Insight on Competency Skills from Job Advertisement for Data Scientists

Nurul Dalia Zulkifli[1], Siti Zanariah Satari[1,*]

[1] Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Persiaran Tun Khalil Yaakob, 26300 Gambang, Pahang, Malaysia

ARTICLE INFO	ABSTRACT
                 	The need for Data Science professionals is expanding rapidly as the Information Technology (IT) sector competes to control the dominion of data. Nevertheless, almost every association has a distinct method for identifying positions in Data Science and its related skills and expertise. Employers, educational institutions, and aspirant Data Science experts are now dealing with a bewildering industry landscape as a result. This research examines the definitions of Data Science and Data Scientist, as well as the Data Science careers in different nations. The data with 169 rows are extracted from LinkedIn and JobStreet which are based in Malaysia from February 2023 to April 2023 by using web crawling and web scraping techniques in Python. Next are the key competencies in a Data Scientist's toolkit which are hard and soft skills that would be analysed through data visualization. For data processing and analysis, proficiency in programming languages such as R, Python, and SQL is required. In addition, proficiency with visualization tools like Tableau or Power BI allows for effective results. Furthermore, communication and visualization skills are essential for effective data storytelling. Data scientists must be excellent communicators to explain complex findings to both technical and non-technical stakeholders. It is easier to effectively share findings when visualizations that are aesthetically pleasing and data-driven are produced. Aspiring Data Scientists can benefit greatly from understanding the abilities and skills listed in job advertisements for Data Scientists. Individuals can position themselves as qualified candidates in this quickly changing sector by honing and demonstrating these talents.

1. Introduction

A job advertisement is an announcement letting individuals know that a specific job opening is offered. It is written in an enticing style and includes details about the employment opportunity, the business, and the perks it provides as mentioned in Talentlyft [1]. Although job advertisements can help job seekers apply for Data Scientist, employability skills can change over time as Data Scientist is a wide career with a variety of related skills. Datamites.com [2] also reports that an IBM analysis predicts a 39% increase in demand from employers for both Data Scientists and Data Engineers.

---

* Corresponding author.
E-mail address: zanariah@umpsa.edu.my

According to a report by ETHRWorld [3] in 2026 it is anticipated that more than 11 million individuals will work in Data Science and Analytics fields.

Even though the term "Data Scientist" has become a professional title, it seems that every sector, job function, and company is still trying to define the position as written by Fayyad and Hamutcu [4]. Both companies and professionals must incur significant fees because of this complication that was supported by Fayyad and Hamutcu [4]. Employers who want to locate the best people for their company needs must develop a knowledge of the nebulous taxonomy of many Data Scientists as mentioned by Fayyad and Hamutcu [4]. On the other side, workers must sift through a sea of Data Science jobs, only to discover that they are unqualified for most of the positions they apply for that was supported by Fayyad and Hamutcu [4].

The goal of this study is to examine the demands of Malaysian Data Scientists by mining job advertisements to distinguish their skills. We collected 169 job advertisements from LinkedIn and JobStreet that are based in Malaysia from February 2023 to April 2023. We perform web crawling and web scraping to extract the skills and requirements needed for the Data Scientist job market in Malaysia. We also analyze the hard skills and soft skills by using data visualization.

1.1 Job Advertisement

Job advertisements are still useful for understanding skill requirements, but not always for estimating the number of openings in the market because it seems difficult to presume that the crucial aspects of an occupation, such as skills, are likely to be consistent across different sorts of businesses and establishments so, even if all online reported job advertisements are gathered, it is still difficult to assess the representativeness of the data as written by Khaouja *et al.,* [5]. University decision-making on degree programs and career opportunities might be aided by an understanding of market requirements that was supported by Brooks *et al.,* [6]. According to Brooks *et al.,* and Malgwi *et al.,* [6,7], the students are more likely to select a major that they believe will offer them greater work chances. Even though many colleges rely their funding on student credit hours or enrolment having a degree that incorporates a curriculum that leads to increased career chances and being aware of the market for certain fields might have an impact on funding for such universities which were supported by Brooks *et al.,* and Brookshire [6,8].

1.2 Data Science

According to UCB-UMT [9], one of the most promising and in-demand career options for qualified professionals is Data Science, which is still developing. Successful data professionals today are aware that they need to go beyond the conventional competencies of big data analysis, data mining, and programming skills. Data Scientists need to possess a level of flexibility and awareness to optimize benefits at each stage of the process and grasp the complete spectrum of the Data Science life cycle to unearth meaningful intelligence for their organizations. Data Scientists combine mathematics and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with particular subject matter expertise to find hidden actionable insights in an organization's data that was reported by IBM Analytic [10].

1.3 Data Scientist

Data scientists are a new breed of analytical data experts with the technical know-how to address complicated issues and the inquisitiveness to investigate what issues are at stake as reported by SAS

[11]. Data Analyst and Data Scientist duties frequently cross over, particularly when it comes to exploratory data analysis and data visualization which was supported by SAS [11]. The skill set of a Data Scientist, however, is often broader than that of the standard Data Analyst so, Data Scientists, in contrast, use popular programming languages like Python to perform greater statistical inference and data visualization as SAS [11] reported. They also have talents that Data Analysts don't: they know how to create, train, and use machine learning and deep learning models to comprehend data as IBM [12] reported. In the USA, there are more than ten states consisting of more than 2,840 Data Scientists in the said country that was supported by Sharma [13]. This shows that Data Scientists are still in demand for the coming year. Meanwhile, in India, by 2026, there will be over 11 million job openings in data science according to analysts which was written from U.S. Bureau Of Labor Statistics [14].

1.4 Employability Skills

In developing nations, there are about 75 million young people without jobs, and in many countries, youth unemployment rates are two to four times greater than adult unemployment rates as written by Fajaryati *et al.,* [15]. According to Fadhilah *et al.,* [16], employers prefer to employ graduates who are well-rounded and have achieved well in school as well as placing a strong emphasis on teamwork, quality work, commitment, employee ethics, lifelong learning, and professionalism, as these qualities are crucial for ensuring that employees perform their duties to the highest standards and with integrity. In addition, Rahmat *et al.,* [17] concluded that rather than a shortage of jobs, the main cause of unemployment was a lack of employability skills held by job searchers, as was reiterated by Yasin *et al.,* [18]. This viewpoint is supported by Yasin *et al.,* [18], who conducted research and discovered that one of the primary causes of unemployment in Malaysia is the mismatch between employers' needs and graduates' possession of skills that was written by Safraz *et al.,* [19].

In today's competitive labour market, only those who keep up with current knowledge and possess the aptitude to learn new skills will be able to survive which was written by Fadhilah *et al.,* [16]. The high unemployment rate is sometimes attributed to the educational system's failure to produce graduates backed by employability skills and high levels of competitiveness from Fajaryati *et al.,* [15]. If a job seeker's lack of abilities is tied to the educational system, the issue of education quality is also involved as mentioned by Fajaryati *et al.,* [15].

1.5 Hard Skills

Hard skills are certain aptitudes, prowess, and skill sets that a person can have and objectively display. Hard skills are attainable abilities that enable people to carry out task-specific duties or that may be necessary for a particular career by Gillis and Pratt [20]. The intelligence quotient thinking, which has signed such as counting, analysing, designing, thorough knowledge, modelling, and critical thinking, can be used to identify the components of hard skills from Ardian *et al.,* [21]. According to Glassdoor, in a workplace that is changing quickly, maintaining the knowledge and skills necessary to be successful in current and future roles requires ongoing upskilling.

Machine learning has the highest number of jobs listing skills and the rest of the skills such as distributed computing, time series analysis, statistical modelling and usability testing has the same number that resulted in the second-highest number of jobs listing skill that was recorded by Glassdoor in 2021. These technical skills are important for jobs such as Data Scientists, Data Analysts, Machine Learning engineers, as well as Software engineers.

1.6 Soft Skills

According to Pumble Learn [22], the common trend since the COVID-19 pandemic is the inability to effectively collaborate and communicate is seen as the primary reason for workplace failures by 86% of employees and executives. Also, to avoid uncomfortable interactions, 16% of managers would prefer email as well and poor communication is cited by 28% of employees as a factor in their inability to produce work on schedule as written in Pumble Learn [22]. Due to their lack of communication skills, teamwork abilities, and the capacity to manage projects autonomously, many graduates are unprepared for the workplace of today which was supported by Saad Fadhil *et al.,* [23]. Saad Fadhil *et al.,* [23] also mentioned that the absence of soft skills among graduates in Malaysia is one of the major problems faced by the industry. The World Economic Forum [24] predicted that soft skills will be increasingly in demand by 2025.

The aim of this study was to examine the common competency skills needed that are related to the employability of Data Scientists and align learning objectives with the needs of the labour market. The key contribution of this study is to have identified means of systematically mapping skills, experience, and qualifications sought by employers for their Data Scientists, thus providing a data-driven pathway for employability and avoiding skills gaps and mismatches in a profession that is pivotal in Industry 4.0.

2. Methodology

This research used three different methods to achieve the objectives. The data was collected from February 2023 to April 2023. The tools used in this study were Python and R programming. The types of data are secondary data which has been taken on two different job advertisement websites which are LinkedIn and JobStreet.

2.1 Web Crawling and Web Scraping

Web crawling is a bot that downloads and indexes stuff from the internet while web scraping is an automated technique for gathering copious volumes of data from websites that was defined by GeeksForGeeks [25]. Web scraping was suggested as a means for gathering data on jobs to determine job openings in Indonesia by Pillai *et al.,* [26] hence, to understand skill mismatches where there is an imbalance in skill and education web scraping is needed.

This study utilized a Python script to extract the data for the purpose of the study, utilizing selenium, pandas, and time libraries. LinkedIn and JobStreet are reliable sources of information. Relevant job postings that were made between February 2023 to April 2023 are gathered by using web crawling. Malaysia and Data Scientists were the keywords for the search. The collection process was carried out twice, yielding 169 job ads in total after 113 links for LinkedIn and 56 links for JobStreet were discovered. Then, we scraped the job requirement before we stored the data in a CSV (Character Separated Value) file. The dataset comprised five data fields which are job title, company name, work time, company location, and job requirement.

2.2 Text Mining

Text mining or data text mining is the process of converting unstructured text into a structured format with the objective of spotting important trends and innovative perspectives. The study of job profiles has been made easier by recent developments in text mining techniques because it is now

easier to find software for web crawling, web extraction, and text analysis. According to Smaldone *et al.,* [27], text mining integrates information-retrieval and information extraction activities and has become crucial for businesses and organizations in coping with the information explosion which is also due to the increased exposure of IT resources.

In this phase, we used the Job Requirement to extract the competency skills for a Data Scientist using R programming. Keyword extraction is important for this process as it includes text pre-processing operations such as tokenization, filtering, lemmatization, and stemming tasks which was supported by Schedbauer *et al.,* [28].

2.3 Data Visualization

Using visual elements like charts, graphs, and maps, data visualization tools provide a simple method for noticing and analysing trends, outliers, and patterns in data. It also provides a fantastic tool for employees or business owners to communicate facts intelligibly to non-technical audiences from Tableau [29].

3. Results and Discussion

An example of a job posting that has been extracted from LinkedIn and JobStreet is shown in Figure 1. This job posting was collected through web crawling and web scraping. Figure 1(a) shows the job title, company name, work time, and company location. Whereas Figure 1(b) shows the job requirements that needed to be analysed for this study.
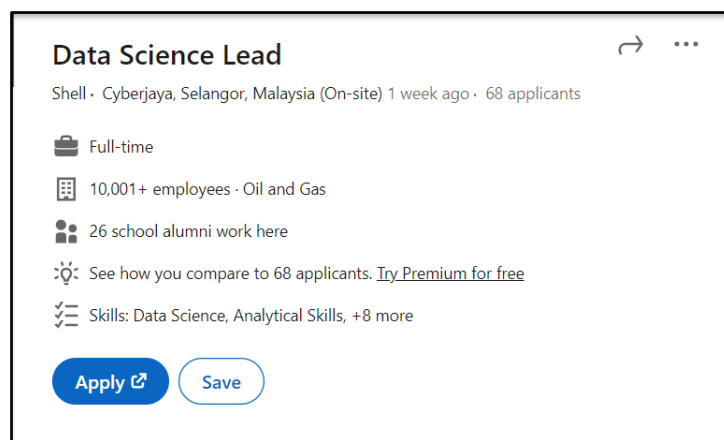
(a)
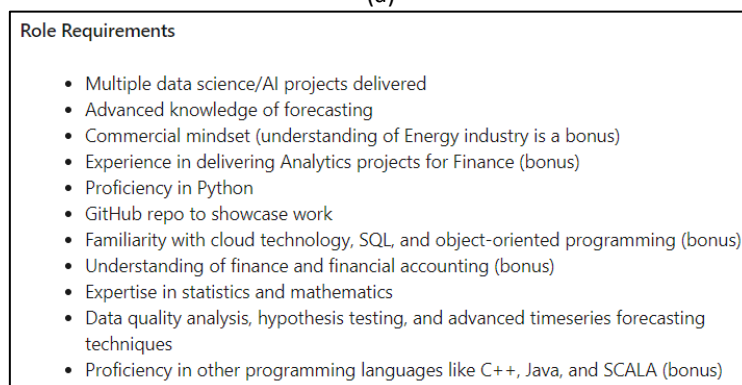
(b)

Fig. 1. Example of job postings that are needed to extract data for the study

Figure 2 visualizes the top 100 most common skills of both hard skills and soft skills in the form of word cloud. There is a clear result of job requirements that have been posted in job advertisements. The most common word or skill is SQL.
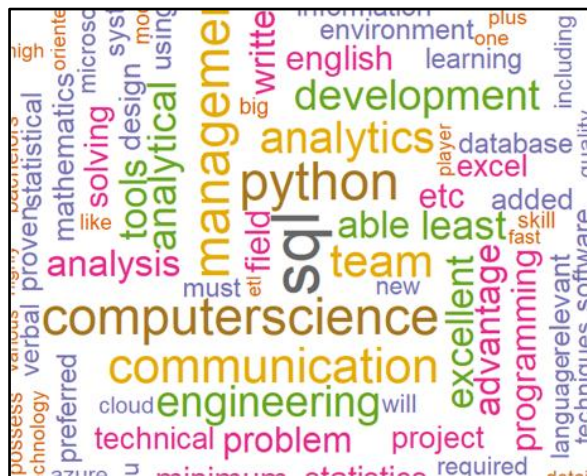

Fig. 2. The word cloud of the most common skills

Then, Table 1 shows the top 6 terms that have been acquired via frequency analysis of job requirements common skills where SQL topped with 110 frequencies followed by Python, Computer Science, Management, Communication, and Team for being a team player.

Table 1
Top 6 terms acquired via frequency analysis of job requirements

Term	n
SQL	110
python	85
computerscience	83
management	81
communication	78
team	73

Next, Figure 3 shows a bar chart of the top hard skills that have been found from the job requirements in the job advertisement that is specifically for Data Scientists. The bar chart shows that SQL is the most wanted skill among the companies in Malaysia, followed by Python, and having a background in Computer Science study.
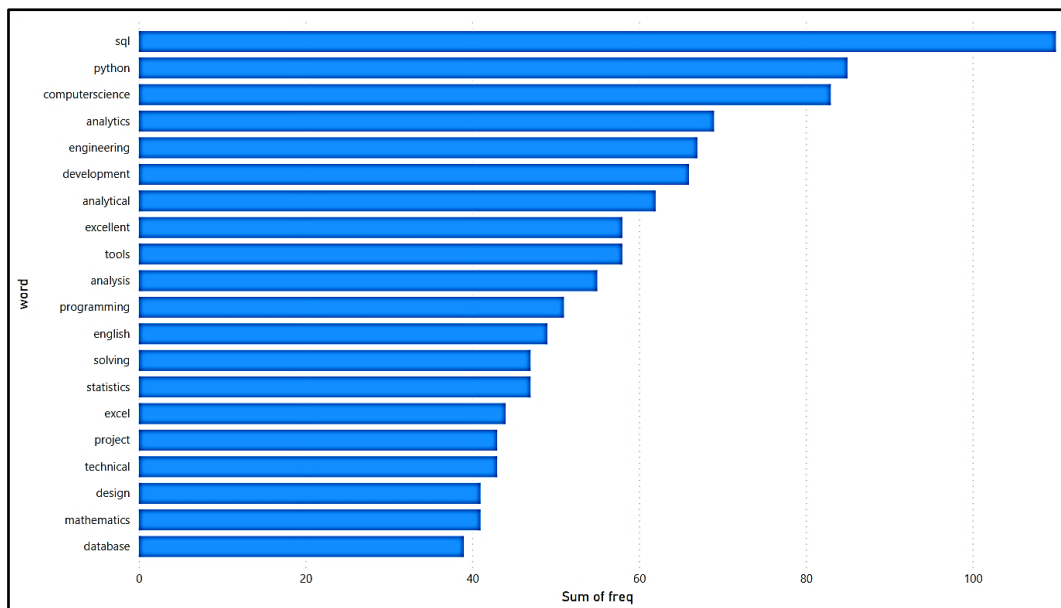
Fig. 3. The hard skills needed for Data Scientists

Then, in Figure 4, the bar chart shows a chart of the soft skills that are needed by job seekers. The result shows that having good management skills is what is needed the most, followed by having good communication skills.
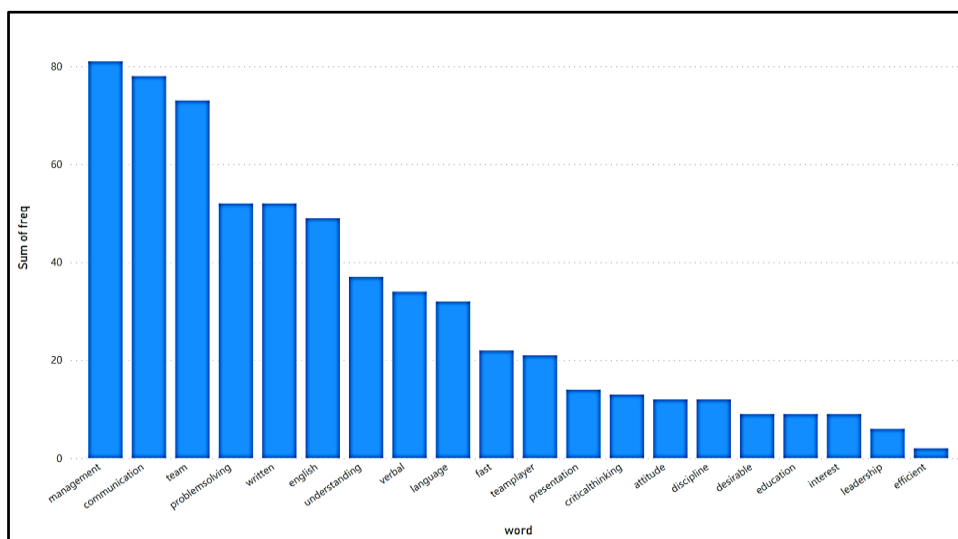

Fig. 4. The soft skills needed for Data Scientists

Figure 5 shows a pie chart of which location in Malaysia has the most Data Scientist job. The result showed that Kuala Lumpur has the most job openings for Data Scientists, which covers more than 50% of the job advertisement.
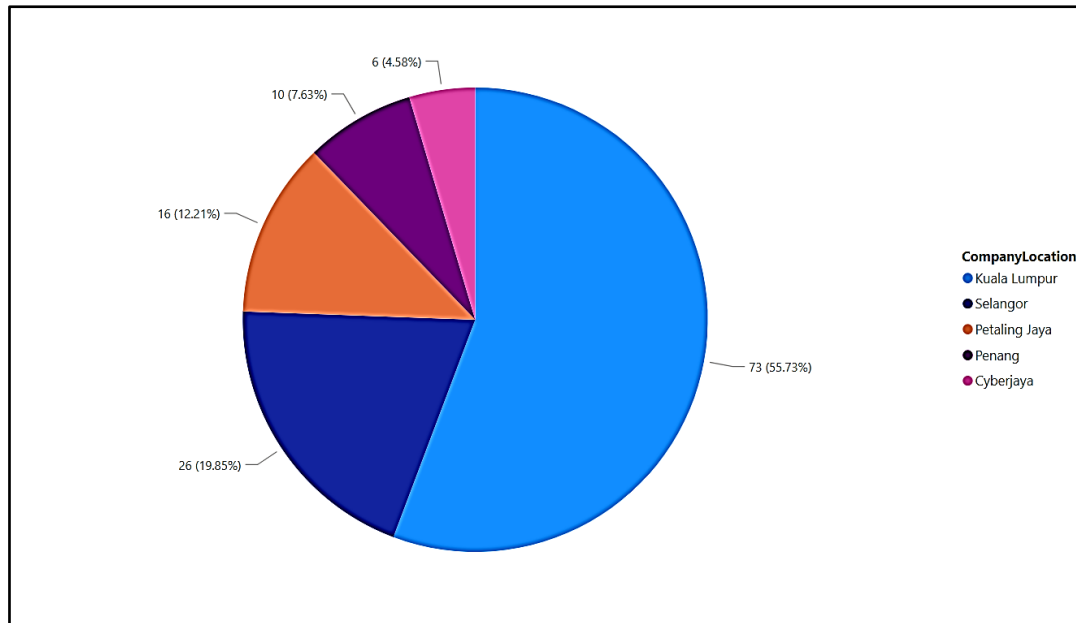
Fig. 5. The location of companies in Malaysia advertised Data Scientist job

According to Krastev [30], a Data Scientist in India should be proficient in Microsoft Excel, have a solid understanding of mathematics and statistics, be comfortable using Power BI and Tableau, and have familiarity with programming languages such as Python and R. Meanwhile, MDS @ Rice [31] stated that Data Scientists in the United States should be proficient in complex arithmetic and statistics, computer languages such as Python, R, and SQL, and the ability to create visual frameworks using Tableau or Python visualization.

Table 2 depicts the comparative skills required by the three countries. According to the results of this study, the competency skills of a Data Scientist in Malaysia are comparable to those of Data Scientists in the United States and India.

Table 2
The comparison skills needed of the three countries

Feature	Malaysia	USA		India	
Hard Skills	SQL, Python, Excel	Python, R, SQL		Excel, Python, R	
Soft Skills	Management, Communication, Team Player	Communication, Business Acumen	Curiosity,	Communication, Business Acumen	Curiosity,
Data Visualization	Power BI, Tableau	Python, Tableau		Tableau, Power BI	

4. Conclusions

This study presents a procedure of web crawling and web scraping job advertisements, extracting data using text mining and visualizing the results of the common skills of Data Scientists in Malaysia. As mentioned, Data Science is still vague among companies, our study could help universities and recruiters to standardize the role of a Data Scientist while job seekers could enhance the skills required to become a data scientist.

The study is significant as it will help job seekers, who are eligible, to apply as Data Scientists in Malaysia with all the skills needed to be employed. It will also help the companies to understand better what the most significant skills a Data Scientist needs to look for. Next, businesses can be more advanced with the skills of Data Scientists to help them with analytical approaches as well as helping

leaders in decision-making. Our findings could also help schools, universities, as well as companies regulate the competency skills of Data Scientists.

Acknowledgement

References
[1] TalentLyft. "What Is Job Advertisement ?" Recruiting and Hiring Resources, n.d. https://www.talentlyft.com/en/resources/what-is-job-advertisement

[2] DataMites. "DataMites Global Training Institute - Data Science, AI, Machine Learning, Python." Datamites Global Training Institute - Data Science, AI, Machine Learning, Python, n.d. https://datamites.com/

[3] ETHRWorld and www.ETHRWorld.com. "11.5 Million Data Science Jobs Will Be Created Globally by 2026: Report." *ETHRWorld.Com*, 2021. https://hr.economictimes.indiatimes.com/news/industry/11-5-million-data-science-jobs-will-be-created-globally-by-2026-report/87800256

[4] Fayyad, Usama, and Hamit Hamutcu. "Toward foundations for data science and analytics: A knowledge framework for professional standards." *Harvard Data Science Review* 2, no. 2 (2020).

[5] Khaouja, Imane, Ismail Kassou, and Mounir Ghogho. "A survey on skill identification from online job ads." *IEEE Access* 9 (2021): 118134-118153.

[6] Brooks, Nita G., Timothy H. Greer, and Steven A. Morris. "Information systems security job advertisement analysis: Skills review and implications for information systems curriculum." *Journal of Education for Business* 93, no. 5 (2018): 213-221.

[7] Malgwi, Charles A., Martha A. Howe, and Priscilla A. Burnaby. "Influences on students' choice of college major." *Journal of education for business* 80, no. 5 (2005): 275-282.

[8] Brookshire, Robert G. "Letter from the editor: Strategies for fighting declining enrollments." *Information Technology, Learning, and Performance Journal* 24, no. 1 (2006): I.

[9] UCB-UMT. "What Is Data Science? | The Data Science Career Path," 2023. https://ischoolonline.berkeley.edu/data-science/what-is-data-science/

[10] IBM. "*The Data Science Skills Competency Model*." New York, United States of America: (2020). https://www.ibm.com/downloads/cas/7109RLQM

[11] SAS. "What Is a Data Scientist?," https://www.sas.com/en_ca/insights/analytics/what-is-a-data-scientist.html

[12] IBM"What Is Data Science?." https://www.ibm.com/topics/data-science

[13] Medium. "Data Analyst vs Data Engineer vs Data Scientist" https://medium.com/@aishwaryasharmaece/data-analyst-vs-data-engineer-vs-data-scientist-7e278505b08

[14] "Tables Created by BLS : U.S. Bureau of Labor Statistics," 2023. https://www.bls.gov/oes/current/oes152051.htm

[15] Fajaryati, Nuryake, Budiyono, Muhammad Akhyar, and Wiranto. "The employability skills needed to face the demands of work in the future: Systematic literature reviews." *Open Engineering* 10, no. 1 (2020): 595-603.

[16] Fadhilah, A. R., M. N. Noraini, A. M. Noreliza, and Kok Hua Chua. "Analysis on the employability skills of UNITEN graduates in engineering program." In *2018 IEEE 10th International Conference on Engineering Education (ICEED)*, pp. 34-38. IEEE, 2018.

[17] Rahmat, Normala, Abdul Rahman Ayub, and Yahya Buntat. "Employability skills constructs as job performance predictors for Malaysian polytechnic graduates: A qualitative study." *Geografia* 12, no. 3 (2016).

[18] Yasin, Sharifah Nurulhuda Tuan Mohd, Mohd Fauzi Mohd Yunus, Jamil Abd Baser, Nur Bahiyah Abdul Wahab, and Azman Hasan. "Evaluation process dimensions to the diploma in mechatronics engineering." *Procedia-Social and Behavioral Sciences* 228 (2016): 272-277.

[19] Sarfraz, Isra, Diana Rajendran, Chandana Hewege, and Mohan Dass Mohan. "An exploration of global employability skills: a systematic research review." *International Journal of Work Organisation and Emotion* 9, no. 1 (2018): 63-88.

[20] Alexander S. Gillis and Mary K. Pratt, "Hard Skills," *CIO*, 2023. https://www.techtarget.com/searchcio/definition/hard-skills#:~:text=Hard%20skills%20are%20specific%20abilities,required%20for%20a%20specific%20job

[21] Sopa, Ardian, Masduki Asbari, Agus Purwanto, Priyono Budi Santoso, Dhaniel Hutagalung Mustofa, Siti Maesaroh, and Riza Primahendra. "Hard skills versus soft skills: which are more important for Indonesian employees innovation capability." *International Journal of Control and Automation* 13, no. 2 (2020): 156-175.

[22] Pumble Learn. "Statistics on Workplace Communication in 2023," 2023. https://pumble.com/learn/communication/communication-statistics/

[23] Fadhil, Sally Saad, Ramlee Ismail, and Alhamzah Alnoor. "The influence of soft skills on employability: a case study on technology industry sector in Malaysia." *Interdisciplinary Journal of Information, Knowledge, and Management* 16 (2021): 255.

[24] World Economic Forum, "The Future of Jobs Report 2020," 2023. https://www.weforum.org/reports/the-future-of-jobs-report-2020/

[25] GeeksforGeeks. "What Is Web Scraping and How to Use It," *GeeksforGeeks*, 2023. https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/

[26] Pillai, Priam, and Dhiraj Amin. "Understanding the requirements of the Indian IT industry using web scrapping." *Procedia Computer Science* 172 (2020): 308-313.

[27] Smaldone, Francesco, Adelaide Ippolito, Jelena Lagger, and Marco Pellicano. "Employability skills: Profiling data scientists in the digital labour market." *European Management Journal* 40, no. 5 (2022): 671-684.

[28] Schedlbauer, Jürgen, Georgios Raptis, and Bernd Ludwig. "Medical informatics labor market analysis using web crawling, web scraping, and text mining." *International Journal of Medical Informatics* 150 (2021): 104453.

[29] Tableau. "What Is Data Visualization? Definition, Examples, And Learning Resources," https://www.tableau.com/learn/articles/data-visualization

[30] Krastev, Ned. "How to Become a Data Scientist in India: Salary, Job Outlook & Skills." *365 Data Science*, 2023. https://365datascience.com/career-advice/career-guides/data-scientist-india/

[31] "12 Essential Skills Required for Data Scientists | MDS@Rice," https://csweb.rice.edu/academics/graduate-programs/online-mds/blog/data-science-skills#:~:text=Top%206%20Technical%20Skills%20in%20Demand%20for%20Data,.%20.%20%206%206.%20Data%20Visualization%20%26%20Storytelling%20.