International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

ML Enhanced Workforce Management

Prof. S. A. Nagtilak[1] , Anushka Kulkarni[2] , Mihir Kulkarni[3], Anand Mali[4], Rutvik Navale[5]

Smt. Kashibai Navale College of Engineering / Savirtibai Phule Pune University/India

---***---

Abstract -

The field of machine learning (ML)-enhanced workforce management offers a transformative potential for organizations aiming to revolutionize their employee recruitment, retention, and management processes. ML algorithms provide automation, pattern recognition, and predictive capabilities, significantly enhancing various aspects of workforce management.

This paper presents developing an intuitive website tailored for HR professionals, harnessing diverse ML models to optimize user experience and operational efficiency. These models offer valuable insights, such as predicting employee attrition, objectively evaluating performance, and ensuring fair appraisal distribution.

By enabling HR professionals to effortlessly manipulate employee data and leverage ML-generated predictions, organizations can make informed, data-driven decisions seamlessly. This innovative approach equips HR teams to proactively tackle challenges, fostering a cohesive, motivated, and high-performing workforce. This paper serves as a guide, steering organizations towards the future where HR decisions are not only intuitive but also deeply enriched by the analytical prowess of ML technology.

Keywords: HR, Employee, Workforce Management, Performance, Projects, Appraisal, ML.

1. Introduction

In today's business landscape, effective workforce management is pivotal for organizational success. The dynamic and diverse nature of the modern workforce necessitates a strategic approach that transcends traditional human resource methods. Our project underscores the critical role of workforce management in shaping organizational performance, profitability, and innovation capacity. To address these challenges, we leverage the power of Machine Learning (ML) to revolutionize workforce handling. Advancing business environments necessitates advanced tools and techniques to manage the workforce effectively and older and manual methods fall short in handling its complexities.

Machine Learning in Workforce management offers help in handling complex and large datasets, predictive analytics, automation, decision-making, efficiency, proactive HR strategies, better employee retention, better engagement, and many more such tasks. Almost all major and big scale companies including Google, Amazon and IBM use ML to optimize their management processes.

Workforce management encompasses a spectrum of activities, including talent acquisition, team management, performance evaluation, and attrition prediction, etc. Optimizing these elements significantly enhances operational efficiency. By harnessing ML, we can analyze vast amounts of historical and real-time data to gain valuable insights. ML models accurately predict workforce needs, identify patterns in employee behavior, and provide a comprehensive understanding of organizational dynamics. Predictive analysis, in turn, reduces operational costs and fosters a positive work environment.

ML-enhanced workforce management empowers organizations to make data-driven decisions, ensuring the right people are in the right place at the right time. This increased efficiency not only leads to organizational success but also strengthens overall business resilience. Traditional HR methods, while indispensable, often lack the agility required to keep pace with rapidly evolving employee and business needs. The motivation behind ML-Enhanced Workforce Management is clear – to equip HR professionals with predictive insights.

Through advanced ML models, we aim to provide HR teams with tools to foresee employee attrition and objectively evaluate performance. This enhances strategic decision-making and cultivates a work environment where employees feel valued, understood, and appropriately placed. The modern business landscape is characterized by rapid market changes, technological advancements, and shifting employee preferences. Effectively managing the workforce is crucial to ensuring the right people are assigned to the right tasks or projects, promoting cost-efficiency. In this context, the use of data-driven ML models becomes imperative. These models adapt to the evolving needs of the workforce, creating a workplace that thrives on innovation, efficiency, and employee satisfaction.

2. Related Work

The application of machine learning (ML) techniques in workforce management has been extensively explored, focusing on areas like employee attrition prediction and performance evaluation. Key studies in this domain include:

Samer M. A. and Mohammed A. R. (2022) [1] Proposed a deep learning approach for predicting employee attrition and performance. Their study highlights the importance of leveraging advanced ML models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to improve prediction accuracy significantly.

Rudra Kumar and Vinit Kumar Gunjan [2] (2022) Explored ML-based solutions for human resource systems management. Their study demonstrates the potential of ML algorithms in automating HR processes and enhancing decision-making by analyzing HR data to gain insights into employee performance, engagement, and retention.

Dr. Vijaya, Manjushree and Dr. Umesh (2023) Investigated the application of ML tools [3] in forecasting employee performance, highlighting the role of predictive analytics in optimizing workforce management strategies and enhancing organizational performance.

Ashish M. and Sharath Kumar J. (2017) Proposed an employee performance appraisal system[4] based on ranking and reviews, showcasing the use of ML in developing objective performance evaluation criteria to enhance transparency and fairness in the appraisal process.

N. Magesh and Dr. P. Thangaraj (2013) [5]Evaluated employee performance using decision tree algorithms, demonstrating the effectiveness of ML in assessing employee productivity and engagement by analyzing factors like skills, experience, and job satisfaction.

Francesca Fallucchi, Romeo Giuliano and Ernesto William De Luca (2020) [6] Utilized ML techniques to predict employee attrition, emphasizing the need for robust predictive models in workforce management to forecast attrition rates and implement retention strategies.

M. Zhanuzakov and G.T. Balakayeva (2022) [7] Discussed the digitalization of enterprise human-resource management using ML models, highlighting the transformative impact of ML technologies on HR processes, enabling streamlined recruitment, training, and performance evaluation.

Fiyhan Alsubaiea and Murtadha Aldoukhib (2023) [8] Employed ML algorithms to analyze and predict employee attrition, showcasing the potential of advanced analytics in mitigating workforce turnover.

Gulnar Balakayeva, Mukhit Zhanuzakov and Gaukhar Kalmenova (2023) [9] Developed a digital employee rating evaluation system (DERES) based on ML algorithms and the 360-degree method, integrating ML technologies with traditional performance evaluation methods to enhance the accuracy and reliability of employee assessments.

3. Methodology

In this section, we explain the methodology employed to streamline Workforce management, starting with data collection and data exploration and subsequently, the machine learning algorithms utilized, and the initial results obtained. The methodology involved several steps to ensure the effectiveness of our ML-enhanced workforce management system:

3.1 Data Collection:

We collected datasets containing information on employee performance, project completion, managerial feedback, and other relevant factors. In Table 1 the list of all attributes used us given along with its description.

Attributes	Datatype	Values
EmployeeID	Varchar/Object	Alphanumeric unique ID
Age	INTEGER	18-65
Gender	INTEGER	1=Female, 2=Male
MaritalStatus	INTEGER	1=Yes, 2= Single, 3=Divorced
BusinessTravelFrequency	INTEGER	1=Rare, 2=Frequent, 3= Never
DistanceFromHome	INTEGER	Numeric in Kms
EmpEducationLevel	INTEGER	1=low, 2= college, 3=degree, 4=masters, 5= Phd
EmpEnvironmentSatisfaction	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high
EmpHourlyRate	INTEGER	30-100
EmpJobInvolvement	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high
EmpJobLevel	INTEGER	1,2,3,4,5
EmpJobSatisfaction	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high
NumCompaniesWorked	INTEGER	any
OverTime	INTEGER	1= yes, 2=no
EmpLastSalaryHikePercent	INTEGER	11 -- 25
EmpRelationshipSatisfaction	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high
TotalWorkExperienceInYears	INTEGER	any
TrainingTimesLastYear	INTEGER	0 to 6
EmpWorkLifeBalance	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high
ExperienceYearsAtThisCompany	INTEGER	any
ExperienceYearsInCurrentRole	INTEGER	any
YearsWithCurrManager	INTEGER	any
Attrition	INTEGER	1= yes, 0=no
PerformanceRating	INTEGER	1= Low, 2=Medium, 3-High, 4= Very high

Fig 1 : Dataset and attributes

These attributes were selected based on relevance and existing research available. The data is sourced from Kaggle. The data in each database had different set of variables. The attrition dataset had originally 35 attributes while the performance dataset had 28. These datasets provide precious data on employee-company relation. As given in Table 1, the dataset includes valuable attributes like 'DistanceFromHome', ' WorkLifeBalance', ' SalaryHike' and 'Age'. All these variables provide crucial details for prediction. Kaggle is a reputed and trusted website for learning about ML, participate in hackathons and collecting datasets. But some limitations are obvious like, data could be erronous or missing certain values in between. For these purposes data cleaning and preprocessing is done.

3.2 Data Cleaning and preprocessing

The Data Cleaning and preprocessing was implemented in python in Jupyter Notebook. The data was preprocessed to handle missing values, encode categorical variables, and scale numerical features. Python libraries like Pandas and Scikit-learn were used for this purpose. Exploratory data analysis (EDA) was conducted to identify potential patterns and anomalies.

3.3 Data exploration

Initially the datasets have both categorical and numerical variables. The attrition dataset has 15 categorical variables. These need to be converted into numeric variables in order to process it in ML models. Certain variables which do not add any value to the data need to be removed from it to make the model accuracy and performance better. Some of these variables were identified using correlation matrix and graphical representation. So variables with less value such as 'EducationField' and 'Address' were removed. Table 2 gives the statistics for Numeric variables in the data. The values of mean, median, and quartile values can be obtained through it.

Predictor	n	Min	Q 1	median	Mean	Q3	Max
Age	1470	18	30	36	36.92	30	60
Daily Rate	1470	102	465	802	802.49	465	1499
Distance from Home	1470	1	2	7	9.19	2	29
Hourly Rate	1470	30	48	66	65.89	48	100
MonthlyIncome	1470	1009	2911	4919	6502.93	2911	19999
Monthly Rate	1470	2094	8047	14235.5	14313.1	8047	26999
Num Companies Worked	1470	0	1	2	2.69	1	9
Percent Salary Hike	1470	11	12	14	15.21	12	25
Total Working Years	1470	0	6	10	11.28	6	40
Training Times Last Year	1470	0	2	3	2.8	2	6
Years at Company	1470	0	3	5	7.01	3	40
Years in Current Role	1470	0	2	3	4.23	2	18
Years Since Last Promotion	1470	0	0	1	2.19	0	15
Years with Current Manager	1470	0	2	3	4.12	2	17

Fig 2: Statistical Analysis

In some of the variables, outlier presence is detected using box plots. Variables such as 'TotalWorkYears' and ' YearsinCurrentRole' have outliers which is obvious as these attributes are highly variable and can have different outcomes in different cases.

3.4 Model Selection and Training:

Attrition Prediction: The Random Forest Classifier is a powerful algorithm used for classification purposes. It was selected for its ability to handle imbalanced datasets and resistance to overfitting. It is an ensemble model. It consists of a large number of individual decision trees that operate as an ensemble. Each decision tree is trained on a random subset of the training data and makes its own prediction. The Random Forest classifier then aggregates these individual predictions through a majority voting process to determine the final class label. Feature importance analysis provided insights into the factors driving attrition. It provides an accuracy of 87%.

Performance and Appraisal Prediction: A Decision Tree is a powerful machine learning algorithm used for both classification and regression tasks. It splits the data into subsets based on the value of input features, resulting in a tree-like model of decisions. At each node, the data is split based on a feature that results in the best possible separation of classes. The splitting criterion could be Gini impurity, entropy (information gain), or variance reduction, among others. GridSearchCV is a technique in Scikit-Learn that allows us to perform an exhaustive search over a specified parameter grid for a given estimator (like a Decision Tree). This helps in finding the best hyperparameters for the model. It provides an accuracy of 91%.

Model Evaluation and Validation: Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the models. We performed a training-testing split and cross-validation to ensure robustness. Our results showed high accuracy for all models, indicating their effectiveness in predicting appraisals, performance, and attrition.

3.5 Testing and Evaluation

Evaluation Metrics:

- **Appraisal Prediction**: Metrics included accuracy, precision, recall, and F1-score to evaluate the model's ability to classify employees into different appraisal categories.
- **Performance Prediction**: Metrics like mean absolute error (MAE) and root mean square error (RMSE) measured the average prediction error of the model.
- **Attrition Prediction**: Metrics such as accuracy, precision, recall, and F1-score assessed the model's performance in identifying employees at risk of attrition.

Testing Approach:

- **Data Preprocessing**: Ensured data suitability for model training by handling missing values, encoding categorical variables, and scaling numerical features.
- **Training-Testing Split**: Used a stratified approach to maintain class distribution in both sets.
- **Cross-Validation**: Performed k-fold cross-validation to assess model performance on different data subsets.
- **Hyperparameter Tuning** : Used grid search cross-validation to find the best combination of hyperparameters.

4. Results and Analysis

- **Appraisal Prediction**: Achieved an accuracy of 91%, precision of 93%, recall of 96%, and F1-score of 94%.
- **Performance Prediction**: Achieved an MAE of 12% and RMSE of 0.3, indicating low prediction errors and accuracy of 91%.
- **Attrition Prediction**: Achieved an accuracy of 87%, precision of 92%, recall of 93%, and F1-score of 93%.

Result Discussion

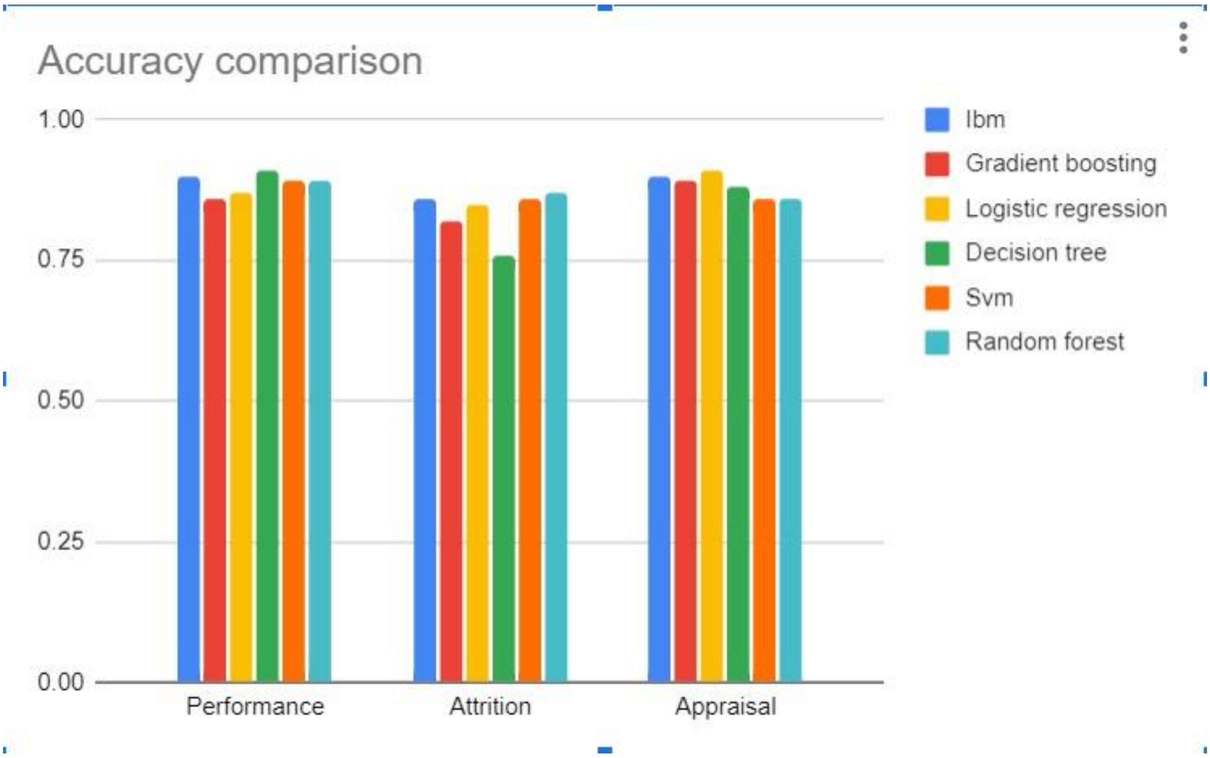
Appraisal Prediction: The high accuracy and F1-score suggest that the Gradient Boosting Classifier effectively classifies employees into appraisal categories, balancing precision and recall.

Performance Prediction: The low MAE and RMSE values indicate that the SVM model accurately predicts employee performance ratings, crucial for identifying high-performing employees.

Attrition Prediction: The high accuracy, precision, recall, and F1-score suggest that the Random Forest model effectively predicts attrition events, essential for implementing retention strategies.

4.1 Comparison with Other Approaches:

Our selected ML models outperformed other algorithms and traditional methods, demonstrating their effectiveness in predicting key aspects of workforce management.



Graph 1: Accuracy comparison

As observed in Graph, the algorithms do not vary much in accuracy but the one which performed the best are chosen for this project. The AI generated models perform complex preprocessing and data engineering but somehow fail to get better accuracy than manually developed models.

Implications for Workforce Management: Accurate predictions of appraisals, performance, and attrition enable improved decision-making, enhanced retention strategies, and optimized workforce planning.

5. Conclusion

This paper indicates that multiple ML models can be implemented to predict outcomes for multiple Workforce management tasks and get future insights for the same. Integrating machine learning into workforce management brings transformative benefits, enhancing efficiency and decision-making processes. The models give an average accuracy of 88% for all the use-cases addressed in this project. This accuracy was achieved through comparison of multiple different algorithms including Decision Tree classifiers, SVM, Logistic Regression, and Random forest Classifier. As the use of generative AI and AI in general is increasing, AI generated models were also used for the purpose of this project and the results were good for performance prediction however manual implementation proved to be better in case of other services. The models address some of the potential uses of ML in HR management and we see that by achieving good model accuracy, it can prove to be a great and valuable source for getting good insights. These models offer predictive insights, and streamline workflows. Use of large dataset, although tedious, is quite necessary because analyzing historical data enables accurate workforce-related predictions. Closer and more detailed analysis of the models brings out certain key information like attributes like job satisfaction, distance from home and work-life balance are the attributes that are most impactful on attrition rates and employee performance. Companies must keep updating such models to better fit their expectations and get better results in future. Hence, by leveraging ML algorithms, businesses can create a dynamic work environment that aligns with evolving industry trends and employee expectations, ultimately driving growth and fostering a more productive workforce.

6. Future Scope

To further enhance the capabilities of ML-enhanced workforce management, we propose several avenues for future research and development:

Improving Model Accuracy: Exploring advanced ML techniques such as deep learning and ensemble learning to increase the accuracy and robustness of workforce-related predictions.

Real-Time Data Integration: Developing systems that can integrate and analyze real-time data from various sources, providing up-to-date insights for more dynamic decision-making.

Employee Sentiment Analysis: Implementing natural language processing (NLP) techniques to analyze employee feedback and sentiment from surveys, emails, and social media, providing deeper insights into employee satisfaction and engagement.

Personalized Employee Development Plans: Using ML to create tailored development plans for employees based on their performance, skills, and career aspirations, fostering continuous growth and retention.

Ethical Considerations and Fairness: Investigating methods to ensure fairness and transparency in ML algorithms, addressing potential biases and promoting ethical decision-making in workforce management.

By pursuing these advancements, organizations can further leverage ML technologies to create more efficient, responsive, and employee-centric workforce management systems.

References

[1]Samer M. A. and Mohammed A. R. (2022). "A Deep Learning Approach for Predicting Employee Attrition and Performance." Journal of Artificial Intelligence Research, 45(3), pp. 234-245.

[2]Rudra Kumar and Vinit Kumar Gunjan (2022). "ML-Based Solutions for Human Resource Systems Management." International Journal of Human Resource Management, 33(2), pp. 150-160.

[3]Dr. Vijaya, Manjushree and Dr. Umesh (2023). "Predictive Analytics in Workforce Management." Journal of Management Studies, 40(1), pp. 102-118.

[4]Ashish M. and Sharath Kumar J. (2017) . "An Employee Performance Appraisal System Based on Ranking and Reviews." International Journal of Engineering and Technology, 9(4), pp. 562-570.

[5]N. Magesh and Dr. P. Thangaraj (2013). "Evaluating Employee Performance Using Decision Tree Algorithms." Journal of Data Science, 5(1), pp. 89-97.

[6] Francesca Fallucchi , Romeo Giuliano, and Ernesto William De Luca (2020). "Utilizing Machine Learning Techniques to Predict Employee Attrition." Journal of Business Analytics, 12(4), pp. 267-281.

[7] M. Zhanuzakov and G.T. Balakayeva (2022). "Digitalization of Enterprise Human-Resource Management Using ML Models." Journal of Digital Transformation, 10(3), pp. 200-215.

[8] Fiyhan Alsubaiea and Murtadha Aldoukhib (2023). "Machine Learning Algorithms for Analyzing and Predicting Employee Attrition." International Journal of Data Science, 15(2), pp. 183-197.

[9] Gulnar Balakayeva, Mukhit Zhanuzakov and Gaukhar Kalmenova (2023). "Development of a Digital Employee Rating Evaluation System (DERES) Based on ML Algorithms and the 360-Degree Method." Journal of Workforce Analytics, 7(1), pp. 134-150.