Comparative Analysis of Machine Learning Models for Employee Performance Evaluation

AKINSIKU, A. M.[1], AKINTOLA K. G[2]
[1]*Computer Science Department, The Federal Polytechnic, Ado-Ekiti, Ekiti State*
[2]*Software Engineering Department, School of Computing, The Federal University of Technology, Akure. Ondo State*

Abstract- Evaluating employees' performance is an important tool used by management of organizations to make decisions related to employee growth, promotions, compensation or renumeration, training, and appraise organizational growth. In this research, machine learning techniques have been applied to staff performance evaluation. The study encompassed the development and evaluation of multiple predictive models, each harnessed to uncover patterns and access the performance of staff. Comparative analysis of selected algorithms, including Naive Bayes, Support Vector Machine (SVM), Decision Tree, Logistic Regression, Neural Network and Ensemble model were carried out to determine which gives the best result. These models were trained and rigorously tested to ascertain their efficacy in predicting staff performance. The outcomes of the study provide valuable insights into the potential of machine learning approaches to unravel staff performance.

Indexed Terms- Human resources, Machine Learning Model, Performance Evaluation, Ensemble

I. INTRODUCTION

Employee performance evaluation is a cornerstone of human resource management, driving decisions related to promotions, compensation, training, and workforce development. Accurate assessments of employee performance enable organizations to recognize talent, identify areas for improvement, and align employee contributions with strategic goals. Journal of Applied Psychology defines "Employee Performance Evaluation as the systematic process of assessing an employee's job performance, usually in terms of predetermined standards and providing feedback to the employee for the purpose of improving future performance".

Performance evaluation typically involves qualitative and quantitative aspects, including technical skills, teamwork, leadership, problem-solving abilities, and communication. These criteria are not always straightforward to measure, as they are influenced by various factors, such as the employee's environment, job role, and interpersonal dynamics. However, conventional performance evaluation methods, such as ranking systems, peer reviews, and supervisor appraisals, often fall short in capturing the full complexity of employee performance. These approaches are frequently criticized for their subjective nature, potential biases, and inability to adapt to dynamic, modern work environments.

Conventional evaluation methods tend to rely on rigid scoring systems, which can oversimplify these multifaceted dimensions and fail to account for the uncertainties inherent in human behavior. Consequently, there is a growing demand for more adaptive and flexible evaluation systems that can handle these complexities and provide more accurate, objective, and transparent assessments. To address these limitations, machine techniques such as Support Vector Machines, Fuzzy Logic, Artificial Neural Networks (ANNs) and other machine-learning models have gained traction as innovative solutions in employee performance evaluation.

II. LITERATURE REVIEW

Employee performance evaluation is a structured process in which an organization reviews and assesses an employee's job performance and overall contributions over a specific time. It typically involves comparing the employee's work behaviors, achievements, and competencies against predefined criteria or goals to determine their effectiveness in the role and to identify areas for improvement or development. The results of these evaluations are used to provide feedback, guide decisions about promotions or compensation, and inform career development strategies.

Armstrong (2017) defines employee performance evaluation as "the formal assessment and rating of individuals by their managers, usually at an annual review meeting." Similarly, Aguinis (2019) describes it as "a continuous process of identifying, measuring, and developing the performance of individuals and aligning their performance with the strategic goals of the organization." making it a critical tool for employee engagement and productivity enhancement. Performance evaluation, therefore, serves administrative, developmental, and strategic purposes, benefiting both the organization and its employees.

The primary purpose of performance evaluation is to assess how well employees perform their assigned duties and to ensure alignment with organizational goals. According to Dessler (2020), performance evaluations are critical in fostering employee engagement, motivation, and productivity. By providing regular feedback, organizations can guide employees toward continuous improvement. This process also ensures that individuals understand the expectations placed upon them and receive the support needed to achieve both personal and professional growth.

Additionally, performance evaluations play a crucial role in decision-making for rewards and promotions. As Aguinis (2019) explains, performance management systems help organizations differentiate high performers from low performers, enabling decisions related to salary increases, promotions, and even layoffs. Furthermore, developmental feedback offered during evaluations helps employees identify their strengths and areas for improvement, which supports long-term career growth and development.

2.1 Types of Performance Evaluation

Employee performance evaluation can be conducted using various methods, each offering unique insights and benefits. The most common ones are;

a. The annual performance review: Under this method, managers assess employee performance against predefined goals. However, this method is often criticized for being retrospective and failing to provide timely feedback.
b. Continuous feedback systems: In response to the limitations of annual performance review, many organizations are adopting continuous feedback systems that allow for more frequent evaluations. Pulakos et al., (2011) argue that continuous feedback helps employees make real-time adjustments to their performance and keeps them aligned with organizational priorities throughout the year.
c. 360-degree feedback: Comprehensive feedback is collected from various sources including peers, subordinates and supervisors. This multi-source feedback method provides a more comprehensive view of an employee’s performance, especially regarding soft skills like communication, teamwork, and leadership. It reduces the impact of individual biases by incorporating multiple perspectives (Grote, 2011).
d. Self-assessment: Employees evaluate their performance against personally set criteria. This is also becoming more common, encouraging employees to reflect on their own performance and take ownership of their development.
e. Peer review: Feedbacks are gathered from an employee’s colleagues and counterparts about their performance. It provides insight into the employee’s interpersonal and teamwork skills and how they contribute to a positive work environment.
f. Behaviorally Anchored Rating Scale (BARS): This is a tool for the evaluation of employees using a set of well-defined performance criteria by comparing their behaviors with specific behavior traits that anchor each performance level to numerical ratings. BARS utilize behavioral statements to explain various stages of performance for each element of performance (Elverfeldt, 2005)

2.2 Key Criteria in Performance Evaluation

To ensure fair and comprehensive assessments, performance evaluations typically rely on a combination of quantitative and qualitative factors. Job knowledge and technical skills form the foundation of most evaluations, as they assess whether employees have the necessary skills to perform their roles. Quality and quantity of work measure how well employees meet their goals, deadlines, and performance standards. Dessler (2020) emphasizes that these are critical indicators of an employee's ability to contribute to the organization's bottom line.

Beyond technical competencies, evaluations also consider soft skills such as communication, teamwork, and problem-solving. According to Armstrong (2017), these are essential for employees who work in collaborative environments or leadership roles. Effective communication and teamwork are often critical for organizational success, while problem-solving and innovation drive improvement and efficiency. Many performance evaluations also assess attitude and behavior, including how well employees align with the organization's values and culture, which is increasingly important in maintaining a positive and productive workplace (Armstrong, 2017).

2.3 Challenges and Limitations of Performance Evaluation

While employee performance evaluation is an important process, it is not without its challenges. Some of the limitations are explicitly highlighted below;

a. Bias: Personal biases, such as favoritism, the halo effect, or recency bias, can distort evaluations and lead to unfair assessments. For example, managers may overemphasize recent successes (recency bias) or allow a single positive trait (halo effect) to influence their overall judgment of an employee (Armstrong, 2017). These biases can reduce the credibility of performance reviews and damage employee morale.
b. Infrequency of evaluations: Particularly in the case of annual reviews, employees may not receive timely feedback that could help them improve throughout the year. As Pulakos and O'Leary (2011) note, continuous feedback systems address this issue by providing more frequent assessments, allowing employees to make real-time adjustments to their performance.
c. Subjectivity of qualitative criteria: Subjectivity of qualitative criteria such as communication and teamwork, can introduce inconsistencies between different managers or departments. This subjectivity makes it difficult to standardize evaluations across an organization, leading to potential dissatisfaction or perceived unfairness (Dessler, 2020).
d. Halo effect: This occurs when an employee's overall performance is rated based on one specific trait, leading to inaccurate assessment.
e. Central tendency: Some appraisers may avoid extreme ratings, this may result into employees been rated as average, which doesn't their true performance.
f. Inadequate feedback: if the feedback provided for employee evaluation isn't specific, employees may not understand how to improve themselves.

2.4 Best Practices for Effective Performance Evaluation

To maximize the benefits of performance evaluations, organizations should adopt several best practices. First, it is essential to set clear, measurable goals that employees can aim for, ensuring that evaluations are objective and aligned with both individual and organizational goals. As Armstrong (2017) points out, SMART (Specific, Measurable, Achievable, Relevant, and Time-bound) goals help eliminate ambiguity and ensure that everyone is on the same page.

Another best practice is to ensure regular feedback rather than relying solely on annual reviews. Continuous feedback systems, as suggested by Pulakos and O'Leary (2011), allow for timely course corrections and help employees remain engaged and motivated. Additionally, using objective data, such as key performance indicators (KPIs) and metrics, can reduce bias and make evaluations more reliable.

Organizations should also involve employees in the evaluation process by incorporating self-assessments and promoting open discussions. According to Aguinis (2019), this encourages employees to take ownership of their development and fosters a sense of collaboration between managers and employees. Finally, training managers on how to conduct fair and consistent evaluations can further reduce bias and ensure that all employees are evaluated according to the same standards.

2.5 Machine Learning

The term "machine learning" is attributed to Arthur Lee Samuel, an AI pioneer, who coined it in 1959. Samuel (1959) defined machine learning as the field of study that grants computers the ability to learn without explicit programming. Machine learning is a multifaceted and dynamic domain, and its definition can vary depending on the specific field of application. Learning lies at the heart of human knowledge and intelligence, and similarly, it plays a pivotal role in constructing intelligent machines. Years of research in AI have demonstrated that trying to build intelligent computers by manually programming all the rules is impractical; automatic learning is indispensable. For instance, humans do not possess innate language comprehension; we acquire it through learning. Consequently, it makes sense to enable computers to learn language rather than attempting to encode it manually.

Machine learning has found extensive applications in data mining, computer vision, natural language processing, biometrics, search engines, medical diagnostics, credit card fraud detection, securities market analysis, DNA sequencing, speech and handwriting recognition, strategy games, and robotics. Andrew, A. M. (2000).

2.5.1 Categories of Machine Learning

Machine learning algorithms are categorized as supervised, unsupervised, semi-supervised, and reinforcement learning. According to Patel et al. (2020), the following categories of machine learning were identified.

a. Supervised machine learning algorithms: They use previous knowledge acquired from labeled dataset to make predictions about future event. Examples in this category are: decision trees, support vector machines, and neural networks, Bayesian classification and logistic regression.
b. Unsupervised machine learning algorithms: They work on data without labels. Unsupervised learning explores how systems can discern hidden structures from unlabelled data. Unlike supervised learning, it doesn't predict specific outputs but rather explores the data to uncover hidden structures. Unsupervised models are typically used for clustering analysis (group assignment) and dimensionality reduction (compressing data into a lower-dimensional representation).
c. Semi-supervised machine learning algorithms: They utilize both labeled and unlabeled data during training. Typically, a small amount of data is labeled, while a large portion remains unlabeled. This approach significantly enhances learning accuracy without requiring extensive labeled data. Notable algorithms in this category include Laplacian support vector machines.
d. Reinforcement machine learning: This involves an interactive learning method where a system interacts with its environment, takes actions, and learns from rewards or errors. It relies on trial and error search and delayed rewards. This approach enables machines and software agents to autonomously determine optimal behavior within a specific context to maximize their performance. The reinforcement signal, often in the form of reward feedback, guides the agent to learn which actions are best. Typical applications include games (e.g., chess, Go, Atari video games) and various robotics applications such as drones, warehouse robots, and self-driving cars.

2.6 Review of Related Work

Sohara, *et al.* (2023) presents Machine Learning Algorithm to Predict and Improve Efficiency of Employee Performance in Organizations. The research was carried out to address the challenges organizations face in evaluating and improving employee performance by developing and implementing machine learning algorithms that can predict and improve the process. Collected dataset were trained on Models like: Logistic Regression, Naïve Bayes, Random Forest, K-Nearest Neighbor and XGBoost. XGBoost not only achieved the highest AUC scores but also consumed less memory and has a faster run-time compared to other models. However, the hybridization of the models would probably have birthed a model with higher accuracy and efficiency.

Satya (2024) presents Machine Learning in Employee Performance Evaluation. This study is to address the limitations of traditional employee performance evaluation methods, which often suffer from biases, subjectivity, and inefficiencies. A combined quantitative data collection through questionnaires with statistical analysis and predictive modeling was employed. Satya (2024) proposes a comprehensive framework and guidelines for integrating ML into employee performance evaluation processes.

Dhivya *et al.* (2023) present Employee Performance Prediction for Workforce Planning Using Ensemble Hybrid Model. This paper introduced a novel approach by combining the strengths of ensemble learning techniques, specifically XGBoost and Random Forest, into a hybrid model (EXGBRF), enhancing the accuracy and predictive power of the ensemble model. However, there are concerns about overfitting, limited dataset and scalability.

Adeniyi *et al.* (2022) carried out a comparative analysis of machine learning techniques for the prediction of employee performance. They employed Artificial Neural Network, Random Forest, and Decision tree algorithms to analyze employee performance. Result showed that Artificial Neural network performed better in the prediction of employee performance. Although the research provides insight into the effectiveness of various ML techniques in the context of employee performance prediction, data availability, the generalizability of the findings, and potential biases in the data used for training the models are some of the limitations to note. Patel *et al.* (2022) presents an ensemble model for employee-performance classification to rank and identify low performers in an organization. Machine learning algorithms such as Random Forest, Artificial Neural Network, Decision Tress, and XGBoost were used as the base models. The soft voting technique was used to generate the ensemble model. The system outperformed all the individual models, however, choosing classes by soft voting in the situation where probabilities are the same can be problematic. Also, the model implementation can be complex in terms of time of computation.

III. RESEARCH METHODOLOGY

The proposed system architecture for performance evaluation of the non-academic staff of Federal Polytechnic Ado-Ekiti is presented in Figure 3.1. The architecture consists of two main parts: Data Pre-processing and Model training. It encompasses the sampling, collection, and analysis of data, as well as the model used for developing the performance evaluation. Five base models and an ensemble were considered for evaluation. The base models are Artificial Neural Network, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM) and Decision Tree. All the six models were considered for evaluation and explicit comparative analysis to determine which is most reliable and efficient with the best overall performance.

3.1 Data collection process

Sample data were engineered from the Annual Performance Evaluation Report of non-teaching staff of the Federal Polytechnic, Ado Ekiti, Ekiti State. The data include the biodata, Qualifications both academic and professional, Experience within the polytechnic, Experience outside the polytechnic, other activities within the polytechnic, other activities outside the polytechnic, Publications in recognized journals, and the HOD rating. All the ratings are on a 5-point scale, except for the HOD rating which is based on a 10-point scale.

System Framework

Figure 3.1 shows the workflow of the framework. It consists of data generation to derive the dataset, preprocessing the dataset, training the models, testing the models and evaluating the performance of the models. Figure 3.2 shows the Ensemble model.

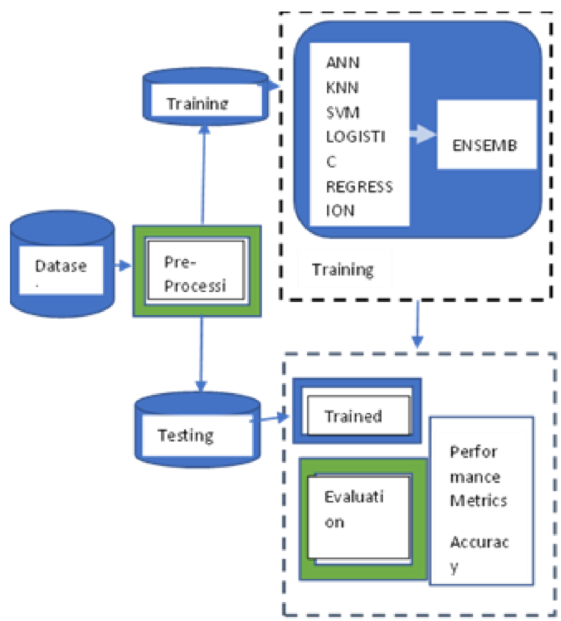

Figure 3.1 The overall system architecture of the evaluation framework.

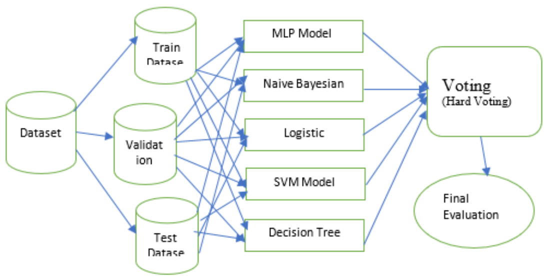

Figure 3.2 Ensemble model for staff performance

IV. RESULTS AND DISCUSSIONS

This research was done using a laptop running 64-bits Windows 10 operating system with Intel (R) Core (TM) i7- 6600U 2.60 GHz processor, and 4GB RAM of memory. The analysis was performed using selected machine learning models. The models were implemented using the Python programming language. A total of 1450 data items was generated. To carry out this experiment, the dataset was split into training set, validation set, and testing set. The training data is the data used to train the models, and the total number of the training data is 1160. The validation set is the data used for the tuning of the model's hyperparameters, and the total number of the validation set is 145. The testing data is the data used for the purpose of testing the models and the total number of the testing data is 145.

Table 4.1: Analysis of the dataset used for each model

SN	Parameter	Neural Network	Logistic Regression	SVM	Bayes	Decision Tree	Ensemble
1	Number of observations	1450	1450	1450	1450	1450	1450
2	Number of training set	1160	1160	1160	1160	1160	1160
3	Number of Validating set	145	145	145	145	145	145
4	Number of test set	145	145	145	145	145	145
5	Number of classes	5	5	5	5	5	5
6	Dimension of input features	8	8	8	8	8	8

4.1 Performance Evaluation

The proposed model is evaluated using the following standard performance metrics; Accuracy, Precision, Recall and F1-score. The outcome of the evaluation of the models are outlined below;

Table 4.2 Performance of the ANN Model

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1.00	1.00	1.00	1.00
Average	1.00	0.99	1.00	0.99
Good	0.93	0.92	0.93	0.93
Very good	0.92	0.92	0.89	0,90
Excellent	1.00	1.00	1.00	1.00

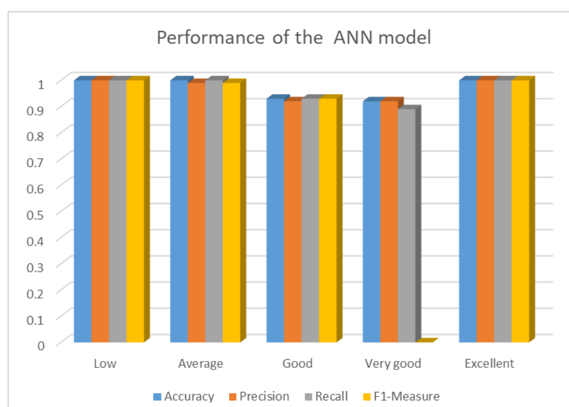

Figure 4.1: Graphical representation of the performance of the ANN model

Table 4.3: Performance of the Naïve Bayes Model

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1.00	0.97	1.00	0.98
Average	0.86	0.99	0.83	0.90
Good	0.87	0.72	0.87	0.79
Very good	0.86	0.80	0,83	0.82
Excellent	0.85	1.00	0.88	0.93

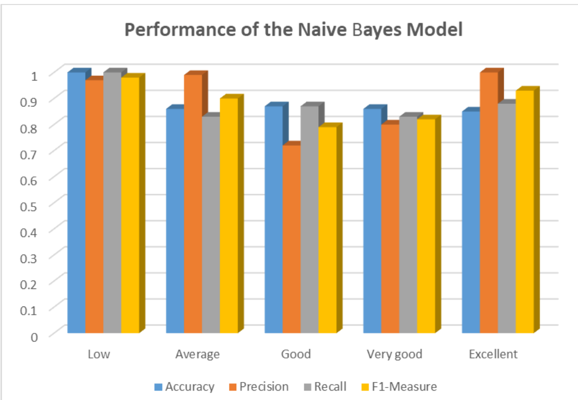

Figure 4.2: Graphical representation of the performance of the Naïve Bayes Model

Table 4.4: Performance of the Logistic Regression Model

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1.00	0.98	1.00	0.99
Average	0.98	0.84	0.86	0.85
Good	0.83	0.82	0.83	0.82
Very good	0.74	0.78	0,74	0.76
Excellent	0.97	1.00	0.97	0.99

Figure 4.3: Graphical representation of the performance of the Logistic Regression model

Table 4.5: Performance of the SVM Model

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1	0.93	1	0.97
Average	0.91	0.99	0.88	0.93
Good	0.97	0.83	0.99	0.9
Very good	0.82	0.94	0.88	0.91
Excellent	0.9	0.99	0.92	0.95

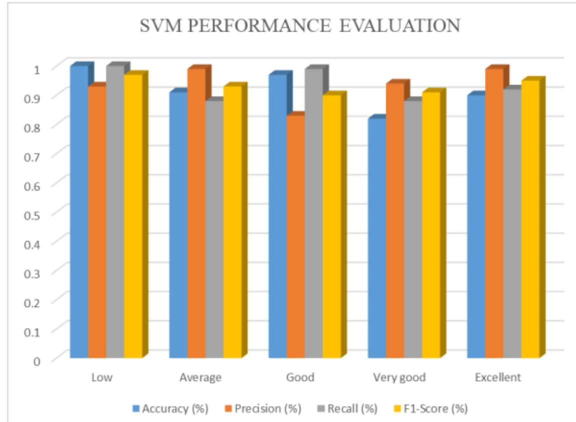

Figure 4.4: Performance of the Support Vector Machine Model

Table 4.6: Performance of the Decision Tree

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1	1	1	1
Average	0.91	0.95	0.91	0.93
Good	0.84	0.83	0.84	0.83
Very good	0.92	0.87	0.92	0.9
Excellent	1	1	1	1

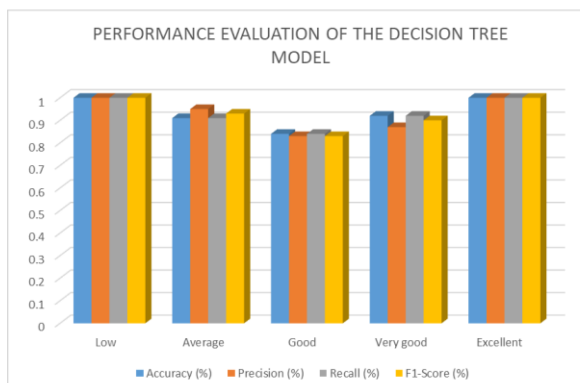

Figure 4.5: Graphical representation of the performance of the Decision Tree Model

Table 4.7: Performance of the Decision Tree

Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Low	1	0.98	1	0.99

Average	0.99	0.86	0.99	0.92
Good	0.79	0.92	0.79	0.85
Very good	0.86	0.96	0.86	0.91
Excellent	1	1	1	1

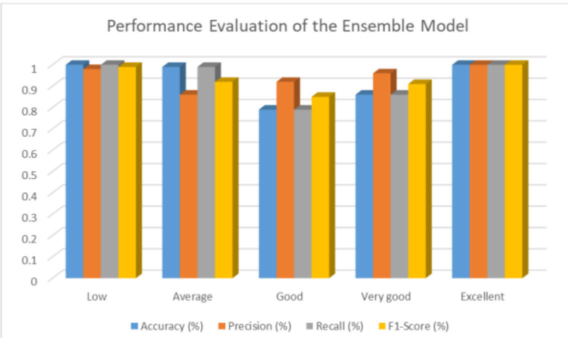

Figure 4.6: Performance of the Ensemble Model

Table 4.8: Comparative Analysis of the Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Model
ANN	0.97	0.97	0.96	0.96	ANN
Naïve Bayes	0.89	0.9	0.9	0.88	Naïve Bayes
Logistic	0.9	0.88	0.92	0.88	Logistic
SVM	0.92	0.94	0.93	0.93	SVM

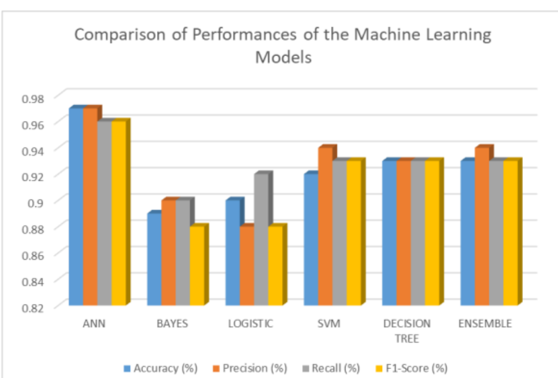

Figure 4.7: Performance of the Models

CONCLUSION

In this research, machine learning techniques have been applied to staff performance evaluation. The study encompassed the development and evaluation of multiple predictive models, each harnessed to uncover patterns and access the performance of staff. The outcomes of the study provide valuable insights into the potential of machine learning approaches to unravel staff performance. By leveraging a combination of machine learning models, a comprehensive comparative analysis of the models using suitability metrics like Accuracy, precision, recall and F1-score helps validate the model with the best overall performance, and most suitable for employee performance evaluation. This shows that ANN possess the ability to provide a more accurate and efficient prediction model for employee performance than other base models analyzed. With this, organizations can create a comprehensive evaluation system that supports data-driven decision-making, and employee development.

REFERENCES

[1] Adeniyi J. K., Adeniyi A.E., Oguns Y.J., Egbedokun G.O., Kehinde Douglas Ajagbe K.D., and Obuzor P.C. (2022). Comparative Analysis of Machine Learning Techniques for the Prediction of Employee Performance. Paradigm Plus vol. 3, No. 3, pp 1- 15, 2022 DOI: 10.55969/paradigmplus. v3n3a1

[2] Aguinis, H. (2019). Performance management (4th ed.). Chicago Business Press. Andrew.A. M. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods by Nello Christianini and John Shawe-Taylor, Cambridge University Press, Cambridge, 2000, xiii+ 189 pp., ISBN 0-521-78019-5 (Hbk,£ 27.50). Robotica, 18(6), 687-689.

[3] Armstrong, M. (2017). Armstrong's handbook of Performance Management (6th ed.). Dessler, G. (2020). Human resource management (16th ed.). Pearson.

[4] Dhivya, R. S., & Sujatha, P. (2023). Employee performance prediction for workforce planning using ensemble hybrid model. In Proceedings of the 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 839-844). New Delhi, India.

[5] Elverfeldt A. V (2005). Performance Appraisal-how to improve its effectiveness. University of Twenty, Enschede.

[6] Grote, D. (2011). How to be good at performance appraisals: Simple, effective, done right. Harvard Business Review Press.

[7] Patel, K., Sheth, k., Mehta, D., Tanwar, S., Florea, B. C., Taralunga, D. D., Altameem, A., Altameem, T., Sharma, R. (2022) Ranker: An AI-Based Employee Performance Classification Scheme to Rank and Identify Low Performers.

[8] Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020) Machine learning methods in drug discovery. Molecules, 25(22), 5277. Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management broken? Industrial and Organizational Psychology, 4(2), 146-164.

[9] Samuel A.L. (1959) Some studies in machine learning using the game of checkers. IBM J Res Dev 3:210–229. https://doi.org/10.1147/rd.33.0210

[10] Satya P. (2024). Machine learning in employee performance evaluation: A HRM perspective. International Journal of Scientific Research and Applications, 10(4), 193-210.

[11] Sohara, B., Nipun, A., Akhil, S., Sobiya, S. & Sai, S. (2023). Machine learning algorithm to predict and improve efficiency of employee performance in organizations. *Journal of Data Science and Applications*, *9*(3), 145-162.