Employee Attrition: Analysis of Data Driven Models

Manju Nandal[1,*], Veena Grover[2], Divya Sahu[3], Mahima Dogra[4]

[1, 2, 3, 4] Assistant professor, Noida Institute of Engineering & Technology, Greater Noida

Abstract

Companies constantly strive to retain their professional employees to minimize the expenses associated with recruiting and training new staff members. Accurately anticipating whether a particular employee is likely to leave or remain with the company can empower the organization to take proactive measures. Unlike physical systems, human resource challenges cannot be encapsulated by precise scientific or analytical formulas. Consequently, machine learning techniques emerge as the most effective tools for addressing this objective. In this paper, we present a comprehensive approach for predicting employee attrition using machine learning, ensemble techniques, and deep learning, applied to the IBM Watson dataset. We employed a diverse set of classifiers, including Logistic regression classifier, K-nearest neighbour (KNN), Decision Tree, Naïve Bayes, Gradient boosting, AdaBoost, Random Forest, Stacking, XG Boost, "FNN (Feedforward Neural Network)", and "CNN (Convolutional Neural Network)" on the dataset. Our most successful model, which harnesses a deep learning technique known as FNN, achieved superior predictive performance with highest Accuracy, recall and F1-score of 97.5%, 83.93% and 91.26%.

*Corresponding author. Email: nandalmanju93@gmail.com

1. Introduction

The competitiveness among organizations and companies hinges significantly on workforce productivity. Creating and sustaining an appropriate environment is the essential factor that fosters stable and cooperative employees. The Human Resource (HR) department plays a pivotal role in shaping such an environment through the analysis of employee database records [1]. A robust workforce contributes to heightened productivity, cost-efficiency, and overall profitability for a company. These benefits are unattainable without the pivotal role played by human resources. When an organization struggles to retain its employees, it can lead to sustained losses over the long term. This phenomenon, often referred to as employee attrition [2], is a phenomenon within an organization wherein employees choose to depart for a variety of reasons. These factors may encompass personal or professional motivations, an unsuitable work environment, excessively long office hours, and inadequate compensation. This deliberate decision to leave, initiated by the employees themselves, is commonly referred to as voluntary attrition. The primary aim of HR departments is to comprehend the underlying reasons for voluntary employee attrition and formulate a corresponding strategy for mitigation. Recognizing and harnessing the existing talent within an organization stands as one of the foremost challenges and critical priorities in talent management. In any organization, human resources assume a pivotal role in shaping strategic decisions. Contented, deeply motivated, and committed employees form the bedrock of a company, subsequently influencing the productivity of the entire organization.

The role of the Human Resources (HR) department in fostering such an atmosphere is pivotal, and it is achieved through the thorough examination of employee database records. This analysis equips the administration with the tools to improve decision-making processes, effectively addressing the challenge of employee attrition. Historically,

inquiries related to employee attrition and retention have been approached through qualitative and anecdotal methods. Typically, HR personnel conduct exit interviews when an employee tenders their resignation, aiming to uncover the underlying reasons behind their departure. In the current age marked by the fourth industrial revolution, powered by advanced technologies such as predictive analytics that employ statistical modelling techniques and machine learning, predicting the likelihood of an employee departing from an organization is now within reach. Organizations utilize machine learning algorithms to forecast the probability of employee attrition and proactively implement measures to prevent such occurrences [3].

Machine learning represents a facet of artificial intelligence (AI) technology that equips systems with the capability to autonomously acquire knowledge and refine their performance through experience, mirroring human-like intelligence without the need for explicit programming [4]. Machine learning (ML) stands as one of the most rapidly advancing research fields, showcasing successful development and application across a diverse array of real-world domains. Due to the expenses associated with hiring employees, providing training, and acquiring intellectual property, it becomes paramount to ensure a minimal attrition rate (employee turnover) within organizations [5]. Employee attrition imposes significant financial burdens on a company, encompassing expenses such as business disruption costs, recruitment and onboarding of new employees, and training of newcomers [6]. While recruiting top talent is vital for organizations, it is equally crucial to ensure their satisfaction and retention. Employees have their unique criteria for selecting and committing to an organization, and if their expectations are not met, they may choose to resign. This can result in employee attrition, often referred to as the phenomenon of employee churn [7]. Lately, leading companies such as IBM, HCL, TCS, and others have grappled with employee attrition challenges. By gathering employee feedback regarding various aspects, including the company's culture, work environment, workload, job satisfaction, and more, organizations can employ statistical methods to predict attrition status.

Hence, attrition must be dealt with utmost importance and measures must be taken by organizations to prevent this [4]. Consequently, forecasting employee attrition and pinpointing the key factors that contribute to attrition emerge as crucial objectives for organizations seeking to bolster their human resource strategies. This paper delves into the application of classification and clustering techniques for analysing attrition. It conducts a comparative assessment to evaluate the accuracy of different data mining algorithms using Weka, a collection of machine learning algorithms employed for data mining purposes. In this study, we employed the IBM Human Resource Analytics Performance dataset and Employee Attrition which is a publicly accessible dataset accessible through the Kaggle Dataset Repository. This dataset was generated by IBM data scientists for research purposes and comprises four primary components: seniority, employee satisfaction, income, and demographic information. Inside the dataset, numerous attributes impact the predictive variable known as 'Attrition'. It consists of a total of 1,470 instances and encompasses 35 attributes, providing a comprehensive dataset for analysis.

2. Related Work

Employee attrition issues were studied by researchers from various viewpoints. Researchers harnessed machine learning techniques to predict employee attrition by analysing data pertaining to the employees. This investigation involved the utilization of several machine learning methods, including Random Forests (RF), Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) while exploring various parameter configurations [8], There are researchers in [9] chose to utilize Classification Trees and Random Forest for the purpose of predicting employee attrition. Their approach commenced with dataset pre-processing, where they excluded less influential variables based on Pearson correlation analysis.

A study utilizing the [4] IBM HR Employee Attrition & Performance dataset revealed an inherent data imbalance issue. During the data exploration phase, the researchers employed correlation plots and histogram visualizations to assess the relationships among continuous variables in the model. Following this analysis, the "Synthetic Minority Oversampling Technique (SMOTE)" was utilized to rectify the imbalance within the Attrition class [10]. To tackle the challenge of predicting employee turnover, we introduced a novel approach: a weighted quadratic random forest algorithm. The algorithm was utilized with a dataset of employees gathered from a branch of a telecommunications company located in China [10]. The researchers presented a comprehensive three-stage framework for predicting attrition. In the first stage, they applied the "max-out" feature selection method to refine the data. Following this, in the second stage, a logistic regression model was trained for prediction. Finally, the third stage involved conducting confidence analysis to enhance the reliability of the prediction model. However, it's worth noting that the system faces challenges, including suboptimal accuracy and elevated complexity due to the preprocessing and postprocessing step [11]. Taylor et al. [7] Tree-based models, specifically light Gradient Boosted Trees and random forests were utilized to make predictions regarding employee attrition. These models demonstrated robust performance, with the light gradient boosted trees exhibiting particularly strong results. The study utilized a custom dataset comprising 5550 samples for their analysis. Machine learning serves a wide array of applications, encompassing tasks from prediction to the classification of various HR data parameters and features [12] the study focuses on the early prediction of employee turnover, considering variables like absenteeism, tardiness, and employee indifference as significant factors influencing employee performance forecasting. Fallucchi et al [13] conducted research and used a variety of machine learning approaches to identify the circumstances that may cause an employee to leave the organization. The best recall value was provided by the

Gaussian Nave Bayes classifier, which contributes to the classifier's capacity to detect positive occurrences. The study [14] provided a hybrid model for anticipating client attrition.

3. Dataset Description

The study used a dataset of 1,470 instances, which comprised detailed information about all employees and 35 features, including the target class. When the gender variable was examined, it was determined that 60% of the employees were male and 40% were female. Surprisingly,

two critical characteristics examined in our trials were job satisfaction and job involvement. Attrition affected approximately 28.29% of employees with low job satisfaction or job involvement. Furthermore, approximately 31.25% of employees with an unfavourable work-life balance left the organization, compared to 17.65% of departing employees with a good work-life balance. Figure 1 shows the correlation between the target variable i.e., employee attrition and other variables Heatmap analysis reveals a strong correlation between job satisfaction, overtime, job level, monthly income, job involvement and age with respect to attrition in the dataset.
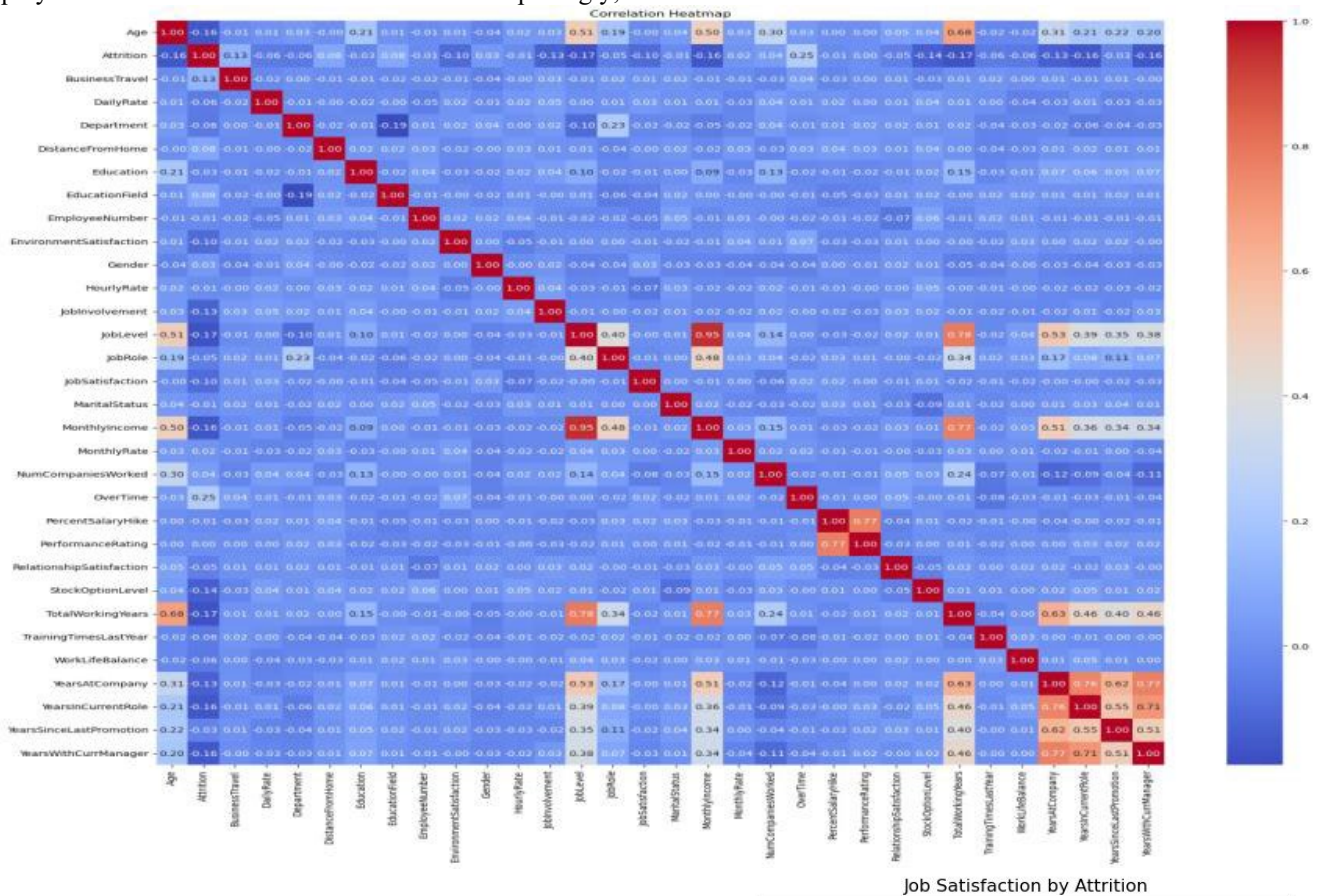

Figure 1: Dataset

4. Data Analysis

Based on heatmap the highly correlated attributes with target variable employee attrition are Overtime, job satisfaction, job level, monthly income, age and job involvement. The Figure 2 shows the bar plot of the correlated variables with the target variable i.e. Employee attrition.

According to the plots employee attrition is higher when overtime is increased, employee attrition is lower when job satisfaction, job level, age and job involvement is higher.
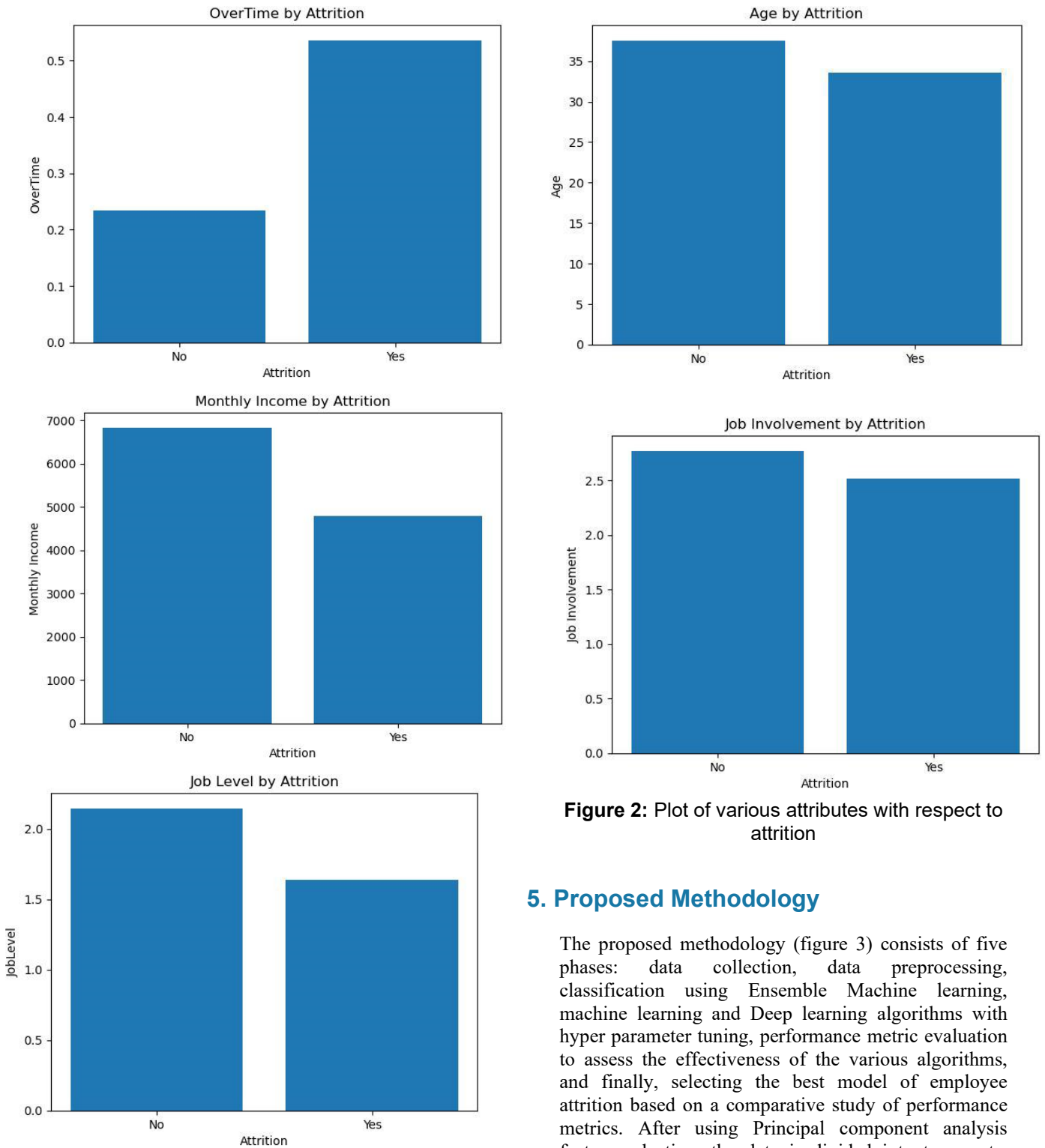

Job Satisfaction by Attrition
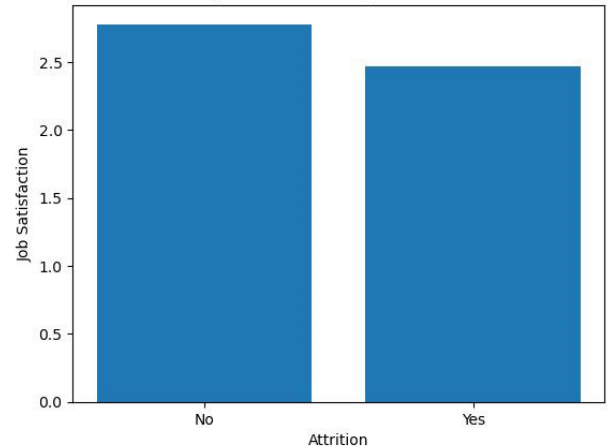
Figure 2: Plot of various attributes with respect to attrition

5. Proposed Methodology

The proposed methodology (figure 3) consists of five phases: data collection, data preprocessing, classification using Ensemble Machine learning, machine learning and Deep learning algorithms with hyper parameter tuning, performance metric evaluation to assess the effectiveness of the various algorithms, and finally, selecting the best model of employee attrition based on a comparative study of performance metrics. After using Principal component analysis feature selection, the data is divided into two parts: 75% for testing and 25% for training.
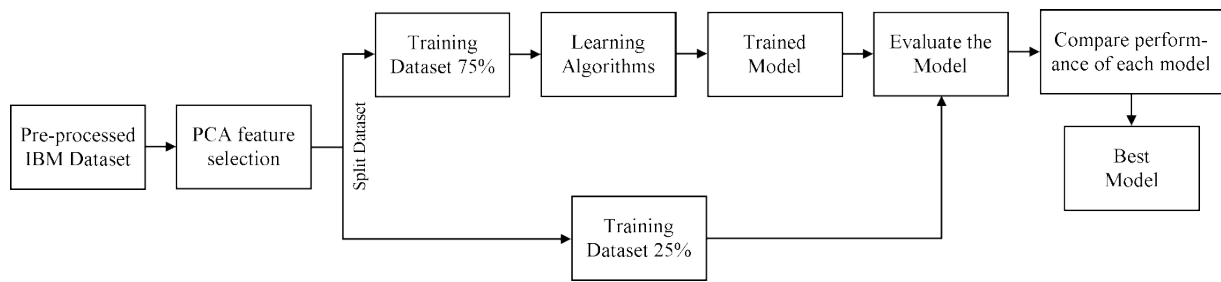
Figure 3: Proposed Methodology

5.1 Data Encoding

To enable machine learning algorithms to process categorical features such as 'Department,' 'Education,' 'Gender,' and 'Work-Life Balance' from the dataset, it was imperative to transform them into numerical representations. To accomplish this, Label Encoding was employed to create numeric representations for these categorical features. Following the feature selection process, these categorical features were further transformed into distinct binary columns containing values of 0 and 1. This expansion of dimensions was achieved by generating a separate column for each unique value present in every original column within our dataset.

5.2 Model Training

Furthermore, the dataset is pre-processed to make it appropriate for model training. Following data pre-processing, the model moves on to the training phase, where the dataset is divided into 75% training and 25% test sets. Following that, data modelling occurs. Support Vector Machines (SVM), Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Gaussian Naive Bayes (GNB), Logistic Regression (LR), and K-Neighbors (KNN) are among the machine learning algorithms used to determine the algorithm with the best accuracy. Multiple graphs were used to provide a more detailed study of the results and to visualize the relationships between different variables.

5.3 Feature scaling

In HR datasets, it is common for features to exhibit varying scales. For instance, in the IBM Attrition dataset, employee ages span a range from 18 to 60 years, while monthly income varies from $2,094 to $26,999. However, such significant disparities in feature scales can impede the efficiency of optimization algorithms like gradient descent. Consequently, feature scaling plays a pivotal role in potentially enhancing both classification performance and learning efficiency in certain machine learning algorithms.

5.4 Dataset Preprocessing

There are 1470 occurrences and 34 attributes with no missing values in the dataset. As dataset contains no missing value so we have applied Min- Max scaling to scale numerical features within a specific range. This technique transforms the data in such a way that it falls within a specified range, often between 0 and 1.
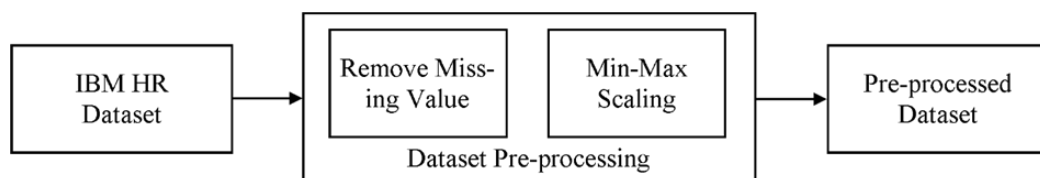

Figure 4: Data Pre – processing Min- Max Scaling

5.5 Machine Learning Models

5.5.1. Support Vector Machine: Support Vector Machine (SVM) is indeed a technique used for classification tasks. It constructs a hyperplane to separate two classes and aims to maximize the margin between this hyperplane and the closest data points from each class. These data points are represented as vectors in a high-dimensional space, and SVM finds the hyperplane that best separates them.

5.5.2 logistic regression: It is a powerful analytical tool for predicting binary outcomes. It transforms the relationship between independent and dependent variables into probabilities, facilitating the assessment of event occurrence likelihood. Moreover, it provides a range of performance metrics that aid in evaluating and fine-tuning the model's predictive capabilities. Some of the key results that can be derived from logistic regression include accuracy, recall, F1 score, ROC (Receiver Operating Characteristic) curve, precision, and the construction of a confusion matrix. These metrics help assess the model's performance, its ability to discriminate between the two classes, and its precision in predicting outcomes.

5.5.3 K-Nearest Neighbor (KNN): It is a straightforward algorithm that relies on the storage of all available data cases to classify new, unseen data points. KNN is often referred to as a "Lazy Learner" because it lacks a discriminative function derived from the training data. Instead, it retains and memorizes the entire training dataset without undergoing a traditional model learning phase.

5.5.4 Naive Bayes: It is a probabilistic machine learning technique used for text classification and classification. It is based on Bayes' theorem, which calculates the likelihood of a specific event occurring based on past knowledge of conditions that may be associated with the event. The chance that a given data point (such as a document or an item) belongs to a specific class or category is calculated using Naive Bayes. It is assumed that the features used for categorization are conditionally independent, which means that the presence or absence of one trait has no bearing on the presence or absence of another. This is a "naive" assumption that simplifies calculations and allows the algorithm to be more tractable.

5.5.6 Decision Tree: It is a graphical representation resembling a tree that helps in the decision-making process. Each branch of the tree represents a potential decision, event, or response. Decision Trees can be employed for both classification and regression tasks. In classification, they are used to categorize data into discrete classes, whereas in regression, they predict numerical or continuous values.

5.6 Ensemble Machine learning

5.6.1 Gradient Boosting: Gradient Boosting Model is a machine learning ensemble technique used primarily for supervised learning tasks, such as classification and regression. It is designed to improve the predictive accuracy of a model by combining the predictions of multiple weaker models (typically decision trees) into a more powerful and accurate ensemble model.

5.6.2 AdaBoost: Ada Boost is a machine learning algorithm that uses boosting to improve the performance of weaker learners [27]. To begin, an initial classifier is trained on the original dataset. Following then, new copies of the classifier are trained over several iterations, each with the explicit objective of fixing errors caused by its predecessor. Various subsets of the dataset are formed during these cycles by assigning variable weights to individual data components. Instances that were misclassified in previous rounds are given higher weights, improving their chances of inclusion in subsequent subgroups. This iterative procedure is repeated numerous times, resulting in the sequential training of several models. To produce a robust classifier, these initially weaker classifiers are integrated using a specified cost function. The accuracy of each individual classifier influences the final prediction, with higher accuracy classifiers bearing greater weight in the ensemble. Random Forest is a classification system based on decision tree concepts. This method, true to its name, creates a forest out of several individual trees. It is under the umbrella of the ensemble algorithm category, which includes techniques that create predictions by combining various algorithms.

5.6.3 Random Forest: Random Forest constructs an ensemble of decision trees from random subsets of the training dataset. This method is done iteratively with different random subsets, with a majority consensus among these trees determining the conclusion.

5.6.4 Stacking model: Stacking model also known as stacked generalization, or stacking ensemble, is an advanced machine learning technique used for improving predictive performance. It combines the predictions of multiple base models (often diverse in nature) by training a meta-model, or "stacker," on top of them. Stacking can significantly improve predictive performance compared to individual base models because it leverages the strengths of different models and combines them to produce a more robust and accurate prediction. It is a powerful technique in machine learning and is often used in competitions and real-world applications where achieving the best possible predictive accuracy is crucial.

5.6.5 XG Boost: XG Boost stands for Extreme Gradient Boosting, is a highly popular and powerful machine learning algorithm that is widely used for both regression and classification tasks. It belongs to the

ensemble learning family and is specifically designed to improve the accuracy and efficiency of decision tree-based models.

5.7 Deep Learning Models

5.7.1 Feedforward Neural Network (FNN): A Feedforward Neural Network (FNN), sometimes known as a Multilayer Perceptron (MLP), is a deep learning artificial neural network. Its architecture is distinguished by many layers of neurons, including an input layer, one or more hidden layers, and an output layer. A Feedforward Neural Network (FNN) is a deep learning model made up of layers of artificial neurons that are coupled. It is intended to process and transform incoming data through a sequence of mathematical operations, creating an output in the end.

5.7.2 Convolutional Neural Network (CNN): "Convolutional Neural Network (CNN)" is a deep learning model specifically tailored for tasks involving visual data, characterized by its use of convolutional layers to automatically learn and extract features from images or other grid-like data. A "Convolutional Neural Network (CNN)" is a deep learning model specifically tailored for tasks involving visual data, characterized by its use of convolutional layers to automatically learn and extract features from images or other grid-like data. The following metrics are examined to determine a model's effectiveness.

To evaluate the effectiveness of a model the following metrics are examined:

Accuracy: It is a performance metric used to assess the model's overall effectiveness when all classes carry equal significance. It is calculated as the ratio of correctly predicted instances to the total number of predictions made. This metric provides a measure of how well the model performs across all classes.

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} (1)$$

Recall, also known as sensitivity or true positive rate, measures the model's ability to correctly recognize positive samples. It is derived by dividing the total number of positive samples by the number of correctly categorized positive samples. A greater recall value suggests that the model accurately identifies more positive samples.

$$Recall = \frac{t_p}{t_p + + f_p} (2)$$

Precision is a performance statistic that assesses the model's ability to categories positive samples properly. It is derived by dividing the total number of positive samples by the number of correctly categorized positive samples. that the model predicted as positive, whether they were classified correctly or incorrectly. Precision gauges how effectively the model identifies true positives among all positive predictions.

$$Precision = \frac{t_p}{t_p + + f_n} (3)$$

F1-score: The F1 score is a well-defined metric that represents the harmonic mean of precision and recall .

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

Where: tp is correctly predicted, fp is incorrectly predicted instances, tn is negatively predicted instances and fn is the negatively predicted instances.

6. Results

This section provides an analysis of the results obtained from different machines and deep learning classification models. The aim of this study is to evaluate the classification effectiveness of both machine learning and deep learning algorithms when applied to the task of categorizing employee attrition. In this study, a wide array of learning algorithms was utilized and assessed using the employee attrition dataset. The ML algorithms encompassed traditional methods such as SVM, logistic regression, KNN, decision tree, and naive Bayes. Additionally, ensemble machine learning algorithms, including Gradient boosting, XG-Boost, AdaBoost, random forest, and stacking, were employed. Furthermore, the study also incorporated deep learning techniques, specifically "Convolutional Neural Networks (CNN)" and feedforward neural network (FNN). To assess the performance of these models, multiple evaluation metrics were employed namely recall, F1 score, precision, accuracy, area under the ROC and precision-recall curve. The evaluation of results includes the use of performance metrics such as recall (Sensitivity), F1-score, precision, accuracy, and AUC, with the corresponding scores detailed in Table 3.

6.1. Machine Learning Models

Table 1 provides an extensive assessment of the machine learning models. Out of the various models evaluated, the Naïvebayes model demonstrated superior performance, achieving an accuracy and F1-Score of 0.541 % and 0.908% respectively. Furthermore, both the Naïve bayes and Logistic Regression models provide highest precision, reaching at 0.769 and 0.727 respectively.

Table 1: Performance analysis of Machine Learning model for Employee Attrition

Model	Accuracy	Precision	F1-Score	Recall
SVM	0.878	0.636	0.236	0.146
KNN	0.870	0.500	0.200	0.125
Naïve bayes	0.908	0.769	0.541	0.417

Decision Tree	0.861	0.440	0.301	0.229
Logistic Regression	0.897	0.727	0.457	0.333

6.2. Ensemble Machine Learning Models

Table 2 provides a depth analysis of Ensemble machine learning models. Among these models, the Stacking model achieved the highest recall, accuracy, and F1-Score respectively. Additionally, Gradient boosting, AdaBoost and XG- Boost techniques exhibited good accuracy levels.

Table 2: Performance analysis of Ensemble Machine Learning Model for Employee Attrition

Model	Accuracy	Precision	F1-Score	Recall
Gradient Boosting	0.894	0.765	0.400	0.271
AdaBoost	0.883	0.619	0.377	0.271
Random Forest	0.872	1.00	0.41	0.021
Stacking	0.899	0.789	0.448	0.312
XG-Boost	0.886	0.650	0.382	0.271

6.2. Deep learning Models

Table 3 provides a detailed analysis of Deep learning models. Among deep learning models, the FNN model outperforms with highest Accuracy, recall and F1-score of 97.5%, 83.93% and 91.26%, respectively. Furthermore, FNN exhibited the highest precision score of 100%.

Table 3: Performance analysis of Deep Learning Model for Employee Attrition

Model	Accuracy	Precision	F1-Score	Recall
FNN	0.975	1.00	0.9126	0.8393
CNN	0.942	0.8148	0.8000	0.7857

6.3. Performance of the Models based on ROC plot.

The Receiver Operating Characteristic curve shows how the threshold change affects the connection between true positive rate and false positive rate. The ROC curve provides an assessment of the classifier's overall predictive performance. It quantifies the likelihood that the classifier will assign a higher rank to a randomly selected positive instance compared to a randomly selected negative instance. The closer the curve approaches the top-left corner, the more effective the classifier is. The ROC curves for each Machine learning classifier are shown in Figure 5. Ensemble Machine learning in Figure 6. and Deep learning classifier in Figure 7.

In machine learning, logistic regression has the highest ROC score of 0.827. In ensemble machine learning methods, gradient boosting does even better with a ROC score of 0.84. However, in the field of deep learning, the Feedforward Neural Network (FNN) outperforms them all, with the highest ROC score of 0.92.
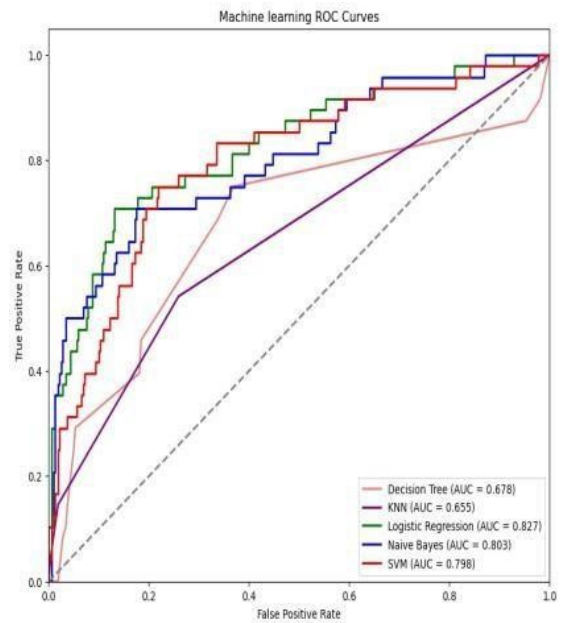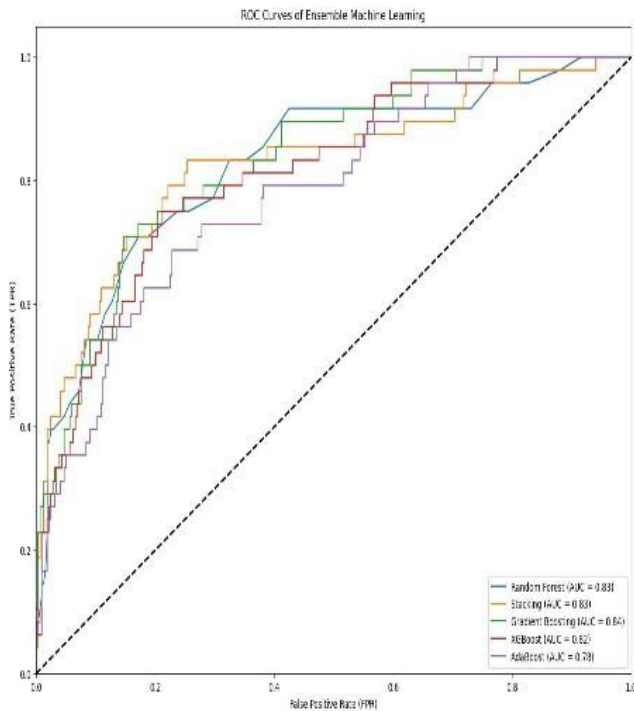

Fig5. ROC Plot for ML models.
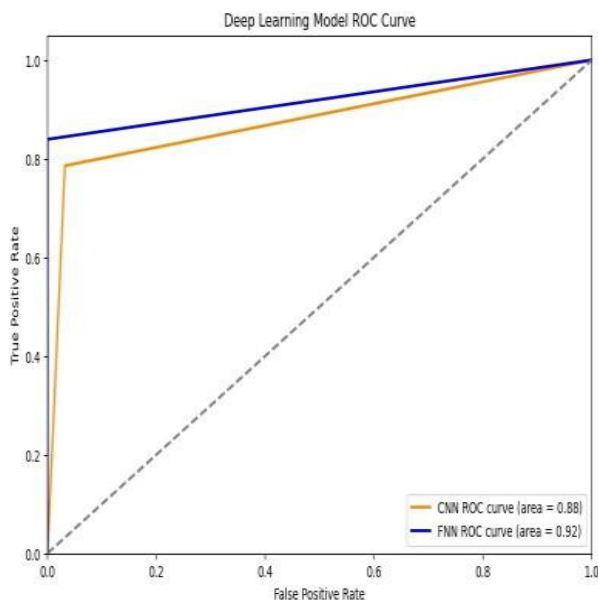
Fig6. ROC Plot for Ensemble Models

Fig7. ROC Plot for DL models.

7. Comparison of existing work

Author	Dataset	Method	Result (Accuracy)	Proposed Work (Accuracy)
[12]	IBM Dataset	KNN	86.00 %	87.00%
[3]	IBM Dataset	Naïve Bayes	80.9%	90.8%
[14]	IBM Dataset	Logistic Regression	79.60	89.7%
[15]	IBM Dataset	Decision tree	82.31%	86.1%
[16]	IBM Dataset	AdaBoost	87.03%	88.3%
[17]	IBM Dataset	Decision tree	84.73%	86.1%

Conclusion:

This paper explored the impact of voluntary attrition on organizations and underscored the significance of predictive modeling in addressing this issue. It provided an overview of various supervised learning classification algorithms employed to tackle the problem of predicting employee attrition, using the IBM HR dataset for evaluation. Initially, five foundational models were trained and assessed. Subsequently, five ensembles were created by leveraging various combinations of these five base models. Two deep learning models were tested. The findings revealed that linear models outperformed others in terms of accuracy, recall, and AUC. Furthermore, deep learning models, particularly the FNN approach, exhibited exceptional accuracy. In contrast, other machine learning models displayed a wider range of accuracy, spanning from 86% to 94%. These results emphasize the potential of both deep learning and ensemble machine learning techniques in achieving high classification accuracy. As a result, the authors recommend employing the FNN classifier for precise predictions of employee attrition within an organization. This approach empowers HR to take proactive measures in retaining employees identified as being at risk of leaving.

References

[1] Mohbey, K.: Employee's Attrition Prediction Using Machine Learning Approaches. Presented at the January 1 (2020). https://doi.org/10.4018/978-1-7998-3095-5.ch005.

[2] Alduayj, S.S., Rajpoot, K.: Predicting Employee Attrition using Machine Learning. In: 2018 International Conference on Innovations in Information Technology (IIT). pp. 93–98 (2018). https://doi.org/10.1109/INNOVATIONS.2018.8605976.

[3] Yedida, R., Reddy, R., Vahi, R., Jana, R., Gv, A., Kulkarni, D.: Employee Attrition Prediction. (2018).

[4] Mansor, N., S Sani, N., Aliff, M.: Machine Learning for Predicting Employee Attrition. Int. J. Adv. Comput. Sci. Appl. 12, 435–445 (2021). https://doi.org/10.14569/IJACSA.2021.0121149.

[5] Abdulkareem, A.B., Sani, N., Sahran, S., Abdi, Z., Alyessari, A., Adam, A., Rahman, A.H.A., Abdulkarem, A.: Predicting COVID-19 Based on Environmental Factors WithMachine Learning. Presented at the (2021).

[6] Pratt, M., Boudhane, M., Cakula, S.: Employee Attrition Estimation Using Random Forest Algorithm. Balt. J. Mod. Comput. 9, (2021). https://doi.org/10.22364/bjmc.2021.9.1.04.

[7] El-rayes, N., Smith, M., Taylor, S.M.: An Explicative and Predictive Study of Employee Attrition using Tree-based Models, https://papers.ssrn.com/abstract=3397445, (2019). https://doi.org/10.2139/ssrn.3397445.

[8] Srivastava, Dr.P., Eachempati, P.: Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi-Criteria Decision-Making Approach. J. Glob. Inf. Manag. 29, 1–29 (2021). https://doi.org/10.4018/JGIM.20211101.oa23.

[9] Yadav, S., Jain, A., Singh, D.: Early Prediction of Employee Attrition using Data Mining Techniques. Presented at the December 1 (2018). https://doi.org/10.1109/IADCC.2018.8692137.

[10] Najafi, S., Shams Gharneh, N., Nezhad, A., Zolfani, S.: An Improved Machine Learning-Based Employees Attrition Prediction Framework with Emphasis on Feature Selection. (2021). https://doi.org/10.3390/MATH9111226.

[11] Gao, X., Wen, J., Zhang, C.: An Improved Random Forest Algorithm for Predicting Employee Turnover. Math. Probl. Eng. 2019, 1–12 (2019). https://doi.org/10.1155/2019/4140707.

[12] Bhatta, S., Zaman, I.U., Raisa, N., Fahim, S.I., Momen, S.: Machine Learning Approach to Predicting Attrition Among Employees at Work. In: Silhavy, R. (ed.) Artificial Intelligence Trends in Systems. pp. 285–294. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-09076-9_27.

[13] Fallucchi, F., Coladangelo, M., Giuliano, R., De Luca, E.: Predicting Employee Attrition Using Machine Learning Techniques. Computers. 9, 86 (2020). https://doi.org/10.3390/computers9040086.

[14] Joseph, R., Udupa, S., Jangale, S., Kotkar, K., Pawar, P.: Employee Attrition Using Machine Learning And Depression Analysis. Presented at the May 6 (2021). https://doi.org/10.1109/ICICCS51141.2021.9432259.

[15] Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R., Alghamdi, H.: Prediction of Employee Attrition Using Machine Learning and Ensemble Methods. Int. J. Mach. Learn. Comput. 11, 110–114 (2021). https://doi.org/10.18178/ijmlc.2021.11.2.1022.

[16] Arqawi, S., Abu Rumman, M.A., Zitawi, E., Rabaya, A., Sadaqa, A., Abunasser, B., Abu-Naser, S.: PREDICTING EMPLOYEE ATTRITION AND PERFORMANCE USING DEEP LEARNING. 100, 6526–6536 (2022).

[17] Kamath, R., Jamsandekar, S., Naik, P.: Machine Learning Approach for Employee Attrition Analysis. Int. J. Trend Sci.Res. Dev. Special Issue, 62–67 (2019). https://doi.org/10.31142/ijtsrd23065.