The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/1934-8835.htm

Machine learningbased approaches for enhancing human resource management using automated employee performance prediction systems

International Journal of Organizational Analysis

Aakanksha Uppal
Symbiosis Law School, Noida,
Constituent of Symbiosis International (Deemed) University, Noida, India

Yashmita Awasthi
School of Commerce, Finance and Accountancy, Christ University, Bengaluru, India, and

Anubha Srivastava
Department of Management, Institute of Technology and Science, Ghaziabad, India

Received 9 July 2024
Revised 28 September 2024
9 October 2024
21 October 2024
Accepted 21 October 2024

Abstract

Purpose – This study focuses on enhancing the accuracy and efficiency of employee performance prediction to enhance decision making and improve organisational productivity. By introducing advance machine learning (ML) techniques, this study aims to create a more reliable and data-driven approach to evaluate employee performance.

Design/methodology/approach – In this study, nine machine learning (ML) models were used for forecasting employee performance: Random Forest, AdaBoost, CatBoost, LGB Classifier, SVM, KNN, XGBoost, Decision Tree and one Hybrid model (SVM + XGBoost). Each ML model is trained on an HR data set covering various features such as employee demographics, job-related factors and past performance records, ensuring reliable performance predictions. Feature scaling techniques, namely, min-max scaling, Standard Scaler and PCA, have been used to enhance the effectiveness of employee performance prediction. The models are trained to classify data, predicting whether an employee's performance meets expectations or needs improvement.

Findings – All proposed models used in the study can correctly categorize data with an average accuracy of 94%. Notably, the Random Forest model demonstrates the highest accuracy across all three scaling techniques, achieving optimise accuracy, respectively. The results presented have significant implications for HR procedures, providing businesses with the opportunity to make data-driven decisions, improve personnel management and foster a more effective and productive workforce.

Compliance with ethical standards.

Conflict of interest: The authors declare no conflict of interest.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Human and animal rights: This article does not contain any studies with the animals performed by any of the authors.

Funding: Not applicable.

Data Availability: Not applicable.

International Journal of Organizational Analysis

1934-8835
DOI 10.1108/IJOA-07-2024-4643

Research limitations/implications – The scope of the used data set limits the study, despite our models delivering high accuracy. Further research could extend to different data sets or more diverse organisational settings to validate the model's effectiveness across various contexts.

Practical implications – The proposed ML models in the study provide essential tools for HR departments, enabling them to make more informed data driven decisions with regard to employee performance. This approach can enhance personnel management, improve workforce productivity and fostering a more effective organisational environment.

Social implications – Although AI models have shown promising outcomes, it is crucial to recognise the constraints and difficulties involved in their use. To ensure the fair and responsible use of AI in employee performance prediction, ethical considerations, privacy problems and any biases in the data should be properly addressed. Future work will be required to improve and broaden the capabilities of AI models in predicting employee performance.

Originality/value – This study introduces an exclusive combination of ML models for accurately predicting employee performance. By employing these advanced techniques, the study offers novel insight into how organisations might transition from a conventional evaluation method to a more advanced and objective, data-backed approach.

Keywords Artificial intelligence, Machine learning, Employee performance, Human resource, Random Forest, Decision making

Paper type Research paper

1. Introduction

The world of business is constantly evolving, and with the advent of artificial intelligence (AI), it has become more exciting than ever before. One of the latest uses of AI in the corporate world is for predicting employee performance in an organisation. Employee performance prediction is the practise of predicting an employee's future performance based on their previous and present performance data, as well as other pertinent aspects. Integrating AI into human resources (HR) practices can help organisations make better decisions in recruitment, training, performance analysis and retention. According to a report (George and Thomas, 2019), cost was a concern for many organisations when it came to implementing AI into HR practices. However, most HR practitioners report welcoming the integration of AI into their HR processes. Another report (New Study, 2024) found that 64% of respondents would trust a robot over their manager for advice.

This technology has the potential to revolutionise the way we evaluate and manage employees in the workplace. AI-powered models can help organisations to identify top performers and offer them opportunities for growth and advancement, while also identifying areas where employees may need additional training or support (Palos-Sánchez *et al.*, 2022). Employee performance prediction has historically depended on arbitrary judgements like management appraisals and annual performance reviews. Although useful, these methodologies were frequently subject to biases, inconsistencies and a lack of objectivity. The application of AI models and machine learning (ML) techniques emerged as a possible remedy as organisations saw the need for more precise and data-driven approaches (Coron, 2022).

ML and predictive modelling are two key technologies that drive AI factors into data analytics. As more data is gathered, ML algorithm examines it and keeps improving by identifying patterns that teach it how the data will behave in different settings (Tambe *et al.*, 2019; Garg *et al.*, 2022; Meijerink, 2020). The success of an organisation and the full potential of its HR are directly impacted by accurate employee performance prediction. Accurate performance prediction enables organisations to make well-informed decisions about talent management, resource allocation and employee development by utilising data-driven insights and cutting-edge analytical techniques (Meijerink *et al.*, 2021). It aids in the identification of high-potential persons, personalisation of development programmes, workforce planning optimisation and promotion of an equitable and open culture. Additionally, precise performance prediction enables businesses to increase efficiency, connect their people resources with strategic objectives and enhance overall business performance (Davies and McDonald, 2018; Potočnik *et al.*, 2021). Traditional approaches for predicting employee performance have run into several challenges, including subjectivity and bias, lack of data-driven insights, need for lot of time and money and lack of flexibility (Nsor-Ambala, 2020; Payne *et al.*, 2009). The adoption of AI models for an improved employee performance prediction system is driven by each of these limitations. With the increasing availability of data and the advancements in AI models, accurate employee performance prediction has become an indispensable tool for organisations to gain a competitive edge and create a thriving work environment.

1.1 Role of artificial intelligence models in improving performance prediction

Employee performance prediction is undergoing a revolution driven by AI models, which provide unmatched capabilities to improve effectiveness and accuracy. AI models can process and analyse enormous volumes of data, find hidden patterns and deliver insightful information about an employee's potential and future performance by utilising cutting-edge algorithms and ML approaches (Budhwar *et al.*, 2022). Using these models, organisations may move past subjective assessments and harness the potential of unbiased, fact-based predictions. AI models also make it easier to combine data from many sources, giving organisations the chance to take a comprehensive look at employee performance. AI models give individualised performance insights, discover important predictors and provide customised recommendations for staff growth by virtue of their capacity to continuously learn and improve (Popkova and Sergi, 2020). AI models are transforming talent management procedures, fostering fairness, transparency and informed decision-making, eventually driving organisational success and unleashing the full potential of the workforce (Kaushal *et al.*, 2021). By using the HR data set to train ML models, we are also striving to unlock their full potential in this work. We aim to identify the optimal model that will accurately predict employees' performance based on the derived feature set from the given data set. In our study, we have used nine ML models i.e. Random Forest, K-NN, LGB Classifier, Decision Trees, extreme gradient boost (XGBoost), AdaBoost, CatBoost, support vector machine (SVM) and Hybrid (SVM + XGBoost) for prediction of employee's performance. All these models are evaluated based on different measures as stated in section. The best model for predicting employee performance is determined using these measures. This model will assist businesses in identifying effective personnel, which will improve their overall performance.

1.2 Contribution of the study

The main contributions of the study are as follows:

- The paper gives a thorough analysis of the research publications that have use ML models for determining employee's performance.
- The study also offers a detailed evaluation of the different ML model for prediction of employee performance.
- All ML models used in our study are capable of accurately predicting the performance of employees and categorising into classes that we have defined in our study. However, Random Forest is dominating all other ML models for giving the best accuracy of 100%.

1.3 Organisation of paper

The remainder of the paper is organised to enhance its legibility. In Section 2, the extant research and literature on predicting employee performance are reviewed, with an emphasis on the development of AI models and their contribution to improving accuracy and efficacy. Section 3 will describe the study's methodology, including data collection and pre-processing, descriptive statistics, feature scaling, training and validation of AI models and evaluation metrics used to predict employee performance. The findings of the study will be discussed in Section 4, which will also provide an analysis of the performance prediction outcomes attained using AI models. It will demonstrate the precision and reliability with which AI algorithms can predict employee performance. Section 5 will conclude by highlighting the significance of precise employee performance prediction, examining the transformative role of AI models in enhancing performance prediction within organisations and summarising the study's key findings.

2. Related work

Business success relies on employee output and productivity, which is compounded at different levels in the workplace. Performance evaluation is essential for businesses because people differ in their skill sets and behavioural traits. A comprehensive review of existing research and literature on employee performance prediction, with a specific focus on the advancement in AI models has been presented in this section. This section attempts to lay a foundation of understanding and identify gaps and opportunities for further investigation by assessing the present level of knowledge in this field.

The study presented by Lather *et al.* (2019) analyses and forecasts employee performance using supervised ML methods. Three output classes are created from the results ranging from low to high for indicating the performance of employees. SVM comes out to be the most effective and scoring highly in validation tests. The employment process has changed as a result of the fourth industrial revolution, which has forced businesses to incorporate AI technology for quick and precise decision-making. Employers may make wise recruiting decisions by using AI to analyse big data for forecasting new candidates' projected performance. To forecast new candidate performance based on past performance and employment conditions, the study proposed by Mahmoud *et al.* (Ali *et al.*, 2019), suggests a follow-up conceptual model that combines performance management with social screening. To get effective and precise findings, this method needs a large amount of historical data, personal information and employment conditions. Decision trees with C4.5 and ID3 along with Naive Bayes are used for prediction. Zhao et.al (Zhao *et al.*, 2018) examined supervised ML methods for predicting employee turnover in organisations using real and simulated data sets. The performance is analysed using statistical methods and guidelines are provided for selecting, using and interpreting these methods. Extreme gradient boosting has been identified as the reliable model for prediction if more HR data sets are available.

Another study proposed by Tambe *et al.* (2019) identifies complexity, limitations, responsibility and employee reactions as the four issues which act as obstacles in applying data science techniques in HR management. The study offers workable answers based on employee input, trials, randomisation, causal reasoning and socially suitable HR management. Line managers need to update their skill set and adopt augmented intelligence, a Bayesian method of incorporating new data into managerial assumptions. Through the identification and prediction of employee performance, HR software can enhance decision-making procedures and talent management. In this paper, the possibility of applying Knowledge Discovery in Databases (KDD) or data mining (DM) for talent forecasting in HR applications is investigated by Jantan *et al.* (2015). The paper discusses talent management applications, prediction methods, DM and prospective HR system architecture for talent forecasting. Another paper for HR data classification as proposed by Yasodha and Prakash (2012) suggested a hybrid approach called CACC-SVM that offers greater accuracy than conventional algorithms. The produced model can forecast potential talent for activities inside an organisation and is compared to other standard techniques and the enhanced CACC-SVM classification algorithm. Sam *et al.* (2023) proposed a new solution for processing application resume with the help of ML. According to their overall performance, candidates are ranked using this system, which optimises their performance in desired competencies. The algorithm examines user skills on course completion certificates.

An employee recommendation system called EmReSys was developed by Jadav (2022), which uses ML for finding potential candidates without human intervention. SVM model is used by the system with an accuracy of 98.9%. Another study proposed by Jayadi *et al.* (2019) emphasises the value of human resource departments in determining if employees are following organisational objectives. This work uses the cross-industry standard process for data mining (CRISPDM) to develop a ML model for employee performance prediction. The accuracy percentage for the Naive Bayes classification algorithm is 95.48%. And it has been found that, the Naive Bayes technique has a low cost and a minor false-positive outcome. Al-Radaideh and Nagi (2012) had developed a classification model for forecasting employee performance using DM approaches. Decision trees served as the primary tool in the CRISPDM DM process. When the model was put to the test using actual data from different organisations, it was discovered that several attributes were useful for predicting performance. Performance prediction was influenced to varying degrees by factors such as job title, university type, degree, grade, marital status, gender, salary, number of prior employers, experiment years and job happiness. Companies and HR departments can use this model to forecast employee performance and avoid employing underperformers. The study presented by Raut (2020) concentrates on developing a categorisation model for forecasting employee performance, especially in Egypt's public sector. The goal of the study is to advance theoretical and applied research on DM in HR to boost efficiency in the public sector. Using DM techniques like association rules, the findings can be used to analyse development programmes for senior staff and find trends affecting teacher and student performance.

Focusing on human capability criteria and improving the performance appraisal process, a prediction model for employee performance forecasting was presented by Kirimi and Moturi (2016). Three distinct DM algorithms, ID3, C4.5 and Naive Bayes, were used in the classification process to determine which approach was the best and most appropriate. The training and development of the classification model were best suited for the C4.5 algorithm because it had the highest accuracy (92.69%). Data from the institute's HR department were used to validate the model. The findings demonstrated that factors such as experience, age, academic achievement, professional training, gender, marital status and the results of past performance reviews had a significant impact on employee performance. The model and its improvements can be used to manage HR departments in schools and prevent risks associated with recruiting underperformers. The work presented by Ancheta *et al.* (2012) used rule-based classification DM approaches, such as sequential covering and hold-out methods, to extract knowledge about the training requirements of recently appointed faculty members. Significant models required for predictive analysis were found using the CRISPDM. The study makes performance predictions for faculty members based on their training requirements, highlighting the significance of professional training for efficient task execution. Education, human resource management and other academic and non-academic fields could all benefit from data mining.

3. Methodology

In this section, the methodology adopted for predicting the performance of employees from the given data set is presented. The pictorial representation of the proposed methodology is presented in Figure 1. The whole methodology has been divided into several steps. The detailed description of these steps has been explained in the following sub-sections.

3.1 Data collection and pre-processing

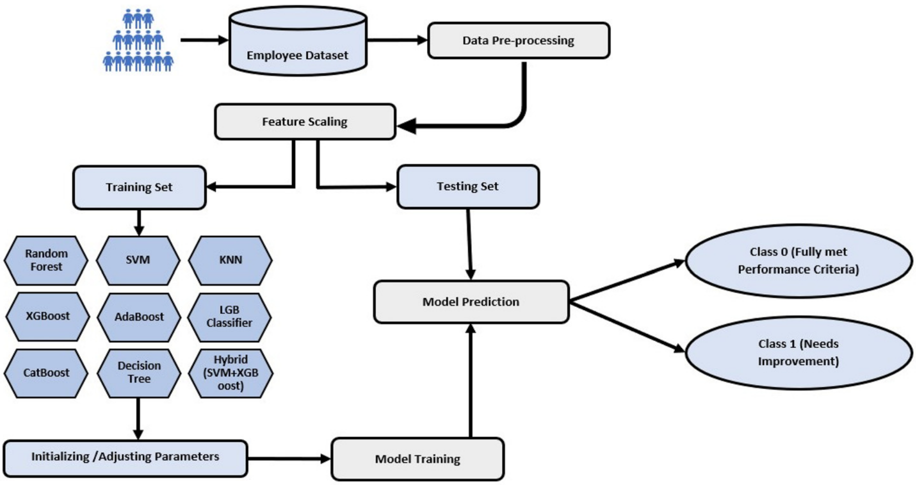
In this paper, we have used the data set designed by Dr Rich Huebner and Dr Carla Patalano (RPubs - HR Dataset Codebook v14, 2024) for predicting employee performance. This data set consists of 36 fields and each year they keep updating the data set to include new features for better prediction. Each field is chosen to gather maximum inputs and generate best possible results. A snap shot of the fields present in the chosen data set is presented in Table 1.

3.1.1 Data cleaning. The data is pre-processed to remove duplicate values, handle missing information, correcting inconsistency, transforming variables etc. All these steps help in making data consistent for input to the AI model. After cleaning all the data, the number of unique values present per column is presented in Table 2.

3.1.2 Descriptive statistics. To gain a general understanding of the data distribution in the data set summary statistics like mean, standard deviation, minimum, maximum and percentiles are calculated. The computed value of all these measures is presented in Table 3.

3.1.3 Data visualisation. For identifying trends, outliers, clusters and relationship between different data variables, a visual representation of the given data set is also presented. Following are the important information and the various relationships that are extracted from the given data set:

- *Count of employee in each position:* The number of employees working in different positions in different organisations is presented in Figure 2. It has been found that a greater number of employees are required as "production technician" in different companies.



Source: Authors own work

Figure 1. Working methodology for predicting performance of employees

Table 1. List of features from the HR data set

S.No.	Data field	Type	S.No.	Data field	Type	S.No.	Data field	Type
1	Employee_Name	object	13	Position	object	25	EmploymentStatus	object
2	EmpID	int64	14	State	object	26	Department	object
3	MarriedID	int64	15	Zip	int64	27	ManagerName	object
4	MartialStatusID	int64	16	DOB	object	28	ManagerID	float64
5	GenderID	int64	17	Sex	object	29	RecruitmentSource	object
6	EmpStatusID	int64	18	MaritalDesc	object	30	PerformanceScore	object
7	DeptID	int64	19	CitizenDesc	object	31	EngagementSurvey	float64
8	PerfscoreID	int64	20	HispanicLatino	object	32	EmpSatisfaction	int64
9	FromDiversityJobFairID	int64	21	RaceDesc	object	33	SpecialProjectsCount	int64
10	Salary	int64	22	DateofHire	object	34	LatPerformanceReview_Date	object
11	Termd	int64	23	DateofTermination	object	35	DaysLateLast30	int64
12	PositionID	int64	24	TermReason	object	36	Absences	int64

Table 2. No. of unique entries per column in the data set

Data field	No. of unique value	Data field	No. of unique value	Data field	No. of unique value
Employee_Name	311	CitizenDesc	3	EmploymentStatus	3
Position	32	HispanicLatino	4	Department	6
State	28	RaceDesc	6	ManagerName	21
DOB	307	DateofHire	101	RecruitmentSource	9
Sex	2	DateofTermination	96	PerformanceScore	4
MaritalDesc	5	TermReason	18	LatPerformanceReview_Date	137

Source: Authors' own work

- *Count of employees in each state:* The data set chosen for our study is for the states of USA. This visualisation helps in determining the state with maximum number of employees within USA. The visualisation presented in Figure 3 clearly shows that Massachusetts has highest number of employees, i.e. 276 followed by Connecticut with 6 employees.
- *Department distribution in the state with highest count of employees:* The state of Massachusetts has the highest number of employees and the Figure 4 gives the departmental distribution of the employees in Massachusetts. It has been found that Production department has highest number of employees followed by IT/IS, Software Engineering and administration. While the least number of employees is present in the executive and sales department.
- *Gender ratio within the company*: The gender ratio plays an important role in determining employee satisfaction, organisational culture, diversity and culture, workplace culture and overall performance of the company. The pie chart presented in Figure 5 depicts the male female ratio in the organisation. The female workforce is more with 56.6% as compared to male workforce with 43.4% of total population.
- *Employee count under each manager*: Another important aspect is determining the work force under each manager to understand organisation management structure, productivity and overall organisational effectiveness. The chart presented in Figure 6 depicts the no. of employees under each manager. The managers are presented by the manager ID. Manager having id 12, 16, 18 and 20 have the maximum number of employees where as manager having id 3 and 30 have least number of employees.
- *Department wise count of employees:* The department wise count of employees is important in determining the department performance evaluation, workload distribution and planning, resource allocation. Figure 7 depicts the allocation of employees to each department. Majority of the employees have been assigned to production wing.

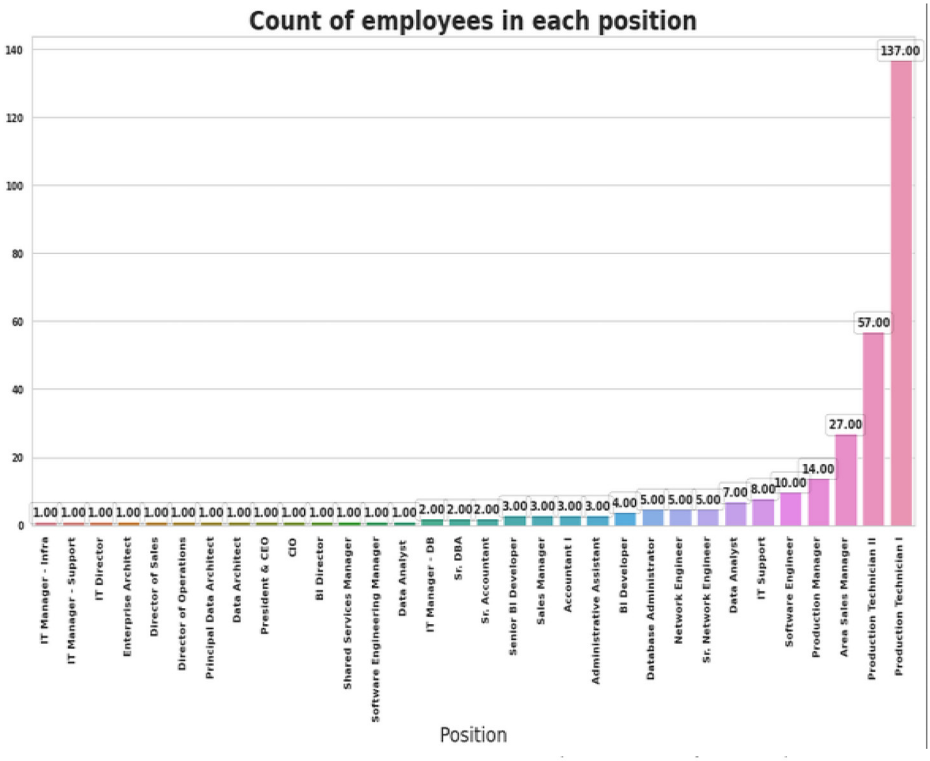
3.2 Feature scaling

To make the data to be given as input to the ML models, feature scaling is applied. It helps in normalising and standardising the features of the given data set. The feature scaling methods used in our data set is as follows:

Table 3. Calculated statistics of important features from the HR data set

Data field	Count	Mean	Std	Min	25%	50%	75%	Max
EmpID	311.0	10156.0000000	89.922189	10001.00	10078.50	10156.0	10233.50	10311.0
MarriedID	311.0	0.3987140	0.490423	0.00	0.00	0.0	1.00	1.0
MaritalStatusID	311.0	0.8102890	0.943239	0.00	0.00	1.0	1.00	4.0
GenderID	311.0	0.4340840	0.496435	0.00	0.00	0.0	1.00	1.0
EmpStatusID	311.0	2.3922830	1.794383	1.00	1.00	1.0	5.00	5.0
DeptID	311.0	4.6109320	1.083487	1.00	5.00	5.0	5.00	6.0
PerfScoreID	311.0	2.9774920	0.587072	1.00	3.00	3.0	3.00	4.0
FromDiversityJobFairID	311.0	0.0932480	0.291248	0.00	0.00	0.0	0.00	1.0
Salary	311.0	69020.6848870	25156.636930	45046.00	55501.50	62810.0	72036.00	250000.0
Termd	311.0	0.3344050	0.472542	0.00	0.00	0.0	1.00	1.0
PositionID	311.0	16.8456590	6.223419	1.00	18.00	19.0	20.00	30.0
Zip	311.0	6555.4823150	16908.396884	1013.00	1901.50	2132.0	2355.00	98052.0
ManagerID	303.0	14.5709570	8.078306	1.00	10.00	15.0	19.00	39.0
EngagementSurvey	311.0	4.1100000	0.789938	1.12	3.69	4.3	4.70	5.0
EmpSatisfaction	311.0	3.8906750	0.909241	1.00	3.00	4.0	5.00	5.0
SpecialProjectsCount	311.0	1.2186500	2.349421	0.00	0.00	0.0	0.00	8.0
DaysLateLast30	311.0	0.4147910	1.294519	0.00	0.00	0.0	0.00	6.0
Absences	311.0	10.2379420	5.582598	1.00	5.00	10.0	15.00	20.0

Source: Authors' own work



Source: Authors own work

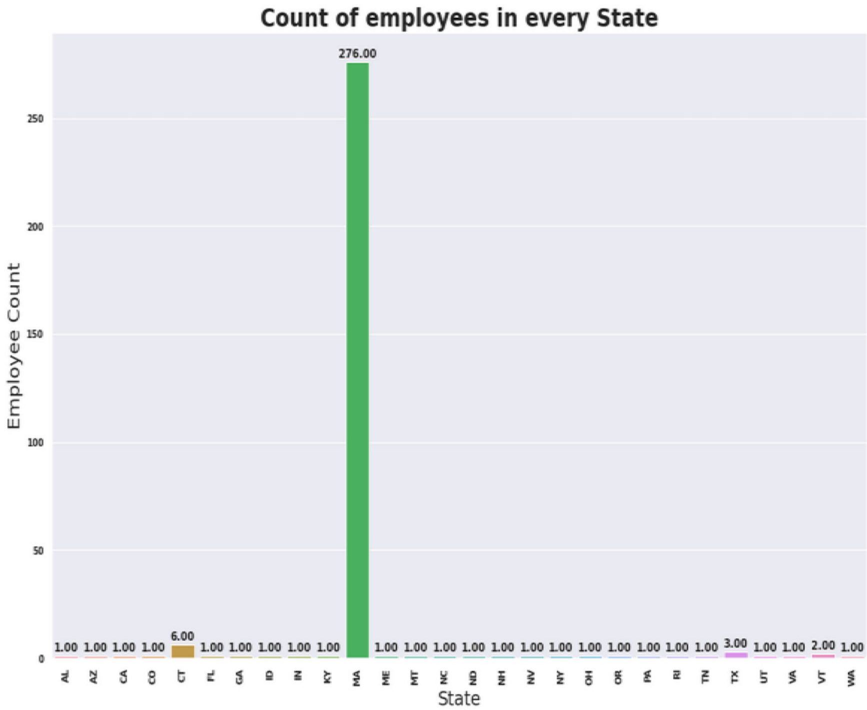
Figure 2. Position-wise distribution of employees

- *Min-max scaling:* Min-max scaling is also referred to as normalisation and is an alternative approach to Z-score normalisation. It scales the feature to a specific range typically between 0 and 1. Min-max scaling preserves the relative relationships between values and is sensitive to outliers. The min-max scaling is calculated using the following equation (i):

$$X_{sc} = (X - X_{min})/(X_{max} - X_{min}) \tag{i}$$

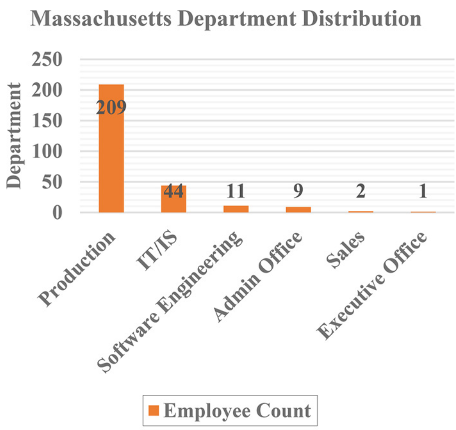
Min-max scaling is advantageous when the magnitude or range of values for various traits or variables differ greatly. It ensures a fair comparison and avoids features with greater values from overpowering the analysis by scaling the data to a common range. In our study, we have used min-max scaling as the feature set used has variable attributes and range of varying values as can be seen in Table 3. The results produced for various ML models used in the study for min-max scaling is presented in Table 4.

- *Standard Scaler:* Standard Scaler is another feature scaling technique which that performs standardisation by transforming the features to have zero mean and unit variance. This ensures that the features are centred around zero and have a standard



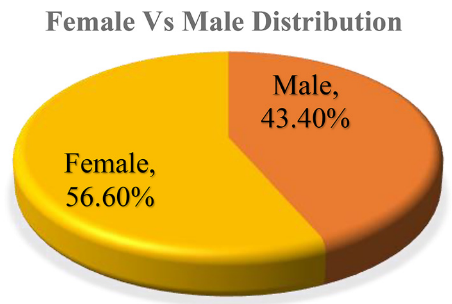
Source: Authors own work

Figure 3. State-wise distribution of employees



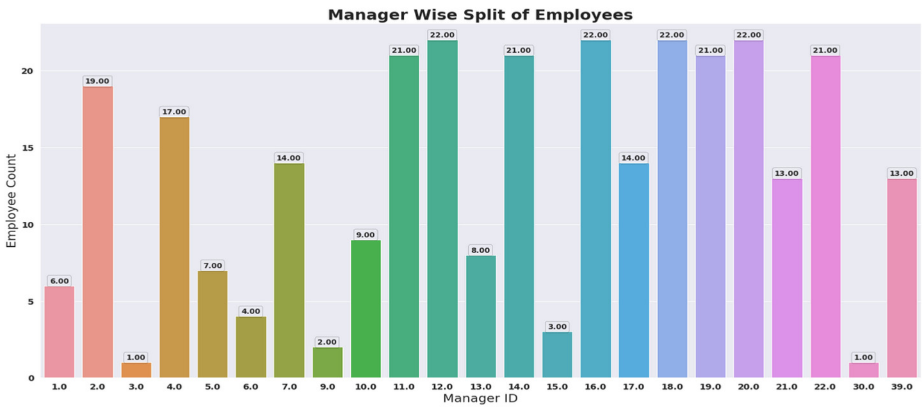
Source: Authors own work

Figure 4. Department-wise distribution of employees



Source: Authors own work

Figure 5. Female–male ratio in the organization



Source: Authors own work

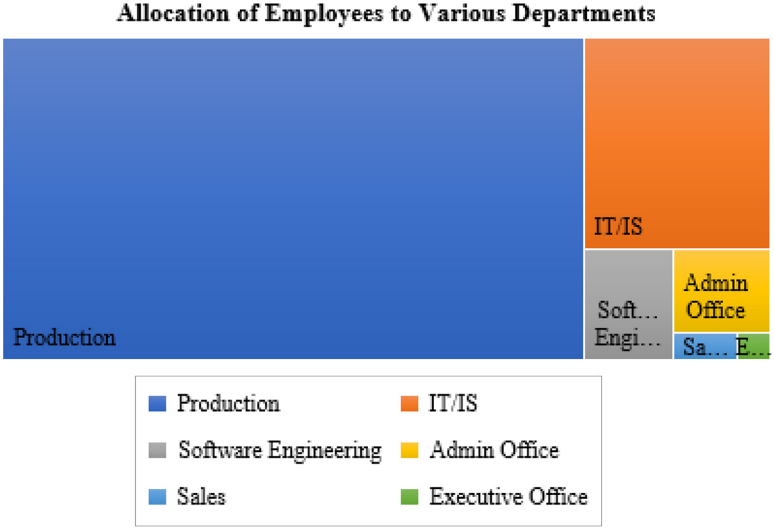
Figure 6. Visual representation of employee distribution under each manager

deviation of 1. The Standard Scaler calculates each feature's values by deducting its mean, and then dividing the result by the standard deviation. The Standard Scaler's standardisation formula is expressed mathematically as follows in equation (ii):

$$Z = (X - \mu)/\sigma \tag{ii}$$

where Z stands for the scaled valued, μ stands for mean and σ stands for standard deviation. Applying the Standard Scaler has the benefit of not altering the data's distribution or form. It guarantees that the data points are on a standardised scale while maintaining the relative relationships between them. And it is one of the major factors for considering Standard Scaler as the second feature scaling option in our study. The results of applying Standard Scaler to the proposed ML model is presented in Table 6.

- *Principal component analysis (PCA):* PCA is a widely used dimensionality reduction technique that allows the transformation of high-dimensional data into a lower-dimensional representation while preserving the most important information. The basic objective of PCA is to identify a set of uncorrelated variables called principal components that best capture the



Source: Authors own work

Figure 7. Employees assigned to various departments

Table 4. Evaluation measures obtained for min-max scaling

ML model	Class 0 Precision	Class 0 Recall	Class 0 F-measure	Class 1 Precision	Class 1 Recall	Class 1 F-measure	Macro average	Weighted average
Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.94	1.00	0.97	0.00	0.00	0.00	0.49	0.92
KNN	0.94	1.00	0.97	0.00	0.00	0.00	0.49	0.92
LGB classifier	0.99	0.97	0.98	0.62	0.83	0.71	0.85	0.96
CatBoost	0.99	1.00	0.99	1.00	0.83	0.91	0.95	0.99
XGBoost	1.00	0.98	0.99	0.75	1.00	0.86	0.92	0.98
Decision tree	1.00	0.87	0.93	0.32	1.00	0.48	0.70	0.90
AdaBoost classifier	0.99	0.97	0.98	0.62	0.83	0.71	0.85	0.96
Hybrid (SVM+XGBoost)	1.00	0.87	0.93	0.32	1.00	0.48	0.70	0.90

Source: Authors' own work

data's overall variance. The calculation of PCA comprises of a series of step starting from standardisation which scale the data within a range so that the output is unbiased. Equation (iii) gives the formula for standardisation of data:

$$Z = \frac{Variable\ Value - Mean}{Standard\ Deviation} \tag{iii}$$

After standardisation, covariance matrix is computed to identify interdependencies between variables and minimising them to improve the performance of the model. Equation (iv) gives the formula for computing covariance matrix:

Table 5. Loss functions and accuracy for ML models with min-max scaling

Algorithms	Mean squared error	Root mean squared error	Accuracy
Random Forest	0.0	0.0	1.0
KNN	0.05	0.23	0.94
SVM	0.07	0.24	0.93
LGB classifier	0.03	0.19	0.96
CatBoost	0.09	0.05	0.99
XGBoost	0.01	0.13	0.98
Decision tree	0.12	0.35	0.87
AdaBoost	0.03	0.19	0.96
Hybrid model (SVM + XGBoost)	0.05	0.23	0.94

Source: Authors' own work

Table 6. Evaluation measures obtained for standard scaler

	Class 0			Class 1			Macro	Weighted
ML model	Precision	Recall	F-measure	Precision	Recall	F-measure	average	average
Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SVM	0.94	0.98	0.96	0.00	0.00	0.00	0.48	0.91
KNN	0.94	1.00	0.97	0.00	0.00	0.00	0.49	0.92
LGB classifier	0.99	0.98	0.98	0.71	0.83	0.77	0.88	0.97
CatBoost	0.99	1.00	0.99	1.00	0.83	0.91	0.95	0.99
XGBoost	1.00	0.98	0.99	0.75	1.00	0.86	0.92	0.98
Decision tree	1.00	0.88	0.94	0.33	1.00	0.50	0.72	0.91
AdaBoost classifier	0.99	0.97	0.98	0.62	0.83	0.71	0.85	0.96
Hybrid (SVM+XGBoost)	0.94	1.00	0.97	0.00	0.00	0.00	0.49	0.92

Source: Authors' own work

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix} \tag{iv}$$

Now, to determine PCA, eigen values and eigen vectors are computed with the help of Covariance matrix. Eigen vectors help in calculation of principal components by identifying the largest variance in the given data set. Finally, principal components are arranged and the component with least significance are eliminated thereby minimising the complexity of the system. In our study, due to variety of features PCA will be helpful in reducing the dimension of the given HR data set thereby yielding appropriate results. The results obtained by applying PCA to the given data set on various ML models is presented in Table 7.

3.3 *Training and validation of ML models*

After applying feature scaling, the data set is divided into two sets i.e. training and testing sets. In our study, we have used 75% of data for training and 25% of data for testing purposes. Top 9 ML models are used in our study for predicting the employee performance from the given data set. All the models used for the study are supervised ML models and classify the data set into two classes i.e. Class 0 and Class 1. Class 0 represents that

Table 7. Loss function and accuracy of different ML models for standard scaler

Algorithms	Mean squared error	Squared mean squared error	Accuracy
Random Forest	0.0	0.0	1.0
KNN	0.07	0.27	0.92
SVM	0.05	0.23	0.94
LGB classifier	0.02	0.16	0.97
CatBoost	0.05	0.09	0.99
XGBoost	0.01	0.13	0.98
Decision Tree	0.11	0.33	0.88
AdaBoost	0.03	0.19	0.96
Hybrid model (SVM + XGBoost)	0.05	0.23	0.94

Source: Authors' own work

performance of the employee is up to the mark whereas Class 1 represents that employee needs improvement. The ML models used in our study are as follows:

- *Random Forest:* Random Forest is an ensemble supervised ML technique used for both regression and classification task. It is based on the idea of decision trees and makes predictions by combining different decision trees. Individual decision tree predictions are combined to make the final prediction. The random forest algorithm creates a "forest" that is trained via bagging or bootstrap aggregation. Bagging helps in improving the accuracy of ML algorithms. The number of trees also affect the decision-making capability of random forest. More the number of trees, more is the accuracy of prediction. In our study, we have used Random Forest as one of the ML models due to its accuracy for classification problem. The pictorial representation of the Random Forest technique used for the Employee database is presented in Figure 8.
- *Support vector machine (SVM):* SVM is a supervised ML technique which is used for regression and classification task but works well for classification task (Schölkopf and Smola, 2018). Because in our study we are also trying to determine the class of employee, hence we have chosen this supervised ML approach. SVM algorithm tries to find the best hyperplane that divides the two classes with the help of statistical approach (Su and Yang, 2008). The equation of the hyperplane is given in equation (v):

$$w.x + b = 0 \tag{v}$$

where w is a normal vector to hyperplane and b is an offset. To define a decision rule, we need to classify the point as negative or positive. The decision rules can be defined as given in equation (vi)-(vii):

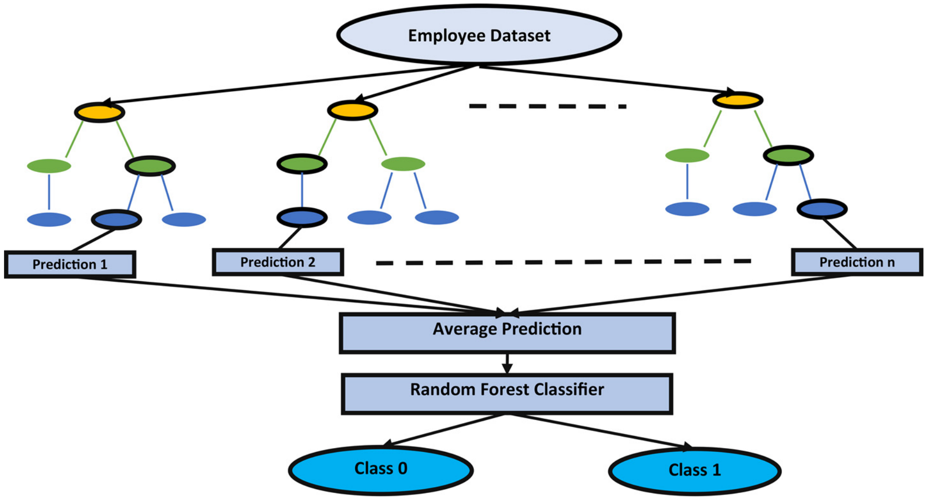
$$\vec{X}.\vec{w} - c \geq 0 \tag{vi}$$

Putting -c as b, we get:

$$\vec{X}.\vec{w} + b \geq 0 \tag{vii}$$

Hence:

$$y = \begin{cases} +1 \, if \, \vec{X}.\vec{w} + b \geq 0 (Positive) \\ -1 \, if \, \vec{X}.\vec{w} + b < 0 (Negative) \end{cases} \tag{viii}$$



Source: Authors own work

Figure 8. Architecture of Random Forest model for employee performance prediction

SVM works well for a small data set with larger number of features. Our HR data set also have 32 features, hence, SVM makes an appropriate choice for solving the problem of performance prediction of Employees.

- *K-Nearest Neighbour:* The K-Nearest Neighbour is a non-parametric supervised ML algorithm which is also referred to as KNN. It is a simple yet powerful ML algorithm for classification tasks. It generates predictions based on the closeness or proximity of the input data to the labelled data points in the training set. It measures the similarity or dissimilarity between data points using a distance metric (Kramer, 2013). Euclidean distance given in equation (ix) and Manhattan distance given in equation (x) are the two most often used distance measures. In a feature space of "n" dimensions, the distance "d" between two data points "x" and "y" can be determined as follows:

$$Euclidean\ Distance : d(x, y) = \sqrt[2]{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots \ldots\ldots + (x_n - y_n)^2} \quad \text{(ix)}$$

$$Manhattan\ Distance : d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \cdots \ldots\ldots\ldots + |x_n - y_n| \quad \text{(x)}$$

Based on the selected distance metric, KNN looks for the "k" closest neighbours to a given input data point. The number of closest neighbours considered for prediction is represented by the value of "k". These closest neighbours are chosen in accordance with how close they are to the input data point. In classification tasks, the input data point's class label is decided by majority vote among its "k" nearest neighbours. The class label with the greatest count among the neighbours is given to the input data point, and each neighbour's class label contributes equally to the final prediction (He *et al.*, 2021). Considering it as a good classification algorithm, KNN is also chosen for employee performance prediction in our study.

- *XGBoost:* A decision tree-based supervised ML technique using the gradient boost framework is referred to as XGBoost (Chen and Guestrin, 2016). Both classification and regression tasks can be solved using this framework. The most important aspect of this framework relies on its optimisation and scalability. It aims to minimize the objective function that measures the error or loss between the predicted and actual values. The objective function comprises of two components: a regularisation term that regulates the model's complexity to avoid overfitting and a loss function that quantifies the difference between forecasts and actual values. Equation (xi) gives the definition of the overall objective function:

$$Objective = Loss + Regularization \tag{xi}$$

where $Loss = \sum_{i}^{n} l(y_i, \widehat{y}_i)$ and Regularisation = $\sum_{k=1}^{K} \Omega(f_k)$; K is the number of trees and f is the functional space. The model's performance is iteratively increased by using the gradient boosting technique which involves training a weak learner to minimize loss function gradients and update weak learner parameters.

- *CatBoost:* CatBoost stands for "Category Boosting" as it works with multiple categories of data ranging from text, audio, images to historical data (Dorogush *et al.*, 2018). It is a supervised ML approach that uses gradient boosting on categorical features. A strong predictive model is produced using an ensemble method that integrates different decision trees. CatBoost uses a shrinkage parameter, also known as the learning rate, to regulate how much each weak learner contributes to the ensemble, just like other gradient boosting methods. Each weak learner's predictions are scaled by the shrinkage parameter, and a lower value results in a more conservative learning process. Different regularisation methods are used by CatBoost to reduce overfitting and boost generalisation. The leaf values undergo L2 regularisation (Ridge regularisation), which increases the loss function's penalty term dependent on the leaf values' magnitudes. The formula for cost function of ridge regularisation including the regularisation term is presented in equation (xii):

$$J(m) = \frac{1}{n}\sum_{i=1}^{n}\left(y_{predicted} - y_{original}\right)^2 + \lambda \sum_{i=1}^{n} {m_i}^2 \tag{xii}$$

where $J(m)$ is the regularised cost function, n is the total number of sample units, λ is the regularisation parameter, also known as the regularisation strength and m_i represents the model's weight. In addition, CatBoost presents a cutting-edge method known as symmetric tree splits that helps balance the regularisation effect across various tree portions.

- *Decision Tree:* Decision trees are supervised ML approach used for both classification and regression applications. They are simple, interpretable models that assign labels or values to the regions of the feature space based on the majority or average of the training samples inside each region to produce predictions. Recursively partitioning the data according to the selected splitting criteria is required to construct a decision tree. Entropy for classification, and mean squared error or variance reduction for regression, are examples of common dividing criteria. Because, ours is a classification problem, entropy is used as the splitting criteria. The formula for calculation of entropy is presented in equation (xiii):

$$K_i = -\sum_{h\epsilon H}^{n} p(i,h) log\ _2 p(i,h) \tag{xiii}$$

where h is the class from H classes, i represent the node of the tree and $p(h)$ is the proportion of data points in the data set sample that are members of the class h. In our case, because the number of classes is two so the value of H = 2.

The data is split at each internal node multiple times also known as recursive partitioning to develop the tree until a halting condition is reached. The primary objective is to increase homogeneity or decrease impurity in the generated divisions.

- *LGB Classifier:* LGB stands for light gradient boosting Model which is a popular supervised learning model using gradient boosting framework with decision trees as weak learners. It employs two different sorts of techniques: exclusive feature bundling (EFB) and gradient-based on-side sampling (GOSS). GOSS leverages big gradients to produce reliable results even with less data sets by excluding tiny gradient data and using the remaining data for information gain estimation. GOSS used the concept of verdict tree whose function can be iterated as "from the input space X to the gradient space G. It is assumed that the training set consists of cases x_1, x_2 and so on upto x_n, where x_i is a vector with dimension s in space X. The negative gradients of the loss function with respect to the model's output are indicated as g_1, g_2, …., g_n in each iteration of gradient boosting. The variance after segregation in this kind of model can be used to gauge data improvement. The following equation (xiv) can be used to describe it:

$$Y = Base_{tree(X)} - lr*Tree1(X) - lr*Tree2(X) - lr*Tree3(X) \tag{xiv}$$

Every instance of a bigger gradient is used in GOSS, whereas the instances of lesser gradients are randomly sampled. On a particular decision tree node, the training data set is specified as O. For this node, the variance gain of splitting measure j at position d is defined as presented in equation (xv):

$$\widetilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{\{x_i\epsilon A_l\}} g_i + \frac{1-a}{b}\sum_{x_{i\epsilon B_l}} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{\{x_i\epsilon A_r\}} g_i + \frac{1-a}{b}\sum_{x_{i\epsilon B_r}} g_i\right)^2}{n_r^j(d)}\right) \tag{xv}$$

where $A_l = \{x_i\ \epsilon\ A{:}x_{ij} \le d\}$, $A_r = \{x_i\ \epsilon\ A{:}x_{ij} > d\}$, $B_l = \{x_i\ \epsilon\ B{:}x_{ij} \le d\}$, $B_r = \{x_i\ \epsilon\ B{:}x_{ij} > d\}$, and the coefficient $\frac{1-a}{b}$ is used for normalising the sum of the gradients over B (Ahamed and Arya, 2021). While EFB effectively eliminates mutually exclusive features without considering non-zero values, it also ensures split point precision without sacrificing feature elimination effectiveness.

- *AdaBoost Classifier:* AdaBoost referred to as adaptive boosting is a predictive ensemble model used in ML (Wang, 2012). It combines several weak learners to produce a powerful learner. It operates by incrementally changing the weights of the training examples to concentrate on the incorrectly classified or challenging samples. Decision trees with one level/split, also known as Decision Stumps, are the most popular estimator used with AdaBoost (Su and Yang, 2008; Kramer, 2013; He *et al.*, 2021; Chen and Guestrin, 2016). The performance of the Decision Stump is calculated using equation (xvi).

$$Performance\ of\ Stump = \frac{1}{2} log_{\ e}\left(\frac{1 - TotalError}{TotalError}\right) \tag{xvi}$$

Now, based on the performance of the stump, the weights are updated for final prediction value. Equation (xvii) below presents the formula for updating weights.

$$New\ Sample\ Weight = Old\ Weight * e^{\pm Amountofsay(\alpha)} \tag{xvii}$$

The value of α is negative, when the sample is classified correctly and negative when the sample is not classified correctly.

- *Hybrid model:* For the Hybrid model, we have clubbed two models, namely, SVM and XGBoost for leveraging the strengths of both algorithms to improve predictive performance. First SVM is applied on the data set and the results of SVM are added as input features to the XGBoost model. The learning rate, maximum tree depth and number of estimators are among the pertinent hyperparameters for the XGBoost model that are specified. Finally, these parameters are fine-tuned using methods like cross-validation and Bayesian optimisation. Predictions are made on the testing data used previously for SVM, using the trained XGBoost model. The hybrid model's result is represented by these forecasts.

3.4 Evaluation metrics for performance prediction

All the models stated above have been evaluated based on classification measures like precision, recall, f-measure, macro average, weighted average, mean square error (MSE), root mean square error (RMSE) and accuracy. With the help of these parameters the appropriate ML model can be identified which can accurately predict the performance of the employees in an organisation. The definition of each of these measures is as follows:

- *Precision:* The accuracy of a model's accurate predictions is measured by precision. It measures the percentage of cases that were accurately predicted as positive out of all instances that were forecasted as positive. The formula for calculating precision is defined in equation (xviii):

$$Precision = TP/(TP + FP) \tag{xviii}$$

where TP stands for True Positives and FP stands for False Positive. TP represents the number of cases where the model properly predicts a positive outcome and the outcome is in fact positive. FP represents the number of cases where the model predicted a positive outcome when it was really a negative outcome.

- *Recall:* Recall evaluates a model's capacity to accurately distinguish false positives from true positives in a data set. It calculates the percentage of accurate positive predictions among all actual positive events. It is also known as sensitivity or true positive rate. The formula for calculating recall is presented in equation (xix).

$$Recall = TP/(TP + FN) \tag{xix}$$

where TP stands for True Positive and FN stands for False Negative. FN represents the number of cases where the model mis-predicted the outcome as being negative when it should have been positive.

- *F-measure:* F-measure also referred to as F! score combines precision and recall by considering both the accuracy of positive prediction and the model's capability to identify positive occurrence into a single score. The formula for calculating F1 score is presented in equation (xx):

$$F1\ score = \frac{2*(Precision*Recall)}{(Precision + Recall)} \tag{xx}$$

- *Macro-Average:* In multi-class classification tasks, evaluation metrics are gathered using macro average. It computes the average of a particular measure across all classes, giving each class an equal weight. The formula for calculating Macro-average is presented in equation (xxi):

$$Macro - Average = (M(Class1) + M(Class2) + \ldots\ldots\ldots. + M(ClassN))/N \tag{xxi}$$

where $M(Class\ i)$ represents the value of any of the chosen metric i.e. precision, recall or F1 score for the i^{th} class and N is the total number of instances. In our case we have taken F1 score metric as presented in Table 4–6.

- *Weighted Average:* Weighted average is also used to combine assessment metrics, but it considers the contribution of each class based on its size or prevalence in the data set. In contrast to the macro average, the weighted average takes the class distribution and gives each class a distinct weight based on how prevalent it is in the data set. The formula for calculating Weighted-average is presented in equation (xxii):

$$Weg\,Avg = \frac{M(Class1)*W(Class1) + M(Class2)*W(Class2) + \cdots\ldots + M(ClassN)*W(ClassN)}{N} \tag{xxii}$$

where $M(Class\ i)$ represents the value of any of the chosen metric i.e. precision, recall or F1 score for the i^{th} class, $W(Class\ i)$ represents the weight assigned to i^{th} class and N is the total number of instances. Again, F1 score is used for weighted average as presented in Table 4–6.

- *MSE:* The average squared difference between the predicted and actual values is measured using the evaluation metric known as MSE. It gives an estimate of the typical error size of the learning model. Equation (xxiii) gives the formula for calculating the MSE.

$$MSE = \left({}^{1}/{n}\right) * \sum (y_i - \hat{y_i})^2 \tag{xxiii}$$

where n is the total number of instances, y_i reprevsents the actual value and $\hat{y}_i$ represents the predicted value.

- *RMSE:* The average absolute difference between the values that were predicted and those that were calculated in actual is referred to as root mean square error. It is calculated as the square root of MSE. The formula for calculation RMSE is presented in equation (xxiv).

$$RMSE = \sqrt{MSE} \quad \text{(xxiv)}$$

- *Accuracy:* Accuracy defines the overall correctness of the prediction model by calculating the ratio of correctly classified instance over the total number of instances in the data set. Equation (xxv) gives the formula for calculation of accuracy.

$$Accuracy = (No.of\,CorrectPredictions)/(TotalNo.of\,Instances) \quad \text{(xxv)}$$

3.5 Comparison of different machine learning models

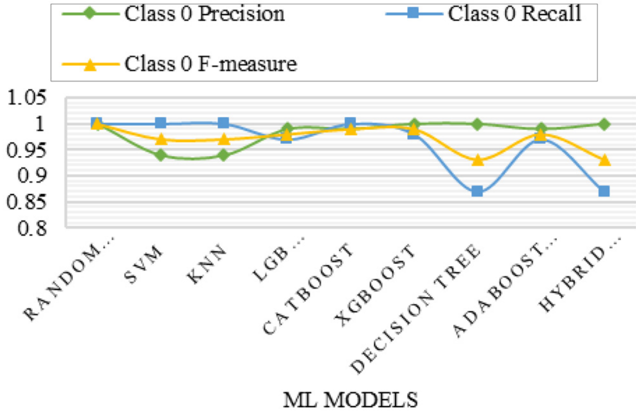
All the models used for employee performance prediction are evaluated on precision, recall, f-measure, accuracy, etc. for determination of the best possible model. Two classes have been created i.e. Class 0 and Class 1. All the proposed models can classify the given HR data set into one of the two classes. Based on feature scaling techniques applied, three sets of results have been generated. One for min-max scaling, another for Standard Scaler and the last one for PCA. The results produced by using these three feature scaling techniques on the ML models as proposed for the study are presented in following sub-sections.

5.1 Results of min-max scaling on human resources data set for machine learning models. By using the min-max scaling technique, two sets of result have been generated. One for the classification report which present result in terms of evaluation metrics like precision, recall, f-measure for both the Classes i.e. Class 0 and Class 1. Combined evaluation measures like accuracy, macro-average and weighted average are also presented as part of classification report. For the visual representation of the results, confusion matrix is also presented for all the ML models used in the study. A confusion matrix is a tabular representation that compares predicted and actual class labels to give a thorough understanding of how well a classification model is performing. Four quadrants representing four possible outcomes of the employee performance prediction problem are presented. The terminology for four quadrants is True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP determines the correctly predicted positive instances whereas TN signifies the correctly predicted negative instances by the ML model. FP and FN represent the incorrectly predicted positive and negative classes respectively, also known as Type 1 and Type 2 error. The results of classification report for min-max scaling indicates Random Forest ML model is outperforming all other models with a perfect accuracy of 100% whereas all other ML models are also generating decent results with an average accuracy of 94%.

The graphical representation of the precision, recall and f-measure values obtained for Class 0 and Class 1 with the application of min-max scaling is presented in Figure 9 and 10, respectively. It is clearly visible from Figure 9 that Decision Trees are the lowest achieving ML model among all the proposed models. From Figure 10, it is evident to note that SVM and KNN are the least precise model for Class 1 classification.

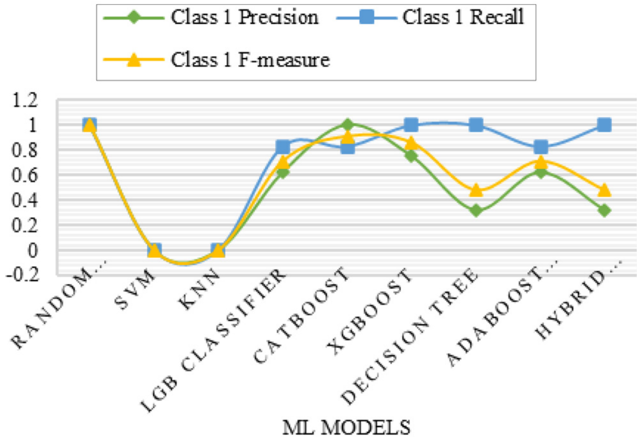
The loss functions i.e. MSE and RMSE are also calculated for all the 9 models along with the accuracy of each model. Table 5 gives the value of loss function and accuracy for each of the above-mentioned ML models used in the study.

The confusion matrix obtained by application of ML model on the HR data set is presented in Figure 11–19. From the confusion matrix results for min-max scaling, as presented in Figure 11–19, it can be interpreted that decision trees are least performing for HR data set with least accuracy of 81.9% whereas 94.29% is the best accuracy obtained across all the applied ML models.



Source: Authors own work

Figure 9. Min-max scaling results of precision, recall and F-measure for class 0

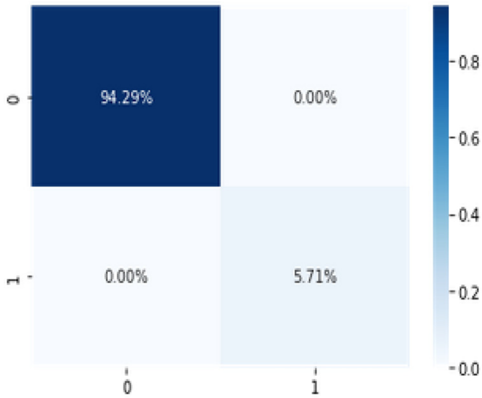


Source: Authors own work

Figure 10. Min-max scaling results of precision, recall and F-measure for class 1

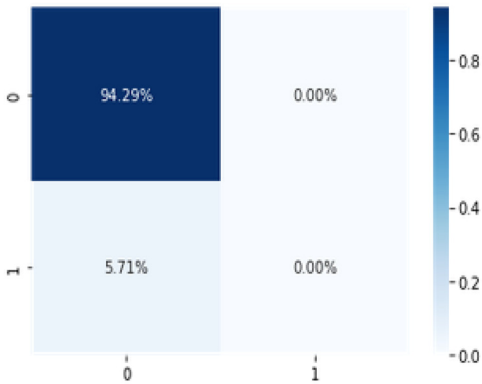
The best accuracy of 94.29% has been obtained by 4 ML models, namely, RF, KNN, SVM and CatBoost. KNN and SVM model seems to be accurate in predicting the positive class, but suggest the weakness of the model in identifying positive instances. CatBoost model has high accuracy in predicting both positive and negative classes and demonstrate good performance in correctly identifying positive instances while maintaining a low false positive rate. Finally, RF model is highly accurate among these four as Type 1 and Type 2 errors are absent in this model; hence, this model can predict both negative and positive instances efficiently with no error rate.

3.5.2 Results of Standard Scaler on human resources data set for machine learning models. Another feature scaling technique used in our study is Standard Scaler, as discussed in Section 3.2. Based on this feature scaling technique, the results for different evaluation



Source: Authors own work

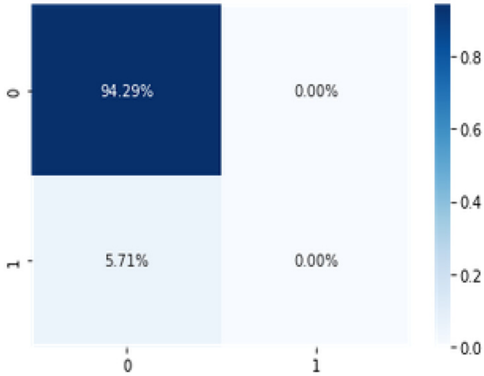
Figure 11. Random Forest



Source: Authors own work

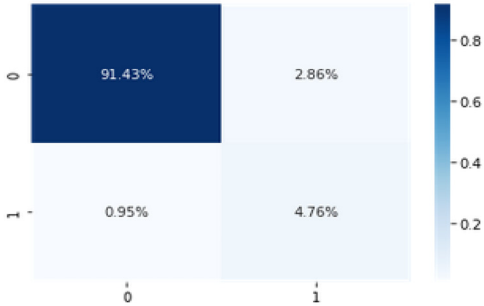
Figure 12. KNN

measures and confusion matrix are presented in following sub-sections. Tables 6 and 7 represent the various evaluation measures and accuracy and loss functions obtained for different ML models proposed in the study using Standard Scaler scaling technique. The Random Forest ML model again outperforms all other models, according to the results obtained for classification report of Standard Scaler with a perfect accuracy of 100%. CatBoost and XGBoost are performing decently with the accuracy of 99% and 98% respectively. The precision value of Random Forest, XGBoost and Decision Tree comes out be a perfect 1. But despite having the best precision value, decision tree is having the least accuracy of 91% of all the proposed models. The graphical representation of precision, recall and F-measure score for all the models for classification into Class 0 and 1 is presented in Figures 20 and 21, respectively.



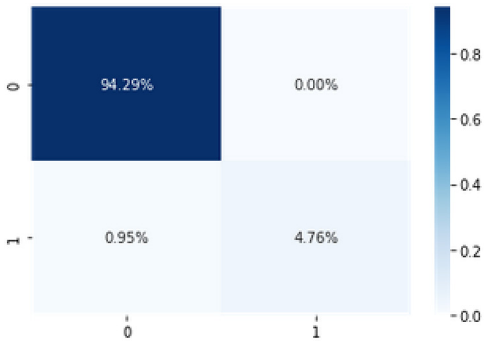
Source: Authors own work

Figure 13. SVM



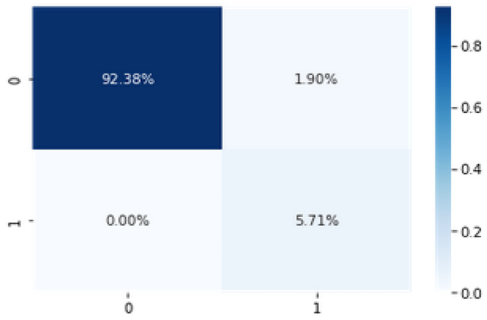
Source: Authors own work

Figure 14. LGB Classifier



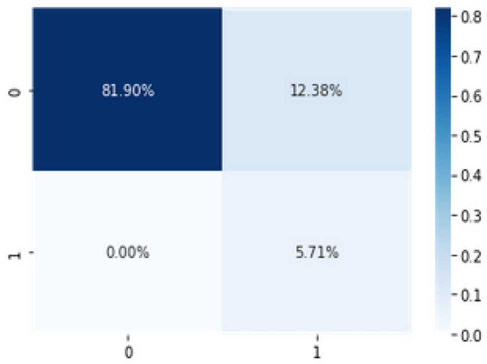
Source: Authors own work

Figure 15. CatBoost



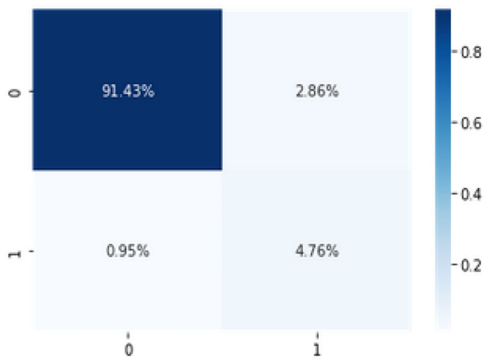
Source: Authors own work

Figure 16. XGBoost



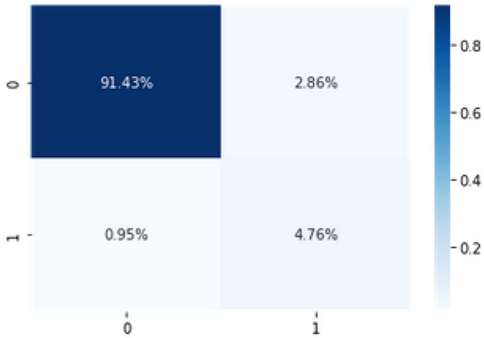
Source: Authors own work

Figure 17. Decision Tree



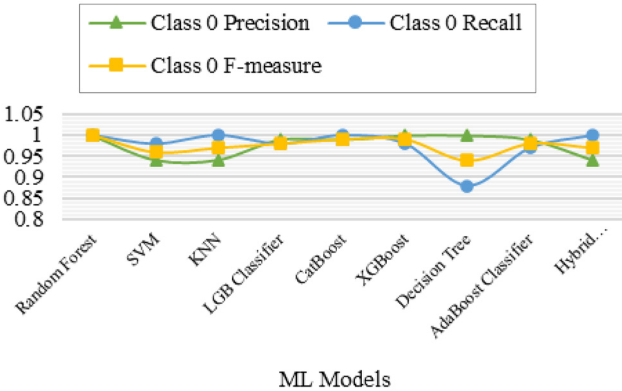
Source: Authors own work

Figure 18. ADABoost



Source: Authors own work

Figure 19. Hybrid (SVM+XGBoost)

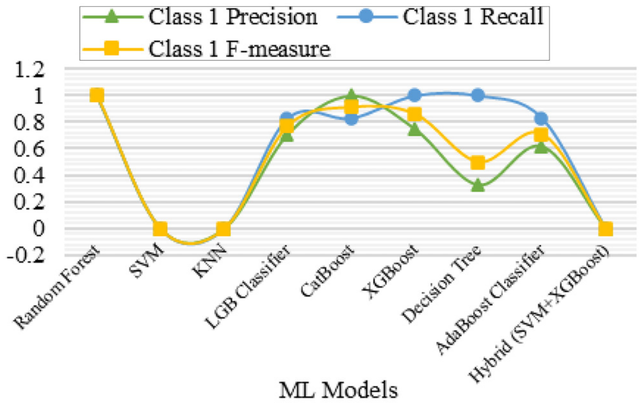


Source: Authors own work

Figure 20. Standard scaler results of precision, recall and F-measure for Class 0

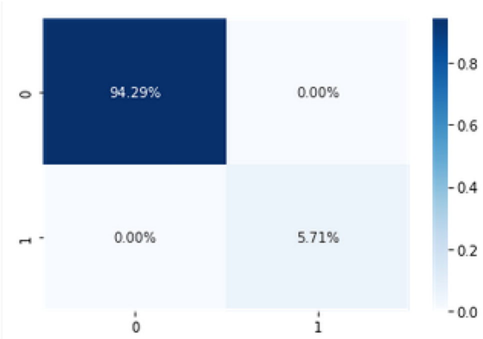
It can be interpreted from the graph that the results produced by min-max scaling and Standard Scaler are quite similar with minor differences in some model performances. The precision of SVM and KNN decision tree model has still not improved for classification of Class 1 followed by decision tree. For Class 0, all the models are performing comparatively better with average precision of 97%.

The confusion matrix for all the ML models with Standard Scaler feature scaling technique is represented in Figures 22–30. With the help of Standard scaling technique, the hybrid model i.e. SVM + XGBoost model is also able to generate best predictive results with highest true positive value of 94.29% along with other ML models i.e. Random Forest, SVM and CatBoost ML models. Among all these only Random Forest model is generating no error rate whereas all other models have incorrectly predicted positive values as negative with an error percent of 5.71%. KNN, LGB Classifier and XGBoost all are performing well with the true positive rate of 92.38% and 1.90% false positive rate.



Source: Authors own work

Figure 21. Standard scaler results of precision, recall and F-measure for Class 1

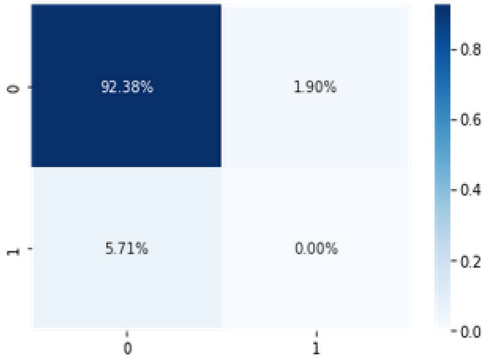


Source: Authors own work

Figure 22. Random Forest

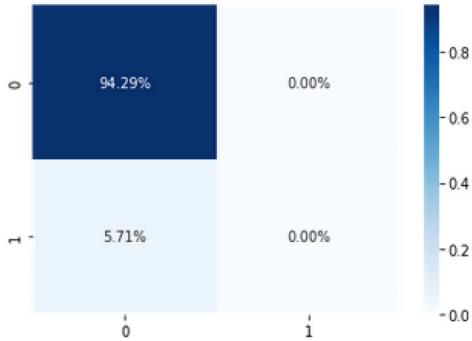
3.5.3 Results of principal component analysis scaling on human resources data set for machine learning models. PCA has been chosen as the third scaling method due to its advantageous dimensionality reduction principle. Because our data set also has around 36 features, identification of important features is a must. For PCA scaling also classification report and confusion matrix for all proposed ML models has been generated as presented in following sub-section.

The classification report for PCA scaling is slightly different from the results obtained for min-max scaling and Standard Scaler scaling technique which is evident from Table 8. In this technique, none of the ML model can achieve 100% accuracy. The highest accuracy percentage obtained by using PCA is of 98% and three ML models were able to achieve this accuracy, namely, Random Forest, LGB Classifier and XGBoost as induced by results seen in Table 9. The least accuracy of 92% is obtained for Decision Tree ML model. For Class 1, the perfect score of precision i.e. 1.0 is obtained for 5 ML models i.e. Random Forest, LGB



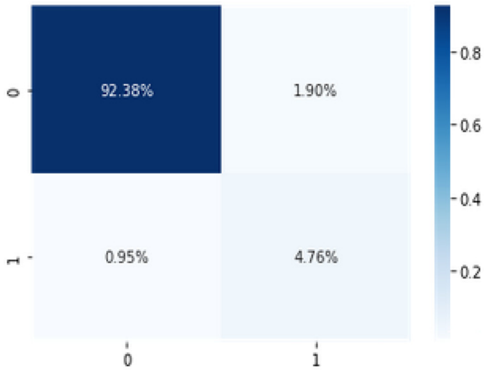
Source: Authors own work

Figure 23. KNN



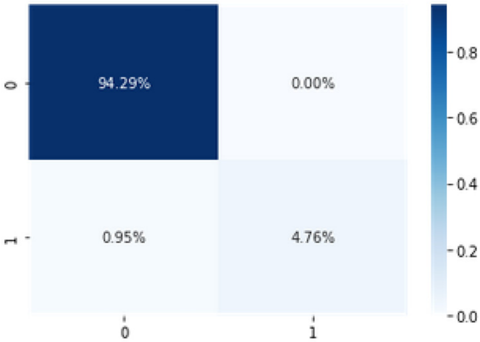
Source: Authors own work

Figure 24. SVM



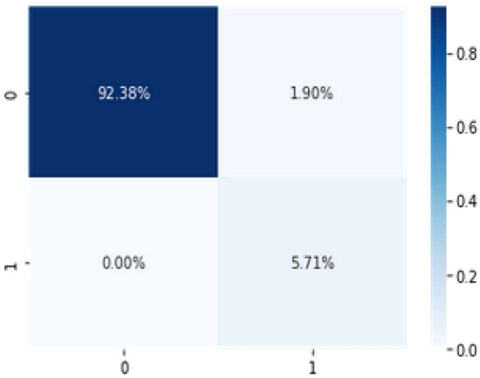
Source: Authors own work

Figure 25. LGB Classifier



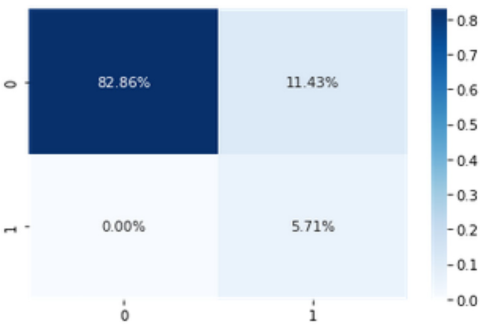
Source: Authors own work

Figure 26. CatBoost



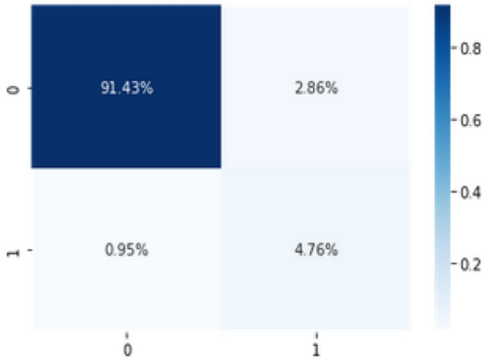
Source: Authors own work

Figure 27. XGBoost

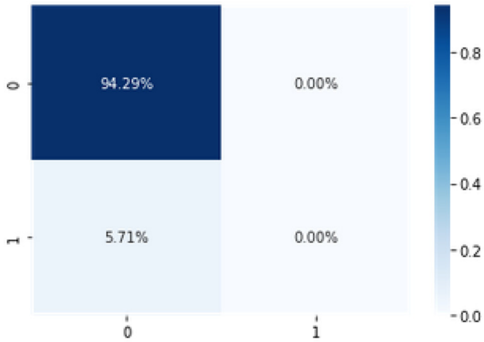


Source: Authors own work

Figure 28. Decision Tree



Source: Authors own work
Figure 29. AdaBoost



Source: Authors own work
Figure 30. Hybrid (SVM+XGBoost)

Table 8. Evaluation measures obtained for PCA

ML model	Class 0 Precision	Recall	F-measure	Class 1 Precision	Recall	F-measure	Macro average	Weighted average
Random Forest	0.98	1.00	0.99	1.00	0.75	0.86	0.92	0.98
SVM	0.94	1.00	0.97	0.00	0.00	0.00	0.48	0.91
KNN	0.94	1.00	0.97	0.00	0.00	0.00	0.48	0.91
LGB classifier	0.98	1.00	0.99	1.00	0.75	0.86	0.92	0.98
CatBoost	0.98	1.00	0.99	1.00	0.75	0.86	0.92	0.98
XGBoost	0.98	1.00	0.99	1.00	0.75	0.86	0.92	0.98
Decision tree	0.98	0.94	0.96	0.43	0.75	0.55	0.75	0.93
AdaBoost classifier	0.98	1.00	0.99	1.00	0.75	0.86	0.92	0.98
Hybrid (SVM+XGBoost)	0.94	1.00	0.97	0.00	0.00	0.00	0.48	0.91

Source: Authors’ own work

Classifier, CatBoost, XGBoost and AdaBoost whereas for Class 0, the highest precision value accounts to 0.98 for 6 ML models.

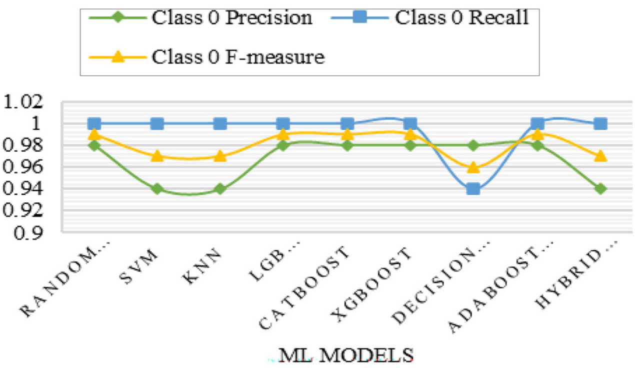
SVM, KNN and Hybrid (SVM + XGBoost) models are struggling to classify the positive instances for Class 1 as all these models are yielding a F-score of 0.00. An F-score of 0.00 indicates high FP and FN i.e. the model is incapable of identifying the correct instances for a given class. Decision Trees also are not performing well and are reporting a F-score of 0.55. All these models require some ramifications for the given data set to yield good results. The graphical representation of all these models on obtained value of precision, recall and f-score for Class 0 and Class 1 is presented in Figure 31 and 32, respectively. Class 1 is showing extreme results i.e. models like RF, LGB, CatBoost, XGBoost and ADABoost are generating a perfect precision score of 1.00 whereas all the remaining models except decision trees are showing 0.00 precision score. Whereas for Class 0 all the models are performing very well with average precision value of 0.96.

The confusion matrix results for all the ML model with PCA scaling is presented in Figures 33–41. The results produced for PCA scaling are quite distinctive with the highest true positive rate of 93.94% for all 8 ML models except the decision tree model which has a

Table 9. Loss function and accuracy of different ML models for PCA

Algorithms	Mean squared error	Root mean squared error	Accuracy
Random Forest	0.01	0.12	0.98
KNN	0.06	0.24	0.93
SVM	0.06	0.23	0.94
LGB classifier	0.01	0.12	0.98
CatBoost	0.05	0.13	0.94
XGBoost	0.01	0.12	0.98
Decision tree	0.07	0.27	0.92
AdaBoost	0.01	0.12	0.97
Hybrid model (SVM + XGBoost)	0.06	0.24	0.93

Source: Authors’ own work

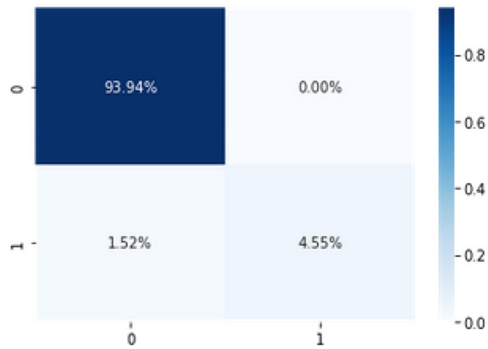


Source: Authors own work

Figure 31. PCA results of precision, recall and F-measure for class 0

Source: Authors own work

Figure 32. PCA results of precision, recall and F-measure for class 1



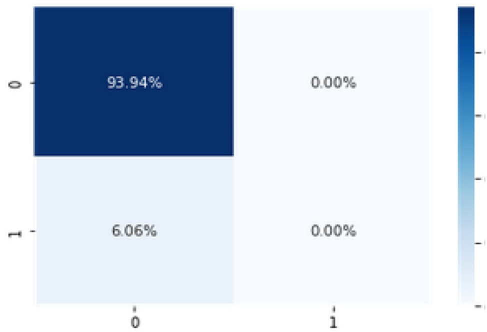
Source: Authors own work

Figure 33. Random Forest

true positive value of 87.88%. KNN, SVM and Hybrid Model (SVM + XGBoost) are giving the same set of results for both true positives and false negative i.e. 93.94% and 6.06% respectively and are the best models with no error rate. Following them is the Random Forest, LGB Classifier, CatBoost, AdaBoost and XGBoost with same set of results for TP, FN and TN i.e. 93.94%, 1.52% and 4.55%, respectively.

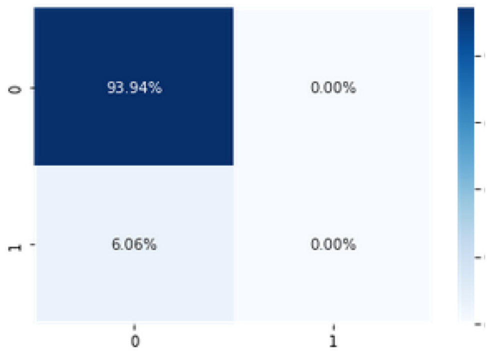
4. Discussion

The key results obtained from the application of all the nine ML models is presented in this section. Two loss functions MSE and RMSE are used to find the discrepancy between values generated by the model and the true target values. Also, accuracy of all the models is also computed. All these measures i.e. loss functions and accuracy are generated for three sets of feature scaling, as mentioned in Section 3.2. The graphical representation of the values obtained using these measures for min-max scaling, Standard Scaler and PCA is presented in



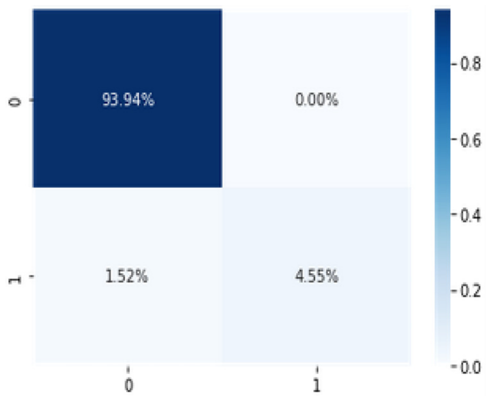
Source: Authors own work

Figure 34. KNN



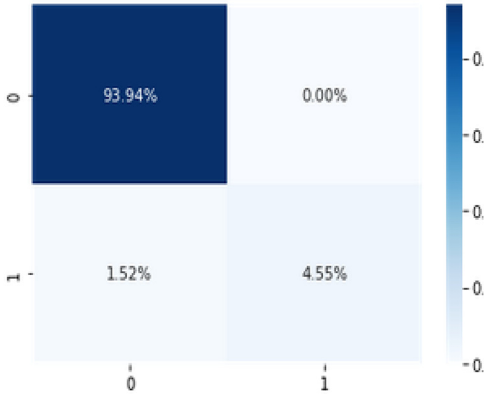
Source: Authors own work

Figure 35. SVM



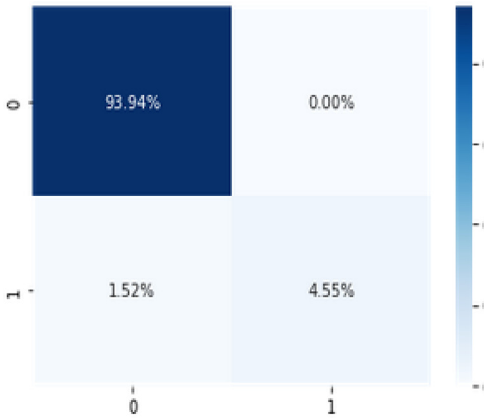
Source: Authors own work

Figure 36. LGB Classifier



Source: Authors own work

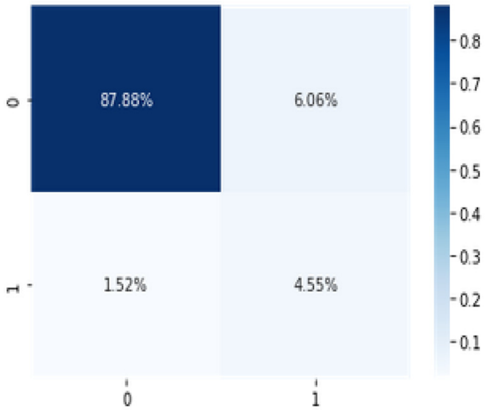
Figure 37. CatBoost



Source: Authors own work

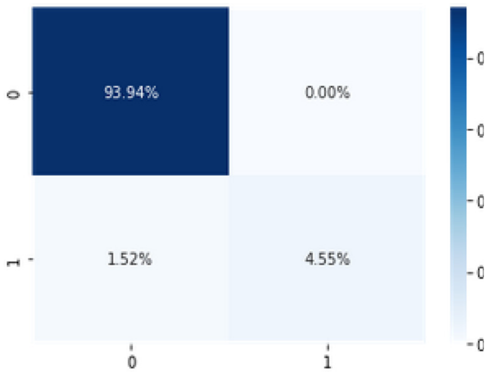
Figure 38. XGBoost

Figures 42, 43 and 44, respectively. Lower MSE and RMSE values represent better predictive performance because they represent less prediction errors and closer proximity between predicted and actual value. The model with lowest prediction error can also be identified using these measures which in our case turns out to be Random Forest with least prediction error in all the three cases of feature scaling. Random Forest is outperforming all other models in terms of accuracy as well. The least performing model according to these measures comes out to be decision trees with highest MSE and RMSE values of 0.12 and 0.35; 0.11 and 0.33; 0.07 and 0.27 for min-max, Standard Scaler and PCA, respectively. CatBoost and XGBoost models are second in line in terms of accuracy with 99% and 98% accuracy for both min-max and Standard Scaler whereas both models are also performing well for PCA also with 94 and 98% accuracy. Min-max scaling and Standard Scaler are



Source: Authors own work

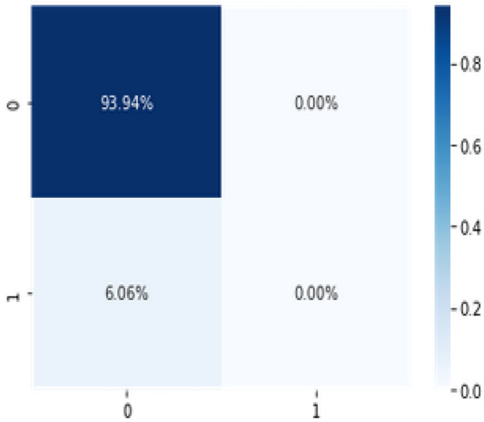
Figure 39. Decision Tree



Source: Authors own work

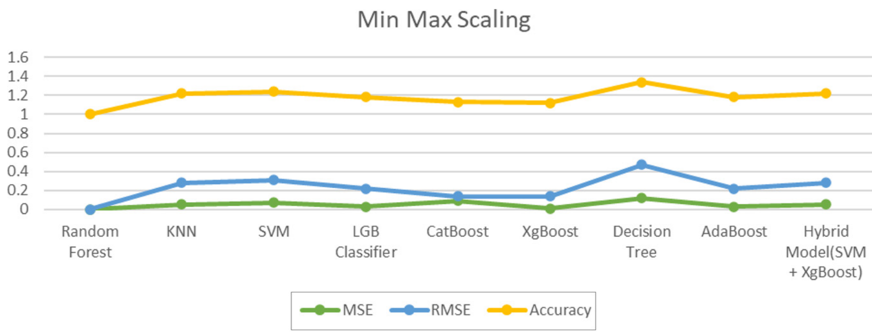
Figure 40. AdaBoost

generating almost similar results with a slight difference of 1% in some models, namely, KNN, SVM, LGB Classifier and Decision Tree. For rest of the models, the results are similar in both feature scaling methods. PCA has generated different set of results for all the models under consideration with highest accuracy of 98% by three ML models, namely, Random Forest, LGB Classifier and XGBoost. The chosen Hybrid model is performing at average level with accuracy of 94% as opposed to standard models which are performing well on the HR data set. It is also significant to note that as per the results obtained Random Forest can be successfully applied to the HR data set and generate desirous results. Also, our problem is a binary classification problem. Hence, results generated are desirous. When the number of classification classes will be increased, there may be variation in results. In future, we tend to check these models with other feature scaling techniques for multi-class classification.



Source: Authors own work

Figure 41. Hybrid (SVM+XGBoost)

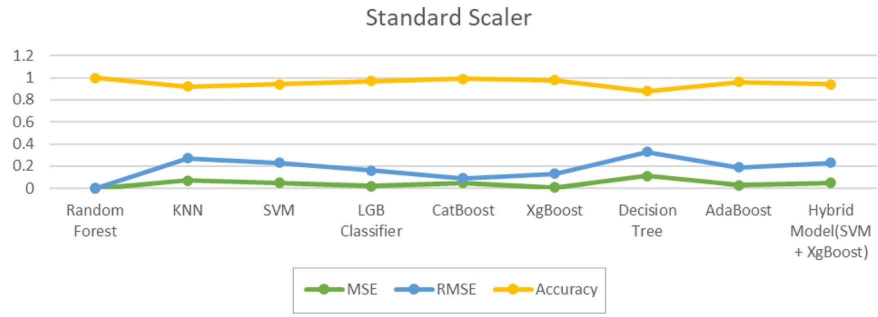


Source: Authors own work

Figure 42. Accuracy, MSE and RMSE values for proposed ML model using min-max scaling

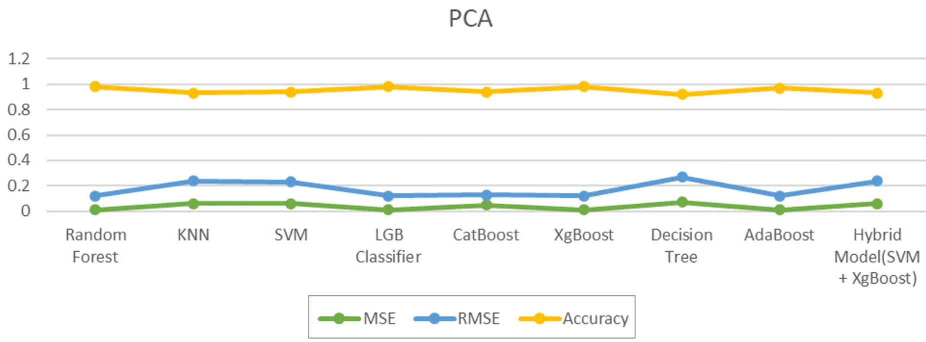
5. Research implications

The findings of the proposed study as presented in Section 3 indicated that AI models have considerable potential for predicting the future talent needed for an organisation to operate effectively. All AI models are quite effective at classifying employee talents and dividing the workforce into two groups. Ones that are helpful and succeeding as well as those that require development to reach their full potential. These types of insights will help companies in identifying patterns, trends and corelations that might not be apparent with the traditional approach. Based on historical data, AI models can forecast future trends and outcomes, enabling organisations to proactively handle possible obstacles or opportunities. Inventory control, resource allocation and marketing tactics can all be improved by predictive analytics. Processes for hiring new employees, onboarding them and evaluating their performance can all be streamlined by AI. Additionally, AI-driven solutions can assist in identifying skill gaps and offer employees individualised programmes for training and development. Organisations will be able to glean insightful information from a variety of sources, including text, photos and audio,



Source: Authors own work

Figure 43. Accuracy, MSE and RMSE values for proposed ML model using Standard Scaler



Source: Authors own work

Figure 44. Accuracy, MSE and RMSE values for proposed ML model using PCA

owing to AI tools' capability to manage complicated and unstructured data. Huge amount of data can be processed with help of AI technology thereby consuming less resources and time. Also, organisations may improve workflows, resource allocation and supply chain management using AI-driven automation, which boosts operational effectiveness. Organisations that adopt AI technology can gain a competitive edge by maintaining a lead in innovation and offering their clients better goods and services.

6. Conclusion and future direction

AI-powered performance management tools can help organisations predict employee potential and take proactive measures to retain valuable talent. By using data analytics and AI, organisations can make more informed decisions, leading to improved employee and organisational success. These tools can provide insights that were previously unavailable, save managers time and effort and create more equitable workplaces. The results of this study have important ramifications for businesses looking to use AI technology to improve the use of their workforce and increase productivity. We have emphasised the significance of precise performance prediction for organisations through our study. The capacity to predict employee performance enables proactive actions to be taken, including customised career planning, performance rewards and tailored training and development programmes. Although AI models have shown promising outcomes, it is crucial to recognise the constraints and difficulties involved in their use. To ensure the fair and responsible use of AI in employee performance prediction, ethical considerations, privacy problems and any biases in the data should be properly addressed. Future work will be required to improve and broaden the capabilities of AI models in predicting employee performance. Future research should focus on expanding data sources, enhancing algorithms and including interpretability criteria.

References

Ahamed, B.S. and Arya, S. (2021), "LGBM classifier based technique for predicting type-2 diabetes", *Eur. J. Mol. Clin. Med*, Vol. 8 No. 3, pp. 454-467.

Ali, I.A.A., Mahmoud, A., Tahani, A.L., Shawabkeh. and Salameh, W.A. (2019), "Performance predicting in hiring process and performance appraisals using machine learning", *10th International Conference on Information and Communication Systems (ICICS),* pp. 110-115.

Al-Radaideh, Q.A. and Nagi, E.A. (2012), "Using data mining techniques to build a classification model for predicting employees performance", *Int. J. Adv. Comput. Sci. Appl*, Vol. 3 No. 2, p. 8.

Ancheta, R.A., Cabauatan, R.J.M., Tanguilig Lorena, B.T. and Rabago, W. (2012), "Predicting faculty development trainings and performance using rule-based classification algorithm", *Asian J. Comput. Sci. Inf. Technol*, Vol. 2 No. 7, pp. 203-209.

Budhwar, P., Malik, A., De Silva, M.T.T. and Thevisuthan, P. (2022), "Artificial intelligence – challenges and opportunities for international HRM: a review and research agenda", *The International Journal of Human Resource Management*, Vol. 33 No. 6, pp. 1065-1097.

Chen, T. and Guestrin, C. (2016), "XGBoost: a scalable tree boosting system", *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min*, Vol. 13, pp. 785-794.

Coron, C. (2022), "Quantifying human resource management: a literature review", *Personnel Review*, Vol. 51 No. 4, pp. 1386-1409.

Davies, W. and McDonald, A. (2018), "The science behind predicting job performance at recruitment".

Dorogush, A.V., Ershov, V. and Yandex, A.G. (2018), "CatBoost: gradient boosting with categorical features support".

Garg, S., Sinha, S., Kar, A.K. and Mani, M. (2022), "A review of machine learning applications in human resource management", *International Journal of Productivity and Performance Management*, Vol. 71 No. 5, pp. 1590-1610.

George, D.G. and Thomas, D.M.R. (2019), "Integration of artificial intelligence in human resource", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 9 No. 2, pp. 5069-5073.

He, C., Ding, C.H.Q., Chen, S. and Luo, B. (2021), "Intelligent machine learning system for predicting customer churn", *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI*, Vol. 2021, pp. 522-527.

Jadav, D. *et al.* (2022), "EmReSys: AI-based efficient employee ranking and recommender system for organisations", 3rd IEEE 2022 Int. Conf. Comput. Commun. Intell. Syst. ICCCIS 2022, pp. 440-445,

Jantan, H., Hamdan, A.R. and Othman, Z.A. (2015), "Knowledge discovery techniques for talent forecasting in human resource application", *World Acad. Sci. Eng. Technol*, Vol. 38, pp. 803-811.

Jayadi, R., Firmantyo, H.M., Dzaka, M.T.J., Suaidy, M.F. and Putra, A.M. (2019), "Employee performance prediction using naïve bayes", *Int. J. Adv. Trends Comput. Sci. Eng*, Vol. 8 No. 6, pp. 8-12.

Kaushal, N., Kaurav, R.P.S., Sivathanu, B. and Kaushik, N. (2021), "Artificial intelligence and HRM: identifying future research agenda using systematic literature review and bibliometric analysis", *Manag. Rev. Q. 2021*, Vol. 732 No. 2, pp. 455-493.

Kirimi, J.M. and Moturi, C.A. (2016), "Application of data mining classification in employee performance prediction", *Int. J. Comput. Appl*, Vol. 146 No. 7, pp. 975-8887.

Kramer, O. (2013), "K-Nearest neighbors," *Dimensionality Reduction with Unsupervised Nearest Neighbors*, Springer, Berlin, Heidelberg, pp. 13-23.

Lather, A.S., Malhotra, R., Saloni, P., Singh, P. and Mittal, S. (2019), "Prediction of employee performance using machine learning techniques", ACM International Conference Proceeding Series.

Meijerink, J. (2020), "Talent management in the gig economy", *Digit. Talent Manag*, pp. 98-121.

Meijerink, J., Boons, M., Keegan, A. and Marler, J. (2021), "Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM", *The International Journal of Human Resource Management*, Vol. 32 No. 12, pp. 2545-2562.

New Study (2024), "64% of people trust a robot more than their manager", available at: www.oracle.com/corporate/pressrelease/robots-at-work-101519.html

Nsor-Ambala, R. (2020), "Applying conventional agency theory to prediction of employee self-reporting performance behaviour", *International Journal of Productivity and Performance Management*, Vol. 70 No. 7, pp. 1728-1750.

Palos-Sánchez, P.R.P., Baena-Luna, A., Badicu, J. and Infante-Moro, C. (2022), "Artificial intelligence and human resources management: a bibliometric analysis", *Appl. Artif. Intell*, Vol. 36 No. 1.

Payne, S.C., Horner, M.T., Boswell, W.R., Schroeder, A.N. and Stine-Cheyne, K.J. (2009), "Comparison of online and traditional performance appraisal systems", *Journal of Managerial Psychology*, Vol. 24 No. 6, pp. 526-544.

Popkova, E.G. and Sergi, B.S. (2020), "Human capital and AI in industry 4.0. Convergence and divergence in social entrepreneurship in Russia", *Journal of Intellectual Capital*, Vol. 21 No. 4, pp. 565-581.

Potočnik, K., Anderson, N.R., Born, M., Kleinmann, M. and Nikolaou, I. (2021), "Paving the way for research in recruitment and selection: recent developments, challenges and future opportunities", *European Journal of Work and Organizational Psychology*, Vol. 30 No. 2, pp. 159-174.

Raut, T., *et al.* (2020), "Employee performance prediction system using data mining", *Int. Res. J. Eng. Technol*, pp. 1552-1554.

RPubs - HR Dataset Codebook v14 (2024), available at: https://rpubs.com/rhuebner/hrd_cb_v14

Sam, D., Ganesan, M., Ilavarasan, S. and Victor, T.J. (2023), "Hiring and recruitment process using machine learning", *Proc. Int. Conf. Artif. Intell. Knowl. Discov. Concurr. Eng. ICECONF 2023*.

Schölkopf, B. and Smola, A.J. (2018), *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, The MIT Press.

Su, C.T. and Yang, C.H. (2008), "Feature selection for the SVM: an application to hypertension diagnosis", *Expert Systems with Applications*, Vol. 34 No. 1, pp. 754-763.

Tambe, P., Cappelli, P. and Yakubovich, V. (2019), "Artificial intelligence in human resources management: challenges and a path forward", *California Management Review*, Vol. 61 No. 4, pp. 15-42.

Wang, R. (2012), "AdaBoost for feature selection, classification and its relation with SVM, a review", *Physics Procedia*, Vol. 25, pp. 800-807.

Yasodha, S. and Prakash, P.S. (2012), "Data mining classification technique for talent management using SVM", 2012 Int. Conf. Comput. Electron. Electr. Technol. ICCEET 2012, pp. 959-963.

Zhao, Y., Hryniewicki, M.K., Cheng, F., Fu, B. and Zhu, X. (2018), "Employee turnover prediction with machine learning: a reliable approach", *Advances in Intelligent Systems and Computing*, Vol. 869, pp. 737-758.

Further reading

Nowaczyk, S., *et al*. (2022), "An efficient AdaBoost algorithm with the multiple thresholds classification", *Appl. Sci*, Vol. 12 No. 12, p. 5872.

Corresponding author
Aakanksha Uppal can be contacted at: aakanksha.uppal@gmail.com