Enhancing Job Recruitment Prediction through Supervised Learning and Structured Intelligent System: A Data Analytics Approach

Abstract

Personnel recruitment processesin various government agencies, ministries, boards, and parastatals encounter challenges in effectively selecting candidates who meet specified requirements for job placement on time. Moreover, human resource (HR) managers face the additional burden of appeasing top government officials while also mitigating issues of nepotism and bias during recruitment. The success or failure of any organization heavily relies on the recruitment and retention of its workforce. Consequently, the decision to select suitable candidates for job positions is of utmost importance to management in every organization. This work develops a structured intelligent system that selects the best machine learning (ML)classification model for predicting applicants' employability based on their attributes using the industry job selection criteria. A dataset of 16240 applicants' records collected from Akwa State Universal Basic Education Board (AKSUBEB) was used to train and test the performance of the ML models. Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), and Decision Tree (DT) classifiers were deployed where results indicate that DT emerged the most effective classifier with a 98% prediction accuracyfollowed by RF with accuracy of 97.59% while LR recorded the least accuracy of 79.43%. This outcome indicates that tree-based ML structures can significantly help HRpersonnel to efficiently select suitable candidates for given job positionswith reduced overhead in the recruitment process.

Keywords: Machine learning, recruitment,applicants, human resource, decision making

1. Introduction

Unemployment in Nigeria experienced a surge, ranking second-highest on a global list, and the rate of joblessness increased from 22.6% to 33.3% between September and December 2021 (Sasu, 2022). This rise in joblessness can be attributed to poor governance, ineffective management of the economy, unfavorable investment climate, lack of employment opportunities, insecurity, inadequate infrastructure, and flawed recruitment processes. The government and private organizations have crucial roles to play in addressing these problems. The recruitment and selection of employees, whether in the public or private sector, is a critical priority because employees are responsible for realizing the vision and mission of any organization. Thus, the efficient utilization of the human resource (HR)management system in the recruitment process is necessary for any organization to attain its objectives (Ekwoaba et al., 2015). This process involves searching and selecting candidates for employment after stimulating them to apply for jobs in the organization (Geetha and Bhanu, 2018).

The recruitment and selection process can be conducted internally or externally in the public service, with each having its merits and demerits. Internal recruitment includes staff promotions and transfers within theorganization while external recruitment involves sourcing candidates from outside the organization through advertisement, referrals, and other sources (DeVaroand Morita, 2013). The recruitment and selection process must be objective to enhance the performance and productivity of the organization (Ekwoaba et al., 2015).

Many organizations outsource the recruitment process to experienced HR personnel, but the deployment of a smart structured recruitment system driven by machine learning (ML)tools has not been effectively utilized. A structured intelligent recruitment system is data-driven, ML-oriented, and utilizes advanced techniques and models to aid HRmanagers in making effective

decision. It incorporatesdata from various sources to process employee quality, enhance employer brand value, and build a talent community (Rifai et al., 2007).

HR management involves developing individuals to achieve the organization's recruitment objectives. HR managers bear the onus of procuring and electing individuals for the institution, whether by internal or external means, performing connecting role between the establishment and the applicants. However, one major hindrance confronting HR managers and the administration is how to select suitable candidates from the large number of job seekers to fill few vacant positions, especially in Akwa Ibom State, where the government is the primary employer of labor. To address this challenge, the government has adopted a web-based recruitment system to ease the recruitment process.However, there stillremains a persistent demand from citizens for a dependable selection process by HRmanagers. This critical undertaking involves the identification of a suitable candidate possessing the requisite skill set as specified by the particular job vacancy. Many companies have resorted to the use of recruitment software, such as Ceridian, Indeed, ZipRecruiter, Workday, Avatar, etc., to select top candidates who typically work in conjunction with, or as a part of HR (Ramachandran and Sharma, 2021). Nevertheless, the HR department is still grappling with challenges that are yet to be resolved, including recruitment selection, the imbalance of interest between management and employees, applicant performance appraisal, dealing with influential board members, and managing ethnicity or nepotism.

Several studies have attempted to address these issues. For example, Saju (2018) examined the intricacies of how ML is affecting job seekers and the employment landscape as a whole through an interactive web application. The author aimed at reducing challenges, such as gender inequality, faced by job seekers but the prediction was based on a limited dataset size, which may not yield an optimal result. Dutta and Bandyopadhyay (2020) utilized supervised ML tools, such as multi-layer perceptron(MLP), K-Nearest Neighbor(KNN), Decision Tree(DT), Naive Bayes(NB), and Random Forest(RF), to investigate scam job recruitment. The results showed that RF classifier yielded an accuracy of 98.27%, outperforming others but the work is constrained by high implementation costs. More et al. (2020) proposed an employee recommendation system using ML algorithms, in which natural language was used to extract applicant details. The results obtained from the ML were evaluated with an accuracy of 84% for the Logistic Regression (LR) and Naïve Bayes algorithms, respectively. However, the system is limited by the use of independent predicting features and suffers from the zero-frequency problem where it assigns zero probability to a categorical variable whose category is not available in the training dataset. Nevertheless, existing works are limited by their insensitivity to the scale of the data and irrelevant features, require high memory for data storage, have long training times and are computationally expensive with poor performance on noisy data. However, job recruitment and selection are non-deterministic combinatorial problems that require to be solved with special algorithms. The problem is to select a set of employees from a set of available vacancies, assign them to the appropriate job position, and take into account the peculiar constraints inherent in the selection process. Each job recruitment process has to be completed within a specific timeline. For efficiency, HR units must perform a minimum number of hours for any allocated task. There is a maximum number of tasks that can be assigned and a maximum number of workers that can be allocated.

There exists a set of procedures that must be followed to achieve effective decision-making by the system. The main objective is to minimize the costs, time, and energy associated with the necessary HRs involved in the recruitment process. However, due to the complexity of the problem, with strong constraints and instances of realistic size exceeding three thousand applicants, it becomes impossible to apply exact methods. Research conducted by Hewitt et al.

(2015) has investigated workforce planning models that contain non-linear models of human learning, which are reformulated as mixed-integer programs. The authors demonstrate that the mixed-integer program is much easier to solve than the non-linear problem. In such a case, the corresponding objective functions and given constraints need to be converted into crisp equivalents and then solved by traditional methods (Ning et al., 2013). Furthermore, uncertain models need to be transformed into an equivalent deterministic form, as shown by Yang et al. (2017). Unfortunately, most of these studies overlook constraints in the recruitment and selection process, simplifying the problems. Nevertheless, the use of ML techniques has proven to be very effective and efficient in solving problems of high complexity (NP-hard) in the field of combinatorial optimization. The advantages ofadaptability, positive feedback for good rapid recovery, not being trapped by local optima, easily identifying trends and patterns, sequences of random decisions, no human intervention, continuous improvement, and handling multi-dimensional and multi-variety data make these techniques ideal. The purpose of this research is to use ML techniques to develop a smart intelligentrecruitment system that will aid HR professionals to predict and select suitable candidates for the right job position in organizations using Akwa Ibom State Universal Basic Education Board (AKSUBEB)as case study.The system can optimize the recruitment process to save costs, time, and manpower utilization. It can also support HRpersonnel, government parastatals, institutions, and organizations in makingaccurate and objective decisions about job recruitment and employees' positions.

The rest of the paper is organized as follows: section 2 reviews related works on job recruitment process and personnel performance evaluation highlighting the strengths and weaknesses in previous studies. Section 3 presents the proposed ML data analytics approach for enhancing job recruitment prediction while section 4 discusses the results and how it can help maintain high-quality placement records for applicants and thus enhance the recruitment and selection process for HR managers. Section 5 concludes the paper with direction for future works.

2. Related Works

Personnel selection is regarded as a complex task where multiple criteria must be considered for good decision making. Multi-attribute decision making methods are often used to solve discrete decision problems. Nevertheless, solving these problems require the construction of an evaluation procedure to rate and rank a set of alternatives in order of preference. For example, the analytical hierarchy process (AHP) mathematically handles multi-criteria decision making through pair-wise comparison of decision makers' subjective judgment of criteria and alternatives with respect to the goal. However, these approaches are computational complex with inherent inconsistencies and uncertainties in the decision process. Asuquo and Onuodu (2016) developed a fuzzy AHP model for selection of university academic staff aimed at utilizing fuzzy logic principles to handle the imprecision, uncertainties, fuzziness and vagueness that exist in the selection process by decision makers.Academic qualification, research experience, and individual factor were used as selection criteria.Mezhoudi et al. (2021) conducted a survey analyzing 20 relevant papers on employability prediction from different academic libraries using ML techniques. The study revealed that gender was the most commonly used feature, while psychometric attributes and economic context were often overlooked, and data completeness and accuracy were identified as significant challenges in the prediction process.

Predicting employability using data mining techniques implies several challenges and limitations that need to be overcome. The major challenge for employability studies is collecting consistent and quality data. Often, ML algorithms perform better with larger datasets, as well as with more information about an applicant or future employee. For example, in Potgieter and Coetzee (2013), the authors mention the need for gathering more attributes such as grades for subjects

taken during the study period, the results of the oral test, and work status. Unfortunately, collecting extra information is expensive and sometimes even legally challenging due to privacy reasons. It isobserved that the work focuses on one cohort and does not address scalability and reproducibility problems. Furthermore, themost used feature is gender. This is not a surprise as the gender gap in employment is a worldwide fact that many governments and organizations are now actively addressing as they seek gender parity. Surprisingly, only a few studies considered psychometric attributes despite their role in employability. These recent studies showed that psychometric attributes contribute to the analysis of youth unemployment. On the other hand, the work did not consider the economic context and features related to the country's economy despite their significance for the job market and employment. It isalso observed that none of the studies collected curriculum vitae (CV), and only two studies collected job adverts, although these are the two main inputs traditionally used by recruiting agents and job seekers. Another major challenge is the completeness and accuracy of gathered data. This is especially the case when the data is collected via web forms as the applicants can quit the webform at any moment or via e-learning activities.

Kulkarni and Che (2019) proposed intelligent software tools for recruitment using artificial intelligence (AI) tools. However, the proposed system did not address the matching criteria problem, and no candidate profile was captured- and does not validate if factors such as company size or industry type play a role in the implementation of AI-based tools. The work in Yiğit and Shourabizadeh (2017) proposed an approach for predicting employee churn using data mining. The authors deployed some data classification tools. The results obtained were evaluated based on accuracy, precision, F-measure, and recall. The work showed that SVM gives 86.7% accuracy performance among all the classification techniques used. This work relied on the simulated data and is computationally expensive.

Warrenbrand (2021) undertook a quasi-experimental analysis of job applicant response to selection processes to deduce whether MLtechniques or human hiring decision-makers influence perceptions of fairness and equity and ultimately organizational attraction to job pursuit intentions. The result shows that the human hiring decision-makers gave more fairness and equity than the ML algorithm. The limitation of the work is that its dataset had few data points collected for analysis and the author did not implement these ML techniques using the dataset butrather performed analysis using statistical tools. As a result, the total sample size prevented their findings from being generalizable to a larger working population expecting recruitment. This work primarily comprises participants identifying as white, while representation from other ethnic identities is notably lower. This study consisted of a survey-only design which makes it difficult to fully understand participant perceptions as compared to a mixed-method study design with qualitative data collection or an experimental design predicting actual behavior. Results only suggest relationships and cannot suggest cause-and-effect nor do they give explanations from participants regarding their responses. Finally, the outputs indicated extremely high values from examining organizational attraction as a mediator between justice perceptions and job pursuit intentions. These high values signify a potential overlap of measures of distributive and procedural justice. Mahmoud et al. (2019) made a follow-up conceptual model of using AI in the hiring process in performance management and social screening to predict a new candidate's expected performance by analyzing historical performances and conditions of employees. The study is aimed at giving HR an additional parameter that assists in the hiring process. A Naive Bayes algorithm was used on the data and 74% performance accuracy was obtained. The study only considered the experience of employers without giving attention to other features.

Faliagka et al. (2012) proposed the application of supervised learning algorithms in automated e-recruitment systems to solve the candidate ranking problem. This work extracted a set of

objective criteria from the applicants' LinkedIn profiles and infers their personality characteristics using linguistic analysis of their blog posts. The objective was to limit interviewing and background investigation of applicants solely to the top candidates. The results of this work show that despite the high degree of uncertainty on applicant personality, the system was able to achieve a correlation coefficient of up to 81% and the senior programmer's position exhibited the lowest consistency, with a Pearson's correlation of up to 73%. The limitation of this work is that very few datasets were used for implementation which may not give a robust system and the system was so reliant on only one main feature which is the extraversion personality of applicants and there was a lot of uncertainty on applicant's profile in the system. Alkhazraji and Buhaliba (2020) proposed MLsoftware in the HRrecruiting process for candidates from the Dubai Police Academy to develop and test a prototype of the system for the functionalities it is meant to perform. The author used a survey research design and triangulates both qualitative and quantitative methods for improving the validity and credibility of the study outcomes. The limitation of the work is that the system relies heavily on specific computer configurations, limiting its portability, compatibility, and future adaptability. Users may face challenges with maintenance, upgrades, and performance variability.

Rifai, et al. (2007) developed an intelligent recruitment system. The system combined both the application form features such as interests, values, skills, talents, objectives, and self-assessment and the interviewing part of the recruitment process. The objective is to give suggestions about whom to hire for a particular job vacancy, based on specific criteria using the hidden Markov model (HMM) and artificial neural network (ANN). The results obtained showed that the speech recognizer produced an accuracy rate of about 80%, smaller words were recognized with a rate close to 95%, while larger and more complex words are recognized with a success rate between 60% and 80%. The work is limited to the fact that the system was constrained to oral interviews and may lead to errors in decision-making. The system cannot produce an ideal candidate profile for each job vacancy.Ekwoaba et al. (2015) investigated the impact of recruitment and selection criteria on organizational performance. The authors used a questionnaire method that was administered to randomly selected respondents and revealed that recruitment and selection criteria have a significant effect on an organization's performance (X^2= 35.723; df= 3;p<0.05), the result shows that the more objective the recruitment and selection criteria, the better the organization's performance (X^2 = 20.007; df= 4; p<0.05) but the work did not address the problem of subjective judgment inherent in the recruitment and selection process. Othman et al. (2018) propose data mining techniques to determine factors that affect the graduate's employability status. Three classification algorithms Support VectorMachine (SVM), ANN, and DT were considered. Their accuracy performance was compared for the best modeland results showed that DT J48 outperformed otherswith an accuracy of 66.0651%. Yadav et al. (2019) highlighted the need for an online job board system for colleges and its effectiveness in bridging the gap between college students and career opportunities. The purpose of the work was to help students in their final year class find better jobs and employers find suitable candidates for the job. The work deployed a questionnaire method to determine the behavior of students toward the use of the e-recruitment system. The results showed that 35% of the students had never used any job portal, only 18.8% had benefitted from the portal for finding jobs, and 78% of students were excited about in campus job portal. This work is not reliable for strong decisions.

Casuat and Festijo (2019) deployed ML for predicting students' employability. Three learning algorithms were used such as DT, RF, and SVM to study how students are recruited. Thesealgorithms were evaluatedbased on performance metrics of accuracy, precision, recall, and F1-score. SVM obtained 91.22% accuracy which gives quality results over DT with 85%, and RF with 84%. Hugo (2018) used five ML models such as LR, Discriminant Analysis (DA), DT,

ANN, and SVM to predict the employment of undergraduate students at graduation. SVM outperformed other models with 87.26% accuracy, followed by ANN with 73.11%, and LR with an accuracy of 72.17%. DT and DA had the least performance accuracy of 71.70% and 69.81% accuracyrespectively. However, a small dataset was used for their analysis which might not give an optimal prediction. Table 1 presents a chronological summary of previous studies on job recruitment systems spanning over a decade, highlighting the study objective, algorithm or model used, evaluation metric, results, strengths, and research gaps.

Rifai et al. (2007) proposed an intelligent recruitment system using HMM and ANN, achieving 80% accuracy for smaller words and 60-80% for larger and complex words; however, the study's limitation of considering only word features for screening resulted in inaccurate job selection, emphasizing the necessity to explore and include additional relevant criteria for improved candidate identification.Ekwoaba et al. (2015) conducted a questionnaire-based study to examine the impact of recruitment and selection criteria on organizational performance, finding a significant positive correlation between more objective criteria and better organizational performance (X^2= 20.007; df< 0.05 = 4; p<0.05). However, the study overlooked the issue of subjective judgment inherent in the recruitment and selection process and lacked the integration of intelligent systems.In their study, Mwakondo et al. (2016) employed ML techniques to predict graduates' skills for industry roles, utilizing three models: Kirkpatrick's Model, the CRESST model, and the theory of cognitive learning model. The results demonstrated an 84% reduction in industry dissatisfaction with graduates' productivity due to improved evaluation of their skills against industry job competence requirements, despite facing challenges of high maintainability and implementation costs.In Saju's (2018) study, an ML tool was utilized to develop a platform for job aspirants to predict job prospect features, achieving a mean squared error of 0.046 and a coefficient of determination of 0.94 as a result. However, the research overlooked addressing issues of nepotism and gender inequality in the job market.Casuat and Festijo (2019) employed an ML approach to predict students' employability, utilizing three learning algorithms (DT, RF, and SVM) and performance measures such as accuracy, precision, recall, F1-score, and support. SVM outperformed DT and RF with the highest accuracy of 91.22%, showing significant improvement. However, the study solely considered students' cumulative grade points and did not include other relevant factors for employability prediction. Asuquo et al. (2020) evaluated the performance of ML models of C4.5, random forest, and Naïve Bayes for employee performance and promotion prediction. RF classifier demonstrated superior performance with the highest prediction accuracy and F-measure of 98.70% and 0.988, respectively. To improve competency and job retention, staff predicted as ‘not promoted’ were recommended by the system for professional training and development support or for normal salary increment. However, the study was limited by using few dataset. Alkhazraji and Buhaliba (2020) proposed predicting employees' sustainability using ML tools, specifically employing a multiple linear regression algorithm based on parameters such as age, previous salary, family size, marital status, experience, and gender to estimate sustainability. The results demonstrated an efficiency of 68%, indicating the system's dependability and precision in measuring employees' sustainability. However, a limitation of the approach is the high cost of implementation.VidyaShreera and Muthukumaravel (2021) used ML tools to predict student careers for job employability, employing Lagrangian formulation for SVM classification. The RF classifier yielded the highest accuracy of 93% among various ML classifiers. However, a limitation of the system is its failure to consider additional features, such as years of experience, for distinguishing between skilled and unskilled workers.

Table1: Summary of Related Works

Author	Objective	Algorithm/model/evaluation metric	Result	Strength	Weakness
Rifai et al. (2007)	Developed an intelligent recruitment system	HMM and ANN models were used	Results indicate that the speech recognizer produced an accuracy of 80% for smaller words while larger and more complex words are recognized with a success rate between 60% and 80%	It gives 72% of applicants word recognition	The system considered only word features for the main screening process which does not give accurate results for job selection
Ekwoaba et al. (2015)	Investigated the impact of recruitment and selection criteria on organizational performance	Questionnaire method used	Result shows that the more objective the recruitment and selection criteria, the better the organization's performance	It discovered that recruitment and selection criteria significantly important	The study did not address the problem of subjective judgment inherent in the recruitment and selection process. The system lacks the utilization of intelligent tools
Mwakondo et al.(2016)	Proposed a model for predictive mapping of graduate's skills to	Used 3 models: Kirkpatrick's Model, CRESST model, and the theory of cognitive	Result shows 84% reduction in growing dissatisfaction by industry	It helps in the reduction of confusion among graduates and increases transparency	High cost of maintainability and implementation

industry roles using ML techniques	learning model. Parameters considered are relevancy, durability, accuracy, and relevancy is given as: $D = t * \frac{12}{T}$ $R = t * \frac{12}{T}$ $A = t * \frac{12}{T}$ where, R, D, A are Relevancy Index, Durability Index, and Accuracy Index respectively. t is the calculated number of requirements while T is the highest possible total	over graduates' productivity as a result of poor evaluation of graduates' skills vis-à-vis industry job competence requirements	during recruitment

Saju (2018)	Used ML tool to create a medium for job aspirants to predict job prospect features like job vacancies, intakes, salary available in different sectors	The variance and chi-square tests were used, given as: $$variance = \frac{\sum(x-\mu)^2}{n}$$ where, x is the value of one observation; μ is the mean value of all observations; n is the number of observations. $$x^2 = \sum \frac{(O_i - E_i)}{E_i}$$ where, x^2 is chi-square test; O_i is the observed value; E_i is the expected value	Accuracy measures of sector-wise prediction, where RMSE (0.046) and R^2 (0.94) were obtained	The study shows that statistical analysis provided in the web application can help job aspirants to understand current and future trends in the job market for different sectors	The work did not address the problem of nepotism and gender inequality

Casuat and Festijo (2019)	Predicting students' employability using the ML approach	Three learning algorithms were used namely, DT, RF, and SVM	Performance measures used include accuracy, precision and recall measures, F1-score, and support measures. SVM obtainedthe highest accuracy of 91.22%compared toDT (85%) and RF (84%)	SVM results obtained 91.22% in accuracy measures which was a significant improvement	The work only considered students' cumulative grade points
Alkhazraji and Buhaliba (2020)	Proposed predicting employees' sustainability using ML tools	Multiple linear regression algorithm based on the following parameters namely, age of the candidate, previous salary, family size, marital status, experience, and gender to estimate the approximate numerical value of sustainability	The result shows 68% efficiency indicating that the system is dependable and precise in measuring employees' sustainability	The system shows a 0.68 efficiency	However, it is limited by the high cost of implementation
Usmani (2020)	Examined recruitment and selection process at the workplace	Both qualitative and quantitative research methods were used. Qualitative data was collected throughan open-ended	The result shows that only physical/facial attractiveness is different across all applicants	The system shows that physical and facial attractiveness is not important in job recruitment	No ML approach was adopted to ascertain the correctness and accuracy of the obtained result

		Questionnaire	and is found to be very insignificant	and selection	
Asuquo et al. 2020	Develops an ML framework for staff performance and promotion prediction	C4.5, Random Forest and Naïve Bayes; Accuracy, precision, recall, F-measure, ROC-curve	RF demonstrated superior performance with highest prediction accuracy and F-measure of 98.70% and 0.988, respectively.	Staff predicted as 'not promoted' were recommended by the system for professional training and development support or for normal salary increment	The study used a few dataset
VidyaShreera and Muthukumaravel (2021)	Used ML tools to predict student careers as a determinant factor in job employability .	Lagrangian formulation, given as: $d(X^T) = \sum_{i}^{l} y_i \alpha_i X_i X^T + b_o$ where, $d(X^T)$ is Lagrangian formulation; y_i represents the label of the call vector X_i; α_i and b_o represent the numeric arguments that are found automatically by	RF classifier yields better accuracy of 93% compared to other ML classifiers	Aids educational institutions take more care of low-grade students and recruiters to select suitable candidates for their companies	The system fails to consider other features like years of experience for skilled and unskilled workers

	the SVM classifier; i is the identifier represented as the total number of support vectors

2.1Machine Learning Approach

ML is a methodology that forms part of AI and performs data analytics by training machines to respond to specific inputs or scenarios based on previously learned inputs. Essentially, it involves granting computers the capacity to learn through statistical techniques. By enabling computers to act without explicit programming, ML minimizes human intervention in machine-dependent problems and scenarios. This approach facilitates the simple and efficient resolution of complex tasks with minimal human labor. ML applications adopt prediction, clustering, classification, or regression techniquesencompassing diverse fields including, image recognition, medical diagnosis, algorithm development, text processing, sentiment analysis, and self-driving cars, among others. The two major ML techniques for addressing complex problems are supervised and unsupervisedlearning. Unsupervised learning entails identifying similarities between data points and their characteristics to create groups without prior knowledge of the final output classes and sets (Awujoola et al.,2021). In contrast, supervised learning involves training an algorithm on a labeled dataset to predict new data.Casuat and Festijo (2019) employed three classification algorithms, including RF, DT, and SVM, to predict the employability status of fresh graduate students. The results indicate that SVM achieved the highest accuracy of 91.22%, outperforming DT and RF algorithms, which had accuracy rates of 85% and 84%, respectively. In a similar study, Hugo (2018) developed five ML models, including LR, DA, DT, ANN, and SVM, to forecast the employment of undergraduate students upon graduation. The results reveal that SVM has the highest accuracy rate of 87.26%, followed by ANN with 73.11% accuracy, and LR with 72.17% accuracy. On the other hand, DTand DA had lower accuracy rates of 71.70% and 69.81%, respectively.

3.Methodology

3.1 Problem Description

The Government of Akwa Ibom State aims to optimize its job recruitment process by harnessing the power of cutting-edge technologies. The government regularly receives a large influx of job applications for various positions within its various departments and agencies, making the candidate selection process time-consuming and resource-intensive.To streamline the hiring process and ensure the selection of many qualified candidates, the State Governmentwants to implement an advanced system that can predict the employability of applicants based on their submitted CVsand other relevant information. The government possesses historical data on past job applicants, including their resumes, educational background, work experience, certifications, and job performance ratings, which are rated on a scale from 1 to 5, with 5 representing the highest performance.The system automatically processes incoming CVs, extract relevant features, and clusters applicants into different groups based on their qualifications and profiles. Utilizing unsupervised ML algorithms such as clustering and dimensionality reduction, the system can uncover patterns in the data and group applicants with similar skill sets and experiences.The system canthen use this clustered information to stimulatepredictive models to

classify each applicant based on job positions. The goal is to predict the potential job performance rating for new applicants based on their similarity to past high-performing candidates within each cluster. Additionally, data analytics will be employed to continuously evaluate and fine-tune the models based on the actual performance ratings of newly hired candidates.The ultimate objective is to increase the efficiency and accuracy of the recruitment process for the State Government, leading to better hiring decisions and ultimately improving the overall performance and effectiveness of the government's workforce. The system's success can be assessedby its ability to reduce the time-to-hire, minimize recruitment costs, and improve the quality of selected candidates while maintaining fairness and transparency in the decision-making process.

3.2 Enhanced Job Recruitment Prediction Design Approach

In this work, a predictive system is designed using ML algorithms to forecast the applicant's best status for a particular job vacancy. The steps used in our methodology, shown in Figure 1, include data collection, data pre-processing, splitting data into training and testing datasets, analyzing and visualizing the data using exploratory data analysis (EDA), designing a conceptualized ML predictive recruitment framework, implementing the framework using Python programming language, evaluating and validating the system performance with metrics like accuracy, precision, recall, and F1-score. The ML classifiers considered are NB, RF, SVM, and LR.

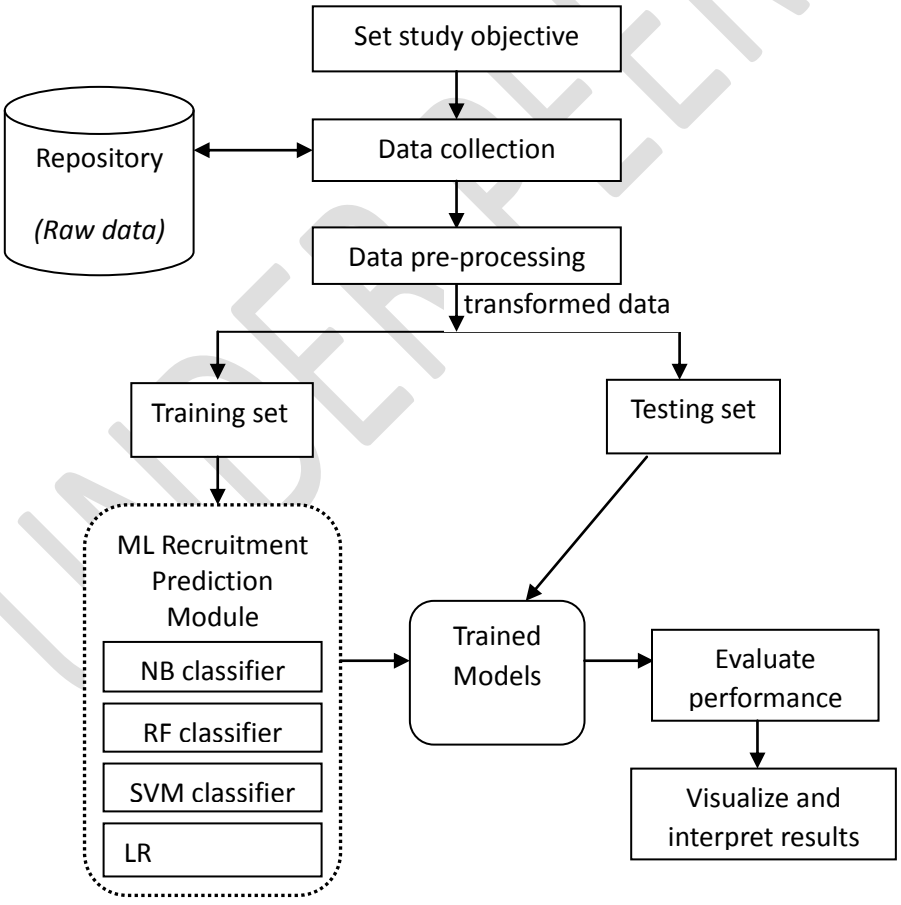

Figure 1:Steps in the proposed methodology

There exists a set of procedures that must be followed to achieve effective decision-making by the system. The main objective is to minimize the costs, time, and energy associated with the necessary HR personnel involved in the recruitment process. However, due to the complexity of the problem, with strong constraints and instances of realistic size exceeding three thousand applicants, it becomes impossible to apply exact methods due tothe limited number of quotasslated for employment. The use of ML techniques has proven to be very effective and efficient in solving problems of high complexity (NP-hard) in the field of combinatorial optimization. The machine learning algorithm is chosen for the development of an intelligent recruitment system due to its numerous advantages, such as adaptability, receiving positive feedback for rapid recovery, avoiding local optima traps, efficient identification of trends and patterns, making sequential decisions without human intervention, enabling continuous improvement, and handling multi-dimensional and multi-variety data. This system is expected to assist organizations in optimizing their recruitment process, leading to cost savings, time efficiency, and better utilization of manpower resources.

The purpose of this research is to use ML techniques to develop an intelligent recruitment system that will aid HR professionals and organizations to predict and selecting suitable candidates for the right job position in Akwa Ibom State Universal Basic Education Board (SUBEB). It will support HR personnel, government parastatals, institutions, and organizations to make accurate and objective decisions about employees' positions.

3.3 Proposed System Architecture

The flaws commonly encountered in existing systems motivated the design of a new applicant recruitment system to aid the recruitment process to effectively and accurately select suitablyqualified applicants for recruitment into AKSUBEB. This study deploysvisualization software tools such as Exploratory Data Analysis(EDA) tools to give a better understanding of the datasets. The system recognizes the number of vacancies, the number of available positions, and the qualifications needed. Features such as referee category, age, gender, and grade are considered. The proposed system architecture for the intelligent recruitment systemis shown in Figure 2 with the following components:

(i) **Data Repository:** Each applicant’s qualifications, age, referees, date of birth as well as computer-based test (CBT) and oral test scores, are collected and stored in the system’s database. Thus, the user-friendliness and the practicality of the system are maintained

(ii) **Data Collection**: Data is collected from the web portal of AKSUBEB. The portal implements the input forms that allow the candidates to apply for a job position. Each candidate is given the option to log into the system using his account credentials, which allows the system to automatically extract all objective selection criteria directly from the user’s profile that is stored in the data repository.

(iii) **Extracting Applicants Data**: The system will automatically extract the submitted applicant's information through a suitable application programming interface (API)and data is now pre-processed using Python libraries.

(iv) **Data Analysis and Visualization:** Raw data that are noisy, incomplete, and inconsistent will be pre-processed using pandas, numpy, and seaborn libraries in Python software and later split into training and testing datasets. Visualization of these pre-processed datasets is done in an exploratory data analysis tool as shown in Figures 3 and 4.

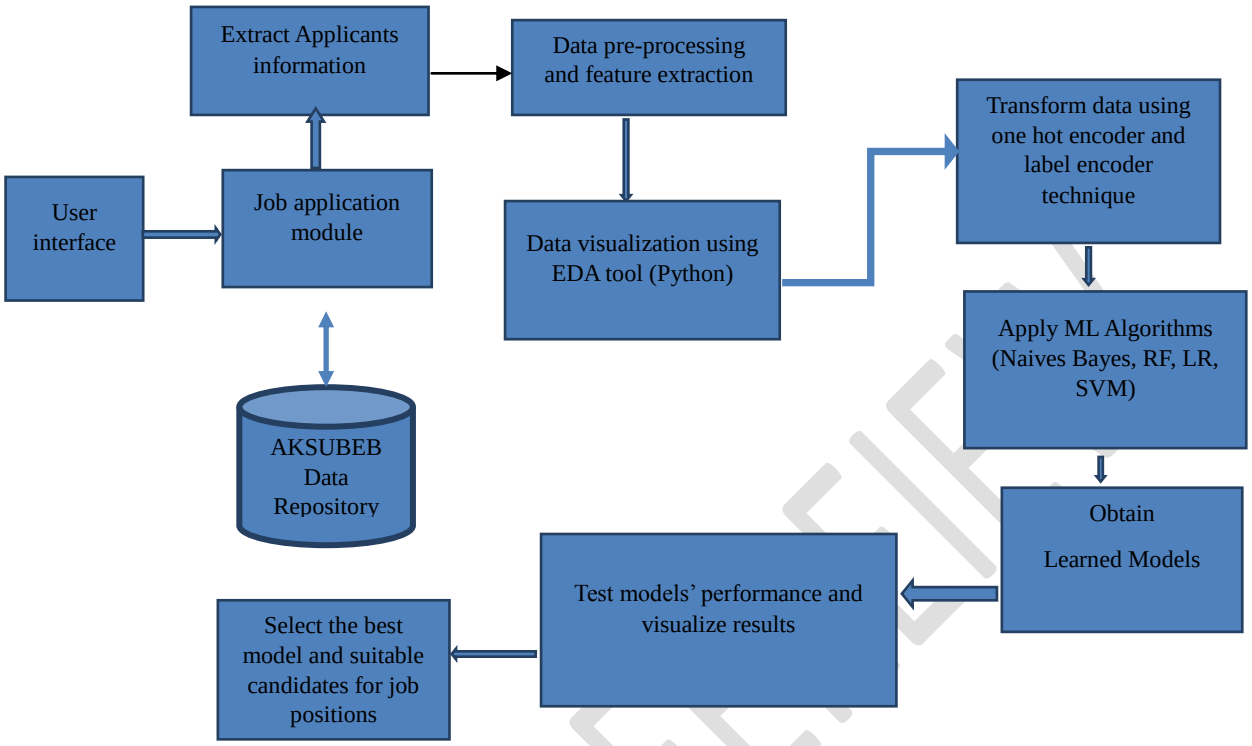

Figure 2:Proposed system architecture for intelligent recruitment system

Figure 3 depicts the percentage of applicants in each of the local governments, the percentage distribution of male and female applicants, the percentage of the referee class category of the applicants, and the percentage of highest qualification grade for each of the applicants while Figure 4 shows the density distribution plot for age, certificate, and scores. These give data profiling and provide direction on the transformation of the dataset for accurate prediction.

Figure 3(a) shows that Oruk Anam Local government has the highest number of applicants with 5.9% percentage, follow by Etinan 5.5% and the least is Eastern Obolo 0.5%. Figure 3(b) shows that 78.5% of female and 21.5% of male candidates applied for the job positions, referee by clergy was 22.1%, follow by top government officials and top politicians with 0.5%, signifying that a greater number of applicants used clergy men as their referees. Also only 0.3% of the applicants had first class, merit (49.5%), distinction (25.2%), credit (15.8%) and pass (9.2%). Figure 4 shows the density distribution of age 0.16, certificate 6.1, computer test scores > 0.01, and oral score >0.01.

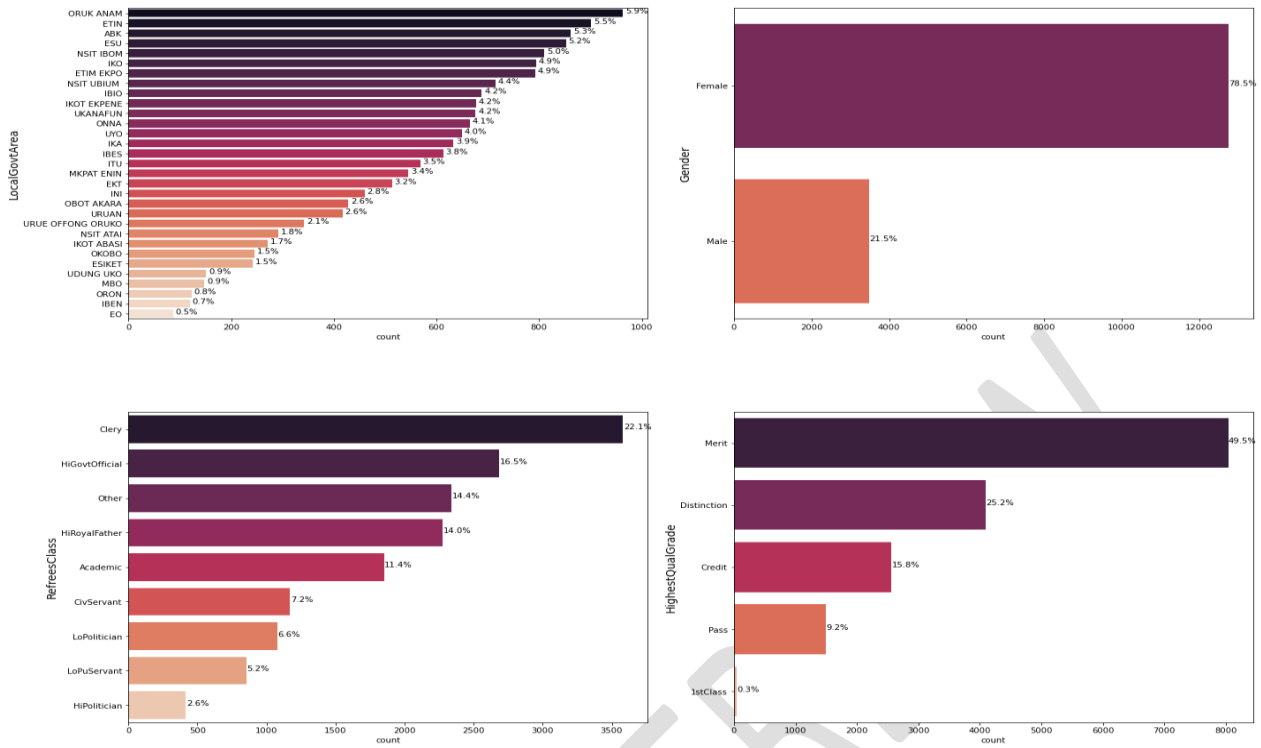

Figure 3: Visualized plot of local govt., reference class, gender, and highest qualification

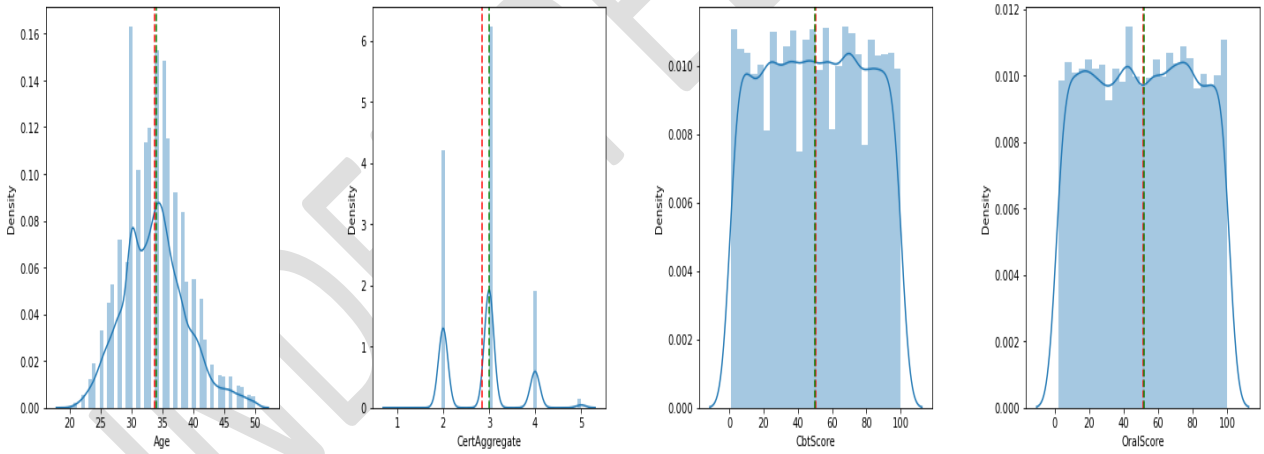

Figure 4: Density distribution plot for age, certificate, and scores

(i) **ML Algorithms**: ML tools such as RF, SVM, NB, and regression analysis were deployed in this study.

a) **Support Vector Machine**

SVM extracts sets of input data and predicts or classifies for each of the given inputs which one of the two possible classes comprises of the input, making the SVM a non-probabilistic binary linear classifier.

b) **Naïve Bayes Algorithm**

It takes an underlying probabilistic model and allows us to capture uncertainty about a model fairly by determining probabilities of the output. It is given as:

$$P(x|c) = \frac{P(x|c)P(c)}{P(c)} \quad (1)$$

where, $P(c|x)$ is the posterior probability of class (c, target) given predictor (x, attributes), $P(c)$ is the prior probability of a class, $P(x|c)$ is the likelihood which is the probability of the predictor given class and $P(x)$ is the prior probability of the predictor (Berrar, 2018).

c) **Logistic Regression**

This is used to predict the probability of a certain event occurring such as the pass or fail, win or lose, alive or dead, or healthy or sick. It is given as:

$$\theta = (X^T X)^{-1}.(X^T Y)^{-1} \quad (2)$$

where, θ is the hypothesis parameters that define the optimal performance of the system., X is the input feature value of each instance (something is missing)and is the output value of each instance (Abdi, 2010).

d) **Random Forest**

It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model**.**It is an ensemble of DTs, and expressed in equation (3) as:

$$RF = 1 - [(P_+)^2 + (P_-)^2] \quad (3)$$

where, P_+ and P_- represent the probability of positive and negative classes respectively.

4.Results and Discussion

The experiment was conducted in multiple stages to enhance prediction accuracy and facilitate result analysis. Furthermore, the evaluation of model accuracy performance was carried out both with and without adjustment of classifiers' parameters.A total of 16240 datapoints were obtained and stored in the data repository for the study. The dataset was split in the ratio of 8:2 for model training and testing.

Table 2showsthat the SVM classifier achieves an accuracy of 84.61%, precision of 84.41%, recall (sensitivity) of 84.61%, as wellas F1-Score of 84.47%.RF classifier achieves an accuracy of 97.59%, precision of97.60%, recall (sensitivity) of 97.59%, and F1-Score of 97.60%.The LR classifier achieved an accuracy of 79.43%, precision of 78.88%,Recall (sensitivity) of 79.43%, F1-Score and Recall of 78.90%, respectively.The DT classifier performed exceptionally well with accuracy, precision, recall, and an F1-Score of 98.49%. The model shows a near perfect classification performance compared to others. The NB classifier achieved a moderate performance with accuracy, precision, recall, and an F1-Score of 85.65%. The model has a good ability to predict positive cases (selected), but it also has more misclassifications compared to RF and DT algorithms.Thus, the highest prediction accuracy was achieved by DT, followed by RF, NB, and SVM while LR achieved the least performance.Figure 5 presents a performance comparison of these classifiers based on accuracy, precision, Recall, and F1-score metrics.

Table 2: Prediction performance of classifiers

Classifiers	Accuracy	Precision	Recall	F1-Score
DT	98.52	98.52	98.52	98.52
RF	97.59	97.60	97.59	97.60
NB	85.65	85.45	85.65	85.32
SVM	84.60	84.44	84.60	84.46
LR	79.43	78.88	79.43	78.90

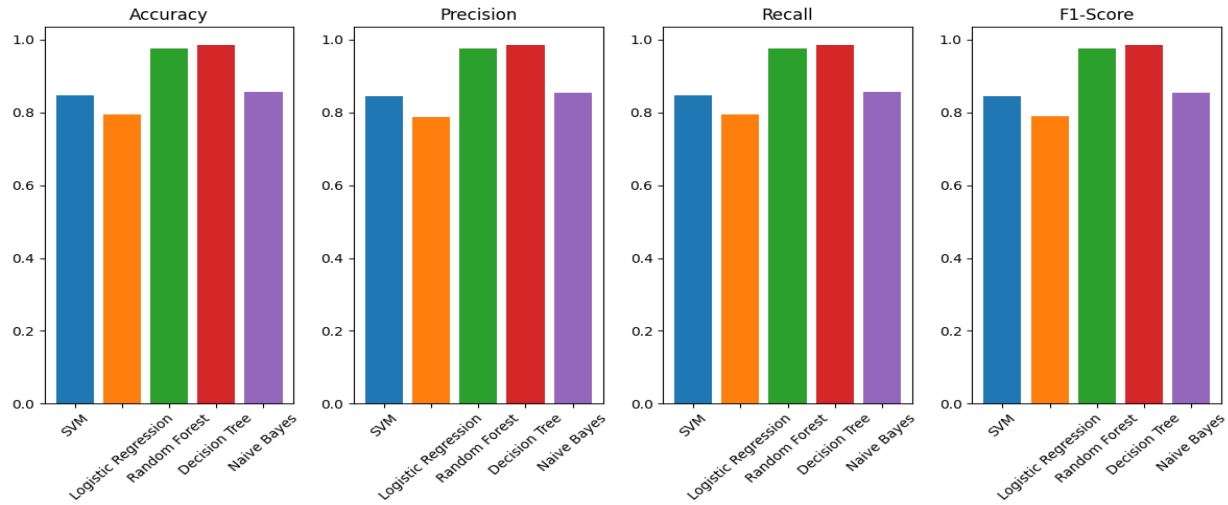

Figure 5: Classifier Performance for Job Recruitment Prediction

Table 3 presents the summary of results generated from the confusion matrix for each classifier. where True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) predictions are made. Results indicate that the DT and RF classifiers outperformed the SVM, NBand LR classifiers in terms of correctly predicting positive and negative instances. The DT classifier exhibited the highest number of true positive and true negative predictions, making it the most accurate classifier among the four.

Table 3: Confusion matrix performance of the classifiers

Classifiers	TP	TN	FP	FN
DT	1023	2176	20	29
RF	1016	2152	44	36
SVM	768	1980	216	284
NB	741	2041	155	311
LR	632	1948	248	420

5. Conclusion

This paper posits the use of ML models to predict the job employment statusof applicants. The study aimedto provide support to the HR department in theproper management of therecruitment process. To identify the most effective model, five different ML models were adopted and performance evaluation indicates that DT outperforms others in the task of predicting suitable job placement for qualified applicants into given job positions, timely and accurately.Thus, the outcome of this study can be deployed to assist HR managers in AKSUBEB and other organizations to make significant contributions to the placement process when analyzing applicants' performance before selection. Moreover, it will enable the government parastatals, ministries, agencies, and boards to maintain high-quality placement records forthe applicants and thus reducenepotism and bias concernsnormally observed during the recruitment and selection process.

References

[1] Abdi, H. (2010). Partial Least Squares Regression and Projection on Latent Structure Regression(PLS Regression). Wiley Interdisciplinary Reviews: Computational Statistics, 2(1), 97-106.

[2] Alkhazraji, I. and Buhaliba, A. S. (2020). Using Machine Learning Software in the Human Resource Recruiting Process for Candidates from Dubai Police Academy. Master Thesis. Rochester Institute of Technology, Dubai.

[3] Asuquo, D. E. and Onuodu, F. E. (2016). A Fuzzy AHP Model for Selection of University Academic Staff, *International Journal of Computer Applications*, 141(1), 19-26.

[4] Asuquo, D. E., Umoh, U. A., Osang, F. B. and Okokon, E. W. (2020).Performance Evaluation of C4.5, Random Forest and Naïve Bayes Classifiers in Employee Performance and Promotion Prediction, *African Journal of Management Information System*, 2(4), 41-55

[5] Awujoola, O., Odion, P. O., Irhebhude, M. E., and Aminu, H. (2021). Performance Evaluation of Machine Learning Predictive Analytical Model for Determining the Job Applicants Employment Status. *Malaysian Journal of Applied Sciences*, *6*(1), 67-79.

[6] Berrar, D. (2019). Bayes' Theorem and Naive Bayes Classifier. Shoba Ranganathan, Michael

[7] Gribskov, Kenta Nakai, Christian Schönbach (eds.), Encyclopedia of Bioinformatics and Computational Biology, Academic Press, 403-412,https://doi.org/10.1016/B978-0-12-809633-8.20473-1.

[8] Casuat, C. D.andFestijo, E. D. (2019). Predicting Students' employability using Machine Learning Approach. In 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 1-5.

[9] DeVaro, J. and Morita, H. (2013). Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis. Journal of Labor Economics, 31(2), 227-269.

[10] Dutta, S. and Bandyopadhyay, S. K. (2020). Fake Job Recruitment Detection Using Machine Learning Approach. International Journal of Engineering Trends and Technology, 68(4), 48-53.

[11] Ekwoaba, J. O., Ikeije, U. U. and Ufoma, N. (2015). The Impact of Recruitment and Selection Criteria on Organizational Performance.Global Journal of Human Resource Management, 8(V) 22-33.

[12] Faliagka, E., Ramantas, K., Tsakalidis, A. and Tzimas, G. (2012). Application of Machine Learning Algorithms to an Oonline Recruitment System. In Proceedings of the International Conference on Internet and Web Applications and Services, 215-220.

[13] Geetha, R. and Bhanu, S. R. D. (2018). Recruitment through Artificial Intelligence: A Conceptual Study. International Journal of Mechanical Engineering and Technology, 9(7), 63-70.

[14] Hewitt, M., Chacosky, A., Grasman, S. E., and Thomas, B. W. (2015). Integer Programming Techniques for Solving Non-linear Workforce Planning Models with Learning. European Journal of Operational Research, 242(3), 942-950.

[15] Hugo, L. (2018). Predicting Employment through Machine Learning. NACE Journal 2019. NACE National Association of Colleges and Employers, Bathelehem, PA.developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and

Engineering Sciences, 374(2065), 20150202.
[16] Kulkarni, S. B. and Che, X. (2019). Intelligent Software Tools for Recruiting. Journal of InternationalTechnology and Information Management, 28(2), 2-16.
[17] Mahmoud, A. A., Shawabkeh, T. A., Salameh, W. A., & Al Amro, I. (2019, June). Performance Predicting in Hiring Process and Performance Appraisals using Machine Learning. In 2019 IEEE 10th International Conference on Information and Communication Systems, 110-115.
[18] Mezhoudi, N., Alghamdi, R., Aljunaid, R., Krichna, G., &Düştegör, D. (2021). Employability Prediction: A Survey of Current Approaches, Research Challenges and Applications. Journal of Ambient Intelligence and Humanized Computing, 1-17.
[19] More, J., Nirmal., B., Bomble, N., Pawar, S. (2020). Implementation of Machine Learning
Algorithms for Employee Recommendation. International Journal of Engineering Research & Technology, 9, 366-368
[20] Mwakondo, F. M., Muchemi, L. and Omwenga, E. I. (2016). Proposed Model for Predictive Mapping of Graduate's Skills to Industry Roles Using Machine Learnining Techniques. The International Journal of Engineering and Science (IJES), 5(4), 15-24.
[21] Ning, Y., Liu, J. and Yan, L. (2013). Uncertain Aggregate Production Planning. Soft Computing, 17(4), 617-624.
[22] Othman,Z., Shan, S. W., Yusoff, I. and Kee, C. P. (2018). Classification Techniques for Predicting Graduate Employability. International Journal on Advanced Science, Engineering and Information Technology, 8(4-2), 1712-1720.
[23] Potgieter, I. and Coetzee, M. (2013). Employability Attributes and Personality Preferences of Postgraduate Business Management Students. SA Journal of Industrial Psychology, 39(1), 1-10.
[24] Ramachandran, S. and Sharma, D. (2021). Analysis of Challenges Facing Human Resources. Management in Current Scenario. International Journal of Innovative Technology and Exploring Engineering, 8, 151-161.
[25] Rifai, S., Haraty, R. A., and Debnath, N. C. (2007). IRS: An Intelligent Recruitment System. In CAINE, 301-306.
[26] Saju, A. G. (2018). Future Job Prediction in Trivandrum using Machine Learning Techniques, DoctoralDissertation, Dublin Business School.
[27] Sasu, D. D. (2022). Forecast Unemployment Rate in Nigeria 2021-2022. https://www.statista.com/aboutus/our-research-commitment/2683/doris-dokua-sasu
[28] Usmani, S. (2020). Recruitment and Selection Process at Workplace: A Qualitative, Quantitative and Experimental Perspective of Physical Attractiveness and Social Desirability. Review of Integrative Business and Economics Research, 9(2), 107-122.
[29] VidyaShreeram, N. and Muthukumaravel, A. (2021). Student Career Prediction using Machine Learning Approaches. In Proceedings of the First International Conference on Computing, Communication and Control System, (I3CAC 2021), Bharath University, Chennai, India, *https://doi.org/10.4108/eai.7-6-2021.2308642*
[30] Warrenbrand, M. (2021). Applicant Justice Perceptions of Machine Learning Algorithms in Personnel Selection. *Masters Theses and Doctoral Dissertations.* https://scholar.utc.edu/theses/691
[31] Yadav, V., Gewali, U., Khatri, S., Rauniyar, S. R and Shakya, A. (2019). Smart Job Recruitment Automation: Bridging Industry and University. In 2019 IEEE Artificial Intelligence for Transforming Business and Society (AITB), 1, pp. 1-6.

[32] Yang, G., Tang, W. and Zhao, R. (2017). An Uncertain Workforce Planning Problem with Job Satisfaction. International Journal of Machine Learning and Cybernetics, 8(5), 1681-1693.
[33] Yiğit, İ. O. and Shourabizadeh, H. (2017). An Approach for Predicting Employee Churn by usingDataMining. In 2017 IEEE International Artificial Intelligence and Data Processing Symposium, (IDAP), 1-4.