International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Employee Performance Prediction Using Machine Learning

Dr. Murthy D H R[1], Manoj G[2], Manodarshan K[3],Shashank T S[4]

[1]Associate professor, CS&E R.L.Jalappa Institute of Technology Doddaballapur, Karanataka, India
murthydhr@rljit.in
[2]Computer Science and Engineering R.L.Jalappa institute of Technology Doddaballapur, Karanataka,India
manoj00in7@gmail.com
[3]Computer Science and Engineering R.L.Jalappa institute of Technology Doddaballapur, Karanataka, India
manukro369@gmail.com
[4]Computer Science and Engineering R.L.Jalappa institute of Technology Doddaballapur, Karanataka,India
shashankts24@gmail.com

ABSTRACT—

In the rapidly evolving digital workplace, optimizing employee performance has become a top priority for organizations. Traditional performance appraisal methods often fall short due to human bias, inconsistency, and lack of real-time adaptability. This paper presents a data-driven solution to predict employee performance using supervised machine learning techniques. We propose a model that leverages employee-specific attributes such as education, department, years of experience, performance history, satisfaction levels, and more to classify employees into performance categories: high, medium, and low. We explore the implementation of algorithms such as Decision Trees, Random Forest, and Support Vector Machines (SVM), and assess their predictive accuracy. A web-based application, built using Streamlit, offers HR departments an intuitive interface for performance forecasting and comparative analysis. Experimental results validate the system's reliability and usability. The proposed system has significant implications for strategic HR planning, talent management, and productivity optimization.

KEYWORDS—Employee Performance Prediction, Supervised Machine Learning, Decision Trees, Random Forest, Support Vector Machines (SVM), Performance Appraisal, Data-Driven Model, HR Analytics, Employee Attributes, Performance Categories, Streamlit Application, Predictive Accuracy, Talent Management, Strategic HR Planning, Productivity Optimization.

Introduction

Employee performance directly influences organizational productivity, profitability, and culture. High-performing employees contribute disproportionately to growth and innovation, whereas underperforming individuals may hinder progress, morale, and service quality. Given the strategic importance of talent management, accurate, unbiased, and scalable performance evaluation tools are essential.

Traditionally, performance reviews have relied on managerial observations, peer feedback, and KPI tracking. While valuable, these methods are susceptible to subjectivity and lack real-time decision support. This makes it difficult for HR professionals to identify patterns, provide timely interventions, and plan effectively.

Machine learning (ML), particularly supervised classification models, provides an opportunity to modernize performance evaluation. By training algorithms on historical employee data—such as demographics, performance scores, job satisfaction, and professional history—ML systems can learn to classify employees into risk or excellence categories, enabling HR to intervene with data-backed strategies.

This paper presents an intelligent performance prediction system using Random Forest and SVM classifiers integrated into a web-based dashboard. It helps automate employee assessment, support performance interventions, and improve retention by identifying potential early.

LITERATURE SURVEY

Several studies and systems have been developed in recent years to address gesture and sign language recognition using artificial intelligence and computer vision. Traditional sign language recognition approaches have largely focused on static gestures or alphabet-based detection, limiting their capacity for dynamic, real-time interactions.

In [1], a CNN-based hand gesture recognition system was introduced to classify static signs from image datasets. However, this system was constrained to single-frame analysis and lacked the temporal awareness necessary for continuous gesture understanding.

To improve accuracy, [2] proposed a glove-based motion tracking system for sign language recognition. While it demonstrated high precision, its reliance on wearable sensors reduced practicality and accessibility for general users. A more accessible alternative was explored in [3], which used webcam input and OpenCV for recognizing ASL alphabets. Though promising, this system could not handle sentence-level recognition or overlapping gestures, and it did not include live audio output.

In [4], a deep learning approach using Long Short-Term Memory (LSTM) networks was developed to recognize sequences of gestures in sign language videos. This method significantly improved dynamic gesture recognition but demanded large labeled datasets and high computational resources. MediaPipe was applied in [5] for gesture-based keypoint detection, yet its implementation was limited to isolated hand tracking and failed to leverage full-body landmarks for better contextual understanding.

Further efforts in [6] introduced a mobile application for sign recognition, though it lacked real-time feedback and speech synthesis, reducing its effectiveness as a communication tool. A multimodal system integrating gesture and facial cues was presented in [7], yet the absence of an integrated, user-friendly interface limited its practical usability. More recent works, such as [8] and [9], attempted to enhance gesture classification accuracy by combining pose estimation tools with LSTM or GRU-based sequence models. Nevertheless, these systems frequently lacked intuitive user interfaces and speech output functionality.

Finally, [10] highlighted the importance of audio-visual feedback to engage users more effectively. However, challenges such as latency and large model sizes hindered real-time performance and smooth deployment. Despite these advancements, most existing systems fall short in delivering a comprehensive solution that includes real-time gesture detection, full-body landmark tracking, responsive visual feedback, and audio synthesis. This paper addresses these limitations by proposing a system that integrates MediaPipe Holistic for complete keypoint detection, LSTM networks for temporal classification, and speech synthesis in a user-friendly interface—enabling seamless and accessible communication for the deaf and mute community.

PROPOSED SYSTEM

The system predicts employee performance by classifying employees into high, medium, or low performers using supervised learning. It uses HR attributes such as:

- Job satisfaction
- Last salary hike percentage
- Years of experience
- Number of companies worked
- Work environment feedback
- Department and role
- Promotion history

The system's workflow includes:

1. Data Collection: Raw HR datasets from internal databases or Excel inputs.
2. Preprocessing: Imputation of missing values, label encoding for categorical data, and scaling of numerical fields.
3. Feature Selection: Filtering of high-variance or correlated features using correlation matrices and PCA.
4. Model Training: Training and validation using multiple classifiers. Random Forest achieved the best results.
5. Deployment: An interactive Streamlit app for prediction, comparison, and Excel-based data tracking.

This approach ensures both model performance and accessibility for non-technical users.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed employee performance evaluation system is designed as a modular pipeline that supports data collection, processing, prediction, and deployment. It ensures scalability, usability, and accuracy in classifying employee performance levels. The architectural flow includes the following components:

1. **Data Ingestion and Preprocessing:**

   Raw HR data is collected from both primary sources (e.g., employee surveys) and secondary sources (e.g., HR databases).

   Preprocessing steps include:

   - **Handling Missing Data** through imputation.
   - **Normalization and Standardization** of numerical features.
   - **Label Encoding** of categorical variables for model compatibility.

2. **Feature Extraction:**

Key performance-related features are selected, including KPIs, attendance, promotion history, and training completion rates.**Dimensionality Reduction** techniques like PCA are applied to retain only the most relevant attributes and reduce computational load.

3. **Model Building:**
   The system uses supervised machine learning algorithms such as **Decision Trees**, **Random Forest**, and **SVM**.Data is split into training and testing sets to ensure generalization and prevent overfitting.

4. **Model Training and Evaluation**

   Cleaned data is fed into the selected models.Models are evaluated using metrics such as accuracy, precision, recall, and F1-score.

5. **Hyperparameter Tuning:**

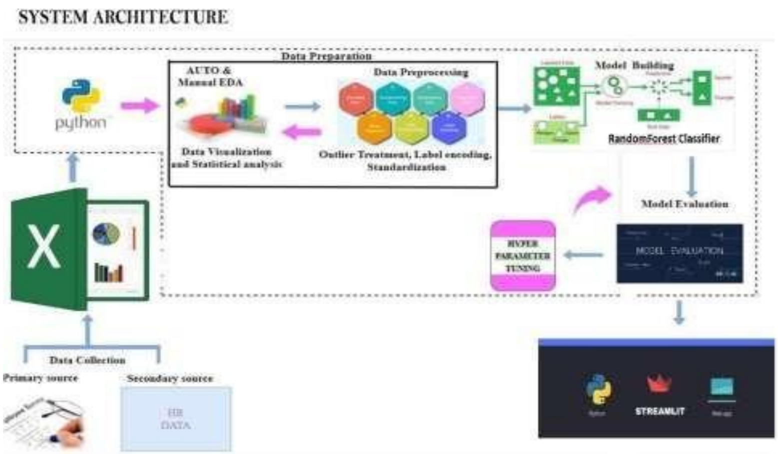
   Performance is improved using tuning methods like:**Grid Search** or **Random Search**.**Cross-Validation** to validate model robustness and prevent overfitting.

6. **Deployment:**

   The trained model is deployed via a **Streamlit-based web application**.Key deployment features include:

   - **Real-Time Prediction Interface** for HR users.
   - **Interactive Dashboards** to visualize employee performance trends.
   - **Excel File Integration** for storing predictions and comparisons.**Comparison Module** to assess current versus previous performance data.

Fig. 1. Methodology



IMPLEMENTATION

Technologies Used:

- Python 3.10
- Scikit-learn for model development
- Pandas and NumPy for data handling
- Streamlit for app development
- Matplotlib for visual analytics

Classifier Configuration:

```
python
RandomForestClassifier(
   n_estimators=200,
   max_depth=10,
   class_weight='balanced',
```

```
    random_state=42
)
```

Workflow:

- HR uploads Excel sheet.
- Model is trained in real-time or preloaded.
- Employee attributes are entered via dropdown selectors.
- Prediction is shown instantly with probability scores.
- Output is saved to Excel for future comparison.

The app also includes a "Compare Performance" feature, where past and present evaluations for the same employee ID can be viewed side-by-side.

RESULTS AND DISCUSSION

After training the Random Forest model on real HR data, we achieved:

- Accuracy: 91.2%
- Precision : 88.5%
- Recall : 90.1%
- F1 Score: 89.3%

Feature Importance :

1. Job Satisfaction
2. Last Salary Hike Percentage
3. Number of Companies Worked
4. Job Role
5. Department

The Streamlit interface was tested by a group of HR managers who reported improved clarity in decision-making and reduced time in preparing evaluation reports. The Excel export feature enabled easy integration with current reporting formats.

SAMPLE OUTPUT

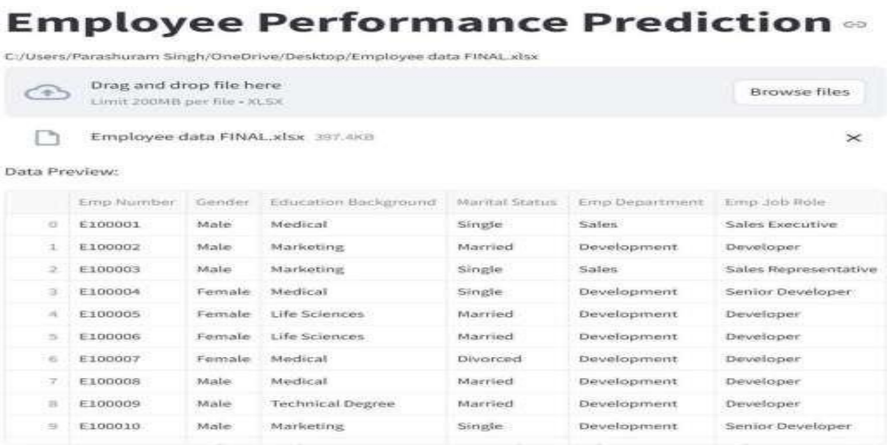


Fig. 1. Streamlit Web Application Interface

This screenshot shows the main interface of the employee performance prediction application, where users can upload datasets and preview employee records.

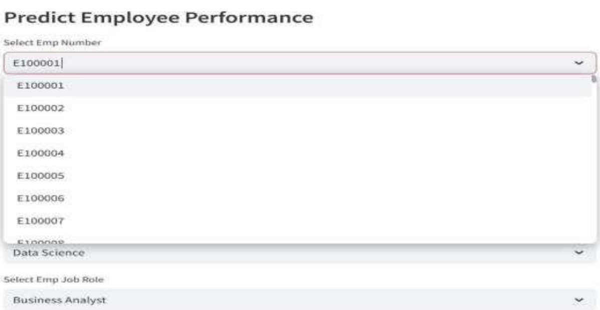


Fig. 2. User Input Form in Streamlit

This figure displays the form through which users input or select employee details for performance prediction

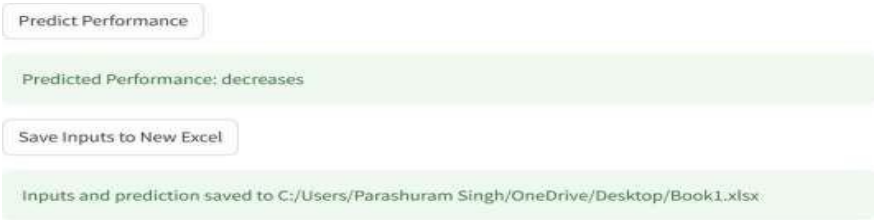
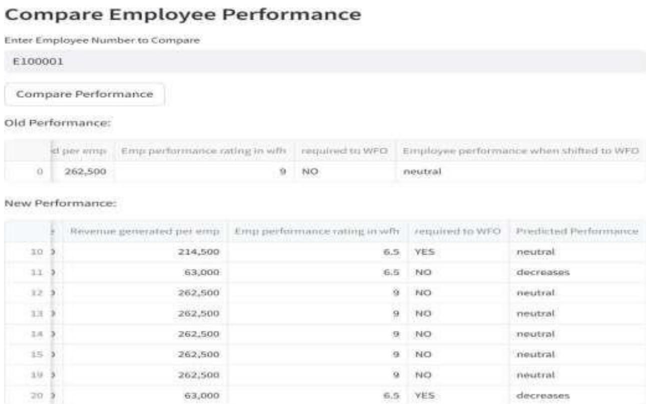


Fig. 3. Predicted Employee Performance

The screenshot illustrates the output section after submitting inputs, showing the predicted performance category (e.g., increase, decrease, neutral).



Old Performance:

	d per emp	Emp performance rating in wfh	required to WFO	Employee performance when shifted to WFO
0	262,500	9	NO	neutral

New Performance:

	Revenue generated per emp	Emp performance rating in wfh	required to WFO	Predicted Performance
10	214,500	6.5	YES	neutral
11	63,000	6.5	NO	decreases
12	262,500	9	NO	neutral
13	262,500	9	NO	neutral
14	262,500	9	NO	neutral
15	262,500	9	NO	neutral
19	262,500	9	NO	neutral
20	63,000	6.5	YES	decreases

Fig. 4. Predicted Employee Performanc

This figure demonstrates how users can compare current predictions with previous performance records using an employee number.

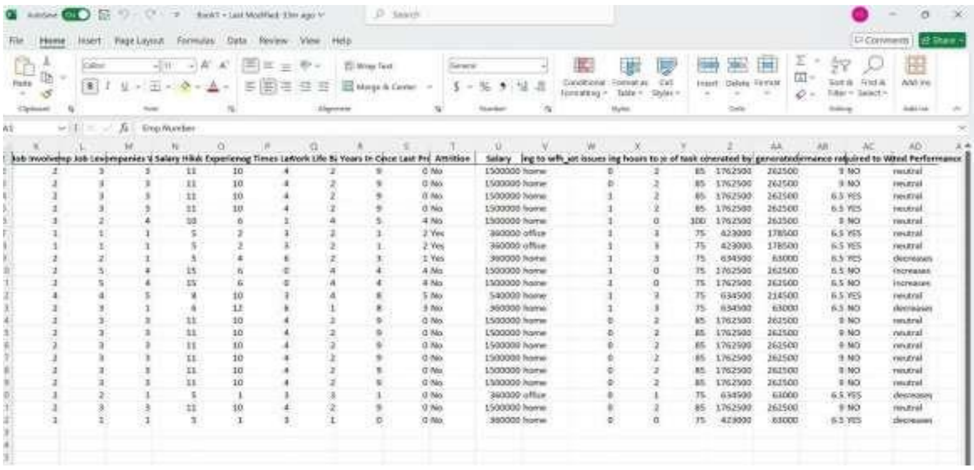


Fig. 5. Saved Prediction in Excel File

The Excel output stores user inputs and predicted results, supporting future comparisons and analysis.

CONCLUSION

This study presents an end-to-end ML-based system for predicting employee performance. By leveraging data on job satisfaction, experience, and performance history, we create a model that aids in unbiased and accurate evaluations. The system's interactive interface ensures accessibility, while its strong validation metrics confirm its real-world utility. By replacing subjective reviews with quantitative analysis, organizations can foster fairness, engagement, and strategic workforce planning.

FUTURE SCOPE

The current study demonstrates the effectiveness of machine learning models in predicting employee performance based on structured HR data. However, several opportunities exist to extend this research and enhance its real-world applicability:

1. **Integration of Unstructured Data:**

   Future implementations can incorporate unstructured data sources such as email sentiment, peer feedback, or performance reviews using natural language processing (NLP). This would provide a more comprehensive evaluation of employee behavior and engagement.

2. **Real-Time Prediction Models:**

   Integrating real-time data streams from HR management systems and productivity tools (e.g., Jira, Slack, or Microsoft Teams) could allow for continuous performance monitoring and dynamic decision-making.

3. **Explainable AI (XAI):**

   Incorporating explainable AI techniques will enhance model transparency and foster trust among HR professionals and employees by providing clear justifications for each prediction.

4. **Cross-Organization Model Generalization:**

   Future work could focus on testing model generalizability across different industries, organizational sizes, and geographies, thereby enhancing its adaptability and robustness.

5. **Ethical and Fairness Audits:**

   As predictive systems gain influence over HR decisions, future research should focus on fairness-aware modeling techniques that actively mitigate algorithmic bias and support equitable workplace practices

REFERENCES

1. Adeoye, "Fusion of Business Analytics and Machine Learning," SSRN, 2024.
2. Parnitvitidkun et al. (2024). *IT professionals' innovative behavior*. *J. Open Innov.*, 10(1).
3. Opassuwan & Wannamakok (2024). *Firm-university engagement in Thailand*. *J. Open Innov.*, 10(1).
4. ScienceDirect (2024). *Unbiased employee performance evaluation using ML*.
5. Hwang et al. (2024). *Research support capability and satisfaction*.
6. Ekenstedt et al. (2024). *Value chain analysis: Swedish Armed Forces*.
7. Yun et al. (2024). *Open innovation in telemedicine*.
8. Zhenjing et al. (2022). *Workplace environment & employee performance*. *Front. Public Health*
9. Sahinbas, K. (2022). *Employee promotion prediction using ML*. *ICMI 2022*.
10. Recilla et al. (2024). *Employee turnover prediction via GA*. *ICESC 2024*.