I. J. Mathematical Sciences and Computing, 2023, 1, 23-36
Published Online on February 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmsc.2023.01.03

Predictive Analytics of Employee Attrition using K-Fold Methodologies

V. Kakulapati*
Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telanganga-501301
Email: vldms@yahoo.com
*Corresponding Author

Shaik Subhani
Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad, Telanganga-501301
Email: subhanicse@gmail.com

Received: 02 October, 2022; Revised: 15 November, 2022; Accepted: 26 December, 2022; Published: 08 February, 2023

Abstract: Currently, every company is concerned about the retention of their staff. They are nevertheless unable to recognize the genuine reasons for their job resignations due to various circumstances. Each business has its approach to treating employees and ensuring their pleasure. As a result, many employees abruptly terminate their employment for no apparent reason. Machine learning (ML) approaches have grown in popularity among researchers in recent decades. It is capable of proposing answers to a wide range of issues. Then, using machine learning, you may generate predictions about staff attrition. In this research, distinct methods are compared to identify which workers are most likely to leave their organization. It uses two approaches to divide the dataset into train and test data: the 70 percent train, the 30 percent test split, and the K-Fold approaches. Cat Boost, LightGBM Boost, and XGBoost are three methods employed for accuracy comparison. These three approaches are accurately generated by using Gradient Boosting Algorithms.

Index Terms: Machine learning, Gradient Boosting Algorithms, K-Fold approaches, Light GBM Boost, and XGBoost.

1. Introduction

Employee attrition denotes to the loss of personnel in any firm as a consequence of employees leaving. Employees are a company's most precious asset. It's essential to determine if the employees are dissatisfied with their jobs or have other reasons to leave. Nowadays days, workers are eager to switch jobs often in quest of better opportunities. But if they abruptly quit their employment, the company might lose a lot of money. It costs time and money to hire new employees, and it takes some time for them to become profitable. Retaining competent and hardworking employees is one of the most difficult problems that many firms face [1].

A significant study on retaining employees has been interested in analyzing the various motivations for individuals' choices to depart companies and how employees decide these things. Organizations may thoroughly investigate why individuals resign by analyzing why they continue and learning how to impact these decisions. The concept of organizational equilibrium can give vital insight into such issues. Including this idea, an employee will remain with an organization. For instance, the motivating factors it provides (such as adequate salary, pleasant work environment, and growth possibilities) are equivalent to or higher than the sacrifices (time, effort) demanded of the employee [2].

The acquisition is the primary aim of each organization and is a critical phase since it involves risk management and risk calculation. It is vital to hire the right individual when hiring staff since hiring the incorrect candidate can result in a slew of difficulties for both the firm and the employee. When it comes to recruiting talented employees, Human resources may be becoming more vulnerable since they are spending valuable time and company funds on training programs for an employee. Thus they make an important choice and they are liable [3].

Recruitment invests much in the training programs of future workers. Sometimes the prospects are not as skilled as the employer anticipated. If staff departs after six months, training is one of the most expensive expenditures for a firm.

There are numerous consequences of employee turnover that might be detrimental to a firm. High turnover rates harm the motivation of retaining employees. If the staff turnover rate is high, this indicates a significantly increased workload, reducing motivation and confidence. Staff turnover has an impact on earnings. Costs, knowledge loss, and decreased production all affect profit [4].

A high attrition rate in an institution results in greater recruitment, employment, and training expenditures. Not just is it expensive, but it is not easy to find a qualified and competent substitution. The top 20 percent of the population produces around 50 percent of their production in most industries [5]. Organizations invest, on average, in new employee training between four weeks and three months. Reducing attrition results in continuous production, lower recruitment costs, continual customer interaction, and better morale for the remaining workers.

Employee attrition may be a massive issue for firms, mainly when highly trained, technically skilled, and essential staff leaves for a better opportunity elsewhere. As a result, a skilled worker cannot replace so fast. So as technology is evolving with the use of the latest technological advancements in the IT industry, we can use Machine Learning and build ML models to predict employee attrition.

1.1 Objective

Employees are valuable to companies that invest in them by offering thorough training and an excellent working environment. They, too, are subjected to intentional attrition as ll as the impacts on the environment. Skilled employees are lost when there is no proper recognition. Hiring is another issue; replacements cost the company money, including hiring, training, and interviewing applicants.

Management is acting more swiftly by changing internal rules and methods to predict staff turnover. Skilled workers on the verge of leaving provide a range of incentives, such as a wage boost or further training, to reduce their odds of leaving the organization. ML techniques are used to forecast member of staff revenue. Analysts may construct and train a machine learning model to predict which employees are leaving using historical data from HR departments.

Recent years have seen remarkable progress in Gradient Boosting algorithms like XGBoost, Cat Boost, and Light Boost. To solve the problem above, find whether the employee is retaining or not, using three Gradient boosting algorithms, namely XGBoost, Cat Boost, and Light Boost. The 70% train to train the model accurately, and 30% test data splits. Then, the K-Fold Validation method for splitting train and test data to compare the accuracy of the corresponding algorithms and select the most accurate trained model. Predictions more accurately depend on the contribution data given to the trained model based on the forecast the company acts accordingly without incurring any loss to their company or businesses.

The article is organized subsequently. Section II provides a brief description of the literature study. Section III examines and contrasts the effectiveness of the three distinct ensemble classifications. Part IV explains the research methods for the dataset's properties, pre-processing and cross-validation, and the selection of measures for comparative precision. The investigation and subsequent discussion's findings are provided in Section V. Section VI ends the research by comparing the turnover forecasting classifications Cat Boost, Light GBM Boost, and XG Boost.

2. Related Work

ML approaches to assess the data of the employees to improve overall organizational status. The Senior managers have not been free to utilize personal specifics in remuneration and work productivity data on revenue rates and unique wages and maintenance records. Utilize RF (random forest) rating to enable worker rating based on their net profits and unstructured data analyzing methods. Utilizing different classifiers to assess employee productivity measures the similarity of performance metrics[6].

Compared with the Naïve Bayes classification and the J48 Decision Tree method, forecast a person leaving the company's probability [7]. Two approaches for each algorithm were studied in particular: Cross-validation 10 times and split 70 %.

A conceptual assessment of voluntary attrition [8] discovered that the most vital indicators of voluntary retention were aging, compensation, and work performance. Numerous research has found that other factors, including workplace environment, career satisfaction, and projected growth, lead to voluntary resignation [9,10].

ML is an application in which AI (artificial intelligence) enables computers to understand knowledge without a program. [11]. The execution of ML algorithms on accessible historical information can anticipate occurrences in the future [12]. Machine learning provides labeled data classifications and may build unlabeled data into hidden structures.

The Ensemble technique trying to boost was implemented in 2014 [13], as a tree-based method. It is also known as the XGBoost. The gradient-boosted trees are adaptable and accurately implemented, expressly intended to improve computing efficiency and modeling effectiveness. XGBoost uses a regularising period to lower the overfitting impact, enhancing forecasting and a quicker computer running time than gradients increase.

Regression analysis is straightforward to develop and performs on continuous classes, one of the most commonly used classification systems[14]. The attrition of employees may be interpreted as leakage or exit from a company from creative capital[15]. Some of the attrition rate studies classify attrition rate either as self-initiated or unwilling.

The probability to forecast staff retention in companies and utilize ML methods to design method attrition. The primary noise problem in the data from HRIS was also addressed, limiting these prediction algorithms' performance. HRIS data compared with the XGBoost classifier and six other supervised classifiers traditionally used to develop turnover models. The data from the worldwide retailer. The findings show that the XGBoost classification method is superior for a substantially greater degree of accuracy, comparatively low runtime, and efficient memory use for turnover prediction[16].

In addition, the XGBoost classification is developing as fault-tolerant for rapid and parallel tree construction under the distributed setting[17]. DMatrix takes data from the XGBoost Classifier. DMatrix is an internal XGBoost data structure designed for memory and training performance. Here, DMatrixes have been developed using NumPy function arrays and classes.

3. Methodology

3.1. Ensemble Learning

Methodologically learning generates and integrates various techniques for addressing a given computer intelligence challenge, such as classifications or analysts. Learning together generally serves to enhance a set of individuals such as classification, prediction, and function approximation.

3.1.1 LightGBM

LightGBM is an open and accessible decentralized gradient boosting framework for machine training created and used based on feature selection methods for grading, categorization, and other ML applications. The process of evolution focuses on efficiency and scalability.

The LightGBM frameworks include GBT, GBDT, GBRT, GBM, MART, and RF methods. Concurrent workouts, varied activation functions, regular bags, and premature stoppages are a few minor benefits of LightGBM over XGBoost. A key differentiation between the two is the form of the trees. Compared to the oldest.

LightGBM does not use the frequently used sorted decision tree technique, seeking the optimal split point on the sorted value of features instead of XGboost and other implementations. LightGBM has used a high-efficiency wavelet tree-learning approach that offers considerable speed and memory saving.

GOSS(Gradient-Based One-Side Sampling) and EFB(Exclusive Feature Bundling) are two unique approaches used in the Light GBM algorithm to run quicker while retaining excellent accuracy. On a cluster of 16 PCs, the distributed version of LightGBM takes just one or two hs to train a CTR predictor on the Criteo dataset, which comprises 1.7 billion records with 67 characteristics. LightGBM supports C++, Python, R, and C# and runs on Linux, Windows, and macOS.

3.1.2 Comparison with other tree-based algorithms

Light GBM is developed upwards, whereas conventional techniques downwards generate trees, stating that Light GBM is boosting tree leaflet by leaflet instead of level by level. Also, with the most significant delta reduction, it grows the leaflet. Blade-specific algorithms can reduce the loss of the same leaf than level-specific methods. The following graphics demonstrate how LightGBM and other enhancement methods are implemented.

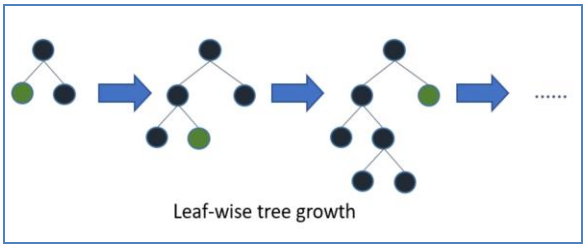

Fig. 1. Tree-growth of LightGBM

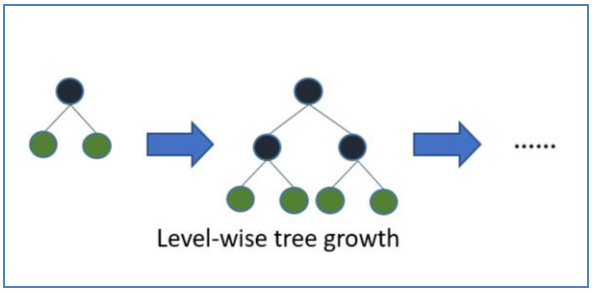

Fig. 2. describes how boosting algorithm works other than LightGBM.

3.1.3 Light GBM getting so much traction

It is difficult for conventional data science methodologies to offer faster discoveries since the amount of the data is expanding at an exponential pace. Light GBM is coming before the term "Light" because of its quick speed. Massive volumes of data may be processed with Light GBM while using less memory. Its emphasis on accuracy is another factor contributing to Light GBM's success. GPU learning is made possible by LGBM. In order to build data science applications, data scientists are using it more often.

3.1.4 XG BOOST

For the Distributed (Deep) ML Community (DMLC) organisation, Tianqi Chen developed XGBoost as a research project. In the beginning, it was a terminal that altered the libsvm settings file. It gained notoriety in the ML competition scene after being used in the winning submission for the Higgs Machine Learning Challenge. XGBoost comprises Java, Scala, Julia, Perl, and more language package implementations, and shortly afterward, the Python and R package building. This increased the library's visibility and popularity among developers and the Kaggle community. Utilizing a considerable number of elements was quickly merged with various programs, making it easy to use in each community. It has now been included in sci-how for Python users and is directly coupled with R users' caret package. A regularly used gradient boosting system for C++, Java, Python, R, Julia, Perl, and Scala is offered by the XGBoost Software Library. It works with Windows, Mac OS X, and Linux. The goal of the project is to provide the distributed, scalable, and flexible "GBM, GBRT, GBDT) Library" as it is defined in the project. On a single server, Apache Hadoop, Apache Spark, and Apache Flink carry out distributed processing.

3.1.5 Cat Boost

CatBoost is the boosting technology for decision tree gradients. Created by Yandex researchers and engineers, it uses searching taxis, recommendation systems, personnel assistance, self-dynamic cars, weather prediction, and many more actions by Yandex and other organizations such as CERN, Cloudflare, and Careem.

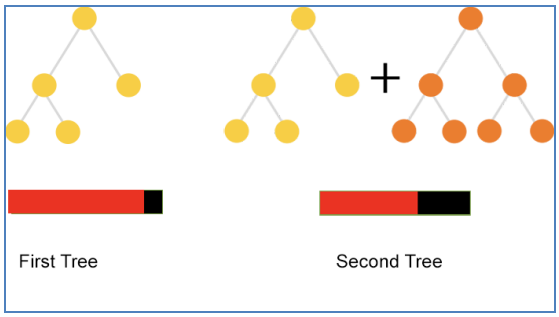

Fig. 3. Significant inaccuracy is frequently present in the first and second trees.

The second iteration is seen in the right-hand figure above, in which the algorithm learns another tree to lessen the inaccuracy caused by the original tree. This technique is repeated until the algorithm achieves a satisfactory quality mode, as seen below:

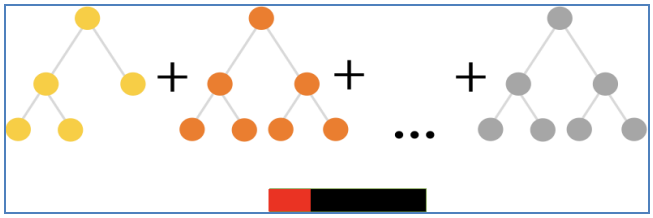

Fig. 4. This technique is repeated until the algorithm has created an excellent quality mode.

3.1.6 Choosing the tree structure

This approach is greedy. Features are selected for replacement in each leaf in order, along with their splits. Based on early split calculations and the transformation of category variables into numerical attributes, candidates are chosen. The tree depth and other structure-choice rules are defined in the initial settings.

How a "feature-split" pair for a leaf is chosen:

1. Candidates ("feature-split pairs") are assigned a leaf when the split is created.
2. Due to the assignment of the candidates produced in step 1 to the leaf, several penalty functions are generated..
3. The split with the least number of penalties is chosen.

The leaf provides the resultant value.

This process repeats for each subsequent leaf (the number of leaves needs to match the depth of the tree).

A random permutation of classifying objects is conducted before each new tree is constructed. The structure of the next tree is chosen using a measure that indicates the direction for further Enhancing the function. For each item, the value is computed sequentially. The item's data is utilized in sequence in which they are placed before the method — the permutation acquired before creating the tree is used in the computation.

4. Results and Analysis

More research in the field of attrition may be found. Because the strategy to forecast employee attrition is quite similar to erosion, it enables us to predict alternative ways. In [19], combining various training previous observations per employee from Training Data improves the predicted performance of retention models compared to using simply the most relevant data. Another issue is that instead of obtaining several samples from the whole term of the individuals, they limit it to a small piece of data, implying that many jobs are once again eliminated.

For implementation analysis, the data set is gathered from the Kaggle database, an open-access repository. Then trained data set Machine Learning models using the k-fold validation methodology, using 70 percent 30 percent dataset splits. CatBoost, XGBoost, and LightGBM Boost are the Machine Learning algorithms employed in research to pick the most accurate model out of all of them and compare their accuracies.

In k-fold cross-validation, we initially rearrange data to ensure that the sequence of the dependent and independent variables is fully random. This process is executed to ensure that none of inputs are skewed. Next, we divided the dataset into k sections. Thus eliminated the overfitting issue, when a classifier is developed utilising all of the data in one brief and gives the greatest prediction performance.

4.1 Pre-processing

The technique of correction or removal from the system is known as data cleaning, inaccurate, damaged, poorly formatted, duplicated, or incomplete. In general, cleaning the dataset removes repeated or irrelevant observations, handles missing data, handles null values, etc. There are f irrelevant columns in the dataset, so that will remove them, and there are some categorical data. So here, will change the categorical data into numerical (integer) data using the Label Encoder from scikit learns.

4.2 Splitting Dataset into Train and Test data

will use two types of tests. The train data split will use 70 percent as train data and the remaining 30 percent as test data. The second is the KFold method to divide the dataset into train and test datasets to compare the accuracy of models built after training and make the best out of the most accurate model trained.

4.3 Training the model

The modeling phase entails selecting projections depending on several machine-learning approaches employed in the experiments. Risk stratification methods based on deep learning Gradient boosting algorithms: Cat Boost, Light, and XGBoost to train model, and so on can be used in forecasting. Every model may be tracked on the set of features in this manner, and the model with the better prediction accuracy can be used for forecasting

After dividing the dataset into train and test data using the two methods described in the above module, will construct a model and train the model using the training dataset. In this work, three different types of Gradient boosting algorithms implementations are implemented. Cat Boost, Light, and XGBoost will train the model. The algorithm for K Folds uses folds values as 3,5,10 to compare the accuracy and pick the best model that is more accurate for predictions.

4.4 Testing the Model

After training the model, will test the trained model using the test data to find the model's accuracy. As have built models using different algorithms and different splitting tests and trained data strategies, can pick the most accurate model out of all models tested depending on the accuracy to predict employee attrition more accurately.

4.5 Predicting using the model

Using the most accurate model from the above module for every given input detail of the employee will predict the employee's attrition.

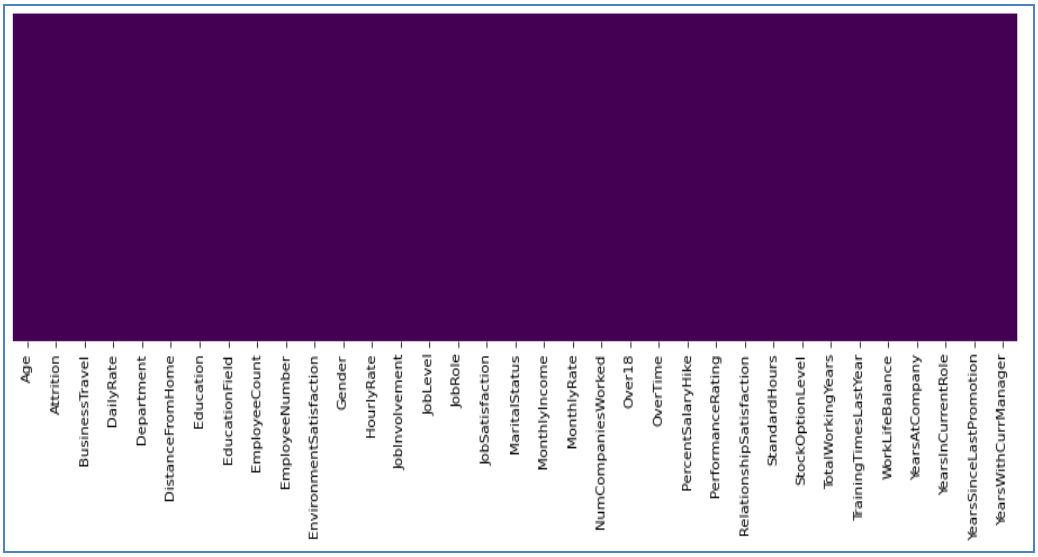

Fig. 5. Heatmap

Firstly, the null values or the missing values are checked in the dataset with specified functions. Then the result is depicted in the form of a heatmap that showcases the null values or missing values in the dataset pictorially.

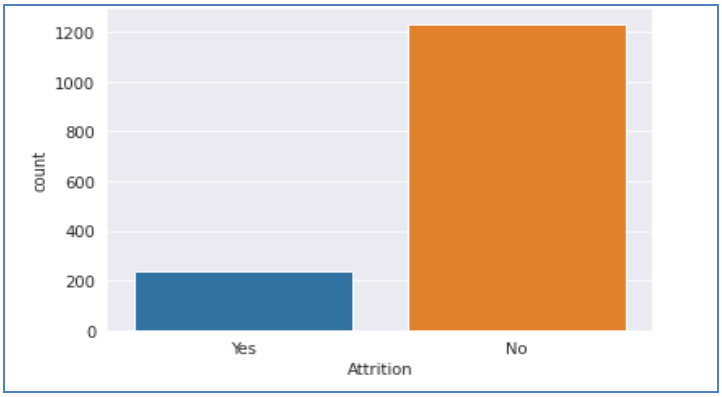

Fig. 6. The above graph represents the count of attritions.

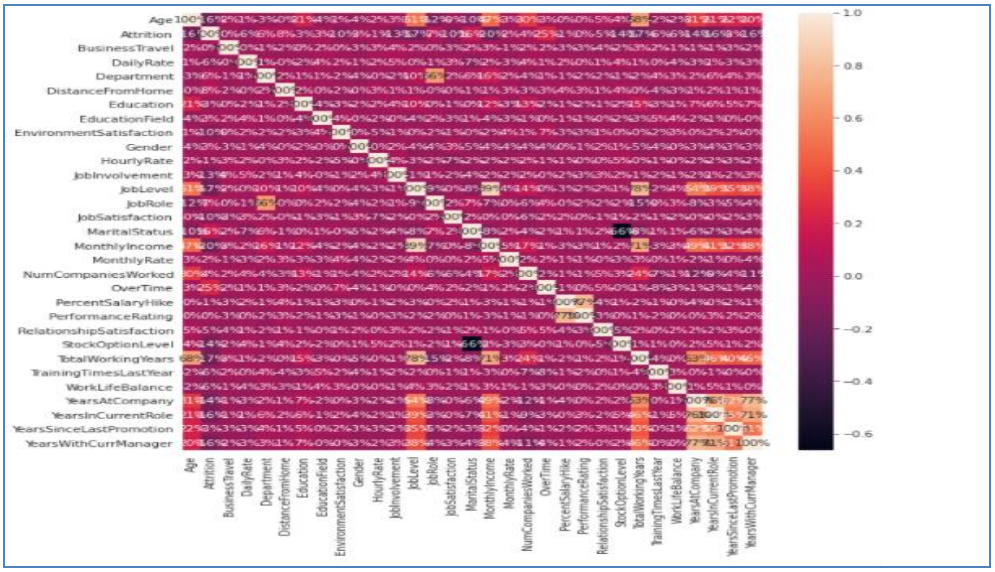

Fig. 7. Correlation matrix

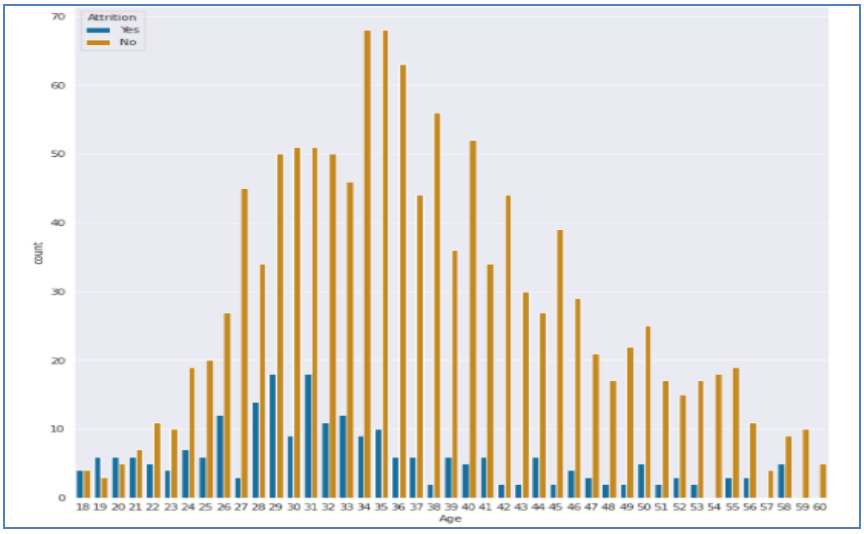

Fig. 8. Relations between Age and Attrition

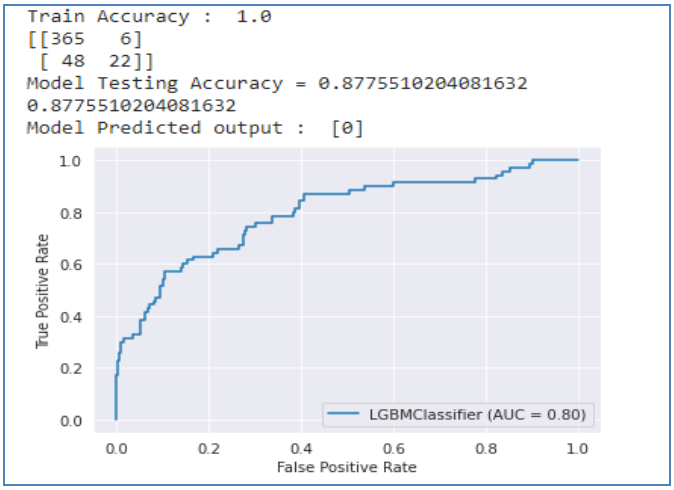

Fig. 9. Predicting using the trained model for given input [0,341,2,9,3,3,1,1,2,2,2,7,3,2,1067,759,1,0,3,0,3,0,10,3,1,10,3,9,7,18]

4.6 Accuracy analysis

Table 1. Accuracy of algorithms for 70 30 split datasets

Model Type	Accuracy (%)	True negatives	True positives	False negatives	False positives
CatBoost	87.52	367	19	51	4
LightBoost	87.75	365	22	48	6
XGBoost	87.30	363	22	48	8

Table 2. Accuracy of algorithms for K fold validation where K=3

Model Type	Accuracy (%)	True negatives	True positives	False negatives	False positives
CatBoost	86.53	411	13	64	2
	85.51	398	21	65	6
	88.97	412	24	50	4
LightBoost	86.32	403	20	57	10
	85.10	394	23	63	10
	88.57	408	26	48	8
XGBoost	86.32	406	17	60	7
	86.32	398	25	61	6
	90.00	411	30	44	5

Table 3. Accuracy of algorithms for K fold validation where K=5

Model Type	Accuracy (%)	True negatives	True positives	False negatives	False positives
CatBoost	86.05	246	7	40	1
	90.13	250	15	28	1
	83.33	232	13	46	3
	88.09	244	15	30	5
	89.11	249	13	30	2
LightBoost	84.35	241	7	40	6
	88.43	246	14	29	5
	83.33	229	16	43	6
	87.41	239	18	27	10
	89.11	247	15	28	4
XGBoost	86.05	244	9	38	3
	89.79	249	15	28	2
	83.33	229	16	43	6
	88.09	239	20	25	10
	88.77	246	15	28	5

Table 4. Accuracy of algorithms for K fold validation where K=10

Model Type	Accuracy (%)	True negatives	True positives	False negatives	False positives
CatBoost	86.39	119	8	18	2
	87.07	126	2	19	0
	89.11	123	8	14	2
	87.07	124	4	17	2
	85.71	116	10	19	2
	79.59	113	4	26	4
	87.75	122	7	16	2
	90.47	124	9	13	1
	86.39	122	5	19	1
	89.11	124	7	12	4

LightBoost	83.67	118	5	21	3
	88.43	125	5	16	1
	88.43	121	9	13	4
	87.07	124	4	17	2
	84.35	114	10	19	4
	77.55	111	3	27	6
	85.71	119	7	16	5
	88.43	122	8	14	3
	86.39	121	6	18	2
	89.79	124	8	11	4
XGBoost	84.35	118	6	20	3
	87.07	126	2	19	0
	89.79	123	9	13	2
	87.75	124	5	16	2
	84.35	114	10	19	4
	80.27	114	4	26	3
	87.75	120	9	14	4
	87.07	120	8	14	5
	87.07	122	6	18	1
	90.47	124	9	10	4

4.7 ROC Curves

The ROC curve depicts the classifier's overall 'tends to cover.' It calculates the predictor's likelihood to score a randomly picked optimistic instance better than a selected randomly unfavorable instance. The classification performed best when the curve was closer to the top left corner.

The area under the receiver operating characteristic curve was used as the model evaluation approach in this work (ROC-AUC). A classifier's AUC is similar to the possibility that the classification would score a selected randomly good model better than a randomly chosen unfavorable instance [19]. Furthermore, the accuracy and F1 scores of the classifications are performed to compare the model results. Both are significant as they demonstrate how well the approach is suited for usage in a specific context.

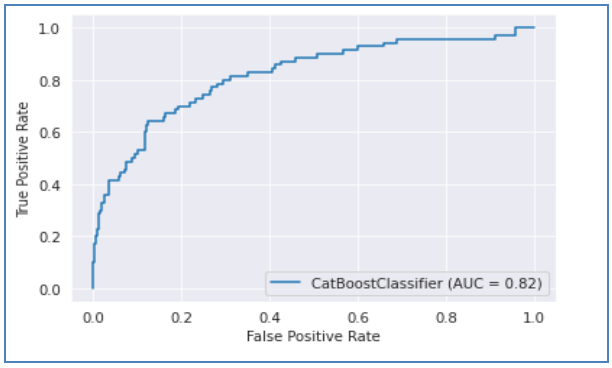

Fig. 10. Cat Boost 70 30 splits with 87.52% accuracy

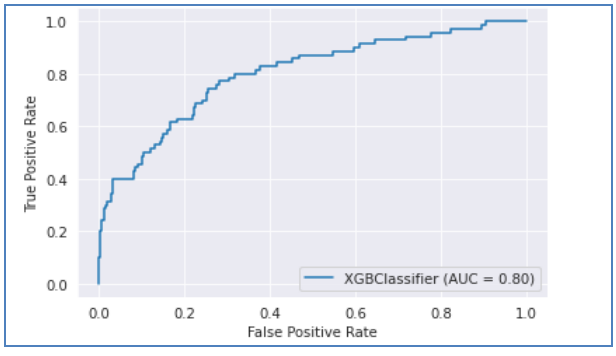

Fig. 11. XGBoost 70 30 split with 87.30% accuracy

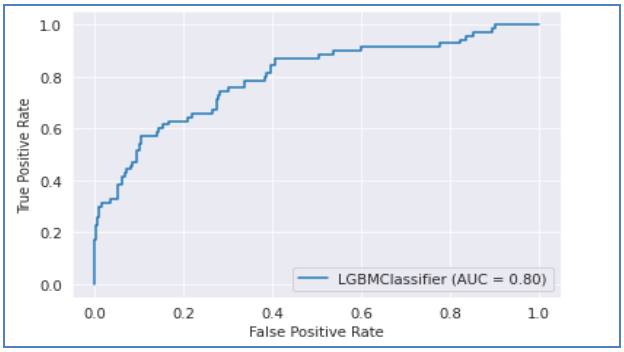

Fig. 12. Light GBM 70 30 splits with 87.75% accuracy

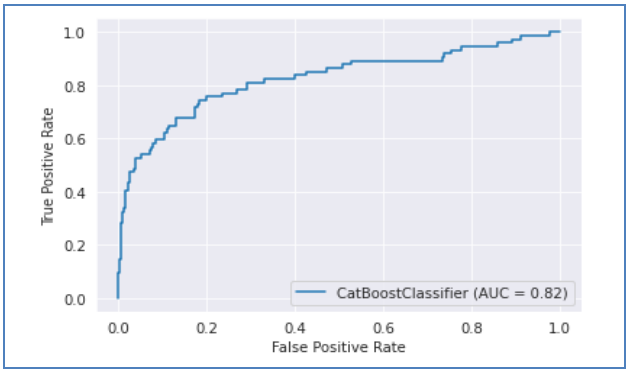

Fig. 13. Cat Boost K 3 split with 88.97% accuracy

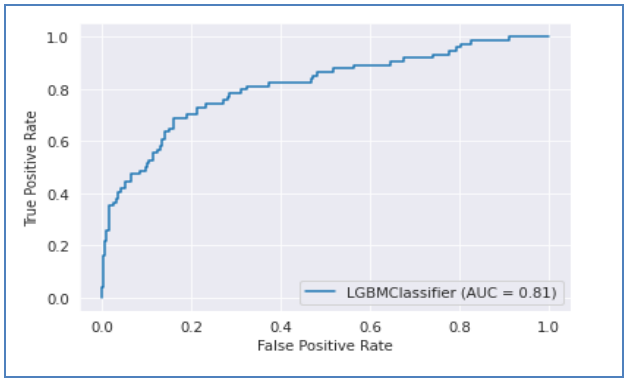

Fig. 14. Light GBM K 3 split with 88.57% accuracy

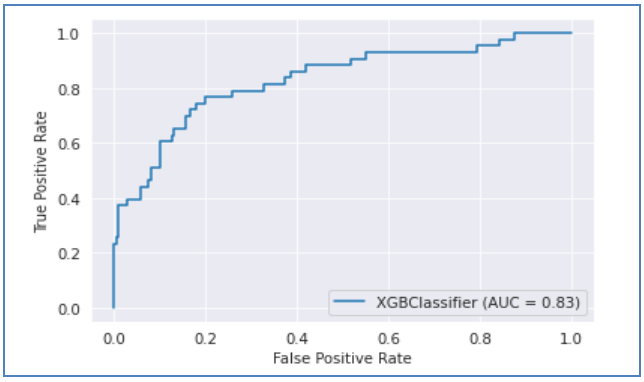

Fig. 15. XGBM K5 split with 89.79% accuracy

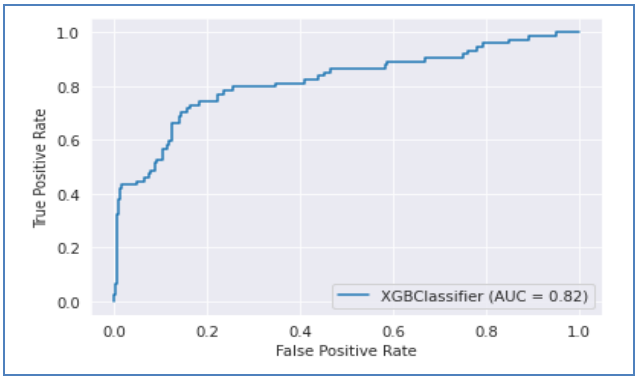

Fig. 16. XGBM K 3 split with 90.00% accuracy

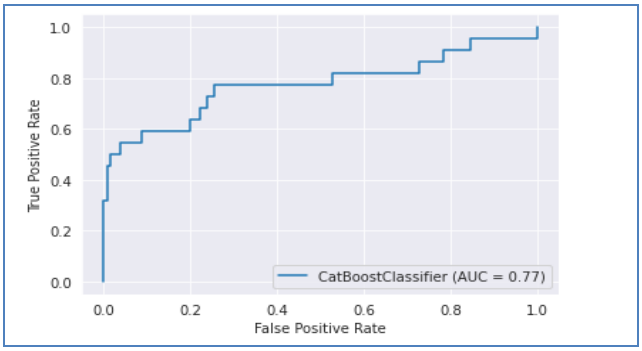

Fig. 17. Cat Boost K 10 split with 90.47% accuracy

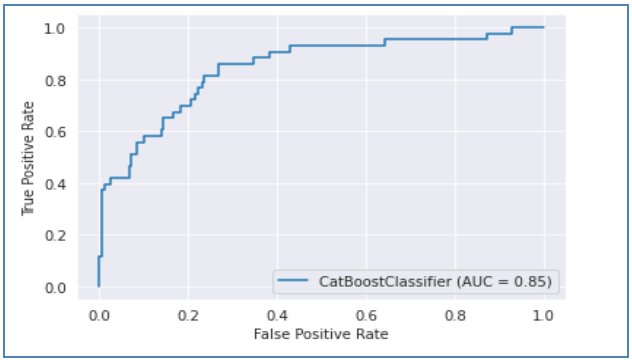

Fig. 18. CatBoost K5 split with 90.13% accuracy

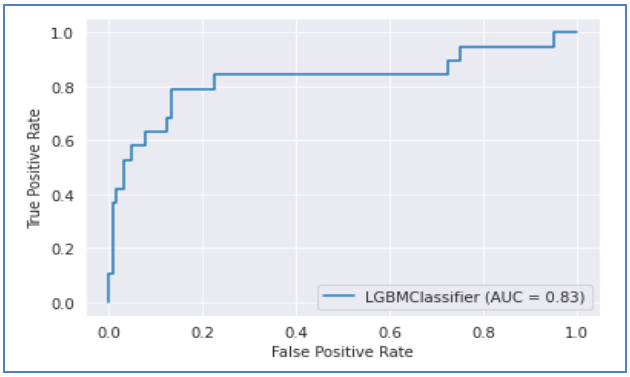

Fig. 19. Light GBM K 10 split with 89.79% accuracy

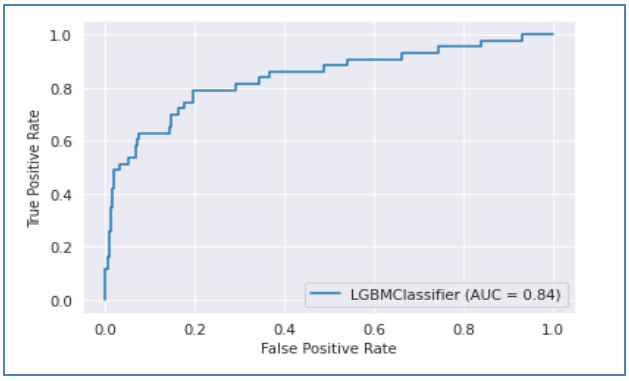

Fig. 20. Light GBM K 5 split with 89.11% accuracy

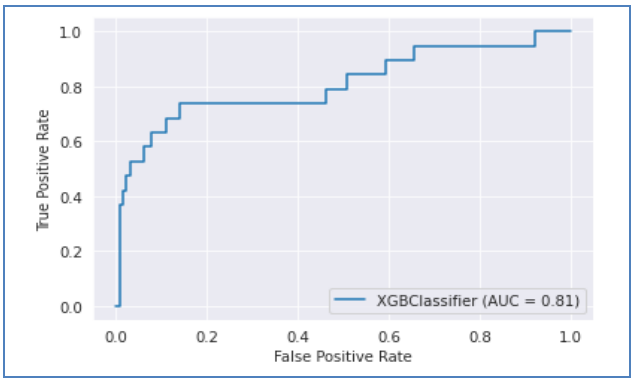

Fig. 21. XGBM K 10 split with 90.47% accuracy

5. Discussion

In the impact of employee turnover, an estimate may be made as to whether or not the person will depart the organization. In this approach, the business can select the individuals who have the highest likelihood of resigning and then provide specific benefits. There may also be situations of misclassification, in which human resource management believes an employee will leave the job within a short amount of time but the individual doesn't really. These blunders may be costly and inconvenient for both workers and talent management, but they are a better bargain for interpersonal growth. On the other side, there can be a short-term problem if an hr department does not motivate staff or raises and subsequently quits the firm. Such talent acquisition error is hazardous to the organization because this not only loses of one worker but also needs to recruit replacement workers and incur the costs of selecting and training. In this context, organizations might categorize the form of help relying on different kinds of staff earnings depending on this circumstance. If an individual is paid well, the institution's appreciation of him could be excessive.

Furthermore, the expense of motivation should be assessed appropriately. The workforce retention prediction issue is focused on people’s beliefs. Also, on the talent acquisition dataset, multiple machine learning algorithms were used in this paper.

6. Conclusion

Using the three boosting algorithms, namely Cat Boost, Light GBM, and XGBoost with 70% train and 30% test dataset split, the Light GBM gave us a more accurate model than the other two algorithms. When K Fold validation is used for algorithms, Cat Boost, Light GBM, and XGBoost have more accurate models with 90.47% accuracy for both the algorithms, namely Cat Boot and XGBoost K=10 in K Fold validation. As the employee is essential to the company, develop a model using Deep Learning and increase the model's accuracy and predictions. Continuous Integration Continuous Deployment, employed for the model, can immediately learn in deployment for the wrong predictions, leading to a more accurate model to predict employee attrition.

7. Future Scope

In the future, the recommended models will be implemented in real situations to discover the causes behind employee sales. Research should also rely on a scientific methodology to observe statistical analysis by applying deep learning methods, a component of HR analytics in an organization.

References

[1] https://www.digitalhrtech.com/employee-attrition.
[2] https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/documents/retaining-talent.pdf.
[3] Pestano, T. (2018). Manila Recruitment. Retrieved from https://manilarecruitment.com/manilarecruitment-articles-advice/risks-in-recruitment-process-and-how-to-mitigate-them/.
[4] Alduayj, S. S. & Rajpoot, K. 2018. Predicting Employee Attrition using Machine Learning. International Conference on Innovations in Information Technology. DOI: 10.1109/INNOVATIONS.2018.8605976.
[5] https://www.clearpeaks.com/predicting-employee-attrition-with-machine-learning-using-knime/
[6] V. Kakulapati, et al. (2020) Predictive analytics of HR - A machine learning approach, Journal of Statistics and Management Systems, 23:6, 959-969, DOI: 10.1080/09720510.2020.1799497.
[7] Usha, P.; et al. Analysing Employee attrition using machine learning. Karpagam J. Comput. Sci. 2019, 13, 277–282.
[8] Cotton, J.L. and Tuttle, J.M., 1986. "Employee turnover: A meta-analysis and review with implications for research" Academy of management review, pp.55-70.
[9] Liu, D., Mitchell, T.R., Lee, T.W., Holtom, B.C. and Hinkin, T.R., 2012. "When employees are out of step with coworkers: How job satisfaction trajectory and dispersion influence individual and unit-level voluntary turnover." Academy of Management Journal, pp.1360-1380.
[10] Heckert, T.M. and Farabee, A.M., 2006. "Turnover intentions of the faculty at a teaching-focused university." Psychological reports, pp.39-45.
[11] https://expertsystem.com/machine-learning-definition/
[12] Tom Mitchell, 1997, Machine Learning, McGraw Hill.
[13] Chen, T., et al. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining, pp. 785–794, ACM (2016).
[14] Raschka, S.: Python Machine Learning. Packt Publishing Ltd, Birmingham (2015).
[15] M. Stoval et al. "Voluntary turnover: Knowledge management – Friend or foe?", Journal of Intellectual Capital, 3(3), 303-322, 2002.
[16] Rohit Punnoose et al. "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms- A case for Extreme Gradient Boosting," International Journal of Advanced Research in Artificial Intelligence, Vol. 5, No. 9, 2016, pp:22-26.
[17] T. Chen and C. Guestrin, "XGBoost: Reliable Large-scale Tree Boosting System, 2015", Retrieved from http://learningsys.org/papers/LearningSys_2015_paper_32.pdf. Accessed 12 December 2015.
[18] Ozden G ¨ur Ali and Umut Arit ¨urk. Dynamic churn prediction framework with more ¨effective use of rare event data: The case of private banking. Expert Syst. Appl., 41:7889– 7903, 2014.
[19] Fawcett, T., 2006. An introduction to ROC analysis. Pattern recognition letters, pp.861-874.

Authors' Profiles

Prof. Vijayalakshmi Kakulapati received a Ph.D. in Computer Science & Engineering in Information Retrieval from JNTU Hyderabad. She works as a Professor in the Department of Information Technology, Sreenidhi Institute of Science and Technology, and has around 25 years of industry and teaching experience. She is a member of various professional bodies like IEEE, ACM, CSTA, LMISTE, LMCSI, IACSIT, FIETE, and professional organizations like big data University, etc. She has more than 170+ publications in international journals and conferences, out of which 50+ are in Springer, 4 ACM, 8 IEEE, and 5 in Elsevier. She has 4 granted patents and 5 published patents. She has authored 4 books and 30 book chapters. She took an active role in National Conferences and International Conferences as Session chair. She received more than 14 awards from different organizations. She was awarded a 3 lakhs project from JNTUIII TEQUIP CRC and a 1 lakh fund for the DST workshop. She serves as a review board member for the Journal of Big Data, IAJIT, *IEEE* Transactions on Computational Social Systems, and many more. She serves as an editorial board member of PLOS One, IJCNS, and more. Currently, she is working with big data analytics, health informatics, Data Science, Artificial Intelligence, Deep learning, machine learning, and the Internet of Things.

Dr. Shaik Subhani received his Ph. D in Computer Science and Engineering from ANU Guntur. His research focus is on Data Mining, Image Processing applications, Image Retrieval Systems, Semantic Analysis, Feedback Systems, Social Network Analysis, Face Recognition Systems, Machine Learning, Database Systems and Data Mining, ,Big Image data analysis, Computer Networks and Security, Mobile Adhoc and Sensor Networks, Green Computing and Communications, Cloud Computing and Information Retrieval Systems. He has published more than 50 Research papers in international journals and 20 papers published in National and International conferences. His areas of interest are Data Mining, Soft Computing and Image Processing. He received best teacher award in 2015-16 academic year

How to cite this paper: V. Kakulapati, Shaik Subhani, "Predictive Analytics of Employee Attrition using K-Fold Methodologies ", International Journal of Mathematical Sciences and Computing(IJMSC), Vol.9, No.1, pp. 23-36, 2023. DOI: 10.5815/ijmsc.2023.01.03