employee attrition. Additionally, SMOTETomek was employed to deal with the imbalanced data issues. After that, the data were separated into a training set and a test set using K-fold cross-validation. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), XGBoost, Random Forest (RF), and SMOTETOMEK were utilised in order to complete the classification process. The . The accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC (ROC) were the performance evaluation that were used to evaluate the predictive efficacy of the model. The actual implementation was done in Python, which was chosen as the programming language of choice.

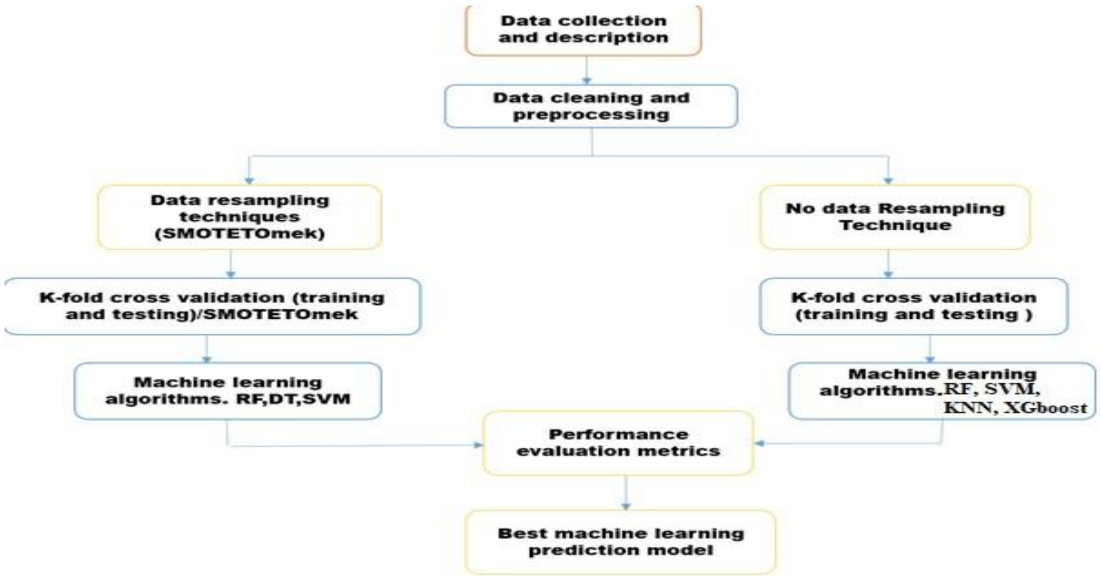

Fig 3.1: Architecture of the proposed model

A. Data collection and description

The Kaggle dataset used for the research was downloaded from https://www.kaggle.com/code/mikimonis/employee-attrition-for-healthcare, which is a synthetic dataset based on the IBM Watson dataset for employee attrition [5] and [22] Employee roles and departments were changed to reflect the health care domain. A total of 1676 samples are available in this publicly available dataset, and its 35 features are relevant to employees' working and personal lives. The dataset has two attrition values (NO,YES), with 1477 employees staying in the organisation (NO) and 199 employees leaving (YES). The goal is to foretell whether or not an employee departed the organisation as indicated by the attribute Attrition (a "NO" or "YES" response, respectively). The features are made up of two data types: integer and string. String data type for sequences of characters, and integer data type for whole numbers only.

B. Model validation

Both the training and the testing of the classifiers relied on the utilisation of k-fold crossvalidation. The model was validated with one-fifth of the data, while the remaining data was utilised for the training phase of the process. Although each run of the Cross-validation procedure used all of the data for training and validation, a fresh one-fifth of the dataset was selected for validation on a per-run basis. Due to the fact that overfitting the training set may easily occur with small datasets, this strategy is utilised in order to prevent it.

C. Classification of the model

Prediction was performed using a Support Vector Machine (SVM), K-Nearest Neighbor (KNN), XGBoost, Random Forest (RF) algorithm with SMOTETOMEK. Each machine learning model was provided with a unique set of hyper parameters. Hyper parameters are those parameters that cannot be determined by the model itself; therefore, they are finetuned manually and given to the model before running them to help obtain the desired accuracy from the model. Previous studies have used the RF, SVM, and DT algorithms for employee attrition, which have shown good performance on the employee attrition datasets [9]; [22]; and [24]. However, in this study we used RF, SVM, K-Nearest Neighbor (KNN), XGBoost and SMOTETOMEK algorithms to perform comparative analysis of the model classification so as to come up with the best model for employee attrition in health care.

D. Performance evaluation metrics

The effectiveness of a classifier is often visualized using a confusion matrix, sometimes called an error matrix, which is a table that shows how well a model can classify labels correctly. This has been used by several researchers such as [9] and [24] to evaluate model performance on employee attrition dataset **Table 3.1** presents the terms used in the confusion matrix for a binary classifier.

TABLE 3.1 Confusion Matrix for Classification

	Actual Negative	**Actual Positive**
Predicted Negative	True Negative(TN)	False Positive(FP)
Predicted Positive	False Negative(FN)	True Positive(TP)

RESULTS AND DISCUSSION

This section presents the results and discussion of the experiments carried out on predictive model for attrition in healthcare. The results in terms of performance accuracy of the predictive model using four different algorithms such as Random Forest, Xgboost, K-nearest Neighbor and Support Vector Machine were presented in Table 4.1a – 4.5 and Fig. 4.2a – 4.5b.

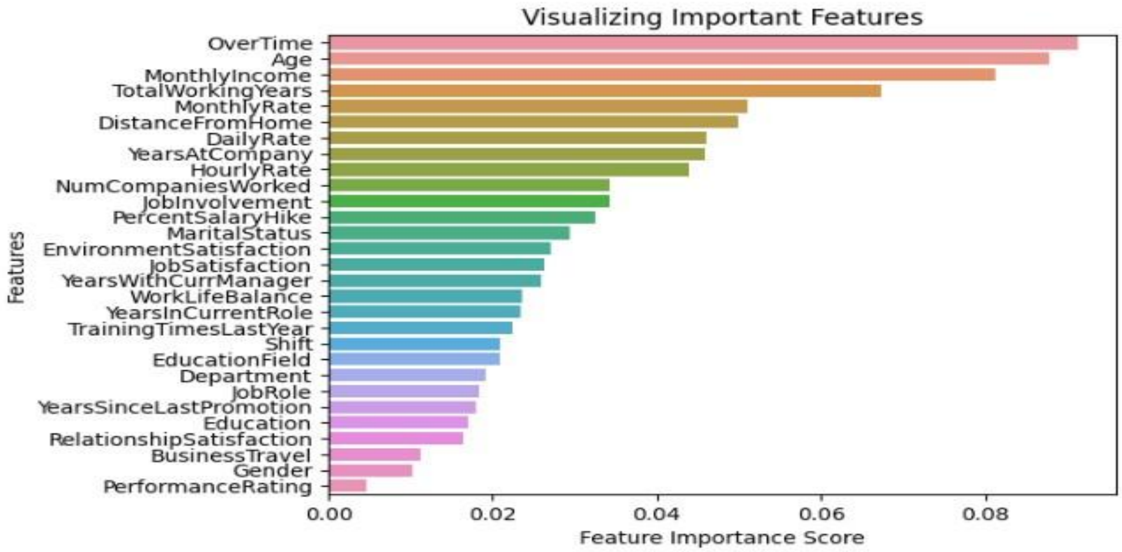

Fig. 4.1: Visualizing Important Features

I. CONFUSION MATRIX FOR RANDOM FOREST ALGORITHM

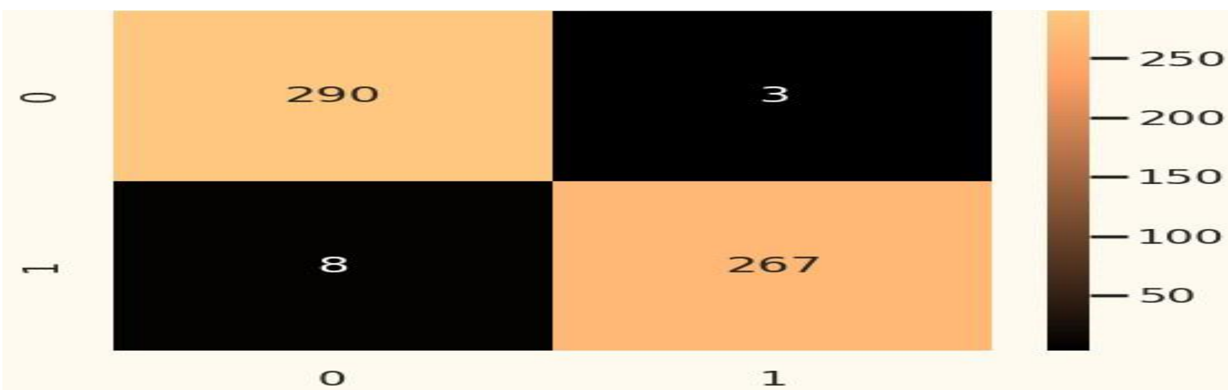

Fig. 4.2a: Confusion matrix for Random Forest Algorithm

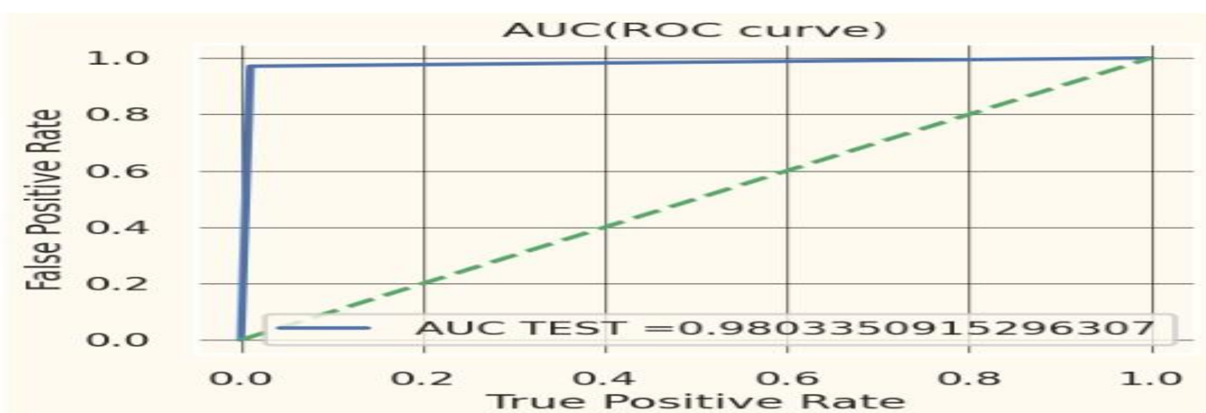

Fig. 4.2b: AUC (ROC curve) for Random Forest Algorithm.

Fig 4.2a represents confusion matrix for Random Forest algorithm. Which contains True Positive, True Negative, False Positive, False Negative. The columns represent predicted NO which is 0 and predicted YES which is 1. In the other hand, the rows represent the actual NO which is 0 and the actual YES which is 1. The classifier made a total number of 568 predictions (that is 568 employees were tested for attrition). Out of those 568 cases, the classifier predicted 270 times for YES and 298 times for NO. True Negative value is greater which shows attrition cases are low.

Table 4.1a: The result for Random Forest Algorithm

	Precision	**Recall**	**F1 score**	**support**
Yes	0.97	0.99	0.98	293
No	0.99	0.97	0.98	275
Accuracy			0.98	568
Macro avg	0.98	0.98	0.98	568
Weighted avg	0.98	0.98	0.98	568

II. CONFUSION MATRIX FOR XGBOOST ALGORITHM

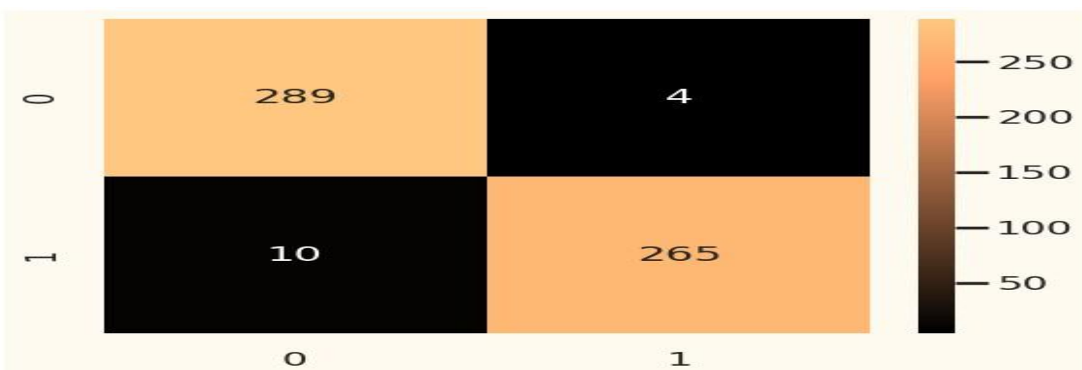

Fig. 4.3a: Confusion matrix for XGBOOST algorithm

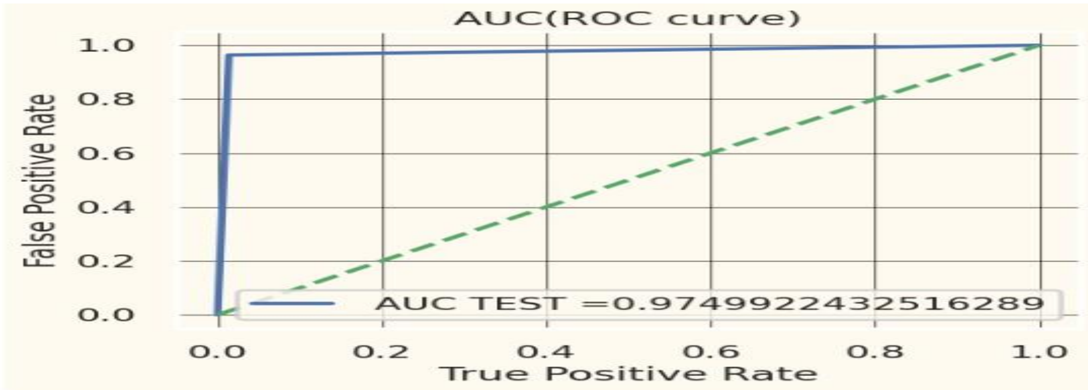

Fig. 4.3b: AUC (ROC curve) for Xgboost algorithm

Fig 4.3a represents confusion matrix for XGBOOST algorithm. Which contains True Positive, True Negative, False Positive, False Negative. The columns represent predicted NO which is 0 and predicted YES which is 1. In the other hand, the rows represent the actual NO which is 0 and the actual YES which is 1. The classifier made a total number of 568 predictions (that is 568 employees were tested for attrition). Out of those 568 cases, the classifier predicted 269 times for YES and 299 times for NO. True Negative value is greater which shows attrition cases are low.

Table 4.2a: The result for XGBOOST Algorithm

	Precision	Recall	F1 score	support
Yes	0.97	0.99	0.98	293
No	0.99	0.96	0.97	275
Accuracy			0.98	568
Macro avg	0.98	0.97	0.98	568
Weighted avg	0.98	0.98	0.98	568

III. CONFUSION MATRIX FOR K-NEAREST NEIGHBOR ALGORITHM

Fig. 4.4a: Confusion matrix for K-nearest Neighbor Algorithm

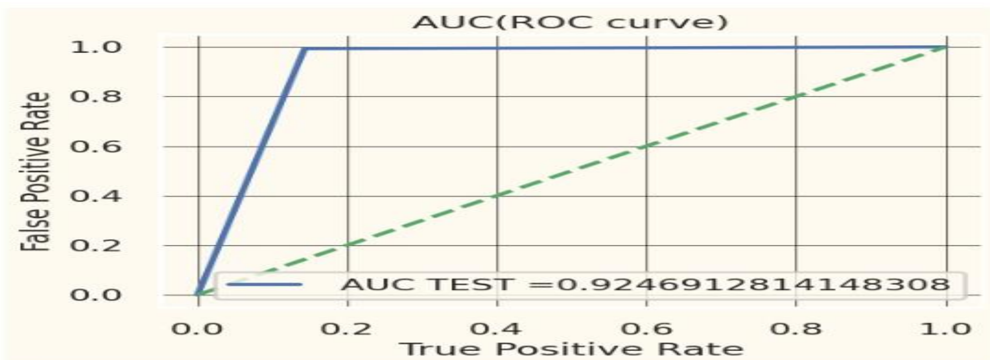

Fig. 4.4b: AUC (ROC curve) for K-nearest Neighbor Algorithm.

Fig 4.4a represents confusion matrix for K-nearest Neighbor algorithm. Which contains True Positive, True Negative, False Positive, False Negative. The columns represent predicted NO which is 0 and predicted YES which is 1. In the other hand, the rows represent the actual NO which is 0 and the actual YES which is 1. The classifier made a total number of 568 predictions (that is 568 employees were tested for attrition). Out of those 568 cases, the classifier predicted 315 times for YES and 253 times for NO. True Positive value is greater which shows attrition cases are high.

Table 4.3a: The result for K-Nearest Neighbor Algorithm

	Precision	Recall	F1 score	support
Yes	0.99	0.86	0.92	293
No	0.87	0.99	0.93	275
Accuracy			0.92	568
Macro avg	0.93	0.92	0.92	568
Weighted avg	0.93	0.92	0.92	568

IV. CONFUSION MATRIX FOR SUPPORT VECTOR MACHINE

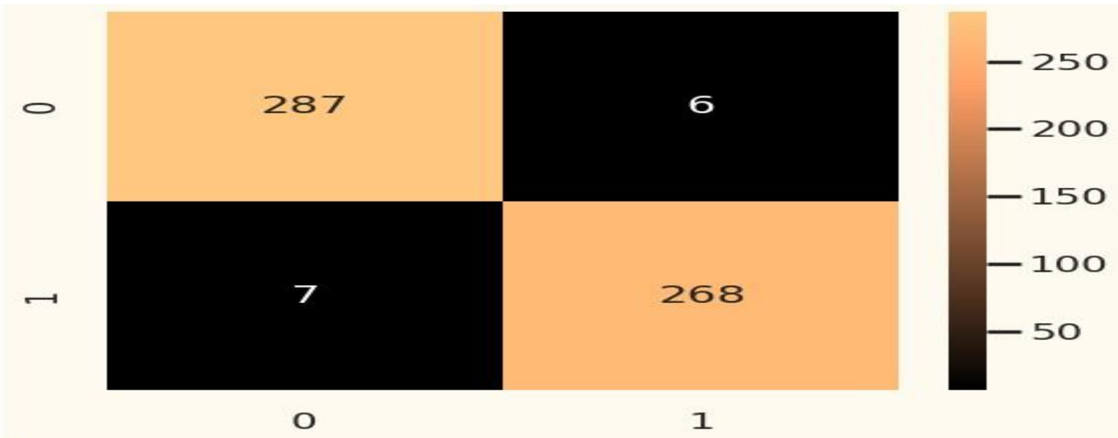

Fig. 4.5a: Confusion matrix for Support Vector Machine

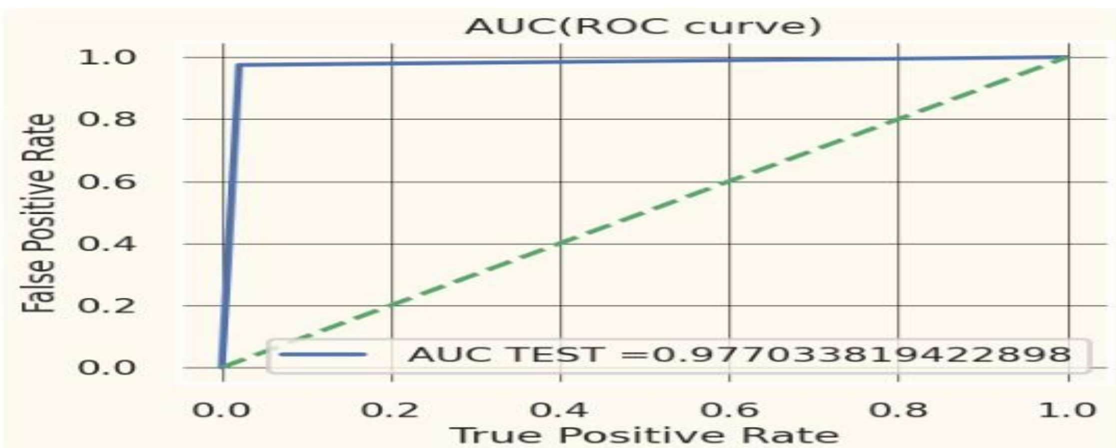

Fig. 4.5b: AUC (ROC curve) for Support Vector Machine.

Fig 4.5a represents confusion matrix for Support Vector Machine algorithm. Which contains True Positive, True Negative, False Positive, False Negative. The columns represent predicted NO which is 0 and predicted YES which is 1. In the other hand, the rows represent the actual NO which is 0 and the actual YES which is 1. The classifier made a total number of 568 predictions (that is 568 employees were tested for attrition). Out of those 568 cases, the classifier predicted 274 times for YES and 294 times for NO. True Negative value is greater which shows attrition cases are low.

Table 4.4a: The result for Support Vector Machine Algorithm

	Precision	Recall	F1 score	support
Yes	0.98	0.98	0.98	293
No	0.88	0.97	0.98	275
Accuracy			0.98	568
Macro avg	0.98	0.98	0.98	568
Weighted avg	0.98	0.98	0.98	568

Table 4.5: Performance Test scores for RF, Xgboost, KNN and SVM Algorithms.

	Accuracy	Precision	Recall	F1-score	AUC
RF	0.98	0.97	0.99	0.98	0.98
XGBOOST	0.98	0.97	0.99	0.98	0.97
SVM	0.98	0.98	0.98	0.98	0.97
KNN	0.92	0.99	0.86	0.92	0.92

Table 4.5 shows that the RF has highest Accuracy of 98% with F1-score and AUC of 98% respectively, Precision and Recall of 97% and 99%. Follow by XGBOOST with 98% classification of accuracy, Precision and AUC of 97% respectively, Recall of 99% and F1-score of 98%. SVM has classification accuracy of 98% with precision, Recall and F1-score of 98% respectively, and AUC of 97%. While KNN has classification accuracy of 92% with F1-score and AUC of 92% respectively, Precision and Recall of 99% and 86%. Therefore, RF outperformed all other algorithms in terms of AUC.

CONCLUSION/FUTURE WORK

The purpose of this study is to investigate the application of machine learning techniques to the issue of employee attrition so that human resources managers in the healthcare business would have a better grasp of the factors that contribute to employee attrition and why it occurs. Therefore, the study tries to develop a machine learning model that makes use of a resampling strategy in order to improve the accuracy with which employee attrition in the healthcare business can be predicted. The study proposes that leveraging ML techniques is essential for addressing the issue of employee attrition in healthcare sector, allowing the development of an early-warning system for this dynamic business environment. After the comparative analysis of different machine learning algorithms used to build the model, the study indicates an impressive overall accuracy of 98% in predicting employee attrition. The importance of using SMOTETOmek techniques for data resampling is clearly supported by the current findings. However, the findings of this study have opened an avenue for further research in the field of employee attrition, such as Application of other machine learning data resampling techniques, such as adaptive synthetic sampling (ADASYN), to hybridize the random forest algorithm.

Lastly, there may be employees that possess life time value to the organization. Such employees may be useful for organizations when making resolutions as whether to allow them to go or retain them. However, this study did not evaluate the characteristics of projected employees' attrition. Therefore, future research can focus on attrition client characteristics.

AUTHORS CONTRIBUTIONS: E.O. JESSICA performs data collection, data analysis and report writing. L. Emmanuel critical review, results interpretation and drafted the manuscript. M.A. Hambali and R.G. Kefas finalized the assessment of the study. All authors have read and agreed to publish this version of the manuscript.

FUNDING: the study has not been funded.

ETHICAL DECLARATIONS: Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT: Not applicable.

DECLARATIONS:

- We declare that there is no conflict of interest.
- Consent for publication: We would want to certify that this paper has not been published or accepted for publication elsewhere, or is undergoing editorial review for publication elsewhere.
- Availability of data and materials: The dataset and the materials used in this study are available from the corresponding author upon request.

RFERENCES

[1] S. S. Alduayj and K. Rajpoot, ―Predicting Employee Attrition using Machine Learning,‖ *IEEE*, Nov. 2018, doi: 10.1109/innovations.2018.8605976.

[2] U. PM and N.V. Balaji, ―Analysing Employee attrition using machine learning‖, Karpagam *Journal of Computer Science*, *13*, pp.277-282, 2019

[3] N. Ocean and C. Meyer, ―Satisfaction and attrition in the UK healthcare sector over the past decade,‖ *PLOS ONE*, vol. 18, no. 4, p. e0284516, Apr. 2023, doi: 10.1371/journal.pone.0284516.

[4] World Health Organization, Healthcare Workforce, 2023 https://www.who.int/health-topics/health-workforce#tab=tab_1 (accessed 3, march 2024)

[5] S. C. Lopes, M. Guerra-Arias, J. Buchan, F. Pozo-Martin, and A. Nove, ―A rapid review of the rate of attrition from the health workforce,‖ *Human Resources for Health*, vol. 15, no. 1, Mar. 2017, doi: 10.1186/s12960-017-0195-2.

[6] A. Mhatre, A. Mahalingam, M. Narayanan, A. Nair, and S. Jaju, ―Predicting Employee Attrition along with Identifying High Risk Employees using Big Data and

Machine Learning,‖ *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, doi: 10.1109/icacccn51052.2020.9362933.

[7] F. Alsubaie and M. Aldoukhi, ―Using machine learning algorithms with improved accuracy to analyze and predict employee attrition,‖ *Decision Science Letters*, vol. 13, no. 1, pp. 1–18, Jan. 2024, doi: 10.5267/j.dsl.2023.12.006.

[8] F. Mozaffari, M. Rahimi, H. Yazdani, and B. Sohrabi, ―Employee attrition prediction in a pharmaceutical company using both machine learning approach and qualitative data,‖ *Benchmarking: An International Journal*, vol. 30, no. 10, pp. 4140– 4173, Dec. 2022, doi: 10.1108/bij-11-2021-0664.

[9] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, ―Predicting employee attrition using machine learning techniques,‖ *Computers*, vol. 9, no. 4, p. 86, Nov. 2020, doi: 10.3390/computers9040086.

[10] A. H. Chowdhury, S. Malakar, D. B. Seal, and S. Goswami, ―Understanding employee attrition using machine learning techniques,‖ in *Lecture notes on data engineering and communications technologies*, 2021, pp. 101–109. doi: 10.1007/978981-16-2937-2_8.

[11] A.Frye, C. Boomhower, M. Smith, L. Vitovsky, and S. Fabricant, ―Employee attrition: what makes an employee quit?‖. *SMU Data Science Review*, 2018,*1*(1), 9. [12] K. M. Mitravinda and S. Shetty, ―Employee Attrition: prediction, analysis of contributory factors and recommendations for employee retention,‖ *IEEE*, Dec. 2022, doi: 10.1109/icwite57052.2022.10176235.

[13] K. Alshehhi, S.B. Zawbaa, A.A. Abonamah and M.U.Tariq, ―Employee retention prediction in corporate organizations using machine learning methods‖. *Academy of Entrepreneurship Journal*, 2021. *27*, pp.1-23.

[14] S. Yang and M.T. Islam, IBM employee attrition analysis. *arXiv preprint arXiv:2012.01286*. (2020).

[15] M. S. Alshiddy, and B. N. Aljaber, ―Employee Attrition Prediction using Nested Ensemble Learning Techniques‖. *International Journal of Advanced Computer Science and Applications*, 2023. *14* (7).

[16] F. Guerranti and G. M. Dimitri, ―A comparison of machine learning approaches for predicting employee attrition,‖ *Applied Sciences*, vol. 13, no. 1, p. 267, Dec. 2022, doi: 10.3390/app13010267.

[17] D. Chung, J. Yun, J. Lee, and Y. Jeon, ―Predictive model of employee attrition based on stacking ensemble learning,‖ *Expert Systems With Applications*, vol. 215, p. 119364, Apr. 2023, doi: 10.1016/j.eswa.2022.119364.

[18] M. S. Alshiddy and B. N. Aljaber, ―Employee Attrition Prediction using Nested Ensemble Learning Techniques‖. *International Journal of Advanced Computer Science and Applications*, 2023. *14*(7).

[19] S. Moyo, T. N. Doan, J. Yun, and N. Tshuma, ―Application of machine learning models in predicting length of stay among healthcare workers in underserved communities in South Africa,‖ *Human Resources for Health*, vol. 16, no. 1, Dec. 2018, doi: 10.1186/s12960-018-0329-1.

[20] M. Subhashini, and R. Gopinath, ―Employee attrition prediction in industry using machine learning techniques‖. *International Journal of Advanced Research in Engineering and Technology,.11*(12), pp.3329-3341. 2020.

[21] F. H. Wardhani and K. M. Lhaksmana, ―Predicting employee attrition using logistic regression with feature selection,‖ *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, vol. 7, no. 4, pp. 2214–2222, Oct. 2022, doi: 10.33395/sinkron.v7i4.11783.

[22] A. Qutub, A. Al-Mehmadi, M. Al-Hssan, R. Aljohani, and H. S. Alghamdi, ―Prediction of employee attrition using machine learning and ensemble methods,‖ *International Journal of Machine Learning and Computing*, vol. 11, no. 2, pp. 110– 114, Mar. 2021, doi: 10.18178/ijmlc.2021.11.2.1022.

[23] N. Mansor, N.S. Sani and M. Aliff, ―Machine learning for predicting employee attrition‖. *International Journal of Advanced Computer Science and Applications*, *12*(11). 2021.

[24] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, ―Predicting employee attrition using machine learning approaches,‖ *Applied Sciences*, vol. 12, no. 13, p. 6424, Jun. 2022, doi: 10.3390/app12136424.

[25] Y. Zhao, M. K. Hryniewicki, F. Cheng, B. Fu, and X. Zhu, ―Employee Turnover Prediction with Machine Learning: A Reliable Approach,‖ in *Advances in intelligent systems and computing*, 2018, pp. 737–758. doi: 10.1007/978-3-03001057-7_56.

[26] R. Jain and A. Nayyar, ―Predicting Employee Attrition using XGBoost Machine Learning Approach,‖ *IEEE*, Nov. 2018, doi: 10.1109/sysmart.2018.8746940.

[27] F. K. Alsheref, I. E. Fattoh, and W. Ead, ―Automated prediction of employee attrition using ensemble model based on machine learning algorithms,‖ *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–9, Jun. 2022, doi: 10.1155/2022/7728668.