IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Job Search and Recruitment Portal Full Stack Python with Flask

K. Aruna Kumari[1], B Laskshmi Manasa[2], V Usha Rani[3], L Nagaraju[4], SK. Jani[5]

[1]HOD & Assistant Professor, Department of Data Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

[2,3,4,5] Students, Department of Data Science Engineering, Chalapathi Institute of Engineering and Technology, Chalapathi Rd, Nagar, Lam, Guntur, Andhra Pradesh- 522034

Email id: aruna.jeshwin@gmail.com[1], manasabogasamudram@gmail.com[2], ushasrinivas888@gmail.com[3], lattupallinagaraju09@gmail.com[4], janishaik527@gmail.com[5]

Abstract:

Resume matching is the process of comparing a candidate's curriculum vitae (CV) or resume with a job description or a set of employment requirements. The objective of this procedure is to assess the degree to which a candidate's skills, qualifications, experience, and other relevant attributes align with the demands of the position. This study compares the effectiveness of various machine learning models to improve recruitment accuracy and efficiency. Using the recruitment data from a major Yemeni organization (2019–2022), we evaluated models including K-Nearest Neighbors, Logistic Regression, Support Vector Machine, Naive Bayes, Decision Trees, Random Forest, Gradient Boosting Classifier, AdaBoost Classifier, and Neural Networks. Hyperparameter tuning and cross-validation were used for optimization. The Random Forest model achieved the highest accuracy (92.8%), followed by Neural Networks (92.6%) and Gradient Boosting Classifier (92.5%). These results suggest that advanced machine learning models, particularly Random Forest and Neural Networks, can significantly enhance the recruitment processes in business intelligence systems. This study provides valuable insights for recruiters, advocating for the integration of sophisticated machine learning techniques in talent acquisition strategies.

Keywords: machine learning, talent recruitment business intelligence systems; random forest; predictive analytics; human resource management

1.Introduction

In today's fast-paced, data-driven business environment, the strategic adoption of technological advancements within business intelligence significantly influences organizational success. The Fourth Industrial Revolution, often referred to as the "Digital Era 4.0", has led to a significant shift towards data-driven decision-making across various industries [1,2]. As companies harness the power of analytics to gain a competitive edge, recruiting individuals with strong analytical skills has become imperative, especially in sectors heavily reliant on business intelligence systems [3]. Adapting and efficiently recruiting top talent is crucial for leveraging analytical capabilities. The traditional recruitment methods, often cumbersome and subjective, are increasingly inadequate given the rapidly expanding data volumes and the complexities of global talent markets [4]. Recent advancements in big data analytics and predictive modeling present novel opportunities to enhance the recruitment processes [5]. Predictive analytics, which uses historical data to forecast future outcomes, has become a focal point for organizations aiming to streamline recruitment and identify optimal candidates efficiently. However, integrating predictive analytics into recruitment processes poses challenges, including data privacy concerns, the complexity of model implementation, and the potential for bias in algorithmic decision-making [6]. This study aims to address these challenges by rigorously evaluating various machine learning algorithms, including K-Nearest Neighbors (KNNs), Logistic Regression, Support Vector Machine (SVM), Naive Bayes, Decision Trees, Random Forest, Gradient Boost Classifier (GBC), AdaBoost, and Neural Networks. The research investigates their performance in enhancing the recruitment efficiency within business intelligence systems. By examining and comparing these models on key performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score, this study tests the hypothesis that machine learning models, when properly tuned and applied, can significantly outperform the traditional methods in predicting the suitability of candidates for specific roles. The main objectives of this research are to identify the most effective predictive model for talent recruitment in BI systems and to provide actionable insights for strategically integrating these technologies into business intelligence frameworks. These insights aim to help organizations optimize their recruitment processes and ensure the acquisition of top analytical talent, thereby enhancing their ability to innovate and achieve strategic objectives. Through rigorous analysis and evaluation, this research concludes that certain models, notably Random Forest and Neural Networks, provide superior accuracy and efficiency in the candidate selection processes. These findings confirm the utility of machine learning in recruitment and highlight the practical steps that organizations can take to implement such technologies effectively.

2. Literature Review

In reviewing the relevant literature, it is evident that predictive analytics and machine learning have emerged as powerful tools in the field of human resource management. These tools have the potential to revolutionize the talent recruitment and acquisition processes by leveraging big data and advanced algorithms. Through the analysis of large-scale talent and management-related data, organizations can gain valuable insights into organizational behaviors and make more informed decisions regarding talent management [7]. Furthermore, machine learning models can enhance the accuracy and efficiency of the recruitment processes by identifying patterns and predicting the success of potential candidates [8]. The use of machine learning and predictive analytics in the field of human resource management has gained significant attention in recent years [9].

2.1 Big Data in Talent Recruitment in the BI

BI systems have become instrumental in managing the complexities of human capital within organizations. The ability to analyze vast datasets for patterns and insights is crucial for identifying the top talent and predicting their job performance [10]. However, the traditional BI systems often fall short in terms of predictive accuracy and real-time analysis, necessitating the integration of advanced analytics and machine learning models to enhance the recruitment processes [11]. Big data analytics has emerged as a powerful tool for HR management, offering the potential to revolutionize talent recruitment. The use of big data can lead to a more nuanced understanding of candidate profiles, thereby facilitating more informed and objective hiring decisions. Studies have shown that big data analytics can improve the efficiency of the recruitment processes and reduce the costs associated with hiring the wrong candidate [12]. However, the implementation of big data in HR also presents challenges, such as ensuring data privacy and managing the complexity of algorithms [13].

2.2 Predictive Modeling in Talent Recruitment

Predictive modeling, which involves the use of statistical algorithms and machine learning techniques to forecast future events based on historical data, has become increasingly popular in talent recruitment. These models can identify the characteristics of the top-performing candidates, thus streamlining the candidate selection process [14]. Various studies have demonstrated the effectiveness of machine learning algorithms like KNNs, SVM, and Random Forest in predicting employee performance and turnover [15,16,17]. However, the choice of the best predictive model often depends on the specific context and the nature of the data available. Ref. [15] developed a model that accurately predicts employee turnover with an 87.43% success rate, highlighting the practical utility of machine learning in reducing turnover and enhancing retention strategies. Ref. [18] explored the impact of big data analytics on business intelligence in China, emphasizing the significance of data quality and the challenges associated with data management and privacy concerns. Refs. [19,20] discussed the broad

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

applications of big data analytics across various sectors, emphasizing its crucial role in strategic decision-making and operational improvements.

3.Methodology

This research adopts a meticulous and systematic approach to optimize talent recruitment strategies within business intelligence systems through predictive analytics. Utilizing advanced machine learning algorithms, the study undertakes thorough data collection, preprocessing, model development, and evaluation stages, as shown in Figure 1, to construct effective predictive models for talent acquisition. These methodological steps are intricately interlinked, fostering a coherent research flow that not only enhances the predictive model robustness but also upholds transparency and reproducibility in methodologies.

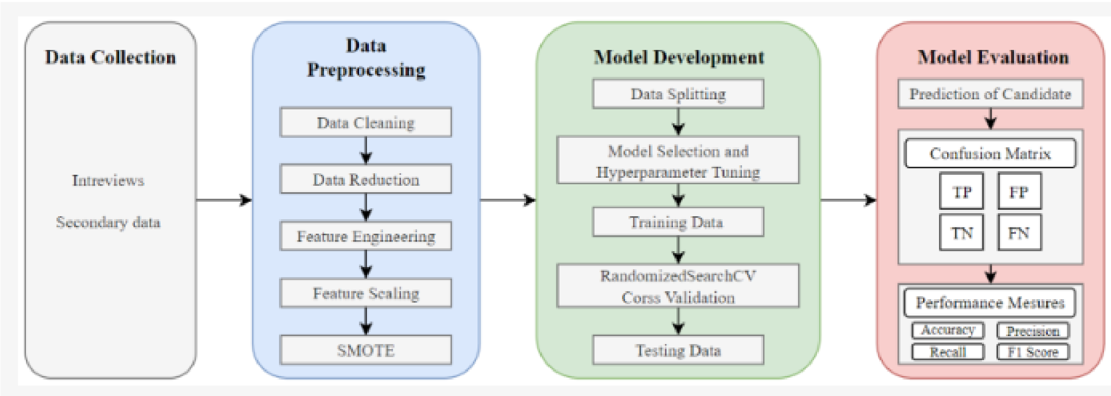

Figure: This figure illustrates the methodological framework adopted for the study

The approach detailed in this section encompasses the entire research process from data collection and preprocessing to advanced predictive model deployment and evaluation. By transparently describing each methodological facet, including the techniques and tools employed for extensive recruitment data analysis, this study ensures reproducibility and reliability in its findings. Through meticulous documentation of methodologies, this research contributes to the scientific rigor and verifiability of the study, providing a foundational framework for future studies exploring similar phenomena.

Data Collection

Data collection is a critical phase in any empirical study as it forms the basis for subsequent analysis and model development. For this research, the data were sourced from a leading recruitment organization in Yemen spanning the period from 2019 to 2022. The initial dataset, comprising 22,487 records, served as the primary material for our analysis and was extracted from the organization's database and comprised a comprehensive collection of recruitment records.

The dataset includes both categorical and numerical features, such as educational background, work experience, employment test results, and candidate success metrics, as shown in table. The inclusion of these variables enables a robust analysis of the recruitment patterns and candidate performance, which are crucial for developing predictive models. The dataset was reduced to 20,245 candidates after data cleaning to ensure high-quality data for analysis.

Data Preprocessing

The preprocessing phase was integral to preparing the dataset for machine learning analysis. It involved several steps to ensure the data were clean, consistent, and suitable for analysis:

Data Cleaning: The data cleaning process commenced with a meticulous review of the dataset to identify and rectify inaccuracies. The initial dataset, encompassing over 20,000 records, was subjected to a thorough examination to detect inconsistencies and address missing values. This rigorous cleaning process culminated in a refined dataset comprising 20,239 records following the elimination of duplicates and rectification of inaccuracies.

Dimensionality Reduction: This study employed Principal Component Analysis (PCA) and Exploratory Data Analysis (EDA) for dimensionality reduction. PCA is a method used to reduce the number of variables in a dataset while retaining its variability, thereby simplifying the model and enhancing its performance. EDA provides insights into the dataset, aiding in the discovery of valuable patterns. Univariate analysis was employed to analyse categorical and numerical data, while bivariate analysis helped to identify correlations between variables.

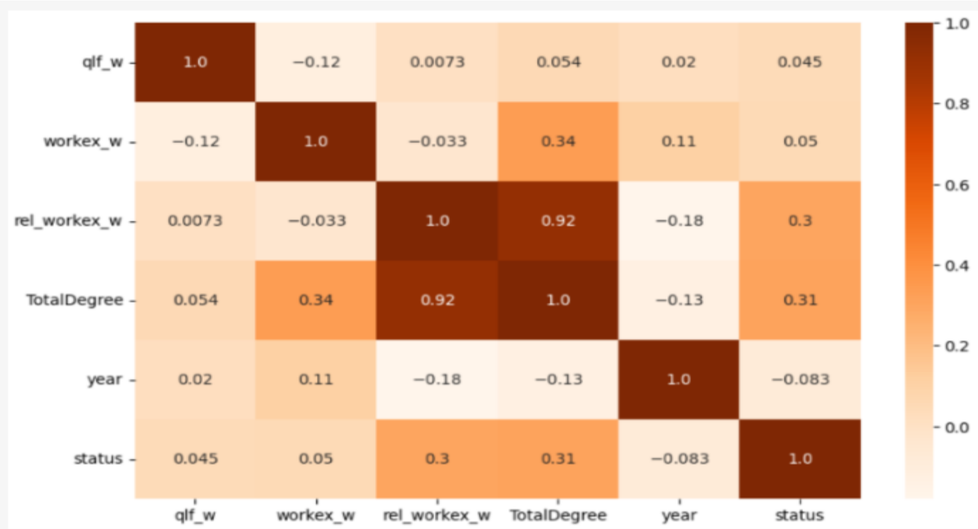

Figure: Correlations between selected features

Feature Selection and Engineering

We conducted a thorough feature selection process to identify the most relevant features that could potentially impact recruitment outcomes. This process involved statistical analysis and domain expertise to ensure that the selected features were both predictive and meaningful in the context of talent recruitment.

Feature engineering efforts included encoding categorical variables using techniques like one-hot encoding and label encoding, ensuring that these variables were represented in a format compatible with machine learning algorithms. The feature engineering process involved selecting relevant features that contribute significantly to the prediction models. Techniques such as one-hot encoding for categorical variables, normalization for numerical variables, and handling missing values through imputation were employed. Feature selection was guided by domain knowledge and statistical methods to retain the most informative features. We also created new features from existing ones to better capture the underlying patterns in the data.

- Encoding Categorical Variables: One-hot encoding was used to convert categorical variables into binary vectors. This technique was particularly useful for handling non-ordinal categorical data. For ordinal data, label encoding was employed to maintain the inherent order in the data.
- Creating New Features: New features were derived from existing ones to enhance the model's predictive power. For instance, a composite feature combining years of experience and qualification levels was created to better represent a candidate's suitability for advanced roles.
- Handling Outliers: Outliers were identified using the Interquartile Range (IQR) method. These outliers were then treated by replacing them with the mean or median values of the corresponding feature to minimize their impact on the model.

4. Results

The analysis of various machine learning models has demonstrated their potential to revolutionize talent recruitment within business intelligence systems. This section presents a detailed analysis of the findings from the application of various machine learning algorithms to enhance the talent recruitment in business intelligence systems. The performance of each model was evaluated based on several key metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). The results are supported by the statistical analysis and visualizations derived from the processed data, enabling a clear comparison of the effectiveness of each predictive model.

Overview of Model Performance

The performance of each predictive model was rigorously evaluated using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and AUC-ROC score. These metrics provide a multi-faceted view of the models' predictive capabilities, capturing their overall correctness, ability to identify positive instances, balance between precision and recall, and discrimination threshold levels, respectively.

Random Forest Model

The Random Forest model emerged as the top performer, achieving an accuracy score of 92.8%. This model demonstrated a high precision of 91.2% and recall of 90.9%, indicating its effectiveness in correctly identifying positive instances without compromising on the breadth of detection. The F1-score of 91.0% further underscores the model's balanced performance in precision and recall. The ROC-AUC score of 0.962 signifies the model's excellent ability to distinguish between successful and unsuccessful recruitment outcomes. To mitigate the risk of overfitting, several strategies were employed. The Random Forest model utilized techniques such as limiting the maximum depth of the trees, setting a minimum number of samples required to split a node, and using cross-validation during the model training. These steps helped to ensure that the model generalized well to unseen data.

Neural Networks Model

The Neural Networks model also exhibited a remarkable performance, with an accuracy score of 92.6%. This model stood out with a precision score of 96%, suggesting its high reliability in identifying suitable candidates. The recall and F1-score of 88% and 92%, respectively, indicate that, while the model is slightly more conservative in its predictions, it maintains a strong balance between precision and recall. The ROC-AUC score of 0.962 is on par with the Random Forest model, reflecting its exceptional discriminative power. For the Neural Networks model, overfitting was mitigated by incorporating techniques such as dropout regularization, which randomly drops units from the Neural Network during training to prevent the co-adaptation of neurons. Additionally, early stopping was used to halt the training once the model's performance on the validation set began to deteriorate, ensuring that the model did not learn noise from the training data.

Other Models

The GBC, Decision Trees, and KNNs models also performed commendably, with accuracy scores of 92.5%, 92.4%, and 92.2%, respectively. These models demonstrated strong predictive capabilities, although they were slightly behind the top performers in terms of precision, recall, and F1-score. The SVM and AdaBoost Classifier models showed solid performance, with accuracy scores of 90.8% and 89.1%, respectively. However, they demonstrated lower precision and recall compared to the top models, suggesting that they may be less effective in identifying the best candidates without a higher risk of false negatives or positives.

The Naive Bayes and Logistic Regression models had the lowest accuracy scores at 86.2% and 78.3%, respectively. These models also had lower precision, recall, and F1-scores, indicating that they may not be as suitable for talent recruitment prediction in business intelligence systems.

Table: Summary of model performance metrics

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
KNNs	92.2%	57%	65%	61%	95.1%
Logistic Regression	78.3%	27%	78%	40%	85.2%
SVM	90.8%	50%	58%	54%	93.3%
Naive Bayes	86.2%	25%	25%	25%	88.6%
Decision Trees	92.4%	58%	63%	60%	95.3%
Random Forest	92.8%	91.2%	90.9%	91.0%	96.2%
GBC	92.5%	60%	56%	58%	95.5%
AdaBoost Classifier	89.1%	42%	48%	45%	91.9%
Neural Networks	92.6%	96%	88%	92%	96.2%

To provide a visual representation of the models' performance, Figure presents a comparative analysis of the accuracy, precision, recall, and F1-score for each model. This visual aid enables a quick and clear comparison of the models' predictive capabilities.

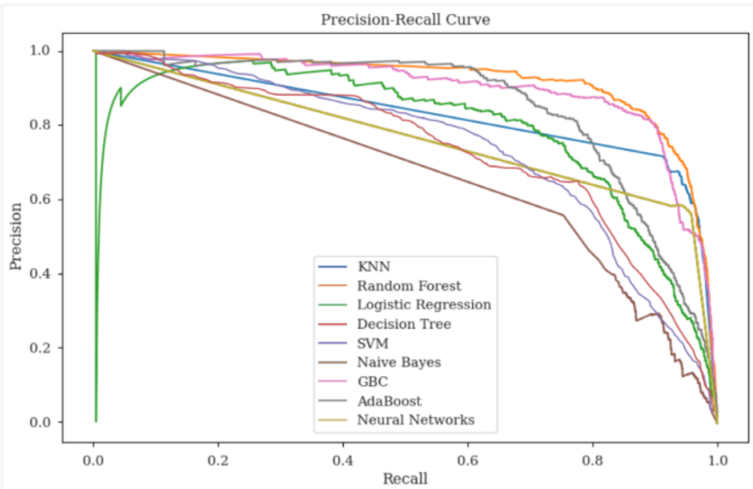

Figure: Precision–recall curves for various models, demonstrating their ability to balance precision and recall effectively

Conclusions

This research systematically explored the efficacy of machine learning models in enhancing the talent recruitment processes within business intelligence systems. The primary objective was to enhance the recruitment accuracy and efficiency through the application of machine learning models.

Our findings indicate that the Random Forest and Neural Networks models outperformed the other algorithms, achieving accuracy scores of 92.8% and 92.6%, respectively. These models demonstrated exceptional precision, recall, and F1-scores, suggesting their effectiveness in identifying the top candidates and minimizing poor hiring decisions. Other models, such as GBC, Decision Trees, and KNNs, showed promising results but were slightly less effective. Conversely, Naive Bayes and Logistic

ISSN 2347–3657

Volume 13, Issue 2, 2025

Regression exhibited lower accuracy, indicating potential limitations in talent recruitment applications. The practical implications of our findings are significant for organizations seeking to optimize their recruitment processes. By leveraging the Random Forest and Neural Networks models, organizations can enhance their recruitment efficiency and effectiveness, leading to better hiring decisions and reduced costs. These models can identify the key variables contributing to successful recruitment outcomes, providing actionable insights for improving the candidate selection and retention rates.

References:

1. Xu, M.; David, J.M.; Kim, S.H. The Fourth Industrial Revolution: Opportunities and Challenges. *Int. J. Financ. Res.* **2018**, *9*, 90.

2. Sarker, I.H.; Colman, A.; Han, J.; Khan, A.I.; Abushark, Y.B.; Salah, K. BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model. *Mob. Networks Appl.* **2019**, *25*, 1151–1161.

3. Nasir, M.; Dag, A.; Young II, W.A.; Delen, D. Determining Optimal Skillsets for Business Managers Based on Local and Global Job Markets: A Text Analytics Approach. *Decis. Sci. J. Innov. Educ.* **2020**, *18*, 374–408.

4. Ozdemir, F.; Coskun, M.; Gezer, C.; Gungor, V.C. Assessing Employee Attrition Using Classifications Algorithms. In Proceedings of the 2020 the 4th International Conference on Information System and Data Mining (ICISDM '20), Hawaii, HI, USA, 15–17 May 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 118–122.

5. Zhang, Y.; Xu, S.; Zhang, L.; Yang, M. Big data and human resource management research: An integrative review and new directions for future research. *J. Bus. Res.* **2021**, *133*, 34–50.

6. Pessach, D.; Singer, G.; Avrahami, D.; Chalutz Ben-Gal, H.; Shmueli, E.; Ben-Gal, I. Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming. *Decis. Support Syst.* **2020**, *134*, 113290.

7. Qin, C.; Zhang, L.; Cheng, Y.; Zha, R.; Shen, D.; Zhang, Q.; Chen, X.; Sun, Y.; Zhu, C.; Zhu, H.; et al. A Comprehensive Survey of Artificial Intelligence Techniques for Talent Analytics. *arXiv* **2024**, arXiv:2307.03195.

8. Koenig, N.; Tonidandel, S.; Thompson, I.; Albritton, B.; Koohifar, F.; Yankov, G.; Speer, A.; Hardy, J.H., III; Gibson, C.; Frost, C.; et al. Improving measurement and prediction in personnel selection through the application of machine learning. *Pers. Psychol.* **2023**, *76*, 1061–1123.

9. Manthena, S.R.L. Impact of Artificial Intelligence on Recruitment and its Benefits. *IJIRMPS—Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.* **2021**, *9*, 58–63.

10. Gomez-Mejia, L.R.; Balkin, D.B.; Cardy, R.L. *Managing Human Resources*, 7th ed.; OCLC: 706965973; Pearson: London, UK, 2012.