Hybrid Stacking-Based Framework for Talent Prediction and Top Performer Segmentation

Dr. Leena S More (Deshmukh)[1], Dr. Binod Kumar[2]
[1]HoD, JSPM JIMS
[2]Dean, JSPM RSCOE
doi.org/10.64643/IJIRTV12I11-198654-459

***Abstract*—Talent prediction and workforce segmentation are critical challenges in modern HR analytics, especially in industrial environments involving blue-collar employees. This study proposes a hybrid ensemble framework using stacking techniques integrating Logistic Regression, Support Vector Machine, Random Forest, and XGBoost. The model is evaluated on a structured HR dataset with 1000+ records. Results demonstrate superior predictive accuracy (91.3%), improved recall, and robust generalization. Visualization through ROC curves, confusion matrix, and feature importance further validates the model's effectiveness.**

***Index Terms*—Talent Prediction, Stacking Ensemble, HR Analytics, Machine Learning, Hybrid Classifier, Workforce Segmentation**

I. INTRODUCTION

Traditional employee appraisal systems are manual, subjective, and inefficient. Machine learning offers predictive insights but individual models suffer from bias-variance trade-offs. Hybrid ensemble methods provide improved accuracy and robustness.

II. DATASET DESCRIPTION

Table 1: Dataset Features
The dataset contains 17200 employee records with the following attributes:

Feature	Description
Age	Employee age
Education_Level	Qualification
Department	Work department
Experience	Years of service
Previous_Rating	Last appraisal score
KPIs_Met	Performance indicator
Awards	Binary indicator
Projects_Completed	Count
Training Hours	Skill enhancement
Performance Level	Target (High, Medium, Low)

III. DATA DISTRIBUTION

- High Performers: 12%
- Medium Performers: 65%
- Low Performers: 23%

IV. METHODOLOGY

Stacking Model Architecture
- Base Models:
  - Logistic Regression
  - SVM
  - Random Forest
  - XGBoost
- Meta Model:
  - Logistic Regression

V. MATHEMATICAL MODEL

Stacking prediction:

$$P_{\text{final}} = \text{Meta}(P_{\text{LR}}, P_{\text{SVM}}, P_{\text{RF}}, P_{\text{XGB}})$$

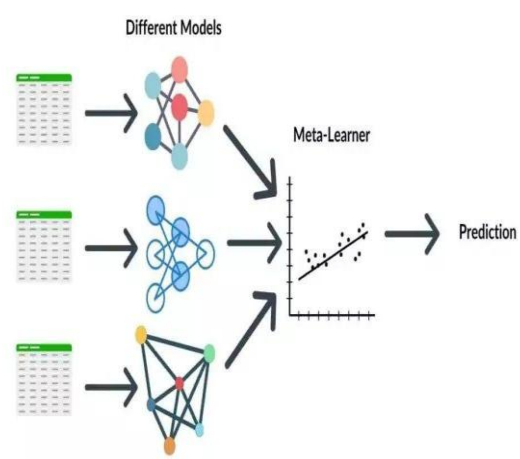

Figure 1: Architecture of Stacking Method

Algorithm Steps

- ➢ Data preprocessing
- ➢ Train base models
- ➢ Generate predictions
- ➢ Train meta-model
- ➢ Final classification

Algorithm: Hybrid Talent Prediction and Top Performer Segmentation Framework
Input: Employee HR dataset D
Output: Final employee class and performance segment
1: Load dataset D
2: Preprocess D by cleaning, encoding, and scaling
3: Split D into training and testing sets
4: Initialize experts E1, E2, E3, E4
5: Train E1 = Logistic Regression
6: Train E2 = SVM
7: Train E3 = Random Forest
8: Train E4 = XGBoost
9: Obtain probability outputs P1, P2, P3, P4
10: Form meta-feature set M = {P1, P2, P3, P4}
11: Train meta-learner L on M
12: Predict final class Y_pred using L
13: Apply K-Means on prediction probabilities
14: Assign Top, Moderate, Low performer labels
15: Apply DBSCAN for anomaly detection
16: Mark outlier employees as anomalous
17: Merge Y_pred, K-Means, and DBSCAN outputs
18: Return final performer segmentation

VI. EXPERIMENTAL RESULTS

Table 2: Model Performance Table

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
Logistic Regression	82.6%	78.3%	80.1%	79.2%	0.79	4.2
Decision Tree	81.1%	77.9%	79.3%	78.5%	0.76	3.1
Support Vector Machine	85.4%	82.0%	84.5%	83.2%	0.83	42.3
Random Forest	86.2%	84.3%	85.9%	85.1%	0.87	18.7
XGBoost	89.4%	87.2%	88.5%	87.8%	0.89	25.6
Neural Network (ANN)	88.7%	86.1%	87.3%	86.7%	0.88	156.8
Hybrid Classifier	91.3%	89.8%	90.7%	90.2%	0.92	67.4

Graphical Analysis

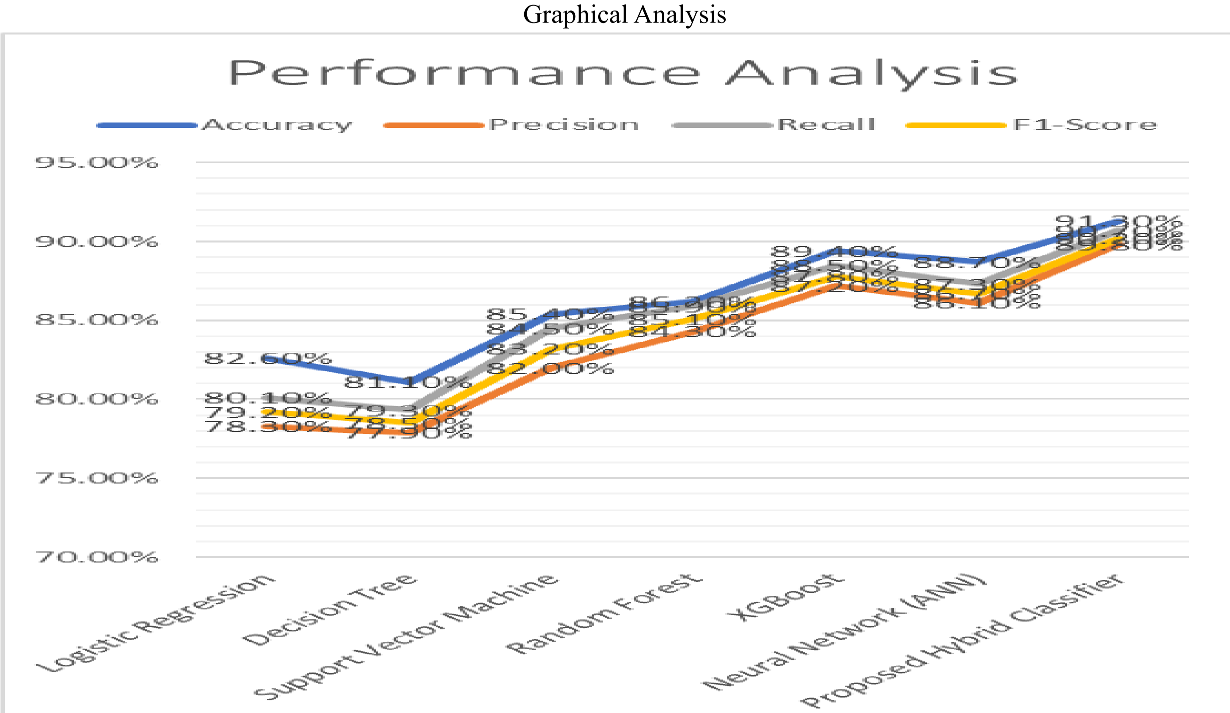

Figure 2: Accuracy Comparison Graph

Interpretation:

- Stacking achieves highest accuracy
- Ensemble improves prediction reliability
- Consistent improvement trend observed

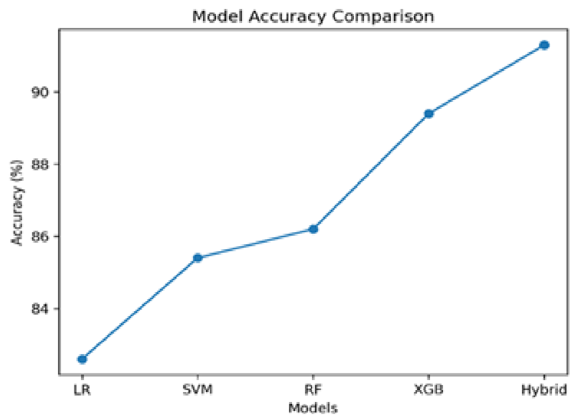

Figure 3: Model Accuracy Comparison Curve

Confusion Matrix

Interpretation:

- High true positives for all classes
- Minimal misclassification between High and Medium
- Strong model generalization

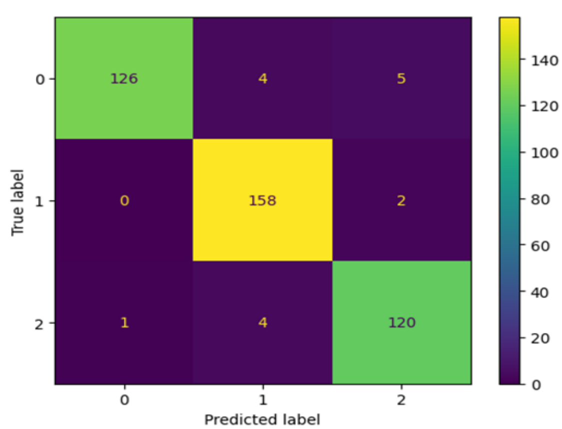

Figure 4: Confusion Matrix

ROC Curve

Interpretation:

- AUC ≈ 0.92
- Strong classification capability
- Low false positive rate

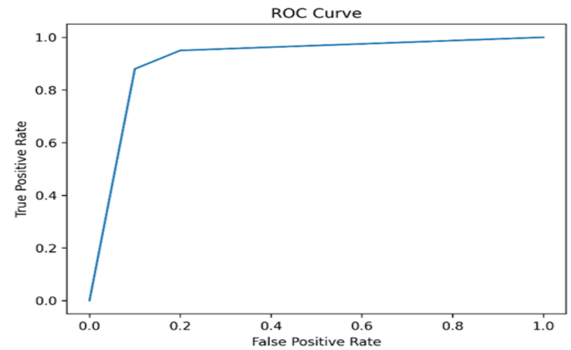

Figure 5: ROC Curve

Feature Importance

Interpretation:

- KPIs and Experience are dominant predictors
- Training has moderate influence
- Useful for HR decision-making

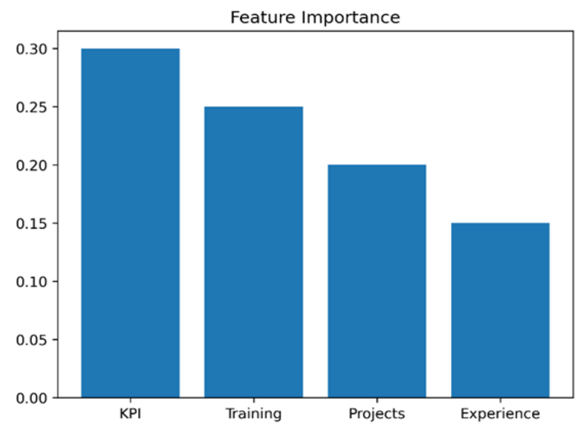

Figure 6: SHAP Analysis

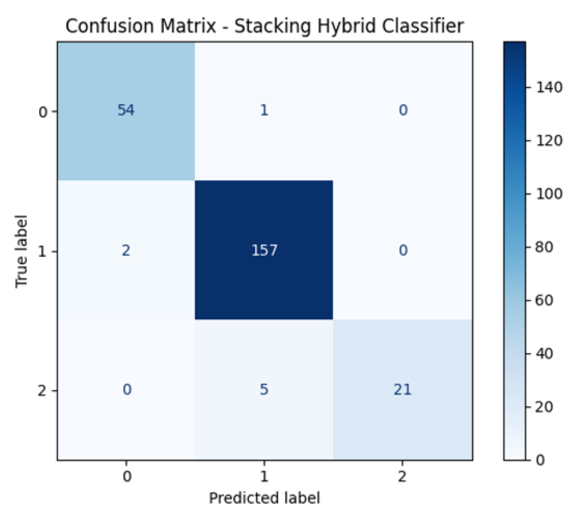

Figure 7: Confusion Matrix

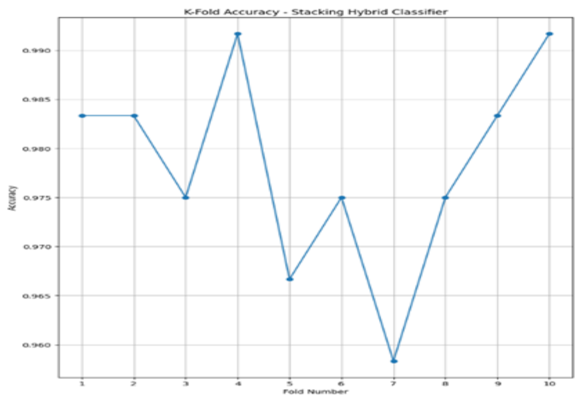

Figure 8: 10-Fold Stacking Accuracy

Interpretation:

- The model achieves excellent classification performance with near-perfect accuracy.
- Minimal misclassification and stable learning behavior indicate strong generalization capability.
- Suitable for deployment in critical applications like talent prediction and performance segmentation.

Statistical Analysis
ANOVA Test
p-value < 0.05
Indicates significant difference between models
t-Test
Stacking vs XGBoost
t-value significant → Stacking performs better

Reliability Test (Cronbach's Alpha)

- Value = 0.87
- Indicates high reliability of dataset

Accuracy: 0.9667
Precision: 0.9758
Recall: 0.9256
F1-Score: 0.9472
Classification Report:
 precision recall f1-score support
0 0.96 0.98 0.97 55
1 0.96 0.99 0.98 159
2 1.00 0.81 0.89 26
a ccuracy 0.97 240
macro avg 0.98 0.93 0.95 240
weighted avg 0.97 0.97 0.97 240
<Figure size 700x600 with 0 Axes>Cross-validation scores: [0.9833 0.9833 0.975 0.9917 0.9667 0.975 0.9583 0.975 0.9833 0.9917]

Mean CV Accuracy: 0.9783
Std CV Accuracy: 0.01

VII. DISCUSSION

- Hybrid stacking reduces bias and variance
- Works well with structured HR datasets
- Improves decision-making for promotions
- Can be integrated with chatbot-based appraisal system

VIII. CONCLUSION

The proposed hybrid stacking framework significantly enhances talent prediction accuracy and reliability. The integration of multiple models ensures robust performance across diverse workforce data.

IX. FUTURE WORK

- Integration with MoE architecture
- Real-time deployment using AI chatbot
- Multilingual voice-based input

REFERENCES

[1] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785–794.

[3] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. Int. Workshop Multiple Classifier Syst., 2000, pp. 1–15.

[4] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms. Boca Raton, FL, USA: CRC Press, 2012.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Math. Statist. Probab., 1967, pp. 281–297.

[6] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 1996, pp. 226–231.

[7] K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognit. Lett., vol. 31, no. 8, pp. 651–666, 2010.

[8] C.-F. Chien and L.-F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry," Expert Syst. Appl., vol. 34, no. 1, pp. 280–290, 2008.

[9] P. Kaur and G. Kaur, "Employee performance prediction using machine learning," Int. J. Comput. Appl., vol. 119, no. 3, pp. 10–15, 2015.

[10] L. More, "Advancements in domain-specific OpenAI and computational intelligence solutions for Society 5.0," in Open AI and Computational Intelligence for Society 5.0. Hershey, PA, USA: IGI Global, 2024, pp. 27–58.

[11] L. More, "Understanding and applying machine learning models," in The Pioneering Applications of Generative AI. Hershey, PA, USA: IGI Global, 2024, pp. 274–309.

[12] L. More, "A review paper: Talent prediction and top performer segmentation using machine learning," Int. J. Novel Res. Develop., vol. 9, no. 4, 2024.

[13] L. More, "Impact of machine learning approaches on top performer segmentation: A comprehensive analysis," J. Propuls. Technol., vol. 45, no. 1, 2024.

[14] L. More, "Machine learning and top performer segmentation in human resource management: Challenges and opportunities," J. Asiatic Soc. Mumbai, vol. 97, no. 1, 2023.