Enhancing Management Strategies: Machine Learning and Creative Performance Insights in Employee Attrition Analysis and Prediction

Neelam Urooj[1], Tasawar Javed[1], Muhammad Usman[2], Muhammad Tahir Mushtaq[3], Ayesha Raza Toor[3]

1. Institute of Business Management and Administrative Sciences (IBM & AS), The Islamia University of Bahawalpur, Bahawalpur, Pakistan
2. Department of Computer Sciences, Quaid e Azam University, Islamabad
3. School of Systems and Technology, University of the Management and Technology, Lahore, Pakistan

Corresponding Author: Muhammad Usman (usmanmaqsood072@gmail.com)

Citation | Urooj. N, Javed. T, Usman. M, Mushtaq. M. T, Toor. A. R, "Enhancing Management Strategies: Machine Learning and Creative Performance Insights in Employee Attrition Analysis and Prediction", IJIST, Vol. 6 Issue. 3 pp 1211-1227, Aug 2024

Received | Aug 01, 2024 **Revised |** Aug 25, 2024 **Accepted |** Aug 26, 2024 **Published |** Aug 26, 2024.

Employee attrition and excessive turnover are major difficulties in today's competitive employment market, affecting many industries. To overcome these difficulties, firms are increasingly relying on artificial intelligence (AI) to forecast staff loss and devise effective retention strategies. This study investigates famous machine learning (ML) models to forecast employee turnover and deliver data-driven solutions. The first section of the study compares various ML models on an imbalanced dataset. The second section introduces the Synthetic Minority Oversampling Technique (SMOTE) for data oversampling and applies ML models to the enlarged dataset. ML can predict employee turnover by examining historical data, employee behavior, and external factors. This early detection enables organizations to respond proactively with targeted retention strategies. The study concludes that the Random Forest model is the best model when combined with SMOTE, achieving performance scores of 0.96 out of 1.

Keywords: Attrition, Prediction, Machine Learning, SMOTE.

Introduction:

In the current era, the labor market is rapidly increasing especially in the field of management and human resources. The main point that is notable in this domain is "involuntary turnover", in which employees leave their positions against the employer's wishes [1]. This change is affecting the business of every industry and country, especially in the pandemic situation. It experienced a significant increase during what has been dubbed the 'Great Resignation,' which began in the spring of 2021 in the US [2]. The enormous rise in employee turnover offers a complicated challenge for companies, resulting in both real and intangible expenses.

Employee attrition rates are an important indicator of an organization's progress. A high attrition rate suggests that employees leave frequently, perhaps resulting in the loss of organizational benefits [3]. There are several types of attrition to understand why employees leave the company.

- **Voluntary Attrition:** It is a situation in which an employee chooses to leave the company on their own.
- **Involuntary Attrition:** It is a situation in which the organization decides to end the employee's employment.
- **External Attrition:** It is the situation in which an employee leaves to work for another organization.
- **Internal Attrition**: It is the situation in which an employee is promoted to another position within the organization.

The employee attrition rate is considered to measure how many workers leave an organization. To figure out the issue, this attrition rate can play an important role. The attrition rate represents the company's progress and retention efforts.

As per the employee attrition data [4], one-third of new employees depart their organizations within six months of starting. Furthermore, according to the Job Openings and Labor Turnover Survey (JOLTS) [5] 3 to 4.5 million employees in the United States quit their positions each month.

The employee attrition rate was 57.3% in 2021 as per the Bureau of Labor Statistics [6]. The report also indicates that the attrition rate is around 19% in several industries. The Society for Human Resource Management (SHRM) estimates the cost per hire for new employees to be $4,129 [7]. A 90% retention rate and less than 10% attrition rate are favorable for the company. This ratio demonstrates that 90% of employees choose to stay in their current organization for a specific period. A low rate of attrition suggests that the company is doing a good job of keeping its employees, which can lead to financial savings, stable operations, and a positive work environment. Since meeting these requirements helps the business succeed as a whole, achieving them is a common goal for human resource management.

Mainly 77% of companies attribute turnover to employees seeking better chances in other industries or organizations, this is not the primary cause. The crises of the last two years have also played an important impact, raising concerns about perceived dangers in specific sectors. As a supplementary factor, 31% of the surveyed companies raised this worry. Furthermore, rising compensation expectations, fueled by recent inflation, and the need for additional flexibility are among the third most commonly cited causes of employee turnover.

ML [8], a subfield of Artificial Intelligence (AI), allows robots to learn from past data and forecast future outcomes. It is currently an essential component of data science work. The primary goal of ML approaches is to provide findings with more accuracy than human capabilities. These models facilitate decision-making by automating the learning process.

Refined data is fed into machines to train them so they can make judgments based on new information. ML models are primarily designed to find and learn from data patterns.

The use of ML in modern technologies is continually expanding. These significant applications include a wide spectrum of real-world domains. ML techniques are used to tackle common problems like image identification [9], audio recognition [10], social analysis [11], stock market trading [12], e-commerce [13], and agriculture [14]. Additionally, ML algorithms are used to forecast staff attrition [15].

The higher turnover can overall impact productivity and innovation. The knowledge gaps and pressure on the staff added due to the loss of experienced staff. Furthermore, the reputation of the organization can be disturbed due to the consistent turnover. With the use of AI and ML for prediction and reducing attrition, organizations can protect their culture and keep a motivated workforce. This study aims to improve management strategies by utilizing explainable AI and delivering innovative performance insights in the analysis and prediction of employee turnover. The study's goal is to use advanced ML techniques to create models that not only properly forecast which employees are likely to depart, but also provide meaningful explanations for these predictions. This method will allow managers to better understand the underlying causes of employee turnover and develop focused tactics to increase retention and overall organizational performance.

- ML algorithms are being applied that not only predict employee attrition but also provide clear, interpretable insights into the basis for these projections.
- The use of innovative performance measures and analytics to provide a comprehensive view of the variables driving employee turnover.
- Actionable methods for managers based on AI model insights to reduce attrition and improve staff retention.
- Performing a thorough assessment of the proposed models against current benchmarks to verify their effectiveness and dependability in predicting staff attrition.
- Interpret the results for balanced and imbalanced datasets.

Literature Review:

An extensive employee attrition prediction literature is presented in this section. The most recent literature is selected and presented in this paper. In recent years, different ML models introduced for the prediction of employee attrition.

Attri, Tanya [16] introduces an optimal hybrid ML model that integrates an oversampling technique (SMOTE) and a feature selection method (SA) with classification algorithms like SVM and LR to identify employees who may leave the organization. The study aims to evaluate the effectiveness of this approach in predicting employee attrition. After testing different models, it was found that the SA-SVM model performed similarly to the Bayesian model. While optimization revealed that SA-SVM was less accurate than other models, it had a good sensitivity score of 80.93%. In contrast, other models, despite having higher accuracy, showed a significant drop in sensitivity to 26.08%. The conclusion drawn is that technical feature selection methods provide more reliable results compared to those suggested by management domain experts.

Duptta et. al., [17] introduced a tool for the prediction of the churn possibilities with a neural network. They achieved an accuracy of 87.01% in their study with the neural network as compared to other classifiers. Kakad et. al., [18] considered the XGBoost model in their study with 12 features to predict employee attrition. The major issues of data duplication were also discussed in their study. They achieved an accuracy of 90% in their study with a machine-learning-based XGBoost model. The multiple cleaning processes method used in the study of Nurhindarto et. al., [19] for performance prediction. In their study, they considered feature selection methods and removed duplicates. They achieved an accuracy of 82.6% with parameters and feature selection methods along the Random Forest model.

Fatma [20] performed a comparative analysis of several ML algorithms to determine which classification algorithm reduces staff attrition rates in a business the most effectively. They used every ML classification algorithm available, including SVM, random forest, Logit Boost, Multilayer Perceptron (MLP), J48, K-nearest neighbors, Linear Discriminant Analysis (LDA), Naive Bayes, Bagging, AdaBoost, and Logistic Regression, on an IBM dataset with 35 features. With an accuracy of 87.14%, Logistic Regression is the most accurate of all of these. Srivastava et. al., [21] explored the importance of deep learning over traditional ML models in their study. For effective and reliable model training, they collected the dataset from the Fast-Moving Consumer Goods (FMCG) industry. The dataset is based on salary, performance rating completed projects within a year, etc. Their experimental results indicate that deep neural networks outperformed gradient boosting and random forest with an accuracy of 91.6%.

An SVM machine-learning model was suggested by Maharjan et al. [22] to predict employee attrition, and it achieved an accuracy of 0.93. The important issue of determining and forecasting the elements influencing retention in numerous businesses is tackled by Karthik Sekaran [23]. To do this, they used IBM employee data to illustrate Explainable AI (XAI) techniques like SHAP and LIME, which lower the risk of churn rate and aid in decision-making when applying ML algorithms. Alsheref et al. [24] introduce an automated algorithm designed to predict employee attrition using various predictive analytics techniques. These techniques were employed to identify the best model from different pipeline configurations. Additionally, an autotuning method was used to determine the optimal combination of hyperparameters for developing the best-performing model. The authors propose an ensemble model to evaluate which model is the most effective based on different assessment metrics. The results of the suggested model indicate that, so far, no single model can be considered perfect and optimal for every business scenario.

A model that tries to predict staff attrition and emotional ratings in a company was developed by Joseph et al., [25]. The information required for analysis was gathered with the aid of a retention questionnaire. This model forecasts and analyzes depression based on these data. Following pre-processing procedures, they used various classifiers, including Random Forests and Decision Trees. As a result, an accuracy of 86.0% is possible. In the study of Raza et. al., [26] the employee attrition analysis and prediction problem are presented. In their study, the four ML models i.e., Extra Tree Classifier, Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree considered for prediction purposes. The explanatory data analysis considered in their study to explore the features that have major impacts on the attrition rate. Their results indicated that the Extra Tree Classifier outperformed other models with an accuracy of 93%. Table 1 presents the overall summary of the literature review in terms of employee attrition prediction.

Table 1. Overall Literature Summary of Employee Attrition Prediction

Study	Model	Methodology	Results
Attri, Tanya [16]	SVM and LR	Feature Selection, SVM, and LR model considered.	80.93%
Duptta et. al., [17]	Neural Network	Neural Network-based model applied for churn possibilities prediction.	87.01%
Kakad et. al., [19]	XGBoost	12 Features considered with XGBoost for the prediction purposes.	90%
Fatma [20]	Logistic Regression	Different machine-	87.14%

		learning models were applied in this study. The considered dataset is based on 35 features.	
Srivastava et. al., [21]	Neural Networks	deep neural networks, gradient boosting and random forest applied.	91.6%
Maharjan et al. [22]	SVM	ML model applied to predict employee attrition.	93%
Alsheref et al. [24]	-	an automated algorithm designed to predict employee attrition using various predictive analytics techniques	-
Joseph et al., [25]	-	Retention Questionnaire	86.0%
Raza et. al., [26]	Extra Tree Classifier	Different ML models are applied for attrition prediction.	93%
Kumar et al., [27]	Logistic Regression	Different ML models are applied for attrition prediction.	87.71%
Gazi et al., [28]	Random Forest	Different ML models are applied for attrition prediction.	86.05%

Methodology:

We concentrated on essential components of the proposed study, such as dataset specifics, feature analysis, and an overview of ML models with employed methodology. These components are essential to understand the techniques and framework underlying our research. The current study focuses on attrition analysis and prediction utilizing ML methods. The overall abstract methodology can be seen in Figure 1.

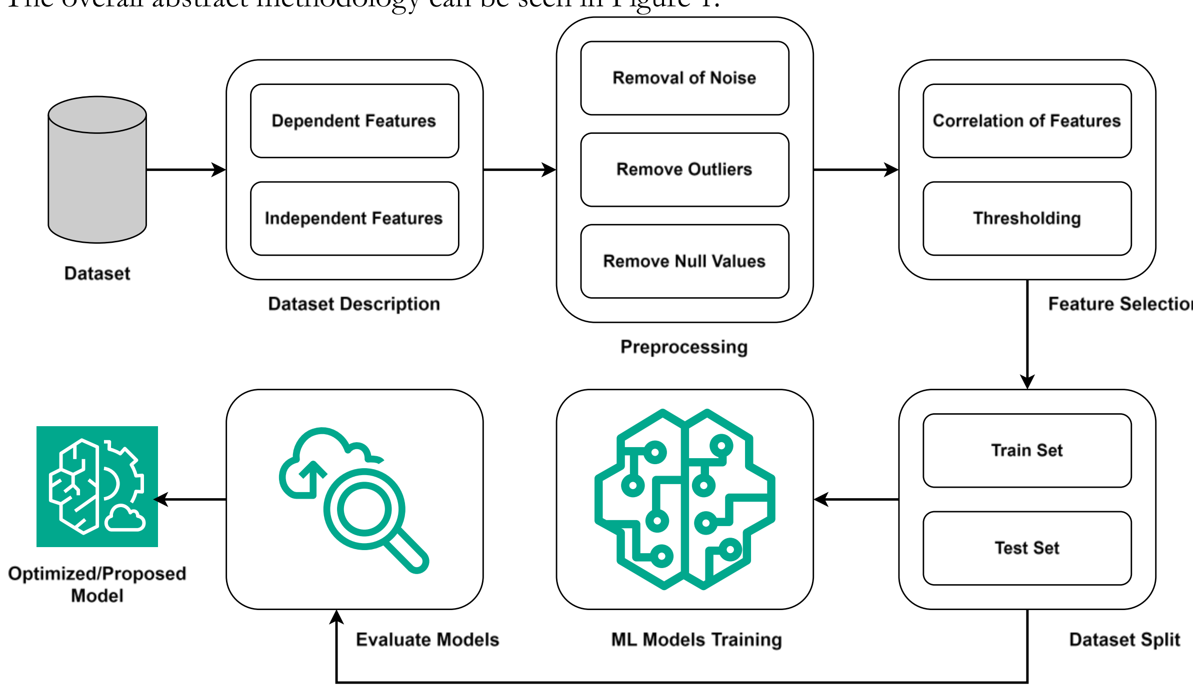


Figure 1. Proposed Study Flow Diagram

Dataset:

We considered the "IBM HR Analytics Employee Attrition & Performance" dataset, which was obtained from Kaggle [29]. This dataset is particularly useful for analyzing employee dynamics in corporate environments since it provides a full understanding of the numerous aspects that contribute to attrition and performance. The dataset consists of 35 columns of which 34 are features/independent variables and the 'Attrition' attribute is the dependent variable.

Feature/Independent Variable Selection:

The dataset under consideration contains 34 explanatory/independent variables (features) and one response/dependent variable. Feature selection is the process of selecting a valuable subset of features to create a successful prediction model. It increases the model's performance in terms of evaluation scores and is necessary for feature reduction. Still, it increases the model's performance in terms of assessment scores while reducing computational model time and cost. Statistical models are being used because of their effectiveness and speed in identifying attributes that have strong correlations with the response/target variable. It is difficult to decide which traits are reasonable in relation. We can tackle this problem using built-in tools such as Rapid Minor [30] [31].

The default feature selection approach utilized by RapidMiner is determined by the platform's unique process or operator. However, a typical default way is a combination of filter and embedding methods. Many RapidMiner processes, particularly those that use machine learning techniques such as decision trees or ensemble approaches, have built-in feature selection mechanisms by default.

ML Models:

ML trains computers using data rather than explicit programming. Instead of adhering to predefined rules, ML algorithms learn from data by recognizing patterns and similarities. These algorithms rely on these patterns to generate predictions or judgments based on the incoming data. Just like people, computers get more adept at learning as they are exposed to more data. Therefore, data play a key role in ML. The availability of data has been substantially facilitated by the internet, allowing for rapid finding and access to huge amounts of data. ML algorithms are typically divided into supervised and unsupervised learning depending on their learning techniques. Supervised learning involves learning from labeled data, where the algorithm is trained on input-output pairs. In the context of this work, various ML algorithms have been employed, including the K-NN, LR, SVM, RF, DT NB, and many more algorithms. These algorithms have been selected based on their suitability for the specific task at hand and their effectiveness in handling the available data.

K-Nearest Neighbors Model:

The K-Nearest Neighbors (KNN) classifier was developed in 1951 [32]. It classifies data points in a multidimensional feature space using a distance measure to determine proximity, then chooses the majority class from their k-nearest neighbors. Because they are critical factors influencing the algorithm's performance, the distance measure choice and k value are important. KNN lacks an easy-to-understand mathematical formula, in contrast to certain machine learning techniques that include explicit formulae. Rather, it functions based on the idea of determining which training set data points are the most comparable to guide classification choices.

Logistic Regression Model

David Cox created the groundbreaking ML technique known as logistic regression in 1958 [33]. It has become one of the most widely used approaches in the area throughout time. Because it makes use of probabilities to both describe and predict outcomes, this method is particularly well-suited to tasks that include categorical classification.

Support Vector Machine Model

Typically employed for classification problems, an SVM is an advanced supervised learning method [34]. It achieves clear segregation between different classes by projecting input data that has been expressed as vectors onto a higher-dimensional space. SVMs are a potent method that works with a range of kernel functions, such as polynomial, radial, linear, and Gaussian kernels. This flexibility makes it possible to handle a variety of datasets effectively. The performance of the classification task is dependent on the choice of kernel function.

Random Forest Model

The ensemble technique family includes the random forest classifier [35]. Instead of relying just on one decision tree, it leverages the collective power of multiple decision trees that serve as base learners. These distinct trees are trained separately, and by averaging the dataset's outcomes, their total predicted accuracy is increased.

Decision Tree Model

Decision trees form the theoretical basis of reinforcement functions. Using modes or means as forecasts for data inside these regions, decision trees use a recursive procedure to divide the feature space into rectangular regions. Because it shows the feature space's division criteria as a tree structure, this method is frequently referred to as the decision tree method. Regression tasks involve grouping data with comparable response values and assigning a fixed value (usually the mean) to each resulting region.

Evaluation Matrices:

The fundamental evaluation measures to measure the effectiveness of ML models considered for this study:

Accuracy: The model's prediction accuracy is measured by dividing the number of correctly classified samples by the total number of samples. However, only accuracy for experiments may be insufficient for evaluation, especially when dealing with imbalanced datasets or scenarios in which different types of errors have varying consequences.

$$Accuracy = (TP + TN)/((TP + FP + FN + TN) \quad (1)$$

Precision: Precision refers to a model's ability to correctly select positive samples from a set of true positives. This metric estimates the ratio of true positives to the sum of true positives and false positives, see Equation (2).

$$Precision = \frac{(TP)}{(TP+FP)} \quad (2)$$

Recall: Recall, also known as sensitivity or true positive rate, measures the model's ability to properly identify positive samples among all genuine positives. It calculates the proportion of genuine positives to the sum of true positives and false negatives. Recall primarily evaluates the comprehensiveness of positive predictions.

$$Recall = \frac{(TP)}{(TP+FN)} \quad (3)$$

F1: The harmonic means of precision and recall. It gives a single statistic that balances precision and recall, making it useful in situations where the class distribution is unequal or where both types of errors are equally important.

$$F1 = \frac{2\times Precision\times Recall}{(Precision+Recall\,)} \quad (4)$$

Results & Discussion:

This section indicates the findings from our experimental examination of several techniques combined with machine learning models for attrition prediction. To improve our analysis, we conducted several experiments both with and without data balancing.

Results on Unbalanced Dataset:

The dataset consists of 1490 samples, from which 237 samples belong to the "No" class whereas 1253 samples belong to the "Yes" class, as can be seen in Figure 2. This situation of unbalancing can be the cause of model overfitting to the majority class.

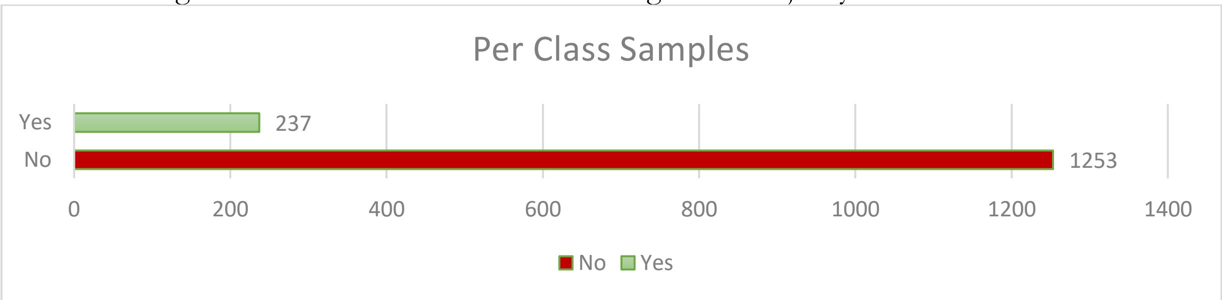


Figure 2: Dataset Class Distribution

We applied different five different machine learning models for this unbalanced dataset and evaluated the models using different evaluation matrices. We developed a thorough statistical analysis to ensure that the classes were appropriate for model training and evaluation. This analysis includes looking at the distribution of classes in the dataset to make sure that each one is sufficiently represented. We also evaluated the class balance and bias to ensure that the model was not skewed towards any specific class during training. In addition, we used statistical tests to evaluate the properties of each class and discover any significant discrepancies that could affect the model's performance. The classification reports of these five models can be seen in Table 2 to Table 6. The models evaluated for 294 samples i.e., test samples.

Table 2. Performance Scores of Logistic Regression Model for Attrition Prediction (Unbalanced Dataset)

	Precision	Recall	F1	Support
No	0.89	0.98	0.93	255
Yes	0.54	0.18	0.27	39
Accuracy			0.87	294
Weighted avg	0.84	0.87	0.84	294

Table 3. Performance Scores of KNN Model for Attrition Prediction (Unbalanced Dataset)

	Precision	Recall	F1	Support
No	0.88	0.96	0.92	255
Yes	0.36	0.13	0.19	39
Accuracy			0.85	294
Weighted avg	0.81	0.85	0.82	294

Table 4. Performance Scores of SVM Model for Attrition Prediction (Unbalanced Dataset)

	Precision	Recall	F1	Support
No	0.87	1.00	0.93	255
Yes	0.00	0.00	0.00	39
Accuracy			0.87	294
Weighted avg	0.75	0.87	0.81	294

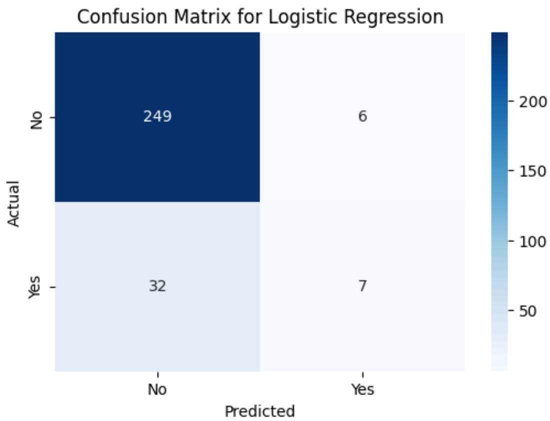
Table 5. Performance Scores of Decision Tree Model for Attrition Prediction (Unbalanced Dataset)

	Precision	Recall	F1	Support
No	0.87	0.84	0.85	255
Yes	0.15	0.18	0.16	39
Accuracy			0.75	294
Weighted avg	0.77	0.75	0.76	294

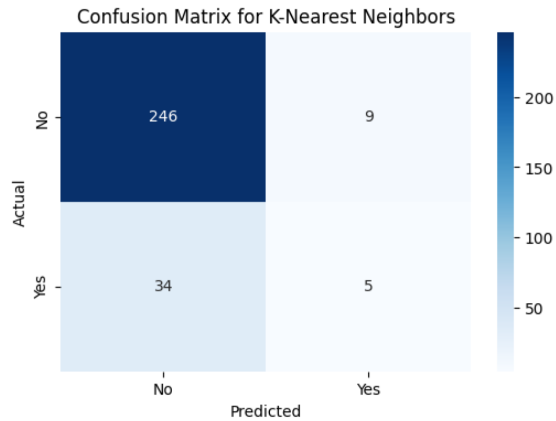
Table 6. Performance Scores of Random Forest Model for Attrition Prediction (Unbalanced Dataset)

	Precision	Recall	F1	Support
No	0.88	0.99	0.93	255
Yes	0.50	0.08	0.13	39
Accuracy			0.87	294
Weighted avg	0.83	0.87	0.82	294

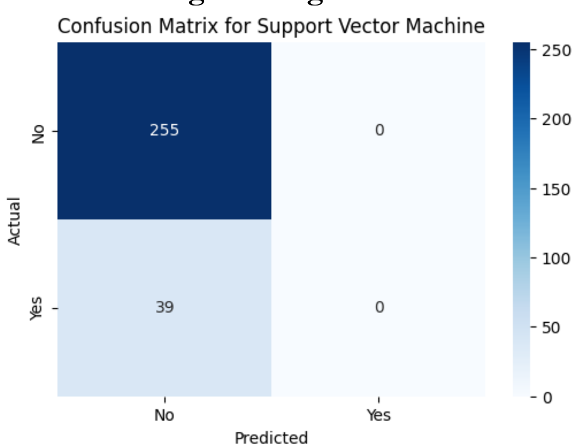
Table 2 to Table 6 concludes that Random Forest, SVM, and Logistic Regression models achieved an accuracy of 0.87 (87%) for the attrition prediction. The confusion matrix for this unbalanced dataset is also presented in this study, see Figure 3.



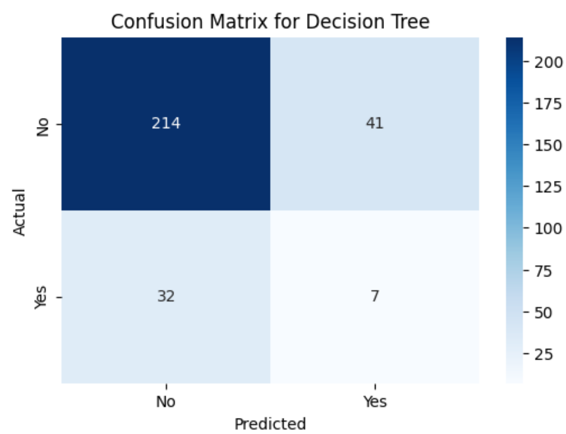
Logistic Regression



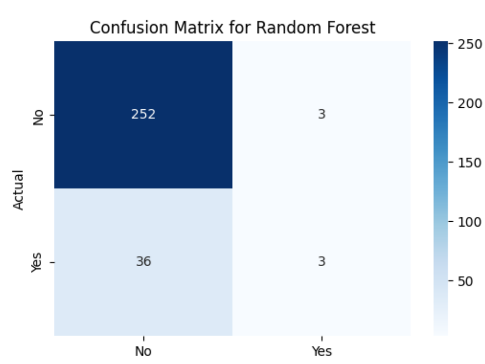
KNN



SVM



Decision Tree



Random Forest

Figure 3. Confusion Matrix of Machine Learning Models for Unbalanced Dataset of Attrition Prediction

The Receiver Operating Characteristic (ROC) [36] curve is an important tool for assessing the performance of classification models, especially in binary classification problems. The ROC curve is a visual classifier's representation, contrasting the True Positive Rate versus the False Positive Rate for different threshold values, see Figure 4.

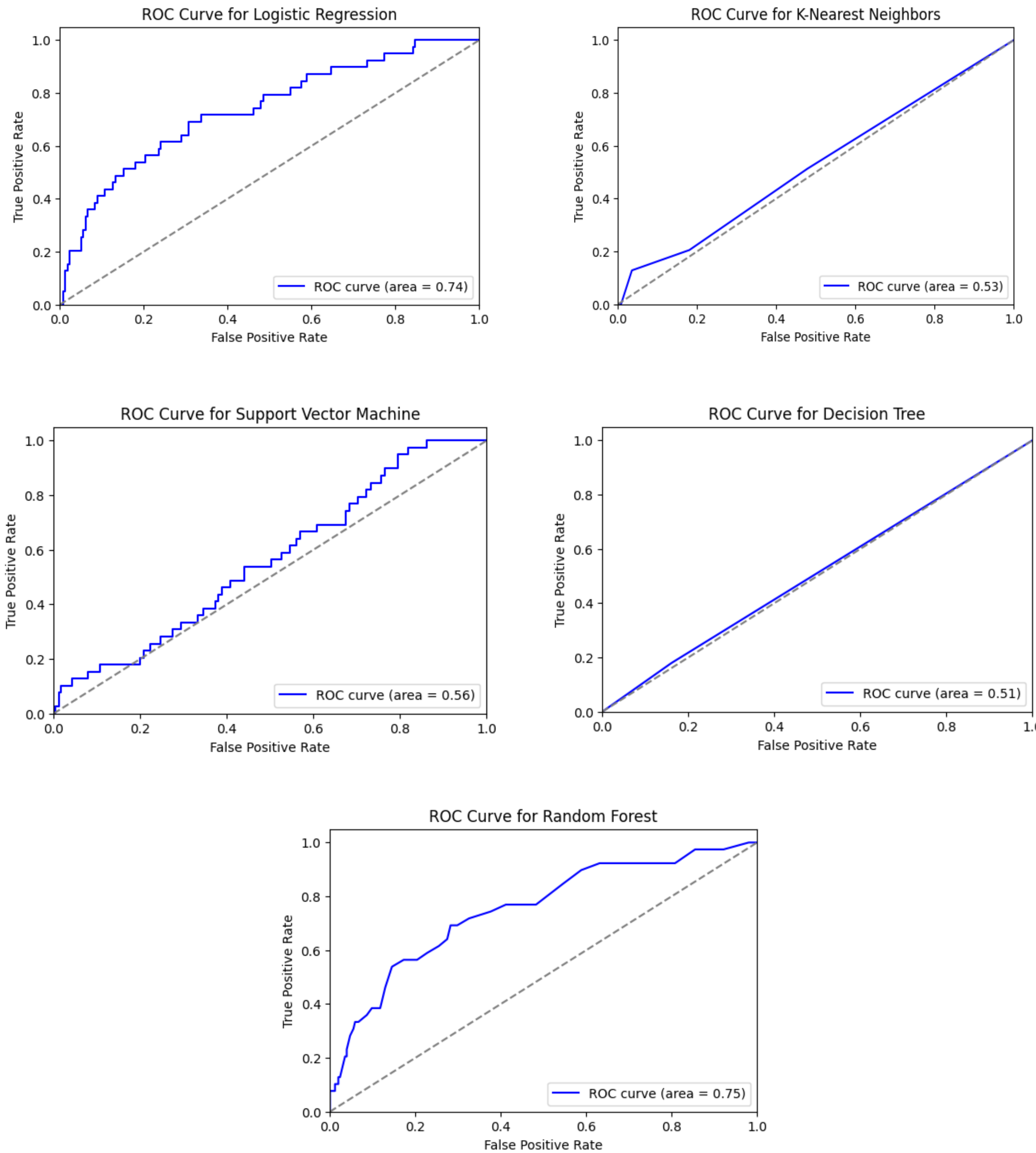


Figure 4. ROC-AUC Curve of Machine Learning Models for Unbalanced Dataset of Attrition Prediction

The importance of machine learning and deep learning models can't be neglected in this era [36] [37] [38] [39] [40] [41] [42], but unbalanced datasets provide substantial difficulty in machine learning, particularly for classification tasks. When one class is considerably underrepresented in comparison to others, biased models might produce negative results for the minority class. We considered the SMOTE (Synthetic Minority Over-sampling Technique) technique [43], which creates synthetic cases by interpolating existing minority class instances, see Figure 5.

Figure 5. Dataset Class Distribution after SMOTE

Table 7. Performance Scores of Logistic Regression Model for Attrition Prediction (Balanced Dataset)

	Precision	Recall	F1	Support
No	0.67	0.67	0.67	250
Yes	0.66	0.67	0.67	244
Accuracy			0.67	494
Weighted avg	0.67	0.67	0.67	494

Table 8. Performance Scores of KNN Model for Attrition Prediction (Balanced Dataset)

	Precision	Recall	F1	Support
No	0.85	0.54	0.66	250
Yes	0.66	0.91	0.76	244
Accuracy			0.72	494
Weighted avg	0.76	0.72	0.71	494

Table 9. Performance Scores of SVM Model for Attrition Prediction (Balanced Dataset)

	Precision	Recall	F1	Support
No	0.62	0.52	0.57	250
Yes	0.58	0.68	0.62	244
Accuracy			0.60	494
Weighted avg	0. 60	0. 60	0.59	494

Table 10. Performance Scores of Decision Tree Model for Attrition Prediction (Balanced Dataset)

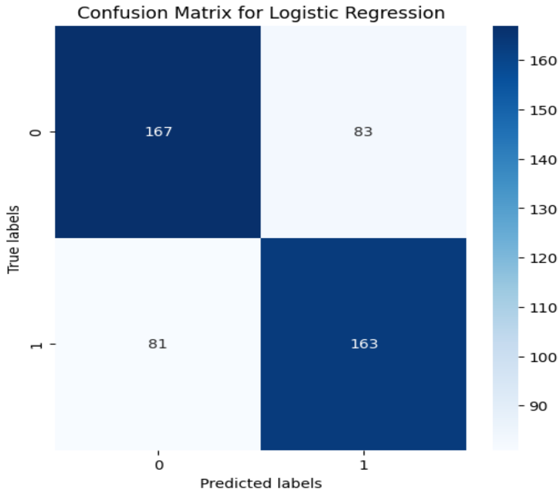
	Precision	Recall	F1	Support
No	0.98	0.84	0.90	250
Yes	0.85	0.98	0.91	244
Accuracy			0.91	494
Weighted avg	0. 92	0. 91	0. 91	494

Table 11. Performance Scores of Random Forest Model for Attrition Prediction (Balanced Dataset)

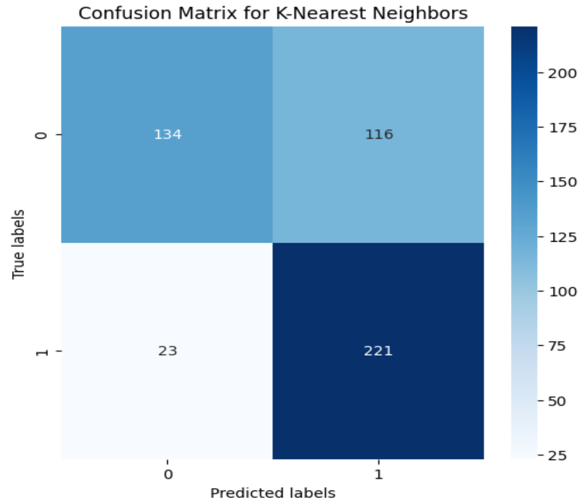
	Precision	Recall	F1	Support
No	0.97	0.95	0.96	250

Yes	0.95	0.97	0.96	244
Accuracy			0. 96	494
Weighted avg	0. 96	0. 96	0. 96	494

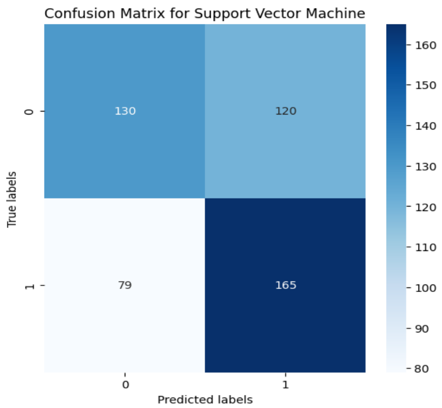
Table 7 - Table 11 concludes that Random Forest outperformed and achieved an accuracy, weighted (precision, recall, and F1) of 0.97 (97%) for the attrition prediction. The confusion matrix for this balanced dataset is also presented in this study, see Figure 6.



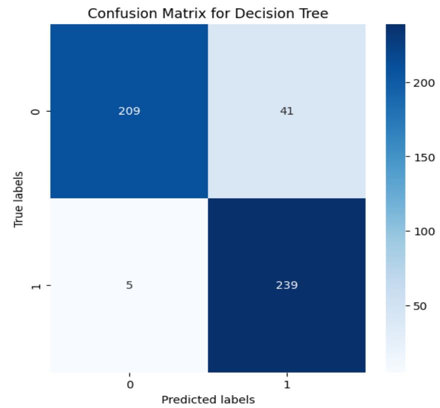
Logistic Regression



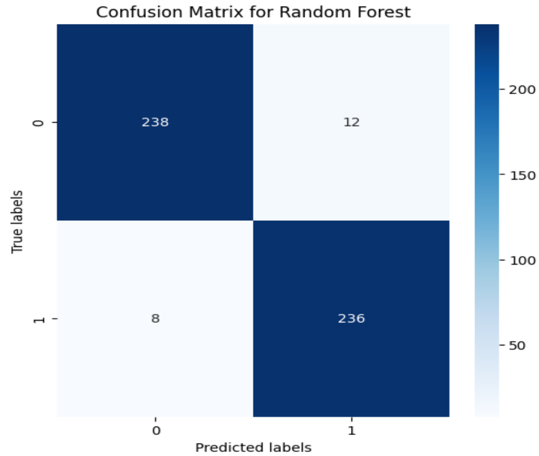
KNN



SVM



Decision Tree



Random Forest

Figure 6. Confusion Matrix of Machine Learning Models for Balanced Dataset of Attrition Prediction

The ROC curves of machine learning models can be seen in Figure 7 for this balanced dataset.

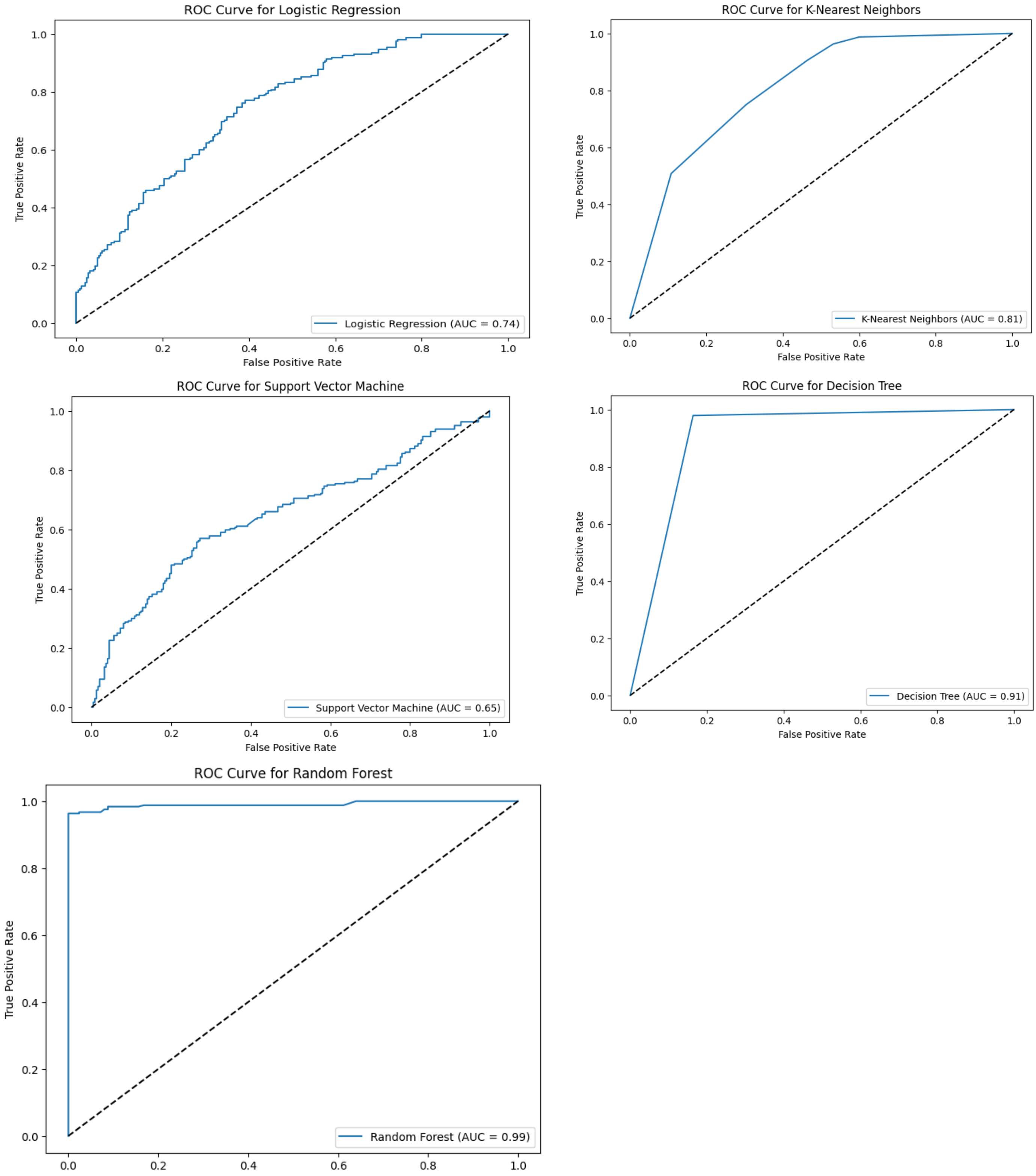


Figure 7. ROC-AUC Curve of Machine Learning Models for Balanced Dataset of Attrition Prediction

The extensive experiments of this study show that random forest performance is effective as compared to the other models with the SMOTE technique. The model is well generalized after this data balancing technique and performs well for each class. The overall results of the experiments can be seen in Table 12.

Table 12. Overall Performance Scores for Attrition Prediction

Models	Dataset Nature	Accuracy	Precision (Weighted)	Recall (Weighted)	F1 (Weighted)	Support
LR	Unbalanced	0.87	0.84	0.87	0.84	294
KNN	Unbalanced	0.85	0.81	0.85	0.82	294
SVM	Unbalanced	0.87	0.75	0.87	0.81	294
Decision Tree	Unbalanced	0.75	0.77	0.75	0.76	294
Random Forest	Unbalanced	0.87	0.83	0.87	0.82	294
LR	Balanced	0.67	0.67	0.67	0.67	494
KNN	Balanced	0.72	0.76	0.72	0.71	494
SVM	Balanced	0.60	0. 60	0. 60	0.59	494
Decision Tree	Balanced	0.91	0. 92	0. 91	0. 91	494
Random Forest	**Balanced**	**0.96**	**0. 96**	**0. 96**	**0. 96**	**494**

The results of the study suggest that AI-driven predictions, particularly using Random Forest combined with SMOTE, can pointedly improve the accuracy of turnover predictions, allowing organizations to implement more effective and targeted retention strategies. This practical data-driven approach is vital for reducing employee attrition and maintaining workforce stability in today's competitive job market.

Conclusion:

In conclusion, this study emphasizes the revolutionary power of artificial intelligence and machine learning in tackling the essential issue of employee attrition in today's competitive job market. Organizations may reliably forecast employee turnover by using machine learning models to assess historical data, employee behavior, and external factors, and then develop proactive, individualized retention measures. The study shows that using strategies like SMOTE to handle imbalanced datasets improves machine learning models' performance dramatically. Among the models tested, the Random Forest model, when combined with SMOTE, performed best, with an amazing performance score of 0.96. This research illustrates the effectiveness of advanced machine learning approaches in providing firms with actionable information to reduce employee attrition and promote a more stable and dedicated staff. To increase the impact and effectiveness of these models, future research should focus on refining them further and investigating their applicability across varied organizational contexts.

References:

[1] S. Rughoobur-Seetah, "The Unprecedented Lockdown: The consequences of job loss," Zagreb Int. Rev. Econ. Bus., vol. 24, no. 2, pp. 1–23, 2021, doi: 10.2478/zireb-2021-0008.

[2] A. Serenko, "The Great Resignation: the great knowledge exodus or the onset of the Great Knowledge Revolution?," J. Knowl. Manag., vol. 27, no. 4, pp. 1042–1055, Mar. 2023, doi: 10.1108/JKM-12-2021-0920/FULL/XML.

[3] "19 Employee Retention Statistics That Will Surprise you (2024)." Accessed: Aug. 27, 2024. [Online]. Available: https://www.apollotechnical.com/employee-retention-statistics/

[4] Peng and Baoyi, "Statistical analysis of employee retention," SPIE, vol. 12163, p.

1216303, Apr. 2022, doi: 10.1117/12.2628107.
[5] "JOLTS Home : U.S. Bureau of Labor Statistics." Accessed: Aug. 27, 2024. [Online]. Available: https://www.bls.gov/jlt/
[6] "U.S. Bureau of Labor Statistics : U.S. Bureau of Labor Statistics." Accessed: Aug. 27, 2024. [Online]. Available: https://www.bls.gov/
[7] "SHRM Survey: Average cost per hire is $4,129." Accessed: Aug. 27, 2024. [Online]. Available: https://www.businessmanagementdaily.com/46997/shrm-survey-average-cost-per-hire-is-4129/
[8] N. Thomas Rincy and R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques:," 2020 IEEE Int. Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2020, Feb. 2020, doi: 10.1109/SCEECS48394.2020.190.
[9] M. E. Korkmaz, M. K. Gupta, E. Çelik, N. S. Ross, and M. Günay, "Tool wear and its mechanism in turning aluminum alloys with image processing and machine learning methods," Tribol. Int., vol. 191, p. 109207, Mar. 2024, doi: 10.1016/J.TRIBOINT.2023.109207.
[10] M. P. Ngoc, T. N. Duy, H. H. Duc, and K. T. Anh, "A Proposed CNN Model for Audio Recognition on Embedded Device," Int. J. Interact. Mob. Technol., vol. 18, no. 08, pp. 116–126, Apr. 2024, doi: 10.3991/IJIM.V18I08.45917.
[11] Y. Rashid and J. I. Bhat, "Topological to deep learning era for identifying influencers in online social networks :a systematic review," Multimed. Tools Appl., vol. 83, no. 5, pp. 14671–14714, Feb. 2024, doi: 10.1007/S11042-023-16002-8/METRICS.
[12] Nisha Gurung, MD Rokibul Hasan, Md Sumon Gazi, and Md Zahidul Islam, "Algorithmic Trading Strategies: Leveraging Machine Learning Models for Enhanced Performance in the US Stock Market," J. Bus. Manag. Stud., vol. 6, no. 2, pp. 132–143, Apr. 2024, doi: 10.32996/JBMS.2024.6.2.13.
[13] T. S. De, P. Singh, and A. Patel, "A Machine learning and Empirical Bayesian Approach for Predictive Buying in B2B E-commerce," ACM Int. Conf. Proceeding Ser., pp. 17–24, Jan. 2024, doi: 10.1145/3647750.3647754.
[14] I. Attri, L. K. Awasthi, and T. P. Sharma, "Machine learning in agriculture: a review of crop management applications," Multimed. Tools Appl., vol. 83, no. 5, pp. 12875–12915, Feb. 2024, doi: 10.1007/S11042-023-16105-2/METRICS.
[15] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, "Predicting Employee Attrition Using Machine Learning Techniques," Comput. 2020, Vol. 9, Page 86, vol. 9, no. 4, p. 86, Nov. 2020, doi: 10.3390/COMPUTERS9040086.
[16] "Why an employee leaves: Predicting using data mining techniques - Google Search." Accessed: Aug. 27, 2024. [Online]. Available: https://www.google.com/search?q=Why+an+employee+leaves%3A+Predicting+using+data+mining+techniques&oq=Why+an+employee+leaves%3A+Predicting+using+data+mining+techniques&gs_lcrp=EgZjaHJvbWUyBggAEEUYOdIBBzc1NmowajeoAgCwAgA&sourceid=chrome&ie=UTF-8
[17] "(PDF) Employee attrition prediction using neural network cross validation method." Accessed: Aug. 27, 2024. [Online]. Available: https://www.researchgate.net/publication/341878934_Employee_attrition_prediction_using_neural_network_cross_validation_method
[18] "(PDF) Employee Attrition Prediction." Accessed: Aug. 27, 2024. [Online]. Available: https://www.researchgate.net/publication/326029536_Employee_Attrition_Prediction
[19] A. Nurhindarto, E. W. Andriansyah, F. Alzami, P. Purwanto, M. A. Soeleman, and D. P. Prabowo, "Employee Attrition and Performance Prediction using Univariate ROC feature selection and Random Forest," Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control, vol. 6, no. 4, pp. 345–350, Nov. 2021, doi:

10.22219/KINETIK.V6I4.1345.
[20] F. Ozdemir, M. Coskun, C. Gezer, and V. Cagri Gungor, "Assessing Employee Attrition Using Classifications Algorithms," ACM Int. Conf. Proceeding Ser., pp. 118–122, May 2020, doi: 10.1145/3404663.3404681.
[21] P. R. Srivastava and P. Eachempati, "Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi-Criteria Decision-Making Approach," https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/JGIM.20211101.oa23, vol. 29, no. 6, pp. 1–29, Jan. 1AD, doi: 10.4018/JGIM.20211101.OA23.
[22] R. Maharjan, "Employee Churn Prediction using Logistic Regression and Support Vector Machine," Master's Proj., Dec. 2021, doi: https://doi.org/10.31979/etd.3t5h-excq.
[23] K. Sekaran and S. Shanmugam, "Interpreting the Factors of Employee Attrition using Explainable AI," 2022 Int. Conf. Decis. Aid Sci. Appl. DASA 2022, pp. 932–936, 2022, doi: 10.1109/DASA54658.2022.9765067.
[24] F. K. Alsheref, I. E. Fattoh, and W. Mead, "Automated Prediction of Employee Attrition Using Ensemble Model Based on Machine Learning Algorithms," Comput. Intell. Neurosci., vol. 2022, no. 1, p. 7728668, Jan. 2022, doi: 10.1155/2022/7728668.
[25] R. Joseph, S. Udupa, S. Jangale, K. Kotkar, and P. Pawar, "Employee attrition using machine learning and depression analysis," Proc. - 5th Int. Conf. Intell. Comput. Control Syst. ICICCS 2021, pp. 1000–1005, May 2021, doi: 10.1109/ICICCS51141.2021.9432259.
[26] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, "Predicting Employee Attrition Using Machine Learning Approaches," Appl. Sci. 2022, Vol. 12, Page 6424, vol. 12, no. 13, p. 6424, Jun. 2022, doi: 10.3390/APP12136424.
[27] P. Kumar, S. B. Gaikwad, S. T. Ramya, T. Tiwari, M. Tiwari, and B. Kumar, "Predicting Employee Turnover: A Systematic Machine Learning Approach for Resource Conservation and Workforce Stability," Eng. Proc. 2023, Vol. 59, Page 117, vol. 59, no. 1, p. 117, Dec. 2023, doi: 10.3390/ENGPROC2023059117.
[28] Md Sumon Gazi, Md Nasiruddin, Shuvo Dutta, Rajesh Sikder, Chowdhury Badrul Huda, and Md Zahidul Islam, "Employee Attrition Prediction in the USA: A Machine Learning Approach for HR Analytics and Talent Retention Strategies," J. Bus. Manag. Stud., vol. 6, no. 3, pp. 47–59, May 2024, doi: 10.32996/JBMS.2024.6.3.6.
[29] "IBM HR Analytics Employee Attrition & Performance." Accessed: Aug. 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset
[30] "Altair AI Studio - Altair RapidMiner Documentation." Accessed: Aug. 27, 2024. [Online]. Available: https://docs.rapidminer.com/latest/studio/
[31] "Data Analytics and AI Platform | Altair RapidMiner." Accessed: Aug. 27, 2024. [Online]. Available: https://altair.com/altair-rapidminer
[32] E. Fix and J. L. Hodges, "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties," Int. Stat. Rev. / Rev. Int. Stat., vol. 57, no. 3, p. 238, Dec. 1989, doi: 10.2307/1403797.
[33] "The Regression Analysis of Binary Sequences on JSTOR." Accessed: Aug. 27, 2024. [Online]. Available: https://www.jstor.org/stable/2983890
[34] T. Han, "Research on Chinese Patent Text Classification Based on SVM," Jul. 2023, doi: 10.4108/EAI.26-5-2023.2334244.
[35] T. K. Ho, "Random decision forests," Proc. Int. Conf. Doc. Anal. Recognition, ICDAR, vol. 1, pp. 278–282, 1995, doi: 10.1109/ICDAR.1995.598994.
[36] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine

learning algorithms," Pattern Recognit., vol. 30, no. 7, pp. 1145–1159, Jul. 1997, doi: 10.1016/S0031-3203(96)00142-2.

[37] H. Younis, M. Asad Arshed, F. ul Hassan, M. Khurshid, H. Ghassan, and M. Haseeb-, "Tomato Disease Classification using Fine-Tuned Convolutional Neural Network," Int. J. Innov. Sci. Technol., vol. 4, no. 1, pp. 123–134, Feb. 2022, doi: 10.33411/IJIST/2022040109.

[38] M. A. Arshed, W. Qureshi, M. Rumaan, M. T. Ubaid, A. Qudoos, and M. U. G. Khan, "Comparison of Machine Learning Classifiers for Breast Cancer Diagnosis," 4th Int. Conf. Innov. Comput. ICIC 2021, 2021, doi: 10.1109/ICIC53490.2021.9692926.

[39] M. A. Arshed, Z. Akram, H. Z. Bajwa, M. H. Azeem, M. A. Dawood, and F. Riaz, "Enhancing Pakistan Rice Plant Disease Detection: A Highly Effective Pre-Trained CNN Model," 2023 25th Int. Multi Top. Conf. INMIC 2023 - Proc., 2023, doi: 10.1109/INMIC60434.2023.10465842.

[40] M. A. Arshed, H. Ghassan, M. Hussain, M. Hassan, A. Kanwal, and R. Fayyaz, "A Light Weight Deep Learning Model for Real World Plant Identification," 2022 2nd Int. Conf. Distrib. Comput. High Perform. Comput. DCHPC 2022, pp. 40–45, 2022, doi: 10.1109/DCHPC55044.2022.9731841.

[41] M. Mubeen, M. A. Arshed, and H. A. Rehman, "DeepFireNet - A Light-Weight Neural Network for Fire-Smoke Detection," Commun. Comput. Inf. Sci., vol. 1616 CCIS, pp. 171–181, 2022, doi: 10.1007/978-3-031-10525-8_14/COVER.

[42] R. A. A. Shahzad, M. A. Arshed, F. Liaquat, M. Tanveer, M. Hussain, "Pneumonia Classification from Chest X-ray Images Using Pre-Trained Network Architectures," VAWKUM trans. Comput. sci., vol. 10, no. 2, pp. 34–44, 2022.

[43] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.