Leveraging on Job Vacancies Advertised Online to Analyse Malaysian Labour Market Using Big Data Analytics

Rabi'atul'adawiah Shabli
Department of Statistics, Malaysia (DOSM)

Abstract

Big data analytics are now a common tool being used by economists and statisticians to address socioeconomic challenges as well as to complement existing labour market information in the country. Based on information retrieved from big data, there are possibilities to obtain huge amounts of multidimensional, diversified and granular data. As a consequence of the rapid usage of the internet, recruiting processes have switched to online applications and as a result, provide accessibility to vacancies data. The aim of this paper is to examine data gathered from job vacancies advertised online for the use of research in labour economics and skills development. Although extracted real-time, the processed data is available on a quarterly basis. In this paper, the information is analysed using several methods, namely descriptive analysis and time-series analysis. In addition, this paper assesses the main characteristics of the labour market data based on scraped vacancies data starting from the first quarter of 2020.

Keywords: big data, analytics, labour market, job vacancies, online job postings

1. Introduction

The use of open source data with new methodologies particularly based on real-time data is one of the current priorities to obtain new sources of data. This is an area of interest with the usage of more innovative data sources and application of big data analytics to produce labour market indicators. Many types of analysis can be done as there are various online tools available to further study on job seekers, occupational categories and job matching. Additionally, given the increasing likelihood of online hiring and the expansion of internet access, there are new opportunities to gather and analyse data on the supply and demand of labour market using the internet. Job advertisement published through the online portal is a direct communication between employers to the potential employees can gather a vast amount of observations. Recently, one of the most prominent approach of online generated data has been the job vacancies advertised online which can be obtained from online job portals.

This paper outlines the new methodology of data collection by the Department of Statistics Malaysia to publish frequent and granular information of job vacancy advertised online by using web scraping to come out with another labour market indicator. During the expansion of the economy, there is normally a high demand for hires which in turn would benefit people who are looking for a job. By using big

data, this will enrich the variables obtained for the analysis of job vacancies as employers usually specify the set of skills required for the job with detailed description of the job responsibilities need from the job seekers.

The growing demand for this labour market information is addressed by the government by looking at categories of occupation, economic activity and state. The information is scraped on a daily basis through several popular recruitment portal in the country and the reporting was published on a quarterly basis focusing on skilled and low-skilled occupations. The first data was released for the third quarter of 2019 vacancies and is available at 4-digit level of occupation using Malaysia Standard Classification of Occupations (MASCO) 2020. The detailed information by types of occupation may help in identifying the skill shortage and current occupation are in demand by the employer. The information can be used to indicate changes and demand of the employers by type of occupation and by location and provide valuable labour market information in supporting policy decision specifically related to human capital.

The advertisement was scraped daily and coded to an occupation classification based on MASCO 2020 by using machine learning and artificial intelligence. The process is automatically coded each job vacancy advertised online to a specific occupation based on key words or phrases obtained in the advertisement. In addition, the machine learning also has been developed to automatically detect and removes duplication between the advertisement according to designated rules.

This paper is divided into several sections; literature review, methodology, results, discussion and limitation.

2. Literature Review

This section reviews several literatures on the usage of scraping information for job vacancies advertised online as one of the official labour market indicators. According to Oxford Learner's Dictionaries (2023), big data refer to a set of information that is too large or too complex to handle, analyse or use with standard methods. In addition, SAS Institute Inc. (2022) stated that big data analytics examine large amounts of data to obtain unstructured insights and enable data analysis using business intelligence solutions. Based on Tableau Software (2023), any organisation can be benefited from big data as it allows more data in multiple formats from various sources to be analyse at a faster rate by providing in cost savings, product development and market insights.

In the European Union, there is one agency namely European Centre for the Development of Vocational Training (CEDEFOP) that was founded in 1975 addressed on the importance to enhance and disseminate skills and qualifications to monitor indicator for skills shortage and skills mismatch by implementing the

adoption of digitalisation. As mentioned by Manyika et al. (2015), the percentage of using online job portals has been increasing over the time. Furthermore, in-depth monthly analysis of online job vacancies has been done by Foundit (2023), through Foundit Insights Tracker to give an overview of the occupation, job responsibilities, salary range and competencies needed by the employers. In 2013, a study was done by Kuhn and Shen relating to online job vacancies on the gender discrimination in the recruitment process in the Chinese labour market by using a web crawler.

Moreover, statistics on job vacancies advertised online is one of the main labour market indicators as Silverstone and Wall (2008) specified employers are using the online recruitment platform to look for potential job seekers. The information also can be used to study the relationship between job vacancies and unemployment as mentioned by Bleakley and Furher (1997).

3. Methodology

Data on online job vacancies is currently available from several sources. Web crawler and web scraping by the individual establishment from recruitment portal is one the most frequent methodology used to gather the information. The daily web scraping to obtain job information is done by using machine learning models and can be classified to national standard classification. **Exhibit 1** shows the overview of the system architecture pipeline to develop the big data analytics for online job vacancies.

Exhibit 1: System architecture pipeline of job vacancies advertised online
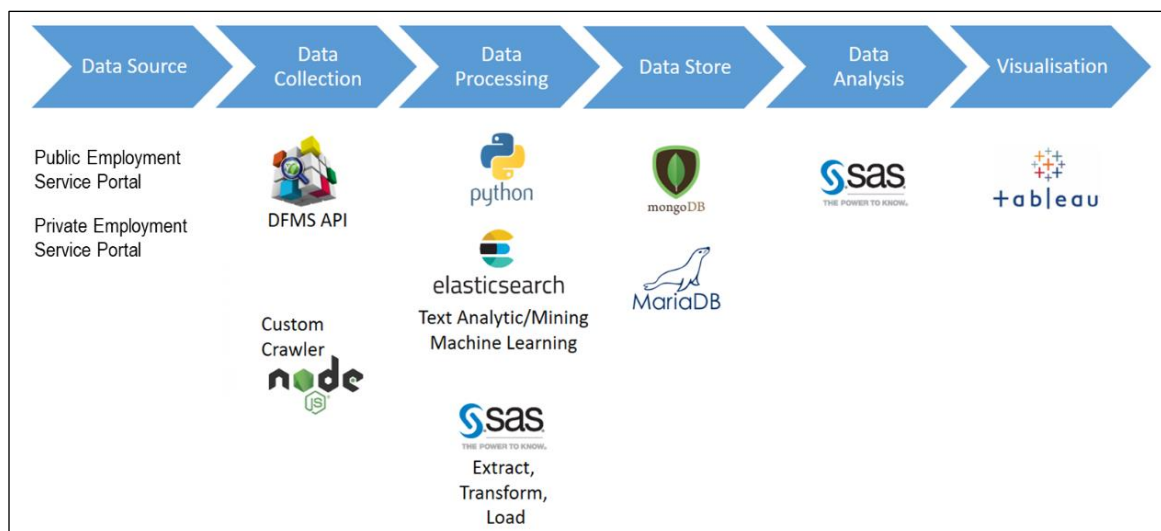
The process to develop the big data analytics for online job vacancies can be simplified as below:

 a. Data Source: To extract job vacancies advertised online from the selected private recruitment portal.

b. Data Collection: The information from the selected private recruitment portal will be extracted by using a web scraper. Each job vacancy advertised online has been stored and coded with a unique identifier.

c. Data Processing: Information about job vacancies gained from the portal is normally unstructured with a lot of noisy data. Extracted data will be going through cleansing, text processing, text mining, library mapping by custom Machine Learning program. The unique identifier will enable the process of deduplication in terms of job vacancies advertised by the same employers to various channels, considering certain rules to remove the same advertisement using machine learning algorithms.

d. Data Store: MongoDB will be used to store the scrapped raw job data while MariaDB will be used to store the final data warehouse.

e. Data Analysis: To use SAS software like SAS Base, SAS EG to perform data exploration, data mining and data analysis. Further aggregated information on the granular data used fuzzy matching based on Python software to identify job categorisation, economic activity, salary and wages, job responsibilities and the competencies required for each type of occupation.

f. Visualisation: Data will be exported to Tableau Desktop for dashboard visualisation.

4. Result

There are five main types of analysis in this study namely:
a. Job vacancies advertised online, Q3 2019 - Q4 2022;
b. Job vacancies advertise online by occupation, Q4 2022;
c. Most in-demand job vacancies advertised online, Q4 2022;
d. Job vacancies advertised online by economic activities, Q4 2022; and
e. The most frequent skill of job vacancies advertised online, Q4 2022.

4.1 Job vacancies advertised online, Q3 2019 - Q4 2022

Chart 1: Job vacancies advertised online, Q3 2019 - Q4 2022
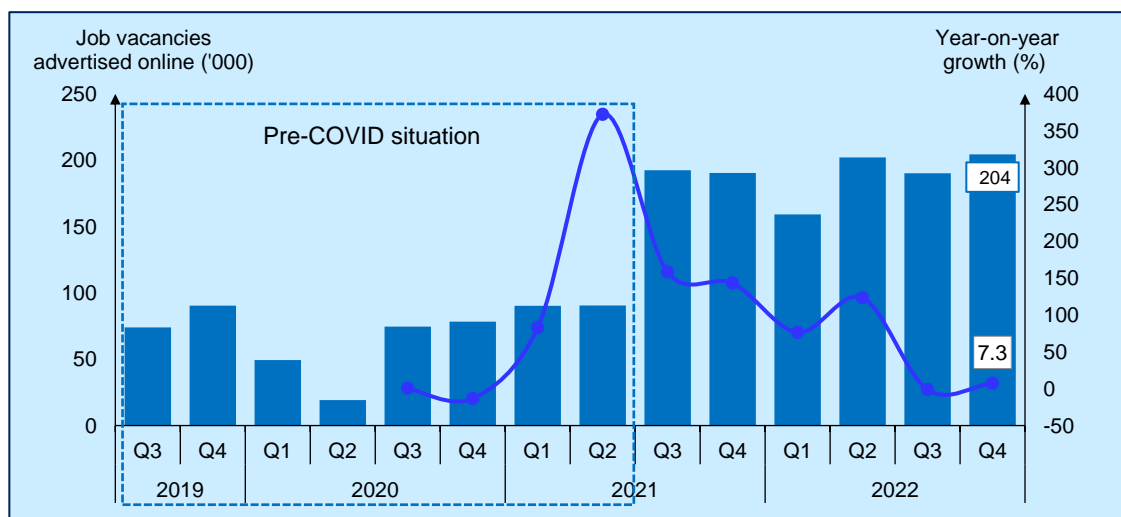
This chart stated the job vacancies advertised online based on a quarterly basis starting from the third quarter of 2019 to fourth quarter of 2022. As stated in **Chart 1**, job vacancies were below 100,000 advertisement posts during the pre-COVID situation. There was a significant increase in the job vacancies advertised online since the third quarter of 2021 which was after the pandemic. Based on the chart, there was a consistent level of job vacancies after the health crisis, indicating an improvement in the labour market scenario. The highest posting was in the latest released during the fourth quarter of 2022 with 204,420 vacancies.

4.2   Job vacancies advertised online by occupation, Q4 2022
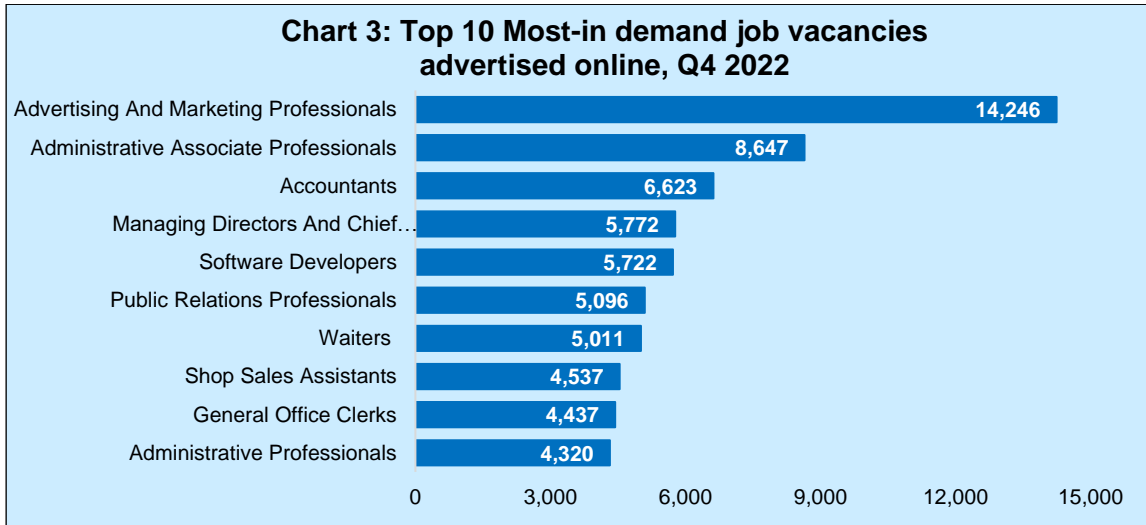

Chart 2: Job vacancies advertised online, Q4 2022
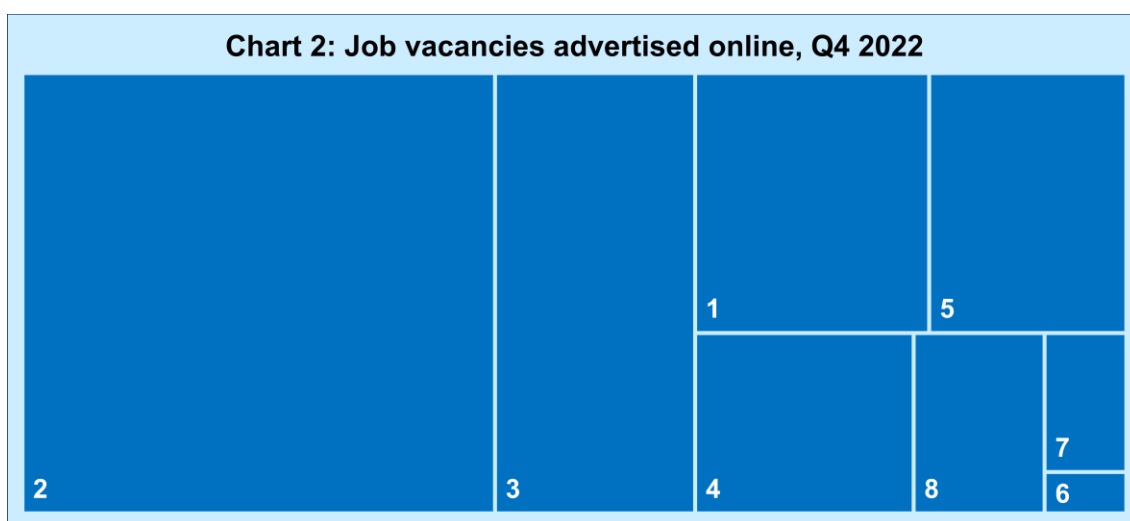
Notes:
1: Managers; 2: Professionals; 3: Technician and Associate Professionals; 4: Clerical Support Workers; 5: Service and Sales Workers; 6: Skilled Agricultural, Forestry and Fishery Workers; 7: Craft and related Trades Workers; and 8: Plant and Machine-Operators and Assemblers.

Chart 2 shows job vacancies advertised online according to the national classification of occupation namely MASCO 2020. As for the second quarter of 2022, the majority of job vacancies were under skilled occupation. Job vacancies in the Professionals made up the largest share at 42.8 per cent or equivalent to 87,494 vacancies. On the other hand, Technician and Associate Professionals category comprised a share of 18.1 per cent with 37,066 vacancies. Managers accounted for 12.5 per cent from overall job advertisement with 25,586 vacancies. However, the lowest most-in demand occupation was in the Skilled Agricultural, Forestry and Fishery Workers category with 1,424 vacancies or 0.7 per cent.

4.3   Most in-demand job vacancies advertised online, Q4 2022

Chart 3: Top 10 Most-in demand job vacancies advertised online, Q4 2022

Occupation	Value
Advertising And Marketing Professionals	14,246
Administrative Associate Professionals	8,647
Accountants	6,623
Managing Directors And Chief…	5,772
Software Developers	5,722
Public Relations Professionals	5,096
Waiters	5,011
Shop Sales Assistants	4,537
General Office Clerks	4,437
Administrative Professionals	4,320

Another way to analyse job vacancies from the web scraping was detailed information by four (4) digit MASCO 2020 which stated the most in demand occupation during the reference period. From **Chart 3**, the most popular occupation was Advertising and Marketing Professionals which are mentioned more than 14,000 times. Administrative Associate Professionals came in second with 8,647 advertisement posts while Accountants and Auditors recorded 6,623 vacancies. In addition, Managing Directors and Chief Executive and Software Developers was mentioned 5,772 and 5,722 times respectively in the advertisement.

4.4   Job vacancies advertised online by economic activities, Q4 2022

Chart 4: Job vacancies advertised online by economic activities, Q4 2022
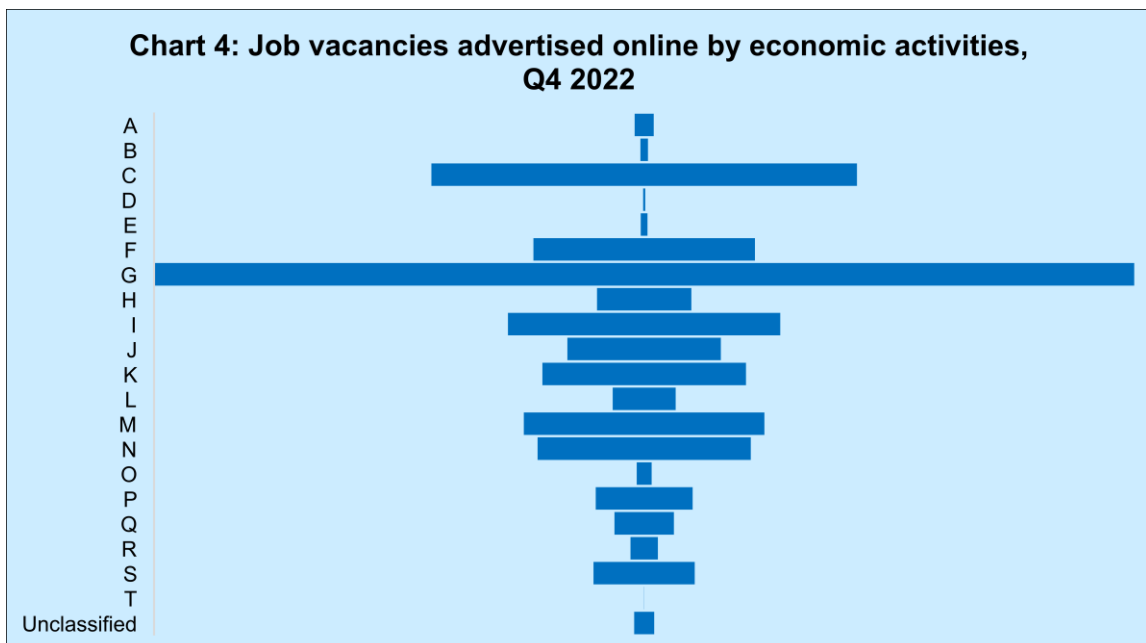
Notes:
A: Agriculture, Forestry & Fishing; B: Mining & Quarrying; C: Manufacturing; D: Electricity & Gas; E: Water Supply & Sewerage; F: Construction; G: Wholesale & Retail Trade; H: Transportation & Storage; I: Accommodation & Food Service Activities; J: Information & Communication; K: Financial & Insurance/

Takaful Activities; L: Real Estate Activities; M: Professional, Scientific & Technical Activities; N: Administrative & Support Service Activities; O: Public Administration & Defence; P: Education; Q: Human Health & Social Work Activities; R: Arts, Entertainment & Recreation; S: Other Service Activities; and T: Activities of Households as Employers.

Looking at the disaggregation of job vacancies advertised online by economic activities, the number of vacancies varied significantly across the industry category. Almost one third of the vacancies were registered in the Wholesale and Retail Trade activities with 62,113 vacancies or 30.4 per cent from overall advertisement. There are 26,974 vacancies (13.2%) posted in the Manufacturing while Accommodation and Food Service Activities stated 17,261 vacancies (8.4%). Subsequently, job vacancies advertised online for Professional, Scientific and Technical Activities and Construction were 15,252 and 14,040 respectively as shown in **Chart 4**.

4.5   Most frequent skill of job vacancies advertised online, Q4 2022

Table 1: Top 5 Most frequent skill of job vacancies advertised online, Q4 2022

No.	Soft skills (130 types)	Hard skill (947 types)
1	English	Mandarin
2	Communication	Analysis
3	Bahasa Malaysia	Microsoft Office
4	Planning	Filing
5	Marketing	Project management

There are two types of skills namely soft skills and hard skills as stated in **Table 1**. Based on job vacancies advertised online by the employers, there is a list of 130 types of soft skills that can be obtained through the portal. As for the soft skills, the employers prefer the employees to have a good communication skill. The most frequently mentioned soft skills in the job advertisement were English, followed by communication, Bahasa Malaysia, planning and marketing. On the other hand, the employers have stated 947 categories of hard skills that are needed from the job seekers. According to the job advertisement, Mandarin was listed as the most required skill. In addition, analysis ranked second followed by Microsoft Office, filing and project management.

5.   Discussion and Conclusion

The information gathered for job vacancy advertisement online by using a web scraping is an administrative dataset that is largely depends on the job postings collected may reduce the administrative costs. By selecting a recruitment portal's that is dominant in the market, it can be regarded as a reliable source that had an impact on the potential of supply and demand for labour market studies. This will enable the researcher or policy makers to analyse the number of job openings in a

particular occupational category based on the information collected. Further analysis might also be possible to predict the unemployment rate using the statistics from job vacancy advertised online.

However, in order to release the statistics as an official data sources in the country, there should be a mixture of analytical process using big data coupled with text mining and expert judgement. There are possibilities of concentration on selected occupations and also by economic activities as there might be tendency for certain establishments that were heavily reliant on the usage of job vacancies advertised through the internet. Online recruitment portals are able to offer a broader variety of options as well as increasingly more sophisticated tools, so they are likely to become an important tool for job matching. This will likely improve coverage of the population of job openings and businesses that participate in it.

In the meantime, there are possibilities of bias over coverage of high skilled and semi-skilled job vacancies for certain occupation or economic activity as some employers rely more on offline sources, since not all job vacancy is advertised through online. As the online job portals also covering low-skilled jobs, it may be useful to report the missing information related to this category in order to have a comprehensive coverage of online job vacancies in Malaysia. However, it should be noted that employers prefer to use traditional methods of recruitment for low-skilled category. This may be the primary cause of certain jobs are overrepresented while possibilities of other jobs being underrepresented in statistics of online job vacancies.

6. Limitation

The major challenges of using online job vacancies data is the unknown population (N) of job vacancies as well as its practical structure. Even though this kind of reporting is common, the majority of hiring still happens privately or through unofficial channels or networks. For instance, large establishments may frequently use internal resources prior to posting a position on the public market. In smaller cities or communities where, labour market participants have stronger connections and relations, the jobs may be offered to recognise applicants first. Competencies profiles based on online job postings would only be biased if it were possible to assume that only specific companies within a given industry tend to post openings and that these companies have very different expectations.

References

Anne Fale, Carmel Tuya and Dafydd Davies (2010), Measuring Job Vacancies in New Zealand Through Jobs Online, Department of Labour, Ministry of Business, Innovation and Employment, New Zealand.

Andrea De Mauro, Marco Greco, Michele Grimaldi, Paavo Ritala (2017), Human resources for Big Data professions: A systematic classification of job roles and required skill sets, www.elsevier.com/locate/infoproman

Lucia Kureková, Miroslav Beblavý, Anna-Elizabeth Thum (2013), Online job vacancy data as a source for micro-level analysis of employers' preferences. A methodological enquiry, First International Conference on Public Policy (ICPP), Centre for European Policy Studies

Oleksii Romanko and Mary O'Mahony (2022), The Use of Online Job Sites for Measuring Skills and Labour Market Trends: A Review, Office for National Statistics
Joseph A. Rios, Guangming Ling, Robert Pugh2, Dovid Becker, and Adam Bacall (2019), Identifying Critical 21st-Century Skills for Workplace Success: A Content Analysis of Job Advertisements, Educational Researcher

Oxford University Press (2023),
https://www.oxfordlearnersdictionaries.com/definition/english/big-data?q=big+data

SAS Institute Inc. (2022), https://www.sas.com/en_my/insights/analytics/big-data-analytics.html

European Centre for the Development of Vocational Training (2023),
https://www.cedefop.europa.eu/en/about-cedefop/what-we-do