IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4334~4343
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4334-4343

Eligibility rate of applicant's LinkedIn account: a naïve bayes classification and visualization

Khyrina Airin Fariza Abu Samah[1], Nurul Athirah Ahmad[1], Anis Amilah Shari[1], Hana Fakhira Almarzuki[2], Zuhri Arafah[1], Lala Septem Riza[3], Amir Haikal Abdul Halim[1]
[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Melaka Branch, Melaka, Malaysia
[2]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia
[3]Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info

Article history:

Received Dec 22, 2023
Revised Mar 13, 2024
Accepted Mar 21, 2024

Keywords:

Classification
LinkedIn
Naïve Bayes
Recruitment
Visualization

ABSTRACT

In the digital era, social media platforms like LinkedIn have become famous for recruitment, and recruiters widely use them to find potential employees. The recruitment process is crucial in organizations, as it involves selecting qualified applicants from a diverse pool. However, the screening process and manual recruitment process entail significant time, high costs, and potential bias. Consequently, it may cause recruiting unqualified applicants and may affect the organizations. Thus, this study aims to classify and generate a list of potential job applicants by analyzing seven attributes of their LinkedIn accounts: title, location, skills, education, language, certification, and years of experience. Data are collected from LinkedIn profiles and then undergo data pre-processing. The naive Bayes (NB) algorithm is implemented as the classification algorithm and sets the classification as "eligible" or "ineligible". The NB model achieved an accuracy testing of 89.8%, indicating good performance in classifying potential job applicants. At the same time, we measure the similarity cosine score to set the mean of the eligibility. The classification results are visualized for the suitable applicants in descending rank, allowing users to choose the applicants' classification status efficiently. For the system usability, we managed to get 90% from the recruitment expert.

Corresponding Author:

Khyrina Airin Fariza Abu Samah
College of Computing, Informatics and Mathematics
Universiti Teknologi MARA Melaka Branch, Jasin Campus, Merlimau, 77300, Melaka, Malaysia
Email: khyrina783@uitm.edu.my

1. INTRODUCTION

Recruitment is a crucial aspect of the human resources (HR) Department as it involves selecting eligible candidates from a vast applicant pool [1]. Employing a variety of strategies is necessary to locate, interview, and recruit individuals for the position. The HR Department's initial task is recruitment to ensure each employee is competitive and contributes to society [2]. Interviewers aim to identify the most suitable candidate who fulfils the employment requirements during this process. Abbas *et al.* [3] emphasized the significance of selecting the appropriate candidate in company operations, stating that competent employees can significantly impact the success or failure of an organization. Lawong *et al.* [4] stated that both the organization and the agents share responsibility for the effectiveness of the hiring process. Recruiters play a crucial role in an organization's performance by implementing effective hiring and recruiting tactics to attract qualified and competent individuals. They accomplish this by conducting research, designing, and implementing these tactics.

In modern times, the world has undergone a digital transformation [5], including social media platforms such as Facebook, LinkedIn, and Twitter, which are more popular for recruitment processes [6]. LinkedIn is the primary site for companies to recruit applicants and emphasizes cultivating professional connections [7], [8] with a usage rate of 77% compared to other platforms [9]. It is the biggest online networking site for professionals, linking more than 900 million individuals in over 200 countries. It enables users to discover job openings, broaden their professional connections, and acquire new skills to achieve success in their careers. According to Wei [10], Malaysia has around 5.79 million LinkedIn users, and numerous recruiters prefer it for recruitment purposes [11].

One step in the recruitment process is the screening of applications [12]. It involves evaluating job candidates to assess their suitability for a position. However, Sivanandam and Mudaliar [13] highlighted that recruiters find it challenging to go through numerous resumes. Examining the applicant's resume is a time-consuming process, leading to delays and ineffective time management [14]. Recruiters must verify and evaluate the minimal credentials for the job to guarantee a successful recruitment process and make an informed selection. Abbas *et al*. [3] stated that a screening process that is not effective could result in generating a roster of inadequately qualified candidates.

Moreover, a manual recruitment method necessitates substantial expenses [15], [16]. Costs and expenses related to recruitment must be considered, such as time to hire, resume screening, and recruiter fees. Furthermore, the expenses would be at their highest if unqualified candidates were selected for the position. Utilizing inexpensive recruitment processes may result in erroneous shortlisted candidates as they do not always ensure the most eligible applications [17]. A manual recruitment procedure might lead to biased outcomes influenced by gender or human perception, affecting the decision-making in the recruitment process [18], [19]. Recruitment bias occurs when the recruiter assesses the applicant only based on their initial impression. It is influenced by human perception, making individuals more inclined to favour a resume with an appealing profile image.

Eight high-demand vocations in Malaysia for 2022 have been determined based on current trends and industry estimates [20]. The list comprises information technology (IT), software development, digital marketing, finance positions, project management roles, business development and sales executives, medical professionals, educators, and customer service executives. IT and software development, digital marketing, and finance jobs are the top three jobs identified by JobStreet and chosen for this research. Job seekers should prioritize reviewing the job requirements information that can provide insight into the specific job requirements for recruiters [21], [22]. Attributes listed in job listings are crucial for recruiting appropriate candidates. Researchers identified 10 key factors for recruiters to consider when evaluating a job opportunity: position, skills, education level and history, languages, years of experience, certification, salary, benefits, location, and working hours [23], [24]. However, this research focuses on seven attributes for classification: title, location, education, years of professional experience, skills, languages, and certificates [25], [26].

A web-based dashboard was created utilizing data extracted from LinkedIn profiles that were scraped from the platform in response to the identified issues. The naïve Bayes (NB) algorithm was utilized to classify and visualize the LinkedIn accounts of applicants who meet the company's job requirements. The system utilizes bar charts and pie charts for visualization. It allows users to see which applications from their LinkedIn accounts meet the requirements for the job. It assists recruiters in identifying the most suitable candidate who meets the job's criteria. The paper is structured as follows: section 1 commences with a concise introduction. Section 2 details the approach, while section 3 presents the results and discussion. Section 4 ends the study and offers a brief review of potential future improvements.

2. RESEARCH METHOD

2.1. Design of the system

System design in research involves developing a framework or structure to investigate and solve a specific research issue. We deliberated on the comprehensive system architecture, system flow, interfaces, and the data pertaining to the system needs. During this stage, we utilize a use case diagram and a flowchart diagram to illustrate the workflow. User interface (UI) is the ultimate stage in the design process. The term pertains to the visual arrangement of the system components that a user can engage with on a website. UI design should be effective and user-centric to guarantee user-friendliness and appeal to potential users. It strives to streamline the user's interaction to efficiently achieve their goals within the system.

2.2. Back-end development

Back-end development includes gathering, preparing, and applying the NB model, as seen in Figure 1. The diagram illustrates the system's operation from data collecting initiation to visualization completion. Data preparation includes collecting and pre-processing data to guarantee its reliability, and then using an NB classifier for categorization. The output consists of both visualized and reported data.

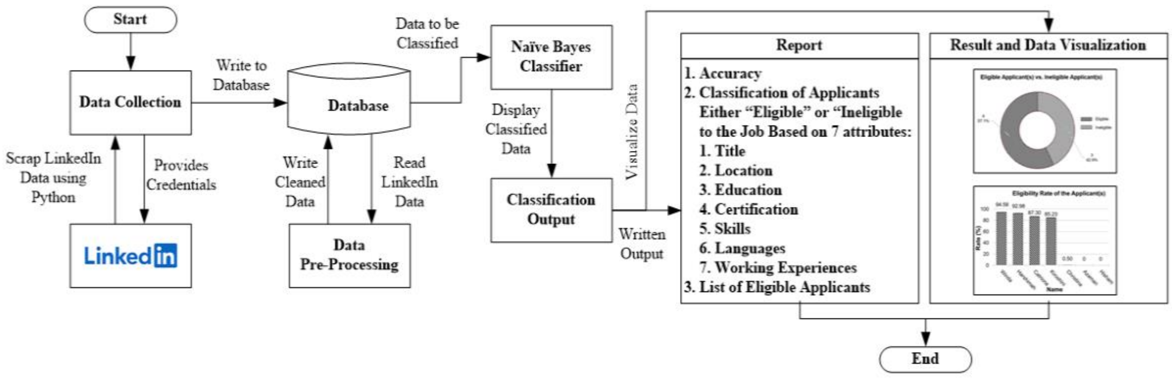

Figure 1. Research design flow diagram for LinkedIn eligibility

2.2.1. Data collection and pre-processing

Data was collected by web scraping from September 1, 2022, to March 31, 2023, utilizing the selenium and beautiful soup packages in Python. Selenium is a potent instrument for managing web browsers and executing browser automation [27]. This study involves using selenium to automate the LinkedIn page to extract applicant data. Subsequently, beautiful soup was utilized to retrieve the LinkedIn profile information. The scraper navigates through all profile pages of the corresponding URLs on the LinkedIn website. The tool retrieves LinkedIn profile information using specific parameters, including title, location, education, experiences, certifications, skills, and languages. The extracted data was saved in a CSV format, and 2006 profiles were gathered for this investigation using scraping.

The data that was gathered was then cleaned up to make high-quality data for testing and training. It was processed to remove symbols, fill in missing numbers, and lowercase all the letters in the dataset. It was changed to the word "none" in the information where the empty list with the symbol "[]" used to be. After that, a similarity score method was used to figure out how similar the applicant's personal data was to the job description data. The cosine similarity function was used to figure out the closeness score. This study looks at titles, skills, and languages and gets the similarity score between them. The equation for cosine similarity score is represented as in (1) [28].

$$Cosine\ Similarity = \frac{(A \cdot B)}{(||A|| \cdot ||B||)} \tag{1}$$

Next, for education attributes, the education level was extracted using regex to match the education level in that job's requirements. If a match was found, it was assigned a value of "1" to indicate a successful match in terms of education level. Afterwards, the duration of the applicant's LinkedIn profile experiences was computed in months to facilitate better comparisons with the minimum duration of experiences required by each job's specifications.

After all the attributes were processed, a few conditions were applied to the data labelled. A new column called "eligibility status" has been added to the dataset. These eligibility status attributes have "eligible" and "ineligible" statuses. The LinkedIn profile was classified as having "Eligibility" status if it fulfilled these three conditions. First, the similarity score between title, skills and languages is higher than the mean of the similarity score between title, skills, and languages. Function mean() calculated the average of the total similarity score for the applicants and was commonly calculated as in (2). The second condition is if the education level that was extracted earlier matches the education level required by the job's description. The last condition is that the duration of the LinkedIn profile experiences in a month should be higher than the duration of the job experiences in a month. Otherwise, if the data does not meet those conditions, it will fall under the "ineligible" category.

$$Mean = \frac{(Cosine\_Similarity(A1,B1) + Cosine\_Similarity(A2,B2) + \ldots + Cosine\_Similarity(AN,BN))}{N} \tag{2}$$

The final dataset has a total of 14 columns for LinkedIn profile: name, title, location, experiences, education, certifications, skills, languages, similarityscores_title_skills_lang, education level, education match, total duration in months, has certification, and eligibility. Finally, the dataset is saved in a CSV file for the next training and testing to develop the model and used for data visualization. Once the pre-processing phase is finished, the final dataset is prepared for the classification method using machine learning.

2.2.2. Naïve Bayes classification model

The NB classifier is a widely used algorithm for classification and prediction analysis [29]–[31]. NB classifier is performed based on the Bayes theorem. It assumes that the features are conditionally independent given the class label. The NB classifier calculates the probability of each class label given the input features. It assumes that the features within the input features are conditionally independent given the class. By leveraging this assumption, the NB algorithm can make predictions accurately and quickly. Bayes' theorem calculates the probability of an event happening based on the probability of another event that has already occurred. According to Brownlee [32], the Bayes theorem was depicted as in (3).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

During training data, the NB classifier calculates the prior probability, which is P(A) from (3) of each class, by counting the number of instances. It also estimates the conditional probabilities by calculating the likelihood, P(B | A), of observing each feature value given the class label. Thus, during testing data, the Bayes theorem was applied to calculate the posterior probabilities of each class given the input features, which is P(A | B) from the equation. The class with the greatest posterior probability is subsequently chosen as the predicted class. The dataset is labelled based on its conditions, where we use a ratio of 80:20 to split the data into 2 parts: training and testing. Four feature columns were selected for training sets, such as similarity scores title skills lang, education match, total duration in months, and has certification. After developing the NB model, the eligibility rate was produced using Bayes's theorem probability.

2.3. Front-end development

HTML, CSS, and JavaScript are three core technologies for developing and designing websites and web applications. A website has been created for users to interact with, and the system can be accessed easily. The web application connects the interface and the Python code using the Flask application. It displays the result of the prediction and offers tools for data visualization, allowing the creation of charts and graphs to represent data visually.

2.4. Testing development

Upon completion of all phases, the final phase is testing, which involves executing test cases on the system. The verification and validation process ensures that the system fulfils the initial requirements and satisfies its aims. Verification evaluates the system's development product, whereas validation assesses the final product. The testing phase is intended to ensure the system's functionality and address any problems or concerns. In this study, the objectives focus on dashboard functionality based on the defined use case and system usability scale (SUS) by HR recruitment professionals.

3. RESULTS AND DISCUSSION

3.1. Model accuracy

Evaluating the NB classifier model involves comparing the predicted values with the actual values in a test dataset to determine accuracy. Figure 2 displays the accuracy testing results of the NB model in this study. The accuracy result is 89.8%. An accuracy score above 90% is considered "excellent" in [33], whereas a score between 70% and 90% is classified as "good." An accuracy score ranging from 60% to 70% is considered "okay," while a score below 60% is labelled as "poor." The system's accuracy ranges from 70% to 90%, which is considered "good." The model achieved an accuracy rate of about 89.8% in classifying events within the test dataset. The confusion matrix provides a detailed breakdown of the true values that align with the model's predictions. The confusion matrix in this study uses the classes "ineligible" and "eligible", which are denoted by "0" and "1" correspondingly.

3.2. Cosine similarity scores result

We use an example using the job's field, "finance," and the job's title, "accountant," to illustrate the result. We obscured the original names with black markings. Then we renamed them using a comparable alphabet to prevent any ethical concerns. We are deviating from the conventional alphabetical sequence because the visualization is arranged according to the descending rate outcome. Cosine similarity scores are computed using the titles, abilities, and languages of seven applicants. Applicants who exceed the average similarity score are deemed to have met one of the eligibility criteria when the rules are applied. Thus,

Figure 3 shows the bar chart of cosine similarity score comparisons where each bar represents an applicant's similarity score. A line with a dotted point represents the mean of the cosine similarity score of 0.26, which indicates the mean of the similarity score across all applicants. In conclusion, only four candidates are eligible since they surpass the mean based on the overall applicant's score.

```
Accuracy: 0.8980099502487562
Confusion matrix:
 [[289  40]
 [  1  72]]

True Positives (TP): 72
True Negatives (TN): 289
False Positives (FP): 40
False Negatives (FN): 1


              precision    recall  f1-score   support

           0       1.00      0.88      0.93       329
           1       0.64      0.99      0.78        73

    accuracy                           0.90       402
   macro avg       0.82      0.93      0.86       402
weighted avg       0.93      0.90      0.91       402
```

Figure 2. Naïve Bayes model accuracy results

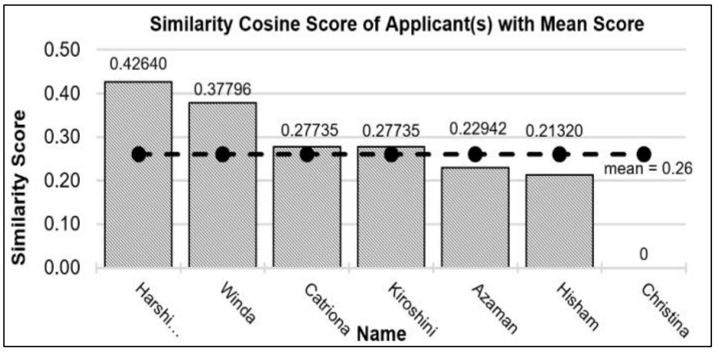

Figure 3. Bar graph of similarity cosine score of applicants with mean score

3.3. Functionality testing of dashboard visualization

The dashboard visualization is related to functionality testing. It is the process where the UI is tested by providing appropriate test input. The output is compared with the expected output. The system's output will be evaluated whether it is successful or not at the end of the test. This testing is mainly concerned with the processing results. The goals are to test and run through all the system functions from the first to the last page of the application. It ensures the application runs, follows the requirements, and displays the desired outcome. It performs a series of test cases covering the application's various functionalities. Thus, functionality testing helps identify bugs, errors, or deviations from the intended behaviour.

Figure 4 shows the "upload file" page where users can select the job's field and title and upload a CSV file containing a list of URLs of the LinkedIn applicant's account that the user wants to perform the classification. Using the same example for the similarity coefficient score, we chose the job's field as "finance" and the job's title as "accountant". Once the button "upload" is clicked, the scraping process will start, and the job requirements selected and the applicant's information will be displayed on the next page, as illustrated in Figure 5. On top of the dashboard, the job's requirement is listed according to the seven attributes chosen for this study. There is also information on the total of applicants that match the requirements.

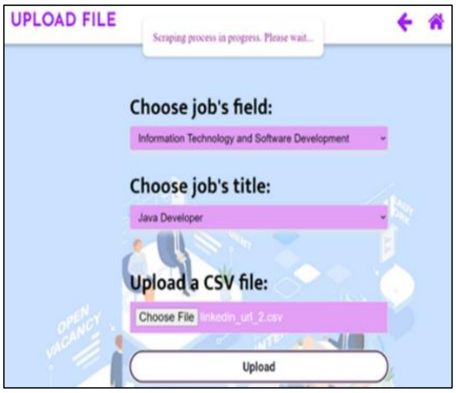

Figure 4. "Upload file" page

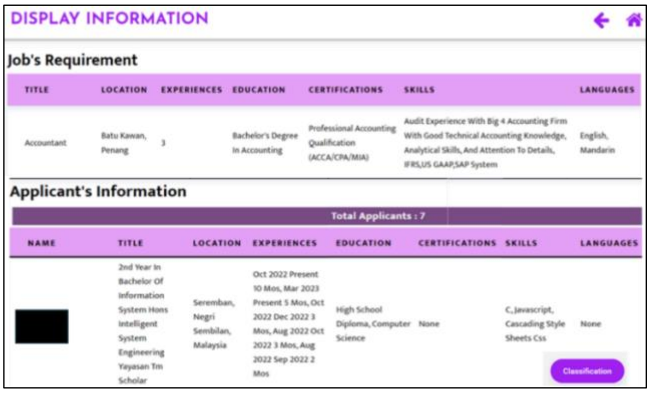

Figure 5. View the "display information" page

The system allows the user to click the "classification" button to show the applicant's information with its eligibility status, either "eligible" or "ineligible" on the right row, as depicted in Figure 6. The user may choose to click "download all applicants" to see the details. The system exhibits the eligible applicants along with their corresponding eligibility rates once the “view eligible applicants” button is selected as shown in Figure 7. At the top of the page, the dashboard visualizes the total number of eligible applicants, which is 4.

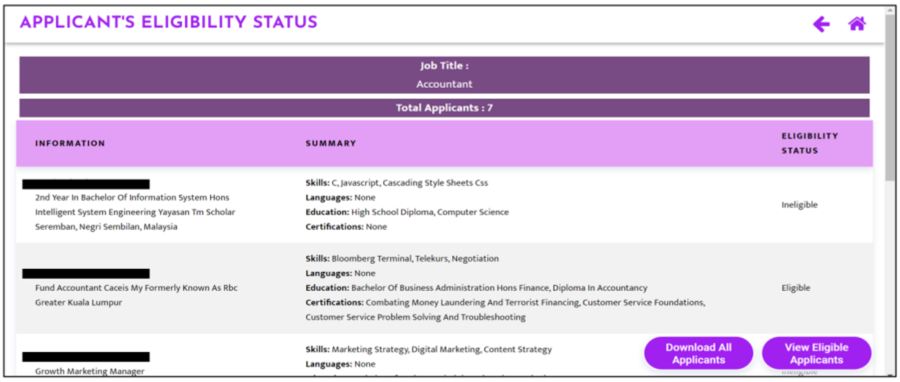

Figure 6. View the “applicant’s eligibility status” page

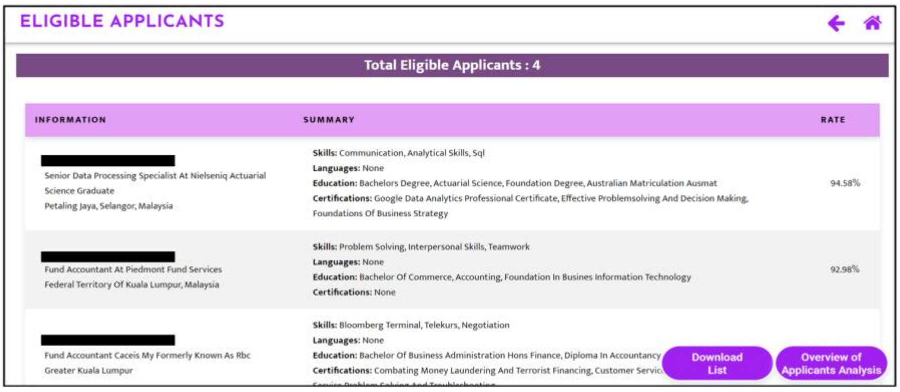

Figure 7. View the “eligible applicants” page

The user has the choice to click the “download list” button to refer to the detailed information. The analysis page was displayed after the user clicked the “overview of applicants analysis” button. Figure 8 shows a pie chart of the total eligible and ineligible applicants. There are four eligible applicants at 57.1% and three ineligible applicants at a percentage of 42.9%.

Following the pie chart, Figure 9 shows the bar chart of the eligibility rate of the applicants plotted in decreasing order to ease the analysis. The y-axis denotes the eligibility rate, whereas the x-axis signifies the applicant’s name. The chart is composed of bars, with each bar representing an applicant and its height corresponding to that applicant’s eligibility rate. With that, the user can quickly identify and analyze the top applicants with the highest eligibility rate.

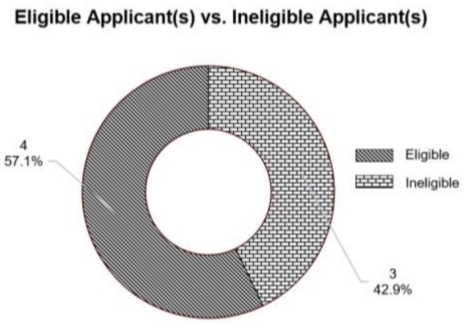

Figure 8. Pie chart of eligible applicant(s) vs. ineligible applicant(s)

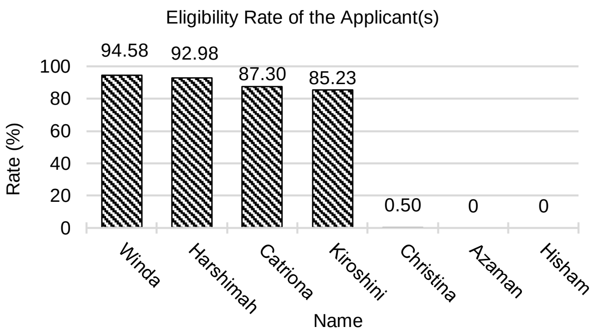

Figure 9. Bar chart of the eligibility rate of applicant(s)

The next page is the "attribute analysis" page. This page visualizes three proportions of total applicants: eligible applicants and ineligible applicants. For each proportion, it analyses seven attributes used for the classification in this research. There will be two visualizations in total to differentiate the classification.

The first analysis was visualized using a stacked bar chart, where each applicant's name and satisfaction status for each attribute were displayed. Figure 10 shows that out of four eligible candidates, Winda fulfils the attributes of certification, experience, and education. Catriona and Kiroshini fulfil attributes for certification and education, while Harshimah only fulfils education attributes.

The second analysis was visualized based on states in Malaysia, as in Figure 11. Figure 11(a) using the pie chart of the total applicants based on the geographical distribution of applicants across different states in Malaysia: WP Kuala Lumpur, WP Putrajaya, Selangor, and Negeri Sembilan. It gives an overview of which states have a higher number of applicants. Then, we split the pie chart in Figure 11(b) to determine which state the four eligible candidates are. As can be seen, there are two eligible candidates from Selangor and WP Kuala Lumpur. Figure 11(c) depicts the pie chart breakdown of the numbers of ineligible candidates that come from WP Kuala Lumpur, WP Putrajaya, and Negeri Sembilan.

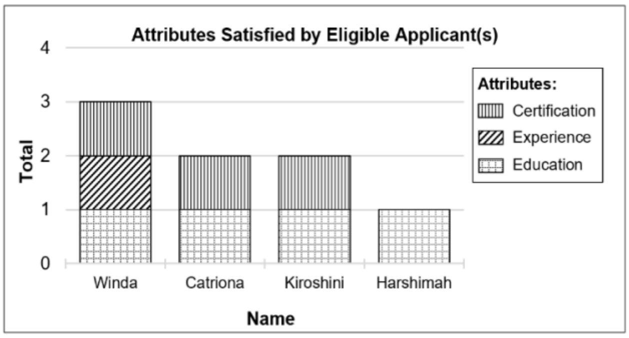

Figure 10. Stacked bar chart of attributes satisfied by eligible applicant(s)

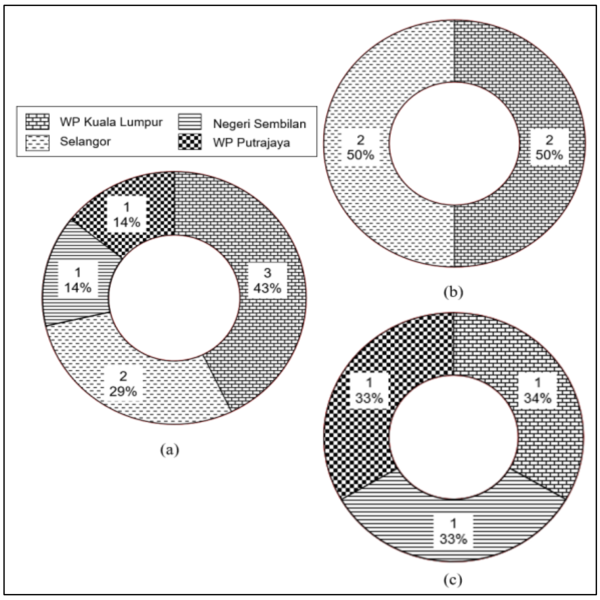

Figure 11. Pie chart of: (a) total applicants based on state distribution in Malaysia, (b) state of four eligible candidates, and (c) breakdown of the numbers of ineligible candidates

3.4. Usability testing

In calculating the SUS score, the responses provided by participants to the SUS questionnaire were used. For this study focus, we conducted interviews with three HR expert personnel who were available for the usability testing due to the time constraints available to complete the study. The questions were divided into two types of statements, where questions with odd numbers (1, 3, 5, 7, 9) are phrased positively. For this type, the respondent's response score from 1 to 5 was deducted by one. The questions with even numbers (2, 4, 6, 8, 10) are phrased negatively. Five points will be deducted from the response score for these questions. The points were tallied together once the adjustments were made for all the questions. The final score was multiplied by 2.5 to convert the total points into a SUS score on a scale of 0 to 100. This calculation method allows for a standardized assessment of the system's perceived usability, with higher SUS scores indicating better usability. Table 1 shows the findings of the SUS with the raw and final score. The respondents predominantly selected a scale of 4, representing "agree", for the positively phrased questions. Conversely, the respondents mainly selected a scale of 1 for the negatively phrased questions.

Table 1. SUS scores findings

Question No.	1	2	3	4	5	6	7	8	9	10	SUS raw score	SUS final score
Expert A	4	1	5	1	4	2	4	1	4	1	35	87.50
Expert B	5	2	4	1	4	1	5	2	5	1	36	90.00
Expert C	4	1	4	1	5	1	5	1	4	1	37	92.50
											Average	90.00

The overall percentage for SUS in this study is 90%, which indicates that the system is considered good. According to Misdan *et al.* [33], the average SUS score is 68, and scores above 85 are associated with "excellent". Scores above average at 71 are presented as "good", and scores at 51 are considered "ok". Based on these criteria, with a SUS score of 90%, the system in this study falls within the excellent rating range, indicating positive feedback from the participants.

4. CONCLUSION

This study aims to help recruiters identify and select the most suitable candidate who meets the job's requirements. Based on the information gathered, we classify and visualize the eligibility rate to find the most suitable applicants that fit the job offered. The NB model that was applied in this study enables the system to do the classification tasks. In addition, the diverse visualizations were analyzed from the information gained in the system application, enabling them to make informed decisions during the hiring process. Clear and intuitive visualization enables the recruiters to retrieve and summarize information about eligible and ineligible applicants. The eligibility rate allows recruiters to efficiently handle and gain a comprehensive understanding of all applicants who meet their criteria in a shorter amount of time. The leveraging power of NB sorts the most qualified candidates for a position and visualization techniques based on the seven attributes for classification: title, location, education, years of professional experience, skills, languages, and certificates. For further study, we recommend adding more functions for contacting the shortlisted applicants from the system and adding more job fields.

ACKNOWLEDGEMENTS

This research was funded by a grant from Universiti Teknologi MARA Cawangan Melaka (TEJA Grant 2023 GDT 2023/1-14).

REFERENCES

[1] N. S. Gill, "Recruitment and selection procedures in human resource management," *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 2, pp. 45–49, 2021, doi: 10.47760/ijcsmc.2021.v10i02.006.
[2] P. A. Hamza *et al.*, "Recruitment and selection: The relationship between recruitment and selection with organizational performance," *International Journal of Engineering, Business and Management*, vol. 5, no. 3, 2021, doi: 10.22161/ijebm.5.3.1.
[3] S. I. Abbas, M. H. Shah, and Y. H. Othman, "Critical review of recruitment and selection methods: Understanding the current practices," *Annals of Contemporary Developments in Management & HR*, vol. 3, no. 3, 2021, doi: 10.33166/acdmhr.2021.03.005.
[4] D. Lawong, G. R. Ferris, W. Hochwarter, and L. Maher, "Recruiter political skill and organization reputation effects on job applicant attraction in the recruitment process: A multi-study investigation," *Career Development International*, vol. 24, no. 4, 2019, doi: 10.1108/CDI-01-2019-0007.
[5] G. Dash and D. Chakraborty, "Digital transformation of marketing strategies during a pandemic: Evidence from an emerging economy during covid-19," *Sustainability*, vol. 13, no. 12, 2021, doi: 10.3390/su13126735.
[6] M. S. Hosain and P. Liu, "Linked in for searching better job opportunity: Passive jobseekers' perceived experience," *Qualitative Report*, vol. 25, no. 10, 2020, doi: 10.46743/2160-3715/2020.4449.
[7] R. Thakkar, "Top 100 hiring statistics for 2022," *LinkedIn*, 2022. Accessed: Nov. 24, 2022. [Online]. Available: https://www.linkedin.com/pulse/top-100-hiring-statistics-2022-rinku-thakkar/
[8] B. T. Janigová, "The genre analysis of job adverts posted on linkedIn," *M.Sc. Thesis*, Department of English and American Studies, Masaryk University, Brno, Czech Republic, 2023.
[9] S. L. -Carril, C. Anagnostopoulos, and P. Parganas, "Social media in sport management education: Introducing LinkedIn," *Journal*

of Hospitality, Leisure, Sport and Tourism Education, vol. 27, 2020, doi: 10.1016/j.jhlste.2020.100262.

[10] K. S. Wei, "LinkedIn in Malaysia: a comprehensive overview of the growing professional network in 2023," *LinkedIn*. 2023. Accessed: May 04, 2023. [Online]. Available: https://www.linkedin.com/pulse/linkedin-malaysia-comprehensive-overview-growing-network-shoo/

[11] G. D. Marin and C. Nilă, "Branding in social media using LinkedIn in personal brand communication: A study on communications/marketing and recruitment/human resources specialists perception," *Social Sciences & Humanities Open*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.1016/j.ssaho.2021.100174.

[12] B. Hmoud and V. Laszlo, "Will artificial intelligence take over human resources recruitment and selection?," *Network Intelligence Studies*, vol. 7, no. 13, pp. 31–30, 2019.

[13] D. P. Sivanandam and M. P. Mudaliar, "A study on scientific screening process in a recruitment consultancy firm," *Journal of Contemporary Issues in Business and Government*, vol. 26, no. 2, 2021, doi: 10.47750/cibg.2020.26.02.053.

[14] E. Fisher, R. S. Thomas, M. K. Higgins, C. J. Williams, I. Choi, and L. A. McCauley, "Finding the right candidate: Developing hiring guidelines for screening applicants for clinical research coordinator positions," *Journal of Clinical and Translational Science*, vol. 6, no. 1, 2022, doi: 10.1017/cts.2021.853.

[15] I. Nikolaou, "What is the role of technology in recruitment and selection?," *Spanish Journal of Psychology*, vol. 24, 2021, doi: 10.1017/SJP.2021.6.

[16] J. L. R. -Sánchez, T. G. -Torres, A. M. -Navarro, and R. G. -Losada, "Investing time and resources for work–life balance: the effect on talent retention," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, 2020, doi: 10.3390/ijerph17061920.

[17] H. S. -Szczapa, "Recruitment of employees—assumptions of the risk model," *Risks*, vol. 9, no. 3, 2021, doi: 10.3390/risks9030055.

[18] G. Erdoğan, "The ethical shortlisting problem," *Computers and Operations Research*, vol. 138, 2022, doi: 10.1016/j.cor.2021.105593.

[19] J. H. Hardy, K. S. Tey, W. C. -Lai, R. F. Martell, A. Olstad, and E. L. Uhlmann, "Bias in context: Small biases in hiring evaluations have big consequences," *Journal of Management*, vol. 48, no. 3, 2022, doi: 10.1177/0149206320982654.

[20] K. A. F. A. Samah, N. S. D. Wirakarnain, R. Hamzah, N. A. Moketar, L. S. Riza, and Z. Othman, "A linear regression approach to predicting salaries with visualizations of job vacancies: a case study of Jobstreet Malaysia," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, pp. 1130–1142, 2022, doi: 10.11591/ijai.v11.i3.pp1130-1142.

[21] M. G. Robinson, "Skills and qualifications for the special library environment in Jamaica: a job advertisement analysis," *Library Management*, vol. 42, no. 1–2, 2021, doi: 10.1108/LM-07-2020-0109.

[22] M. Halinski and J. A. Harrison, "The job resources-engagement relationship: the role of location," *International Journal of Public Sector Management*, vol. 33, no. 6–7, 2020, doi: 10.1108/IJPSM-12-2019-0303.

[23] L. Ronda, C. Abril, and C. Valor, "Job choice decisions: understanding the role of nonnegotiable attributes and trade-offs in effective segmentation," *Management Decision*, vol. 59, no. 6, pp. 1546–1561, 2020, doi: 10.1108/MD-10-2019-1472.

[24] M. Izvercian, S. Potra, and L. Ivascu, "Job satisfaction variables: A grounded theory approach," *Procedia - Social and Behavioral Sciences*, vol. 221, 2016, doi: 10.1016/j.sbspro.2016.05.093.

[25] M. Kaya and T. Bogers, "Effectiveness of job title based embeddings on resume to job ad recommendation," in *2021 Workshop on Recommender Systems for Human Resources, RECSYS IN HR 2021*, Amsterdam, Netherlands, 2021, pp. 1–7.

[26] N. Chaiyama and N. Kaewpila, "The development of life and career skills in 21st century test for undergraduate students," *European Journal of Educational Research*, vol. 11, no. 1, 2022, doi: 10.12973/eu-jer.11.1.51.

[27] A. Taha *et al.*, "Robotic colorectal surgery: quality assessment of patient information available on the internet using webscraping," *Computer Assisted Surgery*, vol. 28, no. 1, 2023, doi: 10.1080/24699322.2023.2187275.

[28] M. M. Öztürk, "Cosine similarity-based cross-project defect prediction," *Bilişim Teknolojileri Dergisi*, vol. 12, no. 3, pp. 159–167, 2019, doi: 10.17671/gazibtd.453436.

[29] L. K. Foo, S. L. Chua, and N. Ibrahim, "Attribute weighted naïve bayes classifier," *Computers, Materials and Continua*, vol. 71, no. 1, 2022, doi: 10.32604/cmc.2022.022011.

[30] K. A. F. A. Samah, N. M. N. Azharludin, L. S. Riza, M. N. H. H. Jono, and N. A. Moketar, "Classification and visualization: Twitter sentiment analysis of Malaysia's private hospitals," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 4, pp. 1793–1802, 2023, doi: 10.11591/ijai.v12.i4.pp1793-1802.

[31] P. N. Mwaro, D. K. Ogada, and P. W. Cheruiyot, "Applicability of Naïve Bayes model for automatic resume classification," *International Journal of Computer Applications Technology and Research*, vol. 9, no. 9, pp. 257–264, 2020, doi: 10.7753/ijcatr0909.1002.

[32] Y. Liu and A. I. Abeyratne, *Practical applications of Bayesian reliability*. Hoboken, United States of America: John Wiley & Sons, 2019.

[33] K. A. F. A. Samah, N. F. A. Misdan, M. N. H. H. Jono, and L. S. Riza, "The best Malaysian airline companies visualization through bilingual twitter sentiment analysis: A machine learning classification," *JOIV: International Journal on Informatics Visualization*, vol. 6, no. 1, pp. 130–137, 2022, doi: 10.30630/joiv.6.1.879.

BIOGRAPHIES OF AUTHORS

Khyrina Airin Fariza Abu Samah is a senior lecturer at the College of Computing, Informatics and Mathematics in Universiti Teknologi MARA (UiTM) Melaka Branch, Jasin Campus. Before joining UiTM, she had 13 years of working experience in the semiconductor industry. She has a Diploma, Bachelor's Degree and Master's Degree in Computer Science and Ph.D. in Information Technology. Her research interests are in artificial intelligence, algorithm analysis, machine learning, data science, optimization, and evacuation algorithms. She can be contacted at email: khyrina783@uitm.edu.my.

Nurul Atirah Ahmad is a Bachelor's student in Computer Science currently undergoing her internship programme. Her research interests are data mining, machine learning algorithms, data analysis and visualization. She is actively involved in the research group at UiTM. Currently, she is pursuing an internship at Amast Sdn Bhd, Selangor, Malaysia. She can be contacted at email: 2020482848@student.uitm.edu.my.

Anis Amilah is a lecturer at the College of Computing, Informatics and Mathematics in Universiti Teknologi MARA (UiTM) Melaka Branch, Jasin Campus. She has a Master of Sciences degree from UiTM Shah Alam. With a solid educational background, she brings expertise and knowledge to her role, contributing to the academic environment. Her research interests revolve around artificial intelligence and emergence algorithms. Her focus on these cutting-edge fields underscores a commitment to exploring innovative solutions and understanding the dynamic interplay between intelligence and emergent phenomena. She can be contacted at email: anisamilah@uitm.edu.my.

Hana Fakhira Almarzuki is a lecturer at the College of Computing, Informatics and Mathematics at Universiti Teknologi MARA (UiTM) Shah Alam, Selangor. She is currently a Ph.D. candidate in the College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Melaka, Jasin Campus, Malaysia. Her research interests are intelligent tutoring systems, artificial intelligence, and machine learning. She can be contacted at email: hanaalira97@gmail.com.

Zuhri Arafah Zulkifli is lecturer at College of Computing, Informatics and Mathematics in Universiti Teknologi MARA (UiTM) Melaka Branch, Jasin Campus. Before joining UiTM, she has been working with HSBC GLC Malaysia. She has a Diploma in Computer Sciences, a Degree in Intelligent Systems from UiTM and a Master of Management Information Systems from UKM. She has a passionate interest in educational technology, knowledge management, intelligent systems, and big data. She can be contacted at email: zuhrizulkifli@uitm.edu.my.

Lala Septem Riza received Ph.D. in Computer Science from Universidad de Granada, Spain, in 2015. He works in the Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia. He teaches machine learning, big data platforms, and statistical data science. His research interests are in machine learning, data science, and education. He can be contacted at lala.s.riza@upi.edu.

Amir Haikal Abdul Halim received a Master's Degree in Computer Science from Universiti Teknologi MARA. He is currently a research assistant and intends to be a Ph.D. candidate in the College of Computing, Informatics, and Mathematics, Universiti Teknologi MARA, Melaka, Jasin Campus, Malaysia. He has a Diploma and a Degree in Computer Science. His research interests are evacuation algorithms, algorithm analysis, artificial intelligence, and machine learning. He can be contacted at email: ameerhaical@gmail.com.