Educational Administration: Theory and Practice
2024, 30(5), 10428-10435
ISSN: 2148-2403
https://kuey.net/

Research Article

Smart Investigation Through Machine Learning: Study On Faculty Selection In A Large Multi-Disciplinary University Of North-Eastern India

Kaberi De[1*], Dr. Prasanjit Dasgupta[2], Dr. Subhrajyoti Deb[3]

[1*]Ph.D Research Scholar, Faculty of Management and Commerce, ICFAI University Tripura, West Tripura, Kamalghat 799210, Email: kaberide@gmail.com
[2]Professor, Faculty of Management and Commerce, ICFAI University Tripura, West Tripura, Kamalghat 799210, Email: prasenjitdasgupta@iutripura.edu.in
[3]Assistant Professor, Faculty of Science and Technology, ICFAI University Tripura, West Tripura, Kamalghat 799210, Email: subhrajyotideb_cse@iutripura.edu.in

Citation: Kaberi De et al. (2024), Smart Investigation Through Machine Learning: Study On Faculty Selection In A Large Multi-Disciplinary University Of North-Eastern India, India, *Educational Administration: Theory and Practice*, *30* (5), 10428 - 10435
Doi: 10.53555/kuey.v30i5.4314

ARTICLE INFO

ABSTRACT

The selection of faculty members in a higher education institute like a university is a difficult task, as these selections impact the learning of the students who join the university with high hopes and aspirations; hence, the essential requirement for selection is to be fair and trustworthy. Using the Support Vector Regression (SVR) model, this paper explores the characteristics of university faculty members' smart rating through various metrics such as self-assessment through the Performance-Based-Appraisal System (PBAS) designed by UGC and Behavioural Competency Evaluation mechanism of the university authority. Through rigorous data pre-processing and feature selection, an SVR is a trained model to classify faculty based on predefined quality labels. Our work demonstrates the ability of SVR to neutrally observe and interpret factors affecting faculty performance in a university-based HR cell setting. The suggested techniques will be widely adopted, particularly for their significant implications in faculty recruitment processes within the HR domain.

Key Words: Machine Learning; Support Vector Regression; Faculty Selection; Performance-Based-Appraisal System; Behavioral Competency Assessment

Introduction:

In today's competitive market, there's a growing demand for tools that help employers find the perfect candidates for open positions. AI-driven technologies streamline recruiters' tasks and enhance the efficiency of the hiring process by replacing traditional, time-consuming methods [1-3]. This paper seeks to recognize and investigate how machine learning (ML) and AI are used to supplement human interventions by a more accurate 'Support Vector Regression' in computing complicated scores of a sustainable competency-based interview process in a large rural university located in the northeast corner of India.

ICFAI University Tripura (IUT), a State Private University, is a large university in Tripura offering 54 Programs in different areas. Its courses range from Science, Engineering, Management, Commerce, Humanities and Social Sciences, Law, Education, Special Education, Allied Health Science, Nursing, Yoga and Physical Education. Its full-time program cater to 6,500 students, starting from undergraduate to doctoral levels, and have 189 full-time faculty members on its payroll. A university running several programs in a rural area of northeastern India needs to work hard to attract and retain high-quality talent. Moreover, the university continuously expands by teaching several courses and increasing the number of programs. The university maintains offices in Hyderabad, Kolkata, Guwahati,Silchar,Manipur and Siliguri and it operates in a rural area; many recruitment functions are initiated in these offices and finalized with the HR Department located at the University Premises at Tripura.

In India, the University Grants Commission (UGC), through its regulations, has enforced the appointment of qualified and experienced faculty members in institutions of higher learning. The performance-based appraisal system (PBAS), recommended by the UGC, serves as a tool for recruitment. In the self-assessment mode, a candidate applying for a teaching position has to fill up the Employment Application Form and send

it to the university's HR department. This score, popularly called API Score, takes the calculation of the total score into account once the behavioral competency-based interview of 40 marks is over. API Score is given 60 percent weightage in the selection process.
The concept of behavioral competency for academia is highlighted as a faculty's inclination to such behaviors, which may help the teacher become effective in class delivery and impart lessons to the students. University authorities identified ten competencies in general required for doing the task of faculty of an HEI. The ten competencies are communication skills, collaboration ability, ethical conduct, mentoring skills, analytical skills, problem-solving ability, multiskilling ability, helping attitude, institution-building attitude, and willingness to disseminate learning to a larger population.
The competencies identified by the authority and used in interviews through questions aimed at understanding such competencies have been evaluated after a year for their reliability based on job performance and retention. In 2023, out of 42 recruits, 40 remained in the job, and 94% have been rated outstanding by their students. It indicates the reliability of the existing selection procedure, and the university authority has decided to continue with the system.
The HR Department of the University was in a state of inconvenience in computing the ratings manually due to lots of criteria. Moreover, often several interviewers evaluate, and each gives their rating. The problem arose when manual calculation caused mistakes and errors in decision-making.

Literature Review:

Arthur Samuel first coined the term "Machine Learning" in 1952. Following this, in 1957, Frank Rosenblatt combined Donald Hebb's brain cell interaction model with Samuel's work, making the perceptron at the Cornell Aeronautical Laboratory. In 1967, the nearest neighbour algorithm was designed, keeping the early stages of essential pattern recognition. This algorithm found applications in mapping routes and translating the travelling salesperson's issue by specifying the most efficient route. The 1960s saw a paramount development with the discovery and utilization of multilayer's in neural networks, revealing that incorporating two or more layers in the perceptron significantly improved processing power compared to a single-layer perceptron [4-7].

Russel and Norvig (2003) claim that machine learning originated from pursuing artificial intelligence [8]. In the initial stages of artificial intelligence as an academic discipline, researchers were willing to enable machines to learn from data. The emergence of machine learning as a specific field gained acceleration predominantly during the 1990s. The field shifted its objectives, driving from the ambitious goal of achieving artificial intelligence to managing functional and solvable problems. This shift involved a departure from the symbolic methods inherited from AI towards adopting strategies and models derived from statistics and probability theory.Shalev-Shwartz and Ben-David (2014) defined *machine learning* as automatically identifying meaningful patterns within data [9]. They followed that over recent decades, it has evolved into a ubiquitous tool for tasks involving extracting knowledge from extensive datasets. Russell (2014) defined machine learning as a component of AI that allows systems to improve their performance based on experience rather than relying solely on explicit programming.Asongo et al.(2021) viewed that vital findings and decisions can be arrived at without human intervention using algorithm techniques [8].

Research Question:

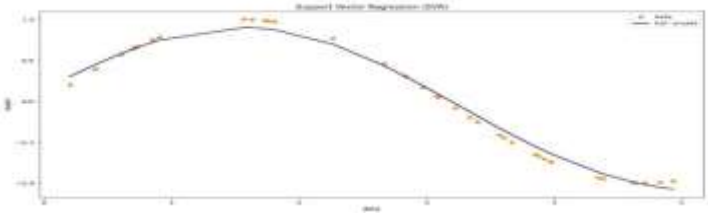
The question arises:
How application of Machine Learning (ML) can be helpful in undertaking the complex computation exercise quickly and without error?
What application of ML is available for doing these exercises?
How are such applications of ML applied in these circumstances?
The study aims to find how machine learning technique assesses the profile of interviewing candidates based on numbers.

Support Vector Regression (SVR):

SVR is a machine learning method that extends Support Vector Machines (SVM) principles to regression studies. In SVR, the preliminary objective is to expect continuous outcomes instead of class labels, which is typical in classification assignments [11]. The notion of SVR revolves around recognizing a hyperplane that effectively represents the relationship between input variables and the interconnected continuous output [12]. Support vectors, pivotal data points that affect the placement and orientation of the hyperplane, play a crucial role in this strategy. SVR tries to find the best fitting line (hyperplane) that minimizes overall error while staying within this margin of patience for most of the data points.

Instead of directly fitting a line through the data in a two-dimensional space, SVR utilizes a kernel function to transform the data into a more complex, high-dimensional space. This transformation allows for a simpler representation of the relationship between the input and output variables with a straight line (hyperplane)

[13]. However, SVR prioritizes minimizing overall error while acknowledging a margin of error (epsilon) around the hyperplane. This flexibility accounts for potential noise or outliers in real-world data. Slack variables are introduced to manage data points that fall outside this margin.While typically applied in higher dimensions, this 2D example showcases how SVR finds the best fitting hyperplane to capture the underlying trend in the data, balancing accuracy with a tolerance for error.



(a)

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.svm import SVR

# Generate sample data
X = np.sort(5 * np.random.rand(40, 1), axis=0)
y = np.sin(X).ravel()

# Fit SVR model
svr_rbf = SVR(kernel='rbf', C=100, gamma=0.1, epsilon=0.1)
svr_rbf.fit(X, y)

# Plot the SVR results
plt.scatter(X, y, color='darkorange', label='data')
plt.plot(X, svr_rbf.predict(X), color='navy', lw=2, label='RBF model')
plt.xlabel('data')
plt.ylabel('target')
plt.title('Support Vector Regression (SVR)')
plt.legend()
plt.show()
```

(b)

Fig 1: (a) SVR Graph (b) SVR Graph generation Python code

This illustration uses a straightforward sinusoidal dataset and an SVR model employing a radial basis function (RBF) kernel.
The line svr_rbf.Predict(X) represents the formulated regression function based on the model. Visualizing more intricate datasets with numerous components poses challenges in a two-dimensional space, often demanding techniques like dimensionality reduction or plotting data slices. Tailoring the visualization technique according to the specific dataset and the number of features involved is crucial. Also, changing parameter values such as C, gamma, and epsilon in the SVR model allows for exploring their impact on the regression outcome.

Methodology:

To optimize faculty recruitment, this approach utilizes SVR, a type of artificial intelligence. By analyzing historical data on candidate qualifications, interview scores, past HR cell structures, and hiring outcomes, the system can predict the most effective HR cell configuration for future hires. After collecting and refining this data, an SVR model is trained and evaluated to ensure its accuracy. Once optimized, the model integrates with the HR system, providing real-time recommendations for optimal HR cell configurations throughout the faculty recruitment process.

Experiment:

This study explores how well faculty ratings can be predicted across different fields using Support Vector Regression (SVR). They considered various factors like teaching experience, student assessment scores, and even faculty behavior. After carefully cleaning and selecting the most relevant data points, they built an SVR model to categorize faculty ratings based on pre-defined quality levels. The next steps of the research are explained in more detail in the following parts.

Data Loading:

Load data from a CSV file using the pandas library.

Feature Extraction:

Extract relevant features (X) from the dataset, including 'UG', 'PG', 'Gate/NET"Exp';' No. of Research Article', 'Ph.D.' 'Communication', 'Collaboration' and Ethical Conduct, and 'Mentoring'

Target Variable Calculation:

Calculate the target variable (y_actual) by applying the calculate_rating function to each row of the dataset. The function assigns scores based on specified conditions for each feature, calculates the total score, and is divided by 20 to obtain the rating.

Feature Scaling:
Scale the features (X) using MinMaxScaler from sci-kit-learn to ensure that all features contribute equally to the model.

SVR Model Creation:
Create an SVR (Support Vector Regression) model using scikit-learn's SVR class with a linear kernel (kernel='linear').

Model Training:
Train the SVR model using the scaled features (X_scaled) and the target variable (y_actual).

Prediction:
Make predictions (y_predicted_svr) for the entire dataset using the trained SVR model and the scaled features.

Rounding Predictions:
Round the predicted ratings to the nearest integer in the case, from 1-5.

Output Display:
Display the relevant columns ('Name' 'UG' 'PG' 'Exp' 'Research Article' 'Ph.D.' 'Communication' 'Collaboration' 'Ethical-Conduct' 'Mentoring' 'Predicted Ratings') of the DataFrame containing both original and predicted ratings.
We used the Support Vector Regression (SVR) framework to estimate the ratings of university faculty members across diverse domains. Metrics such as teaching experience, assessment results, and behavioural study were examined. The details of these procedures are outlined in Algorithm 1.

Algorithm 1: Faculty Rating

```
# Step 1: Data Loading
data = load_data_from_csv("file_path")  # Load data from CSV file

# Step 2: Feature Extraction
X = extract_features(data)

# Step 3: Target Variable Calculation
y_actual = calculate_ratings(data)

# Step 4: Feature Scaling
X_scaled = scale_features(X)

# Step 5: SVR Model Creation
svr_model = create_svr_model()

# Step 6: Model Training
train_svr_model(svr_model, X_scaled, y_actual)

# Step 7: Prediction
y_predicted_svr = predict_ratings(svr_model, X_scaled)

# Step 8: Rounding Predictions
rounded_ratings = round_predictions(y_predicted_svr)

# Step 9: Output Display
display_results(data, rounded_ratings)
```

SVR training was conducted by utilizing the following X and Y parameters for calculation purposes.

X PARAMETERS (Features):
Undergraduate (UG) Marks (ug_marks):

- The academic performance of a teacher during their undergraduate studies indicates their foundational knowledge in the subject area.

Postgraduate (PG) Marks (pg_marks):

- Like UG marks, PG marks reflect the teacher's advanced academic achievements and expertise in their field.

Gate Score (gate):

- GATE (Graduate Aptitude Test in Engineering) evaluates a broad knowledge of engineering and science disciplines through a comprehensive examination.A good GATE score can reflect a teacher's proficiency in their domain.

Experience (exp):

- Teaching experience is often considered valuable as it indicates the teacher's exposure to diverse students and teaching methodologies, contributing to their overall effectiveness.

Research Articles (research_article):

- Published research articles can showcase a teacher's engagement in scholarly activities, contributing to the academic community and indicating a commitment to staying current in their field.

Ph.D. (PhD):

- Holding a Ph.D. degree indicates a high level of expertise and specialization in a particular subject, which can positively influence teaching quality.

Communication Skills (communication):

- Effective communication is crucial for teaching. This parameter may include the ability to convey complex concepts clearly and understandably.

Collaboration Skills (collaboration):

- Collaboration is essential in academic environments. Teachers often work with colleagues, students, and external partners. Strong collaboration skills can contribute to a positive learning environment.

Ethical-Conduct (ethical_conduct):

- Ethical conduct is fundamental in an educational setting. Teachers are expected to adhere to ethical standards in their interactions with students, colleagues, and the academic community.

Mentoring Skills (mentor):

- Mentoring skills are valuable, especially in higher education. Teachers may guide students and junior faculty, impacting their professional development.

Analytical Skills:

It demonstrates the capacity to understand and interpret data. This refers to the skill to analyze information, interpret theories, and develop appropriate and efficient strategies to address situations.

Coordination-Skills:

It is the ability to understand one's role and how it relates to others. It is the ability to deal effectively with all the students to ensure the event is conducted smoothly. For co-ordination, it is essential that the teacher cooperates with others and can work. It means bringing various efforts into a cohesive whole by embracing approaches and methods, regardless of personal opinion, and supporting consensus decisions.

Multi-Tasking:

It refers to the competency to manage two or more tasks within the same time structure, having similar deadlines effectively and smoothly by facilitating the process flow and completion of all tasks equally well and by following a disciplined schedule for each task efficiently.

Problem Solving:

It involves systematically gathering pertinent information and delving into the core of a problem or situation. This entails identifying, scrutinizing, and analyzing circumstances to generate and evaluate various solutions creatively. The selection of alternatives should be reached through a consensus-driven method, and judgment should be employed to determine the most advantageous option.

Y PARAMETER (Target Variable):

Predicted-Rating (rate_score):

- The calculated rating score represents an aggregate measure of a teacher's qualifications, experience, and various skills. It serves as the target variable for the SVR model, aiming to predict overall teacher ratings based on the specified criteria.
- Feature Extraction:
- Extracting the feature values (ug_marks, pg_marks, gate, exp, etc.) from the input row- each feature represents a specific aspect of a teacher's qualifications or skills.
- Assigning Scores for UG and PG Marks:
- Based on a predefined scoring system, assigning scores (ug_value and pg_value) to UG and PG marks. This system categorizes the marks into different ranges and assigns corresponding scores.
- Calculating Total Score:
- To sum up the score- UG and PG marks, along with other features such as experience, research articles, Ph.D., communication skills, collaboration skills, ethical conduct, mentoring skills, and GATE score. This results in the total score representing the overall qualifications and skills of the teacher.
- Calculating Rating Score:

- Dividing the total score by 20 to obtain a rating score (rate_score). This step scales the total score to a reasonable range, making it more interpretable.
- Assigning to y_actual:
- Assigning the calculated rating score (rate_score) to the variable y_actual. This variable represents the actual rating score for a teacher based on the specified criteria.
- Initially, to recognize machine learning models in a way that the model parameters encompass details about the training data sample. The outcomes of the rating are depicted in Figure 2.

(a)

(b) Fig: (a) Samples for rating generation (b) outcomes of the rating

Discussion:

In general, the research responds to the following queries: (1) the application of Artificial intelligence at different phases of the hiring process. (2) Views of stakeholders regarding the use of AI in hiring and (3) Advice on how managers can use AI in hiring. In general, the research paper contributes regardingapplication of machine learning in recruitmentas well asrecommendation of AI for recruitment process in other universities.

Machine learning and the use of Support Vector Regression help to identify the right candidate. It has reduced the chances of error and has added authenticity to the process, apart from computational timewhich has also been reduced to a great extent. In this mode, machine learning algorithms operate under supervision. In supervised learning, the objective is to comprehend the association between input and output data. The input incorporates information conveying individual objects, often named instances or examples, while the output conveys outcomes provided by a supervisor. Classification, a type of supervised learning, concerns creating a mapping or discriminant function to differentiate between mixed-class instances.In machine learning, classes are determined by their output, represented as the class label. The discriminant function is generally comprehended as a classifier or model. A training set is a collection of instances paired with their corresponding class labels. In the classification process, a model is represented by a set of fine-tuned parameters to establish a mapping from samples in the training set to their respective labels. The trained model is, after that, employed to classify or label new instances that have not been previously encountered.

Most practical applications of machine learning mainly depend on supervised learning. In this learning process, there are input variables (x) as well as an output variable (Y), and an algorithm is used to understand and learn the relationship between the input and output through a mapping function.

$$Y = f(X) \quad (1)$$

The goal is to approximate the mapping function in such a way that the algorithm can accurately predict the corresponding output variables (Y) when provided with new input data (X). This process is termed supervised learning because the algorithm's learning method can be likened to a teacher supervising a pupil. In this system, the training dataset serves as a collection of accurate responses, and the algorithm constantly predicts effects on this data, obtaining guidance from the "teacher." The learning repeats until the algorithm reaches a satisfactory level of proficiency.

Application of AI in the Study:

It's important to highlight the application of AI in this context. AI significantly contributes to setting up an SVR-based HR Cell for faculty recruitment by effectively managing and analyzing vast and complex datasets. This enables the discovery of hidden patterns and insights that might be overlooked by human examination. Moreover, AI powered models such as SVR have capability to evolve and refine their predictions with new data, enhancing their forecasting precision and fine tuning HR Strategies in an ever changing recruitment landscape.

We found that incorporating AI into an SVR-based HR system for faculty recruitment can streamline and make decision-making more objective, reducing human bias and improving the accuracy of candidate assessments. Furthermore, AI-enhanced SVR models can adapt to changing recruitment trends, enhancing prediction accuracy and optimizing HR strategies for better faculty hiring outcomes.
Theoretical Contributions: Theoretical progress in setting up an SVR-based HR system for faculty recruitment involves the advancement of using machine learning regression techniques in HR decision-making, particularly in enhancing faculty hiring strategies. Additionally, this approach contributes to a deeper understanding of leveraging predictive analytics to improve efficiency and effectiveness in human resources, paving the way for greater adoption of data-driven recruitment methods in the academic field.
Managerial Implications: Incorporating an SVR-based HR system into faculty recruitment can streamline decision-making, enabling HR managers to make hiring decisions based on data-driven insights, thereby enhancing the efficiency and success of the recruitment process.

Limitation and Scope for Further Research:

A potential drawback of using an SVR-based HR Cell for faculty recruitment is the risk of over fitting, especially if the model is overly intricate or trained on a restricted data set, possibly limiting its applicability to diverse recruitment situation.
Future studies might explore how different SVR Parameters and Kernel types influence the effectiveness and flexibility of HR Cell configurations in faculty recruitment, with the goal of enhancing model performance across a range of recruitment situations.

Conclusion:

Artificial intelligence technology has made recruitment process more efficient and made recruiters' daily tasks easier. Moreover, AI is crucial in every step of the recruitment process, starting from applying online followed by screening, assessment, selection to promotion.Even though recruitment is done through Artificial Intelligence, recruitment stakeholders should be concerned regarding the following factors: whether there was any bias or not in the selection procedure and timely replacement of recruiter's panel. We did the study to simplify the process of computing scores of the Interview Assessment Sheet. Detailed evaluation of candidates for faculty recruitment is required, and the university authority has adopted a matched approach of self-appraisal through -PBAS designed by UGC and competency assessment through its structured analysis form. The calculation and totalling of scores are complicated and often lead to human error. It has resulted in the wrong decision. The use of the ML technique in the form of Support Vector Regression has helped to create a precise result apart from saving time as it would have otherwise occurred for manual calculation.
Future studies could explore AI recruitment's effectiveness at group or executive levels and compare the performance of employees hired through AI tools versus traditional methods. Additionally, research could focus on developing guidelines and training materials to aid users collaborating with AI developers and assessing candidates' perceptions of AI recruitment across different job roles and positions.

References

1. Chen, Z. (2022). Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. Cognition, Technology & Work, 25(1), 135–149. https://doi.org/10.1007/s10111-022-00716-0
2. Samin, H., &Azim, T. (2019). Knowledge Based Recommender System for Academia Using Machine Learning: A Case Study on Higher Education Landscape of Pakistan. IEEE Access, 7, 67081–67093. https://doi.org/10.1109/access.2019.2912012
3. Habeeb Ur Rahiman, &RashmiKodikal. (2023). Revolutionizing education: Artificial intelligence empowered learning in higher education. Cogent Education, 11(1). https://doi.org/10.1080/2331186x.2023.2293431
4. Giddings, R., Joseph, A., Callender, T., Janes, S. M., van, Sheringham, J., &Navani, N. (2024). Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review. The Lancet Digital Health, 6(2), e131–e144. https://doi.org/10.1016/s2589-7500(23)00241-8
5. Southworth, J., Migliaccio, K., Glover, J., Glover, J., Reed, D., McCarty, C., Brendemuhl, J., & Thomas, A. (2023). Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy. Computers and Education: Artificial Intelligence, 4, 100127. https://doi.org/10.1016/j.caeai.2023.100127
6. Abdelwahab, H. R., Rauf, A., & Chen, D. (2022). Business students' perceptions of Dutch higher educational institutions in preparing them for artificial intelligence work environments. Industry and Higher Education, 37(1), 095042222210876. https://doi.org/10.1177/09504222221087614

7. Dempere, J. M., Modugu, K. P., Hesham, A., &Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. Frontiers in Education, 8. https://doi.org/10.3389/feduc.2023.1206936
8. Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 38(11), 5391–5420. https://doi.org/10.1002/hbm.23730
9. ShaiShalev-Shwartz, &Shai Ben-David. (2015). Understanding machine learning : from theory to algorithms. Cambridge University Press.
10. Russell, S., &Norvig, P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
11. Casuat, C. D. (2020). Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System. International Journal of Emerging Trends in Engineering Research, 8(5), 2101–2106. https://doi.org/10.30534/ijeter/2020/102852020
12. Kavitha S, Varuna S, &Ramya R. (2016). A comparative analysis on linear regression and support vector regression. 2016 Online International Conference on Green Engineering and Technologies (IC-GET). https://doi.org/10.1109/get.2016.7916627
13. Mangasarian, O. L., &Musicant, D. R. (2000). Robust linear and support vector regression. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(9), 950–955. https://doi.org/10.1109/34.877518
14. Al-Alawi, A. I., Naureen, M., AlAlawi, E. I., &Naser Al-Hadad, A. A. (2021). The Role of Artificial Intelligence in Recruitment Process Decision-Making. 2021 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/dasa53625.2021.9682320
15. Albert, E. T. (2019). AI in Talent acquisition: a Review of AI-applications Used in Recruitment and Selection. Strategic HR Review, 18(5), 215–221. https://doi.org/10.1108/shr-04-2019-0024