The Bioscan 19(2)S1: 110-115, 2024 www.thebioscan.com
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES

Predicting Employee Promotions: Investigating the influence of Training, KPI Achievement, And Training Score

Guide name: Dr.M.Saravanamuthu

Assistant professor, PhD

Dr. N.Naveen Kumar, Assistant professor

V.Revathi

N.Shaiba sulthana

Y.Ganga Supraja

T.Sulochana

DOI: https://doi.org/10.63001/tbs.2024.v19.i02.S1.pp110-115

KEYWORDS
employee promotion, prediction, HR dataset, imbalanced dataset, data management, ANN, RF, SVM
Received on: 28-02-2024
Accepted on: 28-06-2024

ABSTRACT

When thinking about human resources, one of the most crucial procedures is the promotion process. Managers may inspire their staff and ensure the company's longevity with a well-structured promotion procedure. For many workers, the prospect of a promotion serves as a significant extrinsic motivator. Maintaining the employee's present level of involvement and dedication to the company is a win-win. For the company, it's a key tool for both rewarding employees and keeping tabs on their performance. Promotion candidates are evaluated based on a wide range of criteria, including but not limited to: age, training score, organisational commitment, seniority, performance level, skills, and awards. The purpose of this research is to examine a prediction approach that takes into account the factors used by Machine Learning algorithms like Random Forest, Support Vector Machine, and Artificial Neural Network to assess candidates for promotions. By using the ROS technique, Random Forest attained the best performance, boasting 98% accuracy, 96% precision, 1.0% recall, and 98% f1-score values. Human resources and managers might utilise this research to forecast promotion chances, allowing them to set appropriate criteria for employees.

INTRODUCTION

Promoting an employee is one of the trickiest things to do. A promotion occurs when an individual is moved to a position with more responsibilities, authority, and compensation. Success for the business and employee engagement and loyalty both increase with proper implementation of the promotion process. To progress in one's career, one must have both seniority and qualifications. The effectiveness of the company's career management system relies on the establishment and equitable implementation of a transparent and objective promotion strategy. All employees should be informed in advance of the criteria for promotion, the process for determining who will make the promotions, and the circumstances under which promotions will be made. Job progression is characterised by a rise in responsibilities, tasks, privileges, and power, as well as an increase in these things. Studies have shown that promotions may greatly improve employees' effectiveness on the job. According to Haryano et al. [3], one's performance on the work is positively affected by promotions.

When an employee moves up the corporate ladder, they are promoted from their current job to one with more responsibility, higher prestige, and more authority. Workers will be more motivated to do a good job if they know they would eventually leave the institution in a better position than when they started. For the simple reason that an employee's dedication to the job and productivity can only grow when they have faith in the future of their work and the possibility of advancement. If this holds, output, effectiveness, and quality will all rise. The business will be able to keep making money off of its manufacturing because of this. Consequently, everyone wins: the worker and the boss. Workers are more invested in their jobs and happier overall if they get promotions and better perks. Employees are less likely to leave and more willing to put in extra effort when they are happy in their jobs [4]. Dean and Joseph [5] state that when workers advance in their careers, it's because they are given greater responsibility, better resources, more prestige, higher compensation, and more demanding qualifications. The studies conducted by Tessema and Soeters [6] confirm the existence of a sufficient association between performance methods and employee promotion. Knowles et al. [7] states that promoting employees is all about rewarding those who consistently go above

and beyond. Boost your social standing, earn more money, feel better about yourself, and achieve personal fulfilment. Productivity, self-control, and inspiration in the workplace are all on the upswing. Be open and honest, evaluate staff promotions using evaluation indicators, conduct timely evaluations, and guarantee employee stability. Possibilities for career advancement have far-reaching consequences for companies as a result of emerging professions. Giving workers chances to hone their innovative and creative skills for the company's advantage. Raising workers' levels of expertise and experience inspires them to do their best work. A new organisational structure is being put into place as a result of a mutation in the position. Shahzad et al. [8] found that employees' performance significantly correlates with their chances of being promoted. Since promotions are positively correlated with employee performance and organisational productivity, he said that the company's policies should adhere to certain guidelines when it comes to promotions. Make sure the post doesn't go blank by promoting it to another. Workers whose promotions are well-deserved are more invested in their job and produce better results as a result. So that more people may apply for promotions in their current jobs. In this study, we provide a decision-support system that HR departments may use to determine whether employees are eligible for promotions. The use of unbalanced dataset approaches to address imbalances is the main contribution of the paper. The paper's emphasis on parameter adjustment is another important addition. Human resources will be able to use the study's findings to enhance key performance indicators and identify employees who might be promoted.

(KPI) KPIs in promoted jobs. RF outperformed the other algorithms with 98% accuracy, 96% precision, 1.0 recall and98% f1-score rate obtained among SVM and ANN in this study.
The paper is organized as described below. Section II describes the related work. Section III presents the methods and materials. Imbalanced dataset, ANN, SVM and RF algorithms are explained in detail. Section IV indicates the proposed system overview. Section V gives the experiment results and compare the performance metrics of the algorithms. Lastly, Section VI concludes the paper.

RELATED WORK

The ideal situation, according to McIntyre [9], is to remain with the same business for ten years and to be promoted three or four times during that period. In addition to looking nice on a résumé, those metrics are relevant right now, according to McIntyre. For an employee with a 30-year tenure, he thinks it's fair to promote them 10-15 times on average. In order to identify instances of epilepsy from EEG signals, the study by Hameed et al. [10] combined principal component analysis (PCA) with variable adaptive momentum (BPVAM) backpropagation, which can improve classification accuracy. They also compared their results to those of several automated methods. Using the BUPA and ILPD datasets, Mutlu et al. [11] presented a model for the detection of liver illness that is based on Convolutional Neural Networks (CNNs). An evaluation of CNN's efficiency was conducted in comparison to that of other machine learning methods, including NB, SVM, KNN, and LR. With an accuracy of 75.55% in the BUPA dataset and 72.00% in the ILPD dataset, CNN was shown to be successful in diagnosing liver illness in their study. In order to anticipate the early stages of Parkinson's disease, Rasheed et al. [12] put forth a model. They first classified the dataset using ANN's variable adaptive moment-based backpropagation method (BPVAM), and then they used the size reduction strategy by combining BPVAM with PCA. Their research shown that BPVAM-PCA outperforms BPVAM. According to Ufuk's [13] regression analysis model, there is a robust relationship between the three types of commitment—emotional, continuing, and normative—and the way individuals see marketing tactics. An organisation utilised the suggested algorithm to rank candidates for promotions and choose which ones to promote. A clarification technique was created using α-shear and the optimism index to clarify fuzzy weights, and linguistic variables were used to assess prospective persons according to criteria [14]. This research looked at the banking industry in Pakistan to see how factors like job happiness, performance reviews, and development opportunities affected workers' productivity. Smart PLs used a SEM analytical approach to sample 280 bank workers. Private bank executives would do well to prioritise recognition and incentive programmes in light of the results, as workers have come to anticipate financial compensation for their efforts [15]. In both cases, a higher grade or a salary rise are associated with employee promotion. Acknowledged as a stepping stone to a promotion, it is an ongoing process dependent on professional qualifications and duration of service. Human resource activities, including compensation, promotion, and performance assessment, are positively and significantly correlated with workers' productivity on the job, according to the author [16].

METHODS AND MATERIALS

Dataset

In this study, a data set of Kaggle's publicly accessible employee values was used [17]. In Table 1, attributes in the analysis is indicated.

TABLE I.: ATTRIBUTES USED IN ANALYSIS

Attributes	Explonation
employeeid	The employee ID
department	Employee's department
region	Employement region
education	Education Level
gender	Employee Gender
recruitmentchannel	Channel of recruitment for employee
no of trainings	no of other trainings completed in the previous year on soft skills, technical skills, etc.
age	Age of Employee
previous yearrating	Employee Rating for the previous year
length of service	Length of service in years
awards_ won	if awards won during the previous year then 1 else 0
avg training score	Average score in current training evaluations
is_promoted: (Target)	Recommended for promotion

Imbalanced Dataset

The ever-increasing quantity of data has exacerbated the issue of learning from unreliable data. This relates to how well learning algorithms work. Efficient transformation is at the heart of the issue. Since even distribution principles have been implemented, standard algorithms have strayed to approaches like incorrect cost analysis. This issue keeps cropping up. There will be no improvement in prediction accuracy if it is not possible to learn from datasets that are uneven. Internal and external imbalances both contribute to this issue [18]. The use of hybrid strategies to address this problem is on the rise. Data mining may place a greater premium on the minority class than the non-minority class due to the richness and potential impact of the information contained therein [19].

SMOTE

SMOTE is a method that uses synthetic instances to instantiate the Minority class. In order to do oversampling, a representative sample from each minority group is collected and then shown

along line segments that connect any two adjacent neighbourhoods [20]. For the purpose of lessening the issues that In SMOTE, it was decided that, rather than weighing the data points, it would be suitable to establish new minority samples by oversampling. The data is extrapolated from nearby occurrences of the minority class. The idea of feature space is therefore fundamental [21].

ROS

Return on investment (ROI) is the simplest and oldest method for handling imbalanced classification problems. In order to achieve a desirable class ratio during classifier training, the random oversampling (ROS) method symmetrically distributes labels from minority classes to big classes, thereby doubling the samples of labels from minority classes [22]. Now let's pretend that S is the imbalanced dataset, SNeg is the majority class representing negative data, and SPoz is the minority class representing positive data. For each class S, ROS is responsible for randomly adding K samples from the minority/positive class such that both classes have the same number of samples. With differing degrees of equilibrium for the training dataset, different values of K correspond to different numbers of positive class samples, which is equal to SPos ++K in this example. Once SPos==SNeg, the ROS is considered complete.

Support Vector Machine (SVM)

Linear and non-linear classification, as well as regression, are applications of the Support Vector Machine. Finding the best hyperplane is the next step. The decision boundary is the hyperplane. The ability to accurately construct the decision boundary makes it useful for linear classification, but the generation of residual data makes it unusable for non-linear classification, when accurate classification becomes a challenge. For their categorization, the kernel function is used. Applications for it include face analysis, handwriting analysis, and more [23]. One reason SVMs can learn to categorise items is because it can accomplish tasks like handwriting analysis.

Artificial Neural Network (ANN)

At this stage, the brain's functioning is modelled. It offers a Socratic approach to learning. An artificial neural network (ANN) that aims to mimic human information processing is trained using real-world examples. Through learning, prediction, and recognition, neural networks amass enormous power by capturing all information that happens in the data as patterns and forms. Because of the processing that goes into this, ANNs are referred to as PEs; they also manage to keep the inputs and outputs in check. They are also known as connective models because of the system's reliance on connectors. Neural connections between hidden neurons are established in guided learning. At this point, error back propagation is the learning method that is most often utilised. As part of its unsupervised learning process, a neuron may enhance its response to a specific input while dampening the activity of other neurons. When neurons engage in collaborative learning, they reinforce one another's output [24].

Random Forest (RF)

Classification and regression analyses that rely on the construction of multiple decision trees benefit most from this approach. No amount of "ensemble learning" can derail this strategy. This approach to learning is known as the "bagging method," and it involves creating a large number of decision trees by selecting many subsets of the dataset. In particular, random forests are made up of a large number of decision trees that are then merged to produce a decision tree that is offering optimal outcomes due to the execution of the data categorization procedure
[25]. Using RF, one may accurately build a regression or classification rule, or explore the link between potential prediction variables for a given prediction issue. The first scenario is an effort to safeguard this region by taking the classification error rate and mean error square into account. These two objectives, once transferred, may also unite as a single objective. In order to arrive at its predictions, the RF Classifier employs a cascade of CARTs. A subset of the training instances is plotted using a bagging strategy to build trees. You need to know Ntree and Mtry in order to plant trees in a forest. Random Forest has effective computing skills and can handle big datasets swiftly.

Evaluation

After trained data with methods, some metrics withreference to evaluation, success and power of the models areacquired. The metrics are acquired by using confusion matrix table. The confusion matrix is the summary evaluation tablecreated for each model which obtained with the methods inthe classification analysis. [26]. Evaluation metrics andresults which representing to the power and success of themodels are accuracy, precision, recall and specificity [27-28].Accuracy is the ratio of the number all correct classifications to the number of the all classifications. In theother words, it is the result of how many of the data are
correctly classified. The equation is given as follow:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

Precision is the ratio of how many of the data classifiedas positive are actually positive. With the following equation:

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

Recall is the ratio of how many of the data are actually positively will be predicted as positive with the equation 3: result is achieved: TP / TP + FN

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

Specificity is the ratio of how many of data which are actually negative are predicted as negative. Equation is given on the following:

$$\text{Specifity} = \frac{TN}{TN + FP} \quad (4)$$

F1 Score is the result which obtained by combining the calculation of the “recall and precision” values. In addition, this value (f1 score) shows the power and performances of the methods, by giving classification results and ratio of the methods used. Equation 5 shows the F1-Score calculation.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

SYSTEM OVERVIEW

Data for employee promotion is obtained from the Kaggle website. The diagram of the overall architecture is indicated in Figure 1. Firstly, data are balanced with imbalanced data techniques and then passed to the next stage. ANN, SVM andRF are applied to balanced data by SMOTE and ROS. Consequently, the model generates the classes that labeled non-promoted as 0 and promoted as 1.

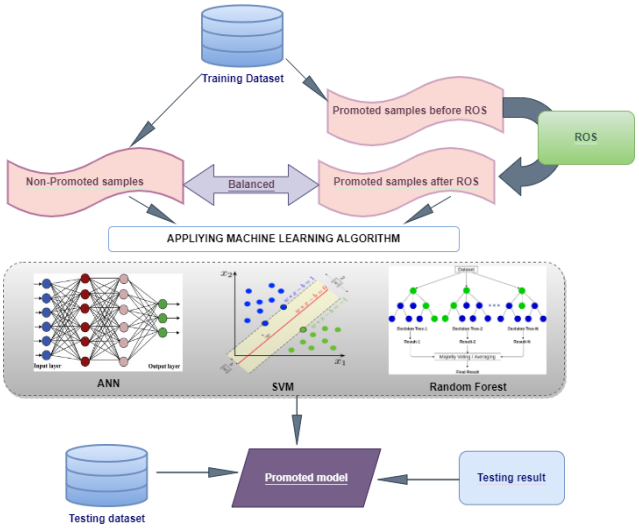


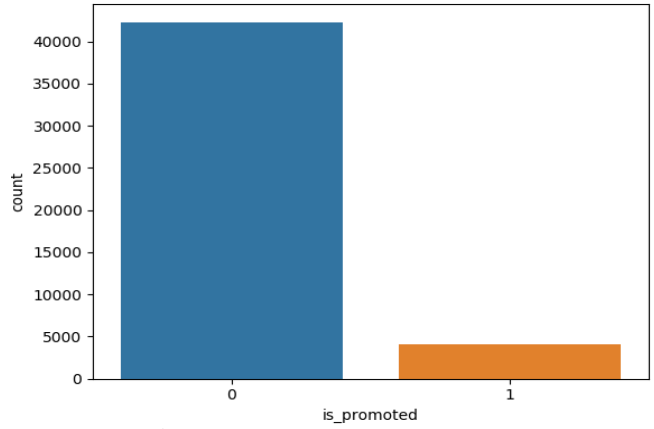
Fig. 1 The proposed system models

EXPERIMENTS AND FINDINGS

In this section, data is balanced by SMOTE and RUS techniques and a model is proposed on the prediction of employee promotion by SVM, ANN, and RF algorithms andthe findings are shown in detail by Python.

A. Imbalanced Data Methods

In this study, the data set consists of 52399 employees. 47853 non-promoted employees are labeled 0 and 4546 promoted employees are labeled 1.



. Fig 2. Number of promoted and not promoted

The distribution of the highly imbalanced dataset is indicated in Figure 2. In the study, SMOTE and Random Oversampling techniques are used to reduce the negative effect of the problem of class imbalance on classification

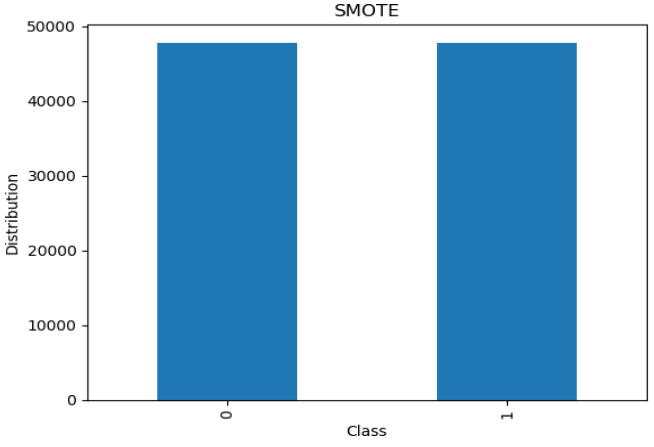


Fig. 3. Dataset after balanced methods

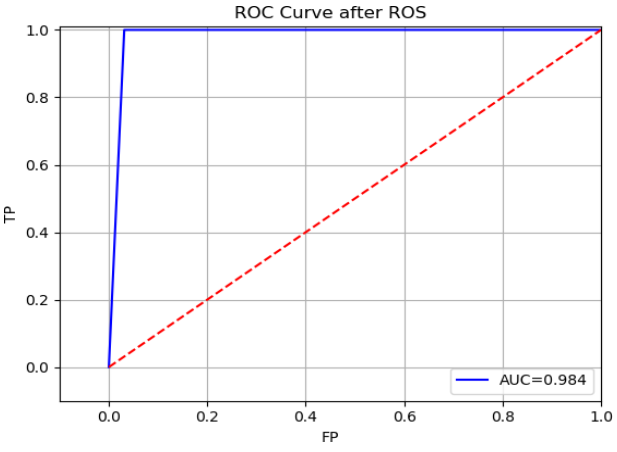


Fig. 4. ROC Curve of RF after ROS TABLE II.

PERFORMANCE METRICS

Algorithms	Performance Metrics			
	Accuracy	*Precision*	*Recall*	*F1-Score*
SVM	0.92	0.867	0.09	0.17
SVM_ROS	0.726	0.756	0.667	0.70
SVM_SMOTE	0.756	0.766	0.736	0.751
ANN	0.9234	0.8266	0.15	0.25
ANN_ROS	0.729	0.788	0.627	0.698
ANN_SMOTE	0.77	0.795	0.728	0.76092
RF	0.9314	0.84	0.247	0.38
RF_ROS	0.9842	0.96	1.0	0.98
RF_SMOTE	0.9157	0.9118	0.92	0.915

This study presents the performance of the SVM, ANN and RF based model in terms of accuracy, precision, recall and F1-Score performance metrics.
The findings from Table II stated that accuracy rate is notenough to select the accurate model for imbalanced dataset. Although accuracy rates are high, recall and f1-score values
are very low without applying any imbalanced techniques. F1- score are evaluated in selection of the model.
The performance of the SVM model was calculated, and the results are shown in Table II. The F1-Score of the SVM, SVM_ROS, SVM_SMOTE are calculated as 0.17, 0.70 and
0.75 respectively.
ANN (0.25), ANN_SMOTE (0.69), ANN_SMOTE (0.76)
scores were obtained according to F1-score performancevalues.
According to the findings obtained as a result of the analysis in Table II, RF model has an obvious advantage for employee promotion prediction and achieve the highest classification performance of predicting promotion with 98%F1-score by using ROS imbalanced technique.

CONCLUSION

In the HR process, promotions have a positive, substantial, and advantageous effect on workers' productivity. Using the RF approach, this research proposes a prediction model for employee promotion.
A method that helps HR departments make decisions on whether workers are eligible for promotions. Methods that are unbalanced in ROS and SMOTE are used. When it comes time to anticipate an employee's promotion, classification algorithms like SVM, ANN, and RF are used. Among SVM and ANN, RF achieved the highest accuracy (98%), precision (96%), recall (1.0), and f1-score rate (98%). According to the results, the F1-score—the harmonic mean of recall and precision—is the one to employ. To prevent picking the wrong model in datasets with unequal distribution, F1 Score is more important than accuracy. Having a measuring tool that include all error costs, not only False Negative or False Positive, is essential, which is why the F1-score is so important. In order to enhance key performance indicator (KPI) KPIs in promoted positions, HR may use this research to assess the time efficiency of their performance. In addition, it might help managers lessen the impact of a mistake in promotion candidate selection on an employee's handicap after a promotion. Using further data balancing approaches, future work will include feature engineering and feature significance into the analysis.

REFERENCES

- O. O. Awosusi and A. O. Jegede, "Nurses' Motivation and Work Performance in the Ekiti State Environment of Nigeria," [1]
- Journal of Pharmaceutical and Bioscience Research, Vol. 2, Issue 2, 2011, pp. 585–595. E. Kiruja and E. Mukuru, "The Impact of Motivation on Performance in Kenyan Public Technical Training Institutions Serving Middle-Level Employees" [2].
- 2018 IJAME.the thirdWork motivation and its consequences on job performance: Evidence from Indonesia, by S. Haryono, S. Supardi, and U. Udin. Letters in Management Science, volume 10, issue 9, pages 2107–2112.
- [4] In Asian Social Science, Vol. 11, No. 9, 2015, M. R. B. Rubel and D. M. H. Kee examined the relationship between nurses' perceptions of organisational commitment, the fairness of performance evaluations, and their intentions to leave their current positions.
- [5] Human Resource Management, 14th Edition, by M. R. W. Dean and M. J. Joseph. Collins Publishers, 2005, UK.
- [6] Drs. M. Tessema and J. L. Soeters Testing the HRM-performance relationship in the Eritrean public service: challenges and practices of HRM in emerging nations. Global Journal of Human Resource Management, 17(1), 86-105, 2006.
- The adult learner, by M. S. Knowles, E. Holton, and R. Swanson, is a seminal work in the fields of adult education and human resource development. It was published in 2005 by Elsevier in Burlington, Massachusetts.
- [8] Impact of HR policies on the perceived performance of university instructors in Pakistan, by K. Shahzad, S. Bashir, and M. I. Ramay. International Review of Business, vol. 4, no. 2, 08, 2008.
- [9] M. G. McIntyre, The dirty little secret of workplace politics: How to get what you want and dominate your boss. Gryphon, St. Martin's, 2005.
- In their paper titled "Classification of Epileptic Seizures using Artificial Neural Network with Adaptive Momentum," A. A. Hameed, A. Jamil, N. Ajlouni, J. Rasheed, A. Özyavaş, and Z. Orman presented at the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), doi: 10.1109/ICDABI51230.2020.9325688, the authors conducted the study.
- [11] “Deep Learning for Liver Disease Prediction” by E. N. Mutlu, A. Devim, A. A. Hameed, and A. Jamil. To be published in Pattern Recognition and Artificial Intelligence by Djeddi, C., Siddiqi, I., Jamil, A., Ali Hameed, A., and Kucuk, İ., the editors. Medical Practise Refresher Image 2021. Volume 15, Issue 15, of the Communications in Computer and Information Science journal, 2022. Published by Springer.
- [12] The following is the referenced paper from the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI): "Application of Adaptive Back-Propagation Neural Networks for Parkinson's Disease Prediction" (pp. 1-5), with the authors' affiliations as J. Rasheed, A. A. Hameed, N. Ajlouni, A. Jamil, A. Özyavaş, and Z. Orman.in
- [13] Ufuk, Bankalarda Terfi Uygulamaların Örgütsel Bağlılık Üzerindeki Etkisini Belirlemeye Yönelik Bir Araştırma. 13(2), 2019, 161-184, BDDK Bankacılık ve Finansal Piyasalar Dergisi.
- [14] The analytical hierarchy process with respect to

personnel and one outcome, M. Dağdeviren. The article was published in 2007 in Gazi University Medical Faculty Journal, volume 22, issue 4.

- [15] In Pakistan's banking industry, A. Kibria, A. S. Kazi, A. A. Banbhan, A. K. Shahani, and I. Junejo conducted research on the correlations between performance reviews, promotions, and contentedness on the job and productivity. Volume 27, Issue 2, Journal of Contemporary Issues in Business and Government, 2021.
- [16] In Asian Social Science, Vol. 11, No. 9, 2015, R. B. R. Mohammad and M., H. K. Daisy examined the relationship between nurses' perceptions of the fairness of their performance reviews, opportunities for advancement, and their intentions to leave their current employers.
- [17] Here is the provided data: https://www.kaggle.com/code/flaviocavalcante/employees- evaluation-for-promotion-eda-ml/data?select=employee_promotion.csv
- [18] "Learning from Imbalanced Data," published in 2009 by IEEE Transactions on Knowledge and Data Engineering, section 21, issue 9, pages 1263–1284, by H. He and E. A. Garcia.
- [19] Synthetic Minority Over-sampling Technique (SMOTE), developed by N. V. Chawla in 2002.
- [20] In honour of the 15th anniversary, A. Fernandez presented SMOTE for Learning from Imbalanced Data: Progress and Challenges in 2018..
- [21] Finding patterns in unbalanced data: current obstacles and potential solutions (B. Krawczyk,year 2016.
- [22]`F. Rayhan, S. Ahmed, A. Mahbub, R. Jani, S. Shatabda & D.M. Farid in their publicationFor unbalanced classification, Cusboost employs cluster-based under-sampling with boosting.The proceedings of the 2017 Second International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 1-5)2017.
- IEEEA survey of decision-tree and support vector machine-based machine learning methods [23] M. Somvanshi2017. IEEE.
- [24] S. Agatonovic-Kustrin, Foundational ideas of ANN modelling and their use in drug discovery research2001 ElsevierOne scalable tree boosting system is XGBoost, developed by T. Chen
- [25].(2016, ACM)When using a random forest, what is the optimal number of trees? (T. M. Oshiro, 2012).S. Umadevi and D. Marseline
- [27], A Comprehensive Review of Data Mining Classification MethodsSignal Processing and Communication: International Conference 2017, Pages 64–268There are 28] A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods, I. A. Zriqat, A. M. Altamimi, and M. Azzeh, 2017, 868-879.