

A linear regression approach to predicting salaries with visualizations of job vacancies: a case study of Jobstreet Malaysia

Khyrina Airin Fariza Abu Samah¹, Nurqueen Sayang Dinnie Wirakarnain¹, Raseeda Hamzah², Nor Aiza Moketar¹, Lala Septem Riza³, Zainab Othman¹

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka Kampus Jasin, Melaka, Malaysia

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Selangor, Malaysia

³Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info

Article history:

Received Nov 9, 2021

Revised Jun 14, 2022

Accepted Jun 21, 2022

Keywords:

Data visualization

Jobstreet Malaysia

Linear regression

Salary prediction

ABSTRACT

This study explicitly discusses helping job seekers predict salaries and visualize job vacancies related to their future careers. Jobstreet Malaysia is an ideal platform for discovering jobs across the country. However, it is challenging to identify these jobs, which are organized according to their respective and specific courses. Therefore, the linear regression approach and visualization techniques were applied to overcome the problem. This approach can provide predicted salaries, which is useful as this enables job seekers to choose jobs more easily based on their salary expectations. The extracted Jobstreet data runs the pre-processing, develops the model, and runs on real-world data. A web-based dashboard presents the visualization of the extracted data. This helps job seekers to gain a thorough overview of their desired employment field and compare the salaries offered. The system's reliability as tested using mean absolute error, the functionality test was performed according to the use case description, and the usability test was performed using the system usability scale. The reliability results indicate a positive correlation with the actual values. The functionality test produced a successful result, and a score of 96.58% was achieved for the system usability scale result, proving the system grade is 'A' and usable.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Khyrina Airin Fariza Abu Samah

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka Kampus Jasin

Merlimau, 77300, Melaka, Malaysia

Email: khyrina783@uitm.edu.my

1. INTRODUCTION

Internet job searching means using the Internet to find employment and job candidates; as such, it is performed by employers and job seekers. Nowadays, many Internet users, especially teenagers, search for their desired job [1]. Job seekers find this easier than traditional job searching methods, which involve newspapers, flyers, and advertisements. Therefore, searching for work on the Internet is faster and offers more options. Moreover, Internet job hunting provides an empirical scenario for understanding the correlation between the quality and amount of information.

This study aims to visualize a vast range of jobs from all over the country, according to Jobstreet Malaysia. Jobs are abundantly available for all graduates from any university or faculty in Malaysia to browse. Jobstreet's job information consists of multiple attributes such as the salary, position, job type and

description, and location. However, far less information is provided in relation to the salary offered, the job scope, and the requirements needed to apply for the job, which graduate students need to find and explore [2]. The Society for Human Resource Management (SHRM) has received many requests from job seekers to identify salaries [3]. Salary information has been a significant problem for job seekers because they tend to compare salaries during their job search. Shlee and Karns [4] supported the evidence that graduate students frequently search for a job with a higher starting salary. They tend to notice that a salary for a vacant position is based on the current salary trends [5], but Jobstreet seems to offer limited information on the salaries offered.

Besides, job seekers may face difficulties comparing and visualizing the demand and trends of the jobs offered on Jobstreet. They are unaware of the higher positions offered on Jobstreet, so they cannot choose a suitable company. Takle [6] claimed that most job seekers do not read the job description details, and [7] proved that some vacancies do not provide the complete requirements, descriptions, expectations, and provisions. According to [8], an effective job search strategy begins with the knowledge of yesterday's, today's, and tomorrow's technologies. Job searchers should pay greater attention to the current employment trends when seeking work. As a result, it is critical to analyze and track the top ten job title categories to monitor the frequency of the most in-demand titles and visualize this information [9].

This study implements a web-based application using Python based on the problems discussed above. It helps users to quickly view the jobs offered on Jobstreet Malaysia in all domains. Therefore, visualizing the data provides a visual representation that allows the user to see more quickly an overview of all the posted jobs that met their criteria. Data extraction and visualization are the two key modules in the study. A vast amount of data with multiple properties can be analyzed more easily with this tool. Hence, this system allows job seekers to analyze jobs on Jobstreet more easily. This paper is organized as follows: it begins with a brief introduction in section 1. Section 2 explains the related work, followed by the research methodology in section 3. Section 4 elaborates on the results and discusses their reliability, functionality, and usability. Finally, section 5 concludes the study and briefly mentions potential future improvements.

2. RELATED WORK

2.1. Jobstreet Malaysia

One of the leading employment information providers in Asia is Jobstreet. It was founded in Malaysia in 1997 and is widely acknowledged to be one of Asia's leading online employment marketplaces. Its main vision is to connect businesses with talent and improve lives by advancing careers. Jobstreet Malaysia entered the market and achieved profitability in early 1999 [10]. As a company, Jobstreet noticed that over 10 million job seekers and over 90,000 employers utilized their services. Despite its many competitors, such as LinkedIn, Monster Inc, and others, job seekers in Malaysia are still utilizing the Jobstreet website because of its ease of use [11]. Most graduates use this website to search for a job. Jobstreet has been in a robust position in Malaysia for almost 23 years. Hence, the Jobstreet website was chosen for this study scope.

This study proposes the development of a web-based application written in Python. Obtained from an online platform, the data was extracted from this website [12] with Python using web scraping. The data scraping from the Jobstreet website was extracted for only six months from November 2020 to May 2021 because Jobstreet only allows clients to advertise their job vacancies. Only information in English was used. However, the jobs offered on the Jobstreet website were not as expected because of the ongoing COVID-19 pandemic. The study scope covers all 13 jobs on Jobstreet, as listed in Table 1.

Table 1. Types of 13 jobs in Jobstreet

No	Types of Jobs
1	Account and Finance Job
2	Admin and Human Resources Job
3	Arts, Media and Communication Job
4	Building and Construction Job
5	Education and Training Job
6	Engineering Job
7	Healthcare Job
8	Restaurant and Hotel Job
9	Computer and Information Technology Job
10	Manufacturing Job
11	Sales and Marketing Job
12	Sciences Job
13	Services Job

2.2. Linear regression

A supervised machine learning method known as linear regression (LR) is often used in mathematical research approaches. It measures the expected effects and models these against numerous input variables [13]. In biological or clinical research, the researcher frequently attempts to comprehend or link two or more independent (predictor) factors to predict an outcome or a dependent variable [14]. A prediction is a forecasted future event [15], and LR is the most fundamental and widely used predictive analysis [16]–[18]. The internet is growing continuously; for example, in recent years, Twitter has been generating 12 terabytes (TB) of data daily, while Facebook has been generating 4 petabytes (PB) of data daily. As a result, collecting, examining, and modeling this massive amount of data is critical to predicting future events in various fields [19].

Simple LR approaches are used in this study to predict each job's salary based on years of experience. The data trend is taken between November 2020 and May 2021 and helps the jobseeker predict the trend, such as salary, using the data from Jobstreet. Analysis of regression estimates shows the connection between one or more independent variables and a dependent variable. The task of fitting a single line through a scatter plot is at the regression analysis's core. The LR formula is in (1).

$$Y = a + bX \quad (1)$$

The explanatory and dependent variables are both given as X . The slope of the line is b and the intercept is a . The best-fit regression line has the smallest sum of the squares. The equation line of regression is obtained by minimizing the sum of the squares created; this is presented as (2). Using the Scikit-learn function, the LR was imported. Table 2 shows the three parameters involved in the LR ensemble process. Using the Scikit-learn, we imported the LR, and Table 2 shows three parameters involved in the ensemble process of LR.

$$m = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \quad (2)$$

Table 2. Parameters in the ensemble process of linear regression

Parameter's Name	Explanation
Fit (x, y)	Fit linear model
Predict (x)	Predict using the linear model
Score (x, y)	Return the coefficient of determination of R2 of the prediction

2.3. Visualization techniques

There is no standard conceptual framework for visualization users across different application areas can explain and share techniques [20]. Computer-to-perceptual representation mapping is called visualization, and it uses encoding techniques to improve human comprehension and communication [21]–[22]. Enhancing trust in modelling machine learning has driven demand increment in better and more effective tools for visualization [23]–[25]. The four visualization techniques used in the study were the line chart, bar chart, treemap chart, and word cloud. This made visualizing the Jobstreet website data more effective [26]. Hence, it became easier to identify the patterns, trends, and outliers in massive datasets. This study intended to transform the data into an effective visualization technique to provide information for job seekers. We use plotly, an open-source interactive graphics library for Python for data visualization. The data was first imported into the pandas' data frames in interactive Python charts and displayed. The data was visualized on the charts to enable rapid comparisons using several visualization techniques.

3. RESEARCH METHOD

Figure 1 shows the overview of the research design for overall system development. The study methodology was divided into three sections for detailed development. It involved the system design, back-end and front-end development.

3.1. System design

The application of a system's product development principles might be considered system design. The design process is facilitated by developing design diagrams. It involved the use case diagrams, flowcharts, and user interfaces, as described in the following subsections.

3.1.1. Use case diagram

Internal and external factors are included in the use case diagram to gather the system requirements. Figure 2 displays the use case diagram for this system as the interactions needed to complete the task. There are eight cases involved directly from the user and four indirect cases for the system.

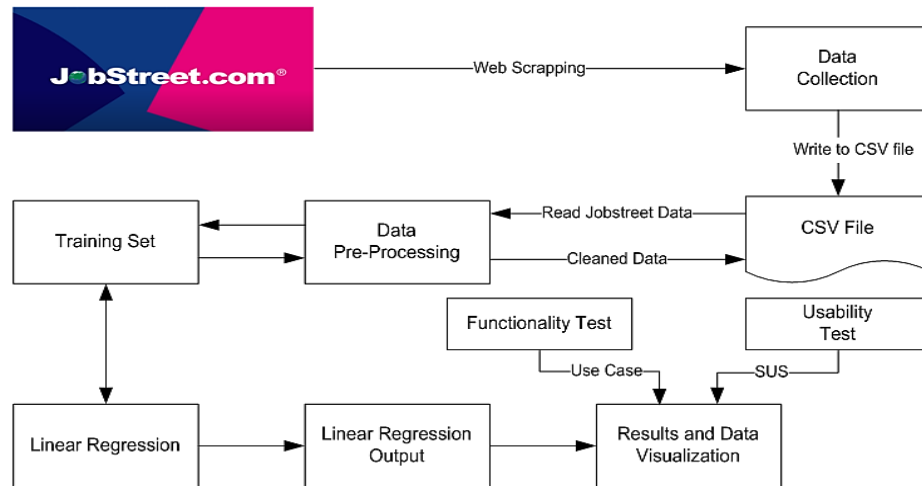


Figure 1. Flow diagram of research design

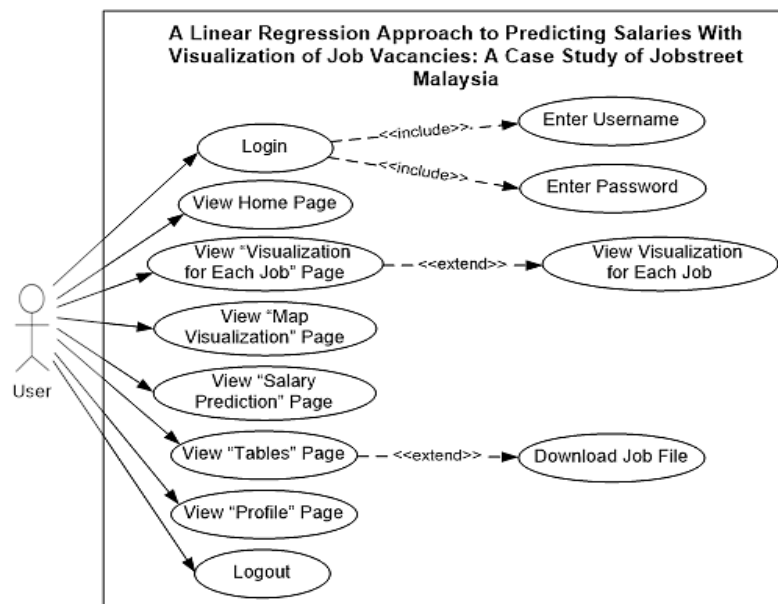


Figure 2. Use case diagram of the system

3.1.2. Flowchart diagram

Figure 3 describes the flowchart of the overall design. Firstly, the user signs in to the system. In this flowchart, users can register a new account or sign in. The user can register a new account if they do not yet have an account. Following the "Visualization of Each Jobs" page, the user must choose a job and visualization type. Next, if the user clicks the "Map Visualization" page, they must choose a job in order to view the map visualization. For the "Tables" page, the user can download each job's comma-separated value (CSV) files. The user can also click on the "Jobstreet Website" page to be redirected to the Jobstreet site. If the user clicks the "Prediction Salary" page, they are redirected to the prediction page, which predicts salaries based on years of experience. Lastly, the user can click on the "Logout" button to log out.

3.1.3. User interface

The system's user interface depicts the system's visual component layout. The user interface was developed according to the use case diagram and the description of the system's features was added. This stage is described in the back-end development subsection.

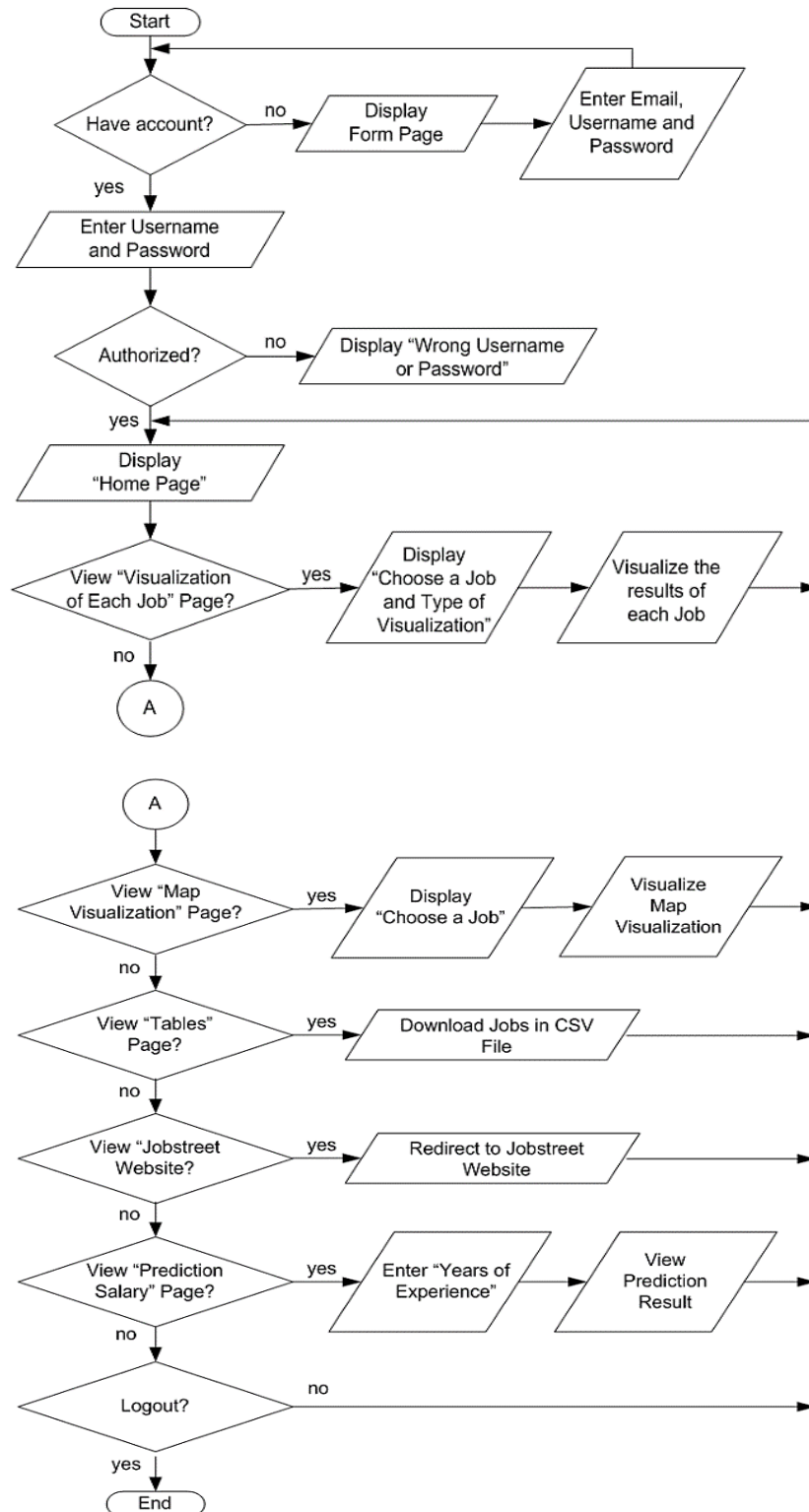


Figure 3. Flowchart diagram for overall system

3.2. Back-end development

The development on the server-side is called back-end development the back-end code aids in transmitting the underlying data to the front-end site. Python was the back-end programming language used for data development and implementation. Data preparation and the prediction model using the LR algorithm are the two of the back-end task that are important for training and testing data:

3.2.1. Data preparation

We collected the dataset from www.jobstreet.com using web scraping, which scrapes all the job categories on the Jobstreet website except the “others” category. The total data scraped produced 22,250 jobs with ten attributes: job title, salary, company, location, description, requirements, qualification, job type, career level, and years of experience. These jobs were scraped from November 2020 to May 2021 because Jobstreet only allows clients to advertise a job vacancy for six months. Besides, the jobs offered on the Jobstreet website were not as expected because of the COVID-19 pandemic. Despite the many ways to scrap data, such as using Rapidminer and Octopurse software, Python is more efficient and could customize the data required for this project. The scraped data for one job took more than 30 minutes to acquire, and limited time was available to discard the data since the Jobstreet website does not allow data scraping during its maintenance periods.

In this project, data cleaning was performed to remove all duplicated data and enter the dataset's null values using Microsoft Excel. The “remove duplicates” function in an Excel file achieved this much faster. The dataset was saved to a working directory after data cleaning. It was then ready to be fed directly into the machine learning algorithm to extract meaningful information in relation to salary predictions.

3.2.2. Prediction model

The data collection was split 80:20, whereby 80% was used for training data and the remainder was used for testing data [27]. LR predicts the salary based on the years of experience specified for each job. This model used four libraries: NumPy, pandas for the dataset, sklearn to implement machine learning functions, and matplotlib to visualize plots for viewing. Five steps were involved in this LR:

- Import the dataset-the dataset was in a CSV-format file, where x represents the years of experience column and y represents the salary column.
- Split the data into two sets: the training and testing sets. The ratio of the testing set for this project was 0.2. A testing set must not be bigger than a training set because it may lack data to train. The random state is the seed for the random number generator, which can be left blank or as 0.
- Visualize the training and testing sets-the LR model was built after the training and testing sets are ready. The code “ $L.fit(xtrain, ytrain)$ ” was used to pass the $xtrain$, which contains the years of experience value, and $ytrain$, which includes the salary values. Thus, the model was formed.
- Initialize and fit the regression model to ensure that the training and testing sets are visualized in the same direction. If so, the model can be considered good to use for the dataset.
- Predict the data based on the years of experience that a specific user wants to predict.

3.3. Front-end development

Front-end development creates the client-side, focusing on what consumers see graphically in their browser or on their application [28]. Front-end languages such as hypertext markup language (HTML), cascading style sheets (CSS), Javascript, and JQuery were used in this project to integrate Flask's back-end framework. After the data processing phases, the Flask web framework and the Python data visualization tools were used to build custom plots and charts in a Python web application context, leveraging the power of both the front-end and back-end development. For each interface, the workflow was depicted in the application with which the user would interact. Also discussed are the correct subsections for each feature of the completed system's interface design, including the home page, visualization for all the job pages, as well as the visualization for each job page, map page, predicted salary page, tables page, Jobstreet website, and profile page.

4. RESULTS AND DISCUSSION

4.1. Reliability testing

A regression model's dependability is assessed by how well its predictions match actual values. Nevertheless, error metrics created by statisticians allow a model's dependability to be evaluated and comparisons of regressions with different parameters [29]. The metrics summarize the data's quality concisely and practically. The mean absolute error (MAE) was used to determine the quality of the model. The MAE is a regression model evaluation statistic. LR was unsuitable for accuracy tests because this project

predicts the salary based on the years of experience required. Hence, it was appropriate to ascertain the quality of the data and the difference between the actual and predicted data [30]. As shown in (3), MAE is the most straightforwardly understood regression error metric; where y is the prediction, \hat{y} is the true value and n stand for total number of data points.

$$MAE = \frac{1}{n} \sum |y - \hat{y}| \quad (3)$$

Moreover, the MAE was the easiest metric to understand because this project evaluated the absolute variation using the actual results and those predicted by a computer model [31]. Using the absolute residual values, the MAE does not indicate whether a model is performing well or poorly (no matter whether the model undershoots or overshoots the actual values). The overall error is proportional to each residual, so larger errors are added linearly. A small MAE suggests a successful prediction, but a large MAE indicates that the model may encounter difficulties in some areas. Table 3 shows the MAE value for each job tested using the testing and prediction data. Figure 4 shows the scatter graph plotted to compare the MAE values with the actual values for two jobs. It can be observed that the predicted values positively correlate with the actual values.

Table 3. Mean absolute error value of each job

Jobs	Mean Absolute Error (MAE)
Account Job	2167.3575862456923
Admin/Human Resouces Job	1856.2874681381668
Arts Job	1477.1881267666222
Building/Construction Job	1952.8038000710462
Education/Training Job	1922.3948270772370
Engineering Job	2082.1180991543570
Healthcare Job	2042.8448050388540
Hotel/Restaurant Job	424.79798113363614
Computer/Information Technology Job	2306.2030678463498
Manufacturing Job	1577.8285877577500
Sales/Marketing Job	1925.2649997237856
Sciences Job	1707.7019373539054

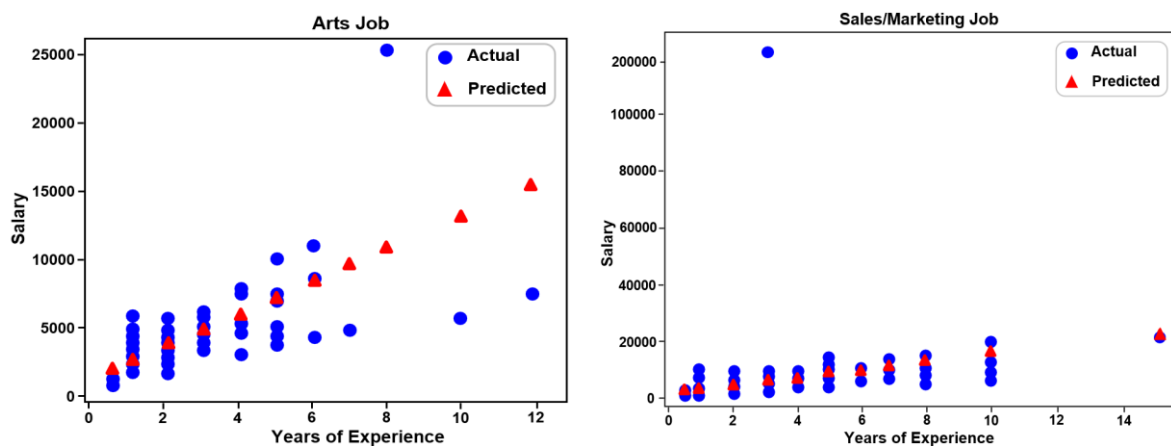


Figure 4. Scatter diagram MAE with the actual value

4.2. Functionality testing

It is critical to test the features of an application to ensure they all work correctly and that any detected errors are corrected. Functionality testing aims to test each function of the visualization application rigorously to determine how closely the specifications match by providing the appropriate input and comparing the output to the functional requirements outlined in previous chapters. The test was conducted using test case scenarios derived from the program specifications with the functionalities tested. The results show that the system ran successfully as planned, without any prompted errors.

The data visualization was also tested to ensure users could view the selected visualization graph. The visualization provides an easily accessible reporting feature so that the user can instantly see and understand the trends and patterns. Figure 5 shows the visualization page menu for all 13 jobs on Jobstreet that are covered in this study. Figure 6 shows the visualization page that visualized four visualizations for the computer/information technology jobs: i) Figure 7 shows the top five locations with the highest number of jobs; ii) Figure 8 shows a line chart of the salaries from each job; iii) Figure 9 shows the qualifications word cloud; and iv) Figure 10 shows job titles word cloud. The user can choose what visualization they want to view. The bar and line chart visualizations are presented interactively. This allows the user to zoom in or out of the chart, download it onto their computer as a png file, pan, box select, lasso select, autoscale, reset axes, toggle the spike lines, show the closest data on hover, and compare data on hover. Figure 11 shows the interface of the “Map Visualization” page, where the user can zoom in and out on the map to view the number of jobs at a specific location. The cluster in red means that the location has many jobs being offered, the green cluster has few jobs, and the yellow cluster means that the location has an average number of jobs being offered. Figure 12 shows the interface that allows users to predict their salary based on experience by entering whichever years of experience they choose.

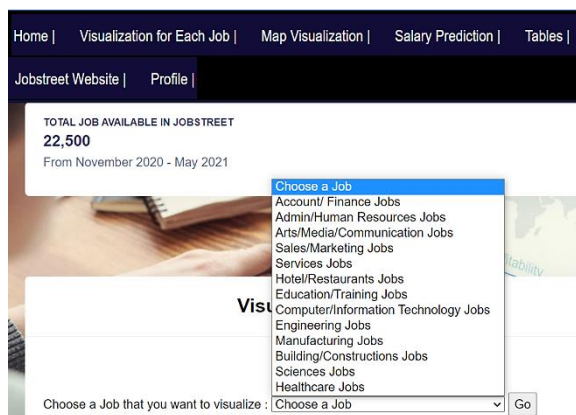


Figure 5. View the “Visualization of Each Job” page

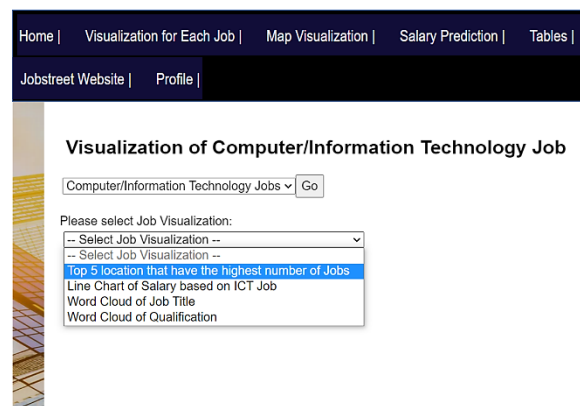


Figure 6. Four visualizations menu for computer/information technology jobs

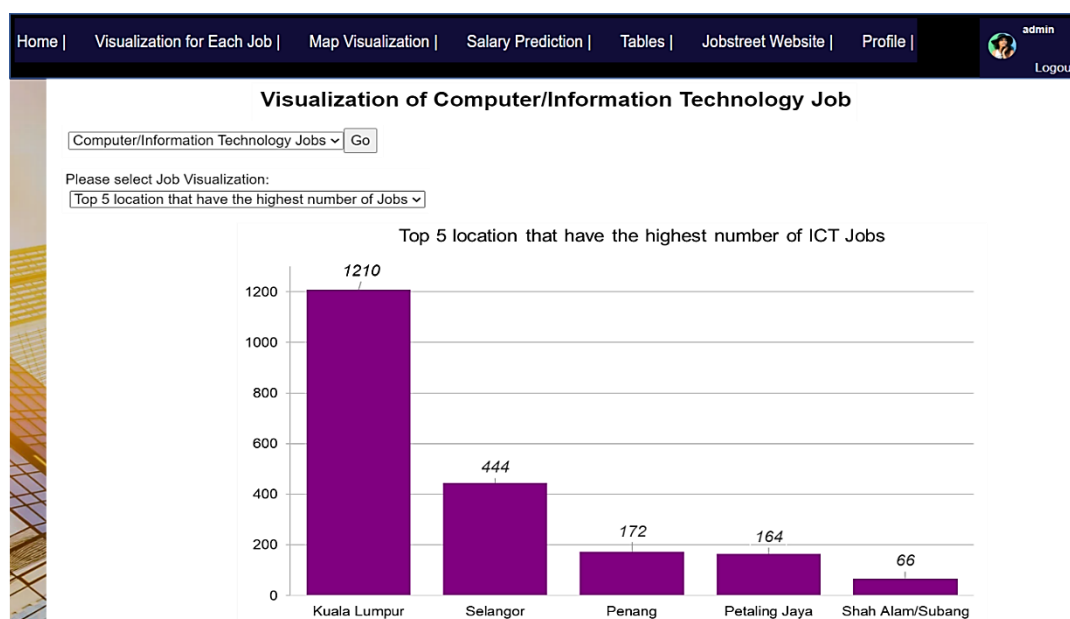


Figure 7. Top 5 locations that have a higher number of computer/information technology jobs

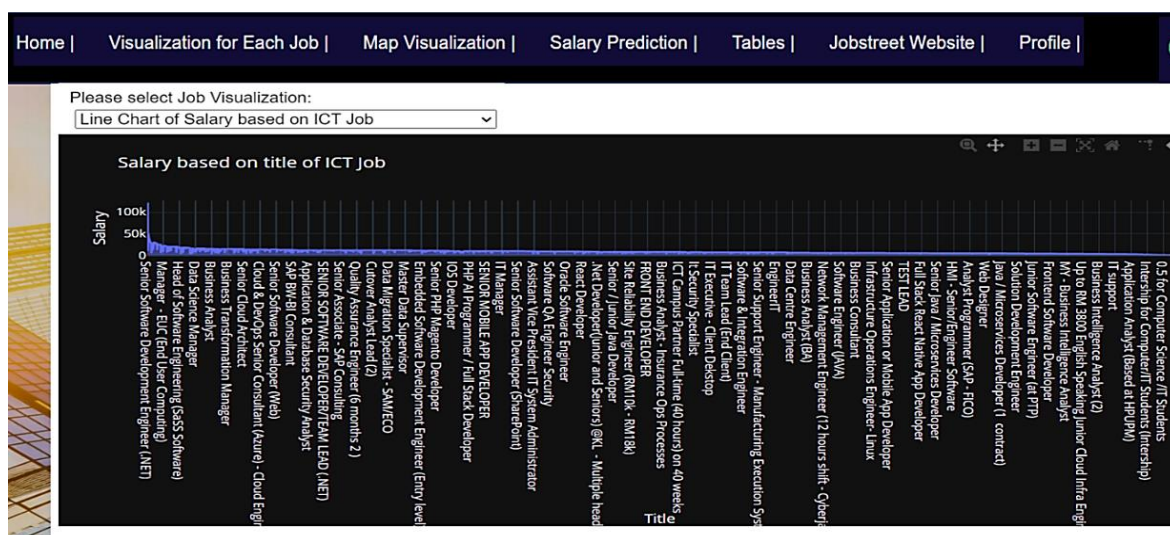


Figure 8. Line chart of salary based on computer/information technology job

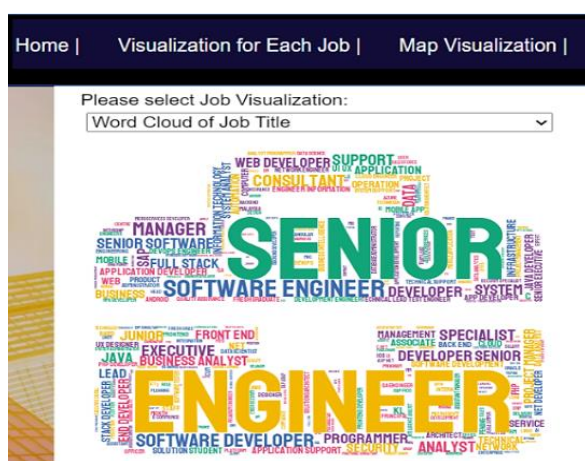


Figure 9. A word cloud for qualification required for computer/information technology jobs

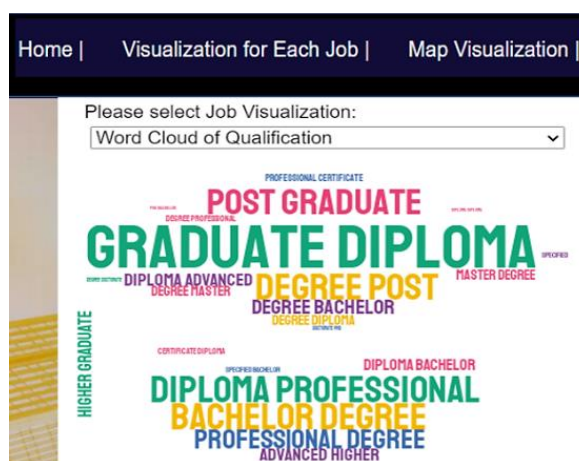


Figure 10. A word cloud for job title offered for
Computer/Information Technology jobs

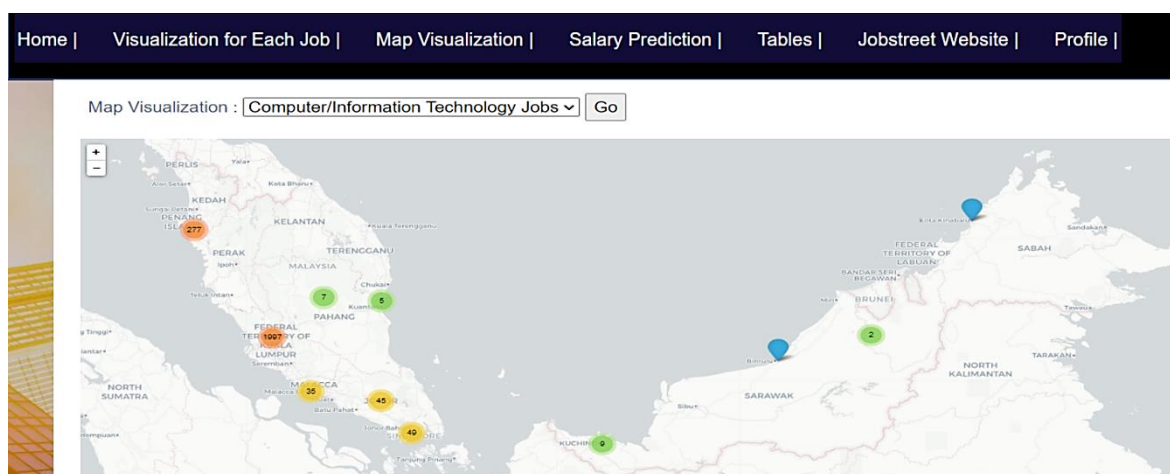


Figure 11. Top 5 locations that have a higher number of computer/information technology jobs

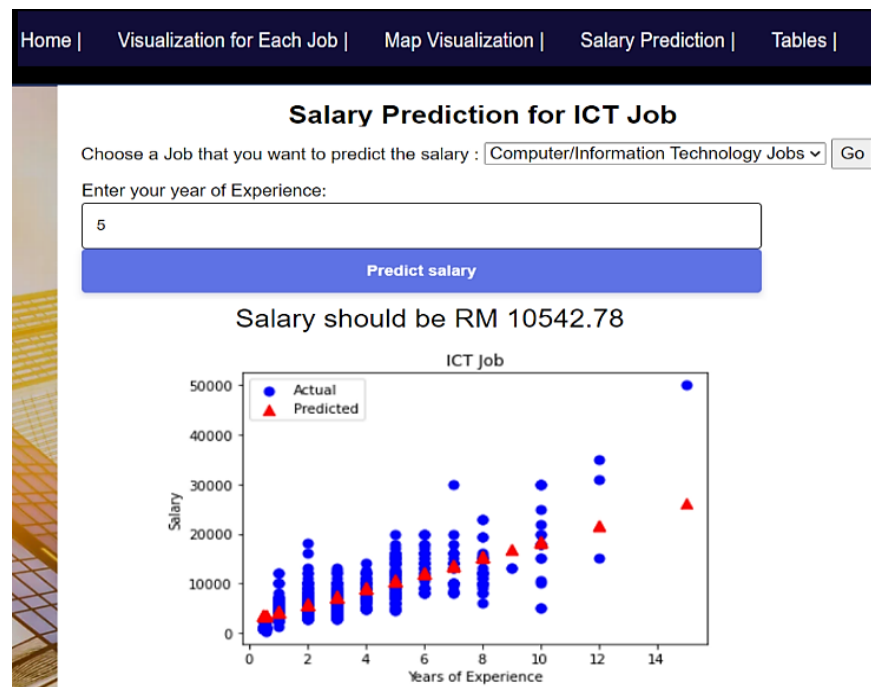


Figure 12. Selection of the job visualization from computer/information technology jobs

4.3. Usability testing

The system usability scale (SUS) was adopted to evaluate the system usability. This was a Likert scale consisting of ten short questions, which were answered by the application users. Based on [32], usability testing is a commonly used method for evaluating user efficiency and their acceptance of products and systems. Usability testing is a technique used in user-centric interface design to assess a product by testing it on users.

Due to the extreme circumstances of COVID-19, in-person usability testing was not possible. The usability tests were conducted individually with the respondents via TeamViewer, with respondents receiving remote access to the program. Thirty respondents seeking jobs were chosen randomly for the usability test and asked to evaluate each system's functionality. Basic information was requested, such as their name, status, and feedback, using the SUS questionnaires on the google form.

The bar chart diagram of the ten SUS statement scores is plotted, as shown in Figure 13. This displays the scales of the SUS statements, as answered by the users in the questionnaire. The graph demonstrates that most respondents agreed with the odd-numbered questions, reflecting affirmative statements, which meant that the users deemed the application to be useful. The respondents also thought the system interface was pleasant and included all the expected features and functionality. This indicates that consumers would not require technical support to use the application's features and navigate other sites. The majority of respondents were pleased with the application.

Figure 14 shows the histogram of the SUS scores. The y-axis (the plotted histogram's vertical axis) illustrates the frequency of users that answered the SUS, while the x-axis (the horizontal axis) shows the percentage range of the SUS scores. The data spread between 90% to 100%, based on the histogram. The plotted graph has a normal distribution with a range starting at 88%, followed by a new range for every increase of 2%. The highest frequency was 96% to 98%, a range that seventeen respondents fell into. The histogram centered on a frequency of 94% to 96%. Ten respondents fell into the category below the central value, and eleven respondents lay in above the central value.

In total, the 30 users who participated in the SUS questionnaire had a SUS score of 96.58%. The baseline for the SUS average score was 68%, indicating the average usability of the system. A score of less than 68% would indicate that the application's usability was likely to cause problems, which would need to be addressed. When the SUS is greater than 80%, the system is entirely usable. Wang [33] claimed that if SUS scores of 80.3% or higher can be obtained, this equates to an 'A' result. Since a high score of more than 80% was obtained, this web-based application was proven usable. It is also worth noting that most respondents gave positive feedback and said they would suggest it to their friends.

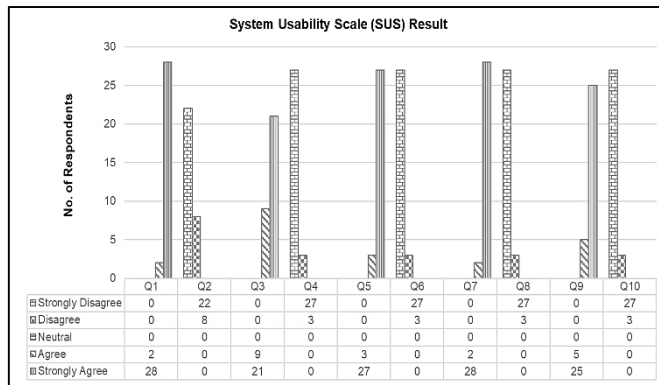


Figure 13. Bar chart of SUS result

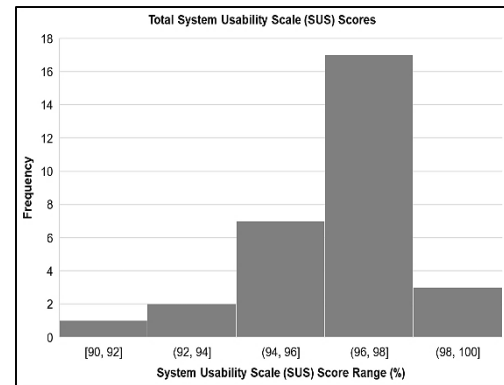


Figure 14. Histogram of SUS result

5. CONCLUSION

The visualization application developed using the dashboard can assist job seekers to easily gain an overview of the jobs offered by identifying the patterns and trends on the Jobstreet website. This visualization changes large and complex data from Jobstreet into understandable and usable data. The researchers have highlighted the drawbacks of using the limited information on the Jobstreet website as single-source data. The application's interactive visualization has changed how users interact with data by focusing on graphic representations. The LR used to predict salaries was incorporated into the application, permitting users to employ the model to predict salaries based on years of experience or similar paths. This would allow the users to gain an overview of the visualization of jobs on the Jobstreet website and predict their salaries according to the years of experience required. Future recommendations concern potential upgrades to the application that could allow the real-time data from the Jobstreet website to be visualized, thus offering job seekers a more rewarding experience. The authors recommended purchasing Jobstreet premium APIs as this allows more data to be extracted, such as the job seeker's details; thus, this application would be suitable for both companies and job seekers.

ACKNOWLEDGEMENTS

The research was sponsored by Universiti Teknologi MARA Cawangan Melaka under the TEJA Grant 2021 (GDT 2021/1-28).




REFERENCES

- [1] E. A. J. van Hooft, J. D. Kammeyer-Mueller, C. R. Wanberg, R. Kanfer, and G. Basbug, "Job search and employment success: A quantitative review and future research agenda," *J. Appl. Psychol.*, vol. 106, no. 5, pp. 674–713, May 2021, doi: 10.1037/apl0000675.
- [2] Y.-L. Liu, "Providing more or less detailed information in job advertisements—Does it matter?," *Int. J. Sel. Assess.*, vol. 28, no. 2, pp. 186–199, Jan. 2020, doi: 10.1111/ijsa.12281.
- [3] B. Sareen, "Relationship between strategic human resource management and job satisfaction," *Int. J. Curr. Res. Life Sci.*, vol. 7, no. 3, pp. 1229–1233, 2018.
- [4] R. P. Schlee and G. L. Karns, "Job requirements for marketing graduates: Are there differences in the knowledge, skills, and personal attributes needed for different salary levels?," *J. Mark. Educ.*, vol. 39, no. 2, pp. 69–81, Jun. 2017, doi: 10.1177/0273475317712765.
- [5] J. Säve-Söderbergh, "Gender gaps in salary negotiations: Salary requests and starting salaries in the field," *J. Econ. Behav. Organ.*, vol. 161, pp. 35–51, May 2019, doi: 10.1016/j.jebo.2019.01.019.
- [6] J. Takle, "Jobviz-A visual tool to explore the employment data," University of New Jersey, 2015.
- [7] S. Chala and M. Fathi, "Job seeker to vacancy matching using social network analysis," in *2017 IEEE International Conference on Industrial Technology (ICIT)*, Mar. 2017, pp. 1250–125, doi: 10.1109/icit.2017.7915542.
- [8] R. B. Mbah, M. Rege, and B. Misra, "Discovering job market trends with text analytics," in *2017 International Conference on Information Technology (ICIT)*, Dec. 2017, pp. 137–142, doi: 10.1109/icit.2017.29.
- [9] L. Samek, M. Squicciarini, and E. Cammeraat, "The human capital behind AI: Jobs and skills demand from online job postings," *OECD Sci. Technol. Ind. Policy Pap.*, vol. 120, Sep. 2021, doi: 10.1787/2e278150-en.
- [10] "Jobstreet Corporation Bhd," *Jobstreet.com*, 2011.
- [11] N. Kamaruddin, A. W. A. Rahman, and R. A. M. Lawi, "Jobseeker-industry matching system using automated keyword selection and visualization approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 3, pp. 1124–1129, Mar. 2019, doi: 10.11591/ijeecs.v13.i3.pp1124-1129.
- [12] "Jobstreet," 2021. <https://www.jobstreet.com.my/> (accessed Mar. 11, 2021).
- [13] H.-I. Lim, "A linear regression approach to modeling software characteristics for classifying similar software," in *2019 IEEE 43rd*




- Annual Computer Software and Applications Conference (COMPSAC)*, Jul. 2019, pp. 942–943, doi: 10.1109/compsac.2019.00152.
- [14] K. Kumari and S. Yadav, "Linear regression analysis study," *J. Pract. Cardiovasc. Sci.*, vol. 4, no. 1, pp. 33–36, 2018, doi: 10.4103/jpcs.jpcs_8_18.
- [15] K. E. Ehimwenma and S. Krishnamoorthy, "Design and analysis of a multi-agent e-learning system using prometheus design tool," *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, pp. 9–23, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp9-23.
- [16] N. Tiraiyari, K. Idris, U. Jegak, and H. Azimi, "Competencies influencing extension workers' job performance in relation to the good agricultural practices in Malaysia," *Am. J. Appl. Sci.*, vol. 7, no. 10, pp. 1379–1386, Oct. 2010, doi: 10.3844/ajassp.2010.1379.1386.
- [17] A. F. Schmidt and C. Finan, "Linear regression and the normality assumption," *J. Clin. Epidemiol.*, vol. 98, pp. 146–151, Jun. 2018, doi: 10.1016/j.jclinepi.2017.12.006.
- [18] S. Liu, M. Lu, H. Li, and Y. Zuo, "Prediction of gene expression patterns with generalized linear regression model," *Front. Genet.*, vol. 10, pp. 1–11, Mar. 2019, doi: 10.3389/fgene.2019.00120.
- [19] S. Rong and Z. Bao-wen, "The research of regression model in machine learning field," *MATEC Web Conf.*, vol. 176, p. 1033, 2018, doi: 10.1051/mateconf/201817601033.
- [20] M. Diamond and A. Mattia, "Data visualization: An exploratory study into the software tools used by businesses," *J. Instr. Pedagog.*, vol. 18, no. 1, 2017.
- [21] H. C. Purchase, N. Andrienko, T. J. Jankun-Kelly, and M. Ward, "Theoretical foundations of information visualization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4950 LNCS, Springer Berlin Heidelberg, 2008, pp. 46–64.
- [22] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko, "Augmenting visualizations with interactive data facts to facilitate interpretation and communication," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 672–681, Jan. 2018, doi: 10.1109/tvcg.2018.2865145.
- [23] C. Xiong, L. Padilla, K. Grayson, and S. Franconeri, "Examining the components of trust in map-based visualizations," in *Proceedings of the EuroVis Workshop on Trustworthy Visualization (TrustVis '19)*, 2019, pp. 19–23.
- [24] A. Chatzimpampas, R. M. Martins, I. Jusufi, K. Kucher, F. Rossi, and A. Kerren, "The state of the art in enhancing trust in machine learning models with the use of visualizations," *Comput. Graph. Forum*, vol. 39, no. 3, pp. 713–756, Jun. 2020, doi: 10.1111/cgf.14034.
- [25] M. Z. A. Razak, S. S. M. Fauzi, R. A. J. M. Gining, and M. N. F. Jamaluddin, "Data visualisation of vehicle crash using interactive map and data dashboard," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 14, no. 3, pp. 1405–1411, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1405-1411.
- [26] Z. Yang and S. Cao, "Job information crawling, visualization and clustering of job search websites," in *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Dec. 2019, pp. 637–641, doi: 10.1109/iaeac47372.2019.8997713.
- [27] N. M. Khan, N. Abraham, and M. Hon, "Transfer learning with intelligent training data selection for prediction of Alzheimer's disease," *IEEE Access*, vol. 7, pp. 72726–72735, 2019, doi: 10.1109/access.2019.2920448.
- [28] M. Kaluža and B. Vukelić, "Comparison of front-end frameworks for web applications development," *Zb. Veleučilišta u Rijeci*, vol. 6, no. 1, pp. 261–282, 2018, doi: 10.31784/zvr.6.1.19.
- [29] A. Botchkarev, "A new typology design of performance metrics to measure errors in machine learning regression algorithms," *Interdiscip. J. Information, Knowledge, Manag.*, vol. 14, pp. 45–76, 2019, doi: 10.28945/4184.
- [30] L. Frías-Paredes, F. Mallor, M. Gastón-Romeo, and T. León, "Dynamic mean absolute error as new measure for assessing forecasting errors," *Energy Convers. Manag.*, vol. 162, pp. 176–188, Apr. 2018, doi: 10.1016/j.enconman.2018.02.030.
- [31] C. Chen, J. Twycross, and J. M. Garibaldi, "A new accuracy measure based on bounded relative error for time series forecasting," *PLoS One*, vol. 12, no. 3, p. e0174202, Mar. 2017, doi: 10.1371/journal.pone.0174202.
- [32] A. M. Wichansky, "Usability testing in 2000 and beyond," *Ergonomics*, vol. 43, no. 7, pp. 998–1006, Jul. 2000, doi: 10.1080/001401300409170.
- [33] Y. Wang, "System usability scale : A quick and efficient user study methodology," *IXD@Pratt*, 2018. .

BIOGRAPHIES OF AUTHORS







Khayrina Airin Fariza Abu Samah    is a senior lecturer at the Faculty of Computer and Mathematical Sciences in Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. Before joining UiTM, she has 13 years of working experience in the semiconductor industry. She has a Diploma, Bachelor's Degree and Master's Degree in Computer Science and PhD in Information Technology. Her research interest is in Artificial Intelligent, Operational Research, Algorithm Analysis, Clustering and Optimization, Evacuation Algorithm, Internet of Things (IoT) and Sentiment Analysis. She can be contacted at email: khayrina783@uitm.edu.my.







Nurqueen Sayang Dinnie Wirakarnain    is a Bachelor's student in Computer Science currently undergoing her internship programme. Her research interests are machine learning algorithms, data analysis and visualization. She is actively involved in the research group at UiTM. Currently, she is under an internship at Petronas Digital Sdn Bhd. She can be contacted at email: 2018276258@isiswa.uitm.edu.my.







Raseeda Hamzah     is currently working as a Senior Lecturer in the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia. She received her bachelor and master degrees from Universiti Teknikal Malaysia Melaka and the University of Malaya. Her PhD in (Computational Linguistics) from UiTM. She is actively researching signal processing, pattern recognition and feature analysis for various areas and different types of data. She can be contacted at email: raseeda@uitm.edu.my.







Nor Aiza Moketar     is a Senior Lecturer in the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. She holds a PhD and MSc in Software Engineering and Intelligence from UTeM and BSc in Information Technology from University Malaya. Before commencing as an academician, she worked as a software developer in a multi-national company. Her research interests are Software Engineering, requirements, testing and Intelligence. She can be contacted at email: noraiza1@uitm.edu.my.



Lala Septem Riza     received PhD 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 platform, and statistical data science. His research interests are in machine learning, data science, and education. He can be contacted at email: lala.s.riza@upi.edu.



Zainab Othman     is a senior lecturer of Computer Science at the Faculty of Computer and Mathematical Sciences (UiTM), Melaka Jasin Campus. She holds MSc in Information Technology from UUM, Diploma and Bachelor's Degree in Computer Science from UiTM and UKM, respectively. She has 24 years of teaching experience in various courses such as Software Engineering, Programming, Data Structures, Algorithm Analysis. Her research interest in Software Engineering, Algorithm, Programming, Game-Based Learning, Augmented Reality. She can be contacted at email: zainab_othman@uitm.edu.my.