Journal of Theoretical and Applied Information Technology
31st October 2021. Vol.99. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASSOCIATION ANALYSIS ON JOB VACANCY CRITERIA

[1]ALBERT HENDERSON, [2, 3]IMAN HERWIDIANA KARTOWISASTRO

[1]Author, Department of Computer Science, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jl. Kebon Jeruk Raya No. 27, Jakarta, 11530, Indonesia

[2]Author. Department of Computer Science, BINUS Graduate Program – Doctor of Computer Science, Bina Nusantara University, Jl. Kebon Jeruk Raya No. 27, Jakarta, 11530, Indonesia

[3]Author. Department of Computer Engineering Faculty of Engineering, Bina Nusantara University, Jl. Kebon Jeruk Raya No. 27, Jakarta, 11530, Indonesia

E-mail: [1]albert.henderson@binus.edu, [2, 3]imanhk@binus.edu

ABSTRACT

IT Based job openings have a various criterion of needed skills. University graduates are required to fulfill the industry needed. The purpose of this work is to help improve university's graduate awareness about criteria mostly need by industries. By using association analysis and utilizing existing job openings data are expected to find any association in the several vacancy group with the expected skills by the industry. The technique will be used is association analysis. The process is carried out in four major steps, consist of Data Collection, Pre-Processing Data, Text Processing, and Association Analysis. Produced rules grouped into four, based on support and lift value. Result of this work shows that only criteria from IT/MIS-Quality Assurance (2016), IT/MIS-Administrator (2017), Engineering-Computer, IT/MIS-Multimedia Designer, IT/MIS-Product Management, IT/MIS-Technical Support (2018), and IT/MIS-Technical / Functional Consultancy (2019) that qualified as mostly appear criteria also often appear together. With the discovered knowledge of this work, related students and university faculties can obtain information in a better way from existing job openings and can fulfill the industry terms and needs based on the needed criteria.

Keywords: *Association Analysis, Association Rule, Data Mining, Job Vacancy*

1. INTRODUCTION

The workforce rate in Indonesia continues to increase yearly. In 2017 the workforce around 131,55 million people. In 2018, the workforce increases 1,82% from 2017 or around 133,94 million people. In 2019, the workforce increases 1,67% from 2018 or around 136,18 million people. The increased workforce is proportional to an increase in the number of universities graduates. In 2017, number of universities graduates are 11,59 million. In 2018, it grows 1,03% to be around 11,71 million. In 2019, it grows 7,69% to be around 12,61 million [1][2].

BINUS University is one of many universities in Indonesia. To ensure and prepare all graduates can be employed in industry, BINUS University has a business unit called BINUS Career Center. BINUS Career Center as a place to bring together companies from various fields of business from local to international scope with students and alumni. Therefore, BINUS University can ensure minimum 80% of all graduates on every period have worked or being an entrepreneur.

Job vacancies provided by the company, require applicants to meet their needed criteria. There is a challenge that must be faced by all graduates including BINUS University graduates. To be able to master it, and answer what the industry mostly needed criteria, it is necessary to find knowledge in existing job vacancies.

The purpose of this work is to help improve university's graduate awareness about criteria mostly need by industries. The technique that will be used is association analysis. We'll be analyzing the pattern of job criteria on each job position. Therefore, the result of job criteria can be used as references for related faculty or college student to gain knowledge what the industry needed. This could make an impact in long term to prepare university graduates to be able to compete in terms of applying for jobs.

2. MATERIAL AND METHOD

Association analysis often used for analyzing customer behavior in the case of a grocery items or also called Market Basket Analysis (MBA). It works by analyzing pattern through each item of the transaction, to understand purchase behavior of items that always bought together [3]. Association analysis also can be used as a support decision-making process from hidden relation in wide datasets [4] Association rules are one of major technique in data mining by finding frequent pattern, association, correlation among sets of item or object in transactional database. Rule mining also helped many sectors on decision making regarding selection of many applications [5] The percentage relation produced between one item to the other gives a new knowledge for the determiner to make a decision [6].

One of the most popular research when it comes to association analysis method is Market Basket Analysis. The objective of market basket analysis is to find things that often bought together by the history of transaction. The discovered pattern able to provide knowledge for the store to make an arrangement on the product placement close to each other that could lead to increased sales of the correspondent products and directly increase the store revenue. This also can be a signal too for the store to stock up the corresponding products to avoid shortages [7]. Another research conducted at a Muji brand sales shop has also been carried out in Japan. The researchers conducted associations between types of stores to sales ratio, the total sales of each store that was widespread. This could lead to buyer behavior on each branch sales shop and most desirable item in each store location [8].

In Thailand, a study was conducted to detect vehicles that were indicated to have committed crimes. In conventional systems the vehicle is only identified by using the color, brand, and type of vehicle, making it difficult to distinguish. By using association analysis, the data is combined with the Department of Land Transport (DLT) database to get the vehicle frame number. With so, detecting illegal vehicles or vehicles used by a criminal can be more easily found [9].

Association analysis can be applied on medical purpose. Based on the existed historical patient information dataset like age, gender, heart rate, blood pressure, etc., it can identify what variable on a certain age matter the most to the patient that can lead to cardiovascular disease (CVD). This could help the insurance company to determine the cardiovascular system status of the applicant [10]. Similar research also has been conducted to be an early detection for patient that had a risk of gestational diabetes mellitus. Produced rules based on several basic risk factor of pregnant ladies that can lead to gestational diabetes mellitus [11].

Unemployment and overqualified jobs are becoming serious problem [12]. Based on the research that has been done by others, association analysis could be implemented on to many uses with the same purpose to find itemset that often appear together from dataset. That way we present in this paper is an association analysis based on job vacancy to find out job criteria mostly needed by the industry. Job vacancy criteria provided by the company are stored in long and unstructured text. This would consume lot of time if were studied and analyzed each one of the vacancies. To be able to gain knowledge of the most needed criteria also happened to appear together from existing job vacancy, association analysis method is utilized. Association analysis method able to study patterns based on job vacancy criteria. Before the pattern can be found, text processing is needed to simplify unstructured text of vacancies criteria. This will be left with important keywords of criteria to be found the pattern using association analysis. We'll be analyzing the pattern of job criteria on each job position. In order to obtain more comprehensive and detailed information, the association analysis process needs to be done by each year of the vacancies.

The method we used to achieve those result is consisted of four major steps. Method workflow can be seen on Figure 1.

Journal of Theoretical and Applied Information Technology
31st October 2021. Vol.99. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

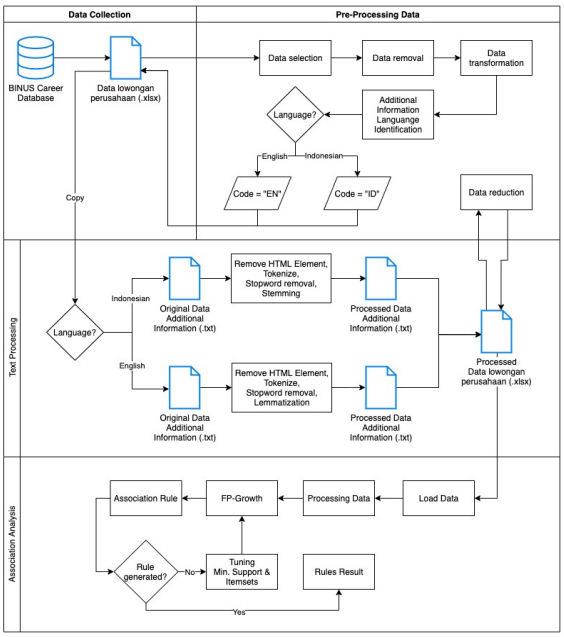

Figure 1: Workflow of job vacancy association analysis

2.1 Data Collection

Job vacancy data were submitted by registered companies throughout the year and stored on SQL database. Job vacancy will be used are ranging from 2016 until 2019. Then, data then stored as Excel file form. Total obtained vacancy data as many as 14,425 rows of data.

Table 1: Detail Row of Data 2016-2019

Year	Total Row of Data
2016	3314
2017	2021
2018	7501
2019	1589

Vacancy data consist of 23 variables, as follows: *Employer ID, Employer Name, Industry, AP ID, Title, Position, Job Type, Work Location, Employment Status, Jobseeker Gender, Min Salary, Max Salary, Work Experience, Minimum Education Degree, City, Country, Additional Information, Min GPA, Max Age, Number Employee Needed, Major of Employee Needed, Expired Date, Posted Date*. Not all variables will be used on association rule, some of them will be used only on pre-processing stage, and other will be used for text processing stage. Example of unprocessed job vacancy can refer on table 2 as written in English and table 3 as written in Indonesian.

Table 2: Example of unprocessed job vacancy written in English

Variable	Stored Value
Employer ID	20160005
Employer Name	PT. RST
Industry	Computer/IT
AP ID	2016010022
Title	Senior Programmer/Developer VB
Position	IT/MIS-Programmer
Job Type	Employee
Work Location	Office
Employment Status	Full-Time
Jobseeker Gender	-
Min Salary	3000000
Max Salary	4000000
Work Experience	0
Minimum Education Degree	Bachelor (S1)
City	Jakarta
Country	Indonesia
Additional Information	Qualification : - Age: in between 23 up to 35 years old - Background education: DIII/IV or S1 (Management Informatics/Technical Informatics/System Informatics or equivalent) - Minimum 3 years of working experience - Have knowledge of applications in hospitals or healthcare are preferred - Have understanding of algorithms and programming - Experienced with Microsoft Visual Basic 6, Microsoft SQL Server 2005/2008 database, Crystal Report Version 8 or more and Microsoft Visual Studio 2008 - Has the ability of other programming languages (PHP / Java, etc.) is an advantage - Able to work under pressure, independently or within a team - Able to work hard, meticulous and disciplined - Accustomed to documenting the work well and cleanly - Have good communication skills Please email in details of your work experience, qualifications, present & expected salaries and contact number to recruitment@rst.net We regret that only shortlisted candidates will be contacted
Min GPA	None
Max Age	35
Number Employee Needed	1
Major of Employee Needed	,Computer Science,
Expired Date	04/02/16
Posted Date	07/01/16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Example of unprocessed job vacancy written in Indonesian

Variable	Stored Value
Employer ID	20120330
Employer Name	PT. UVW
Industry	Pharmaceutical
AP ID	2016020115
Title	Programmer Internship
Position	IT/MIS-Programmer
Job Type	Intern
Work Location	Office
Employment Status	Other
Jobseeker Gender	-
Min Salary	2500000
Max Salary	3000000
Work Experience	0
Minimum Education Degree	Bachelor (S1)
City	DKI Jakarta
Country	Indonesia
Additional Information	*Mahasiswa Tahap Akhir/Skripsi Bersedia untuk menjalani masa internship selama 6 bulan Memahami Pemrograman Joomla, Asp.net, Java Script, HTML Penempatan di Head Office (Kawasan Industri Pulo Gadung)*
Min GPA	None
Max Age	None
Number Employee Needed	1
Major of Employee Needed	,Computer Engineering,Computer Science,Information Systems,Information Systems,
Expired Date	31/12/16
Posted Date	10/02/16

2.2 Pre-Processing Data

Pre-processing data consist of five stages namely, data selection, data removal, data transformation, language identification, and data reduction. The data used in this research are vacancies aimed at information technology-based faculties. There are three faculties on BINUS University that considered as information technology-based faculty such as School of Computer Science [13], School of Information System [14], and Faculty of Engineering [15].

Data selection is carried out on the Position variable to select only relevant Positions for this research. Selected Position consist of 27 types, as follow: Engineering-Computer, Engineering-Hardware, Engineering-Network, Engineering-Software, IT/MIS-Administrator, IT/MIS-Application Specialist, IT/MIS-Computer Operation, IT/MIS-Hardware Design & Engineer, IT/MIS-Multimedia Designer, IT/MIS-Network Engineer, IT/MIS-Others, IT/MIS-Product Management, IT/MIS-Programmer, IT/MIS-Project Manager, IT/MIS-Quality Assurance, IT/MIS-SAP / Oracle, IT/MIS-Software Design & Engineering, IT/MIS-Software Testing, IT/MIS-System & Network Support, IT/MIS-System / Business Analyst, IT/MIS-System Engineer, IT/MIS-Technical / Functional Consultancy, IT/MIS-Technical Support, IT/MIS-Technical Writing, IT/MIS-Trainer / Faculty, IT/MIS-Web Developer, and IT/MIS-Webmaster.

Dataset retrieved from database sometimes happened to be duplicate. This is caused by data with relation one to many relationships on the dataset. The duplicate data can affect the association analysis result later. Then, data removal is needed to tidy up whole dataset from duplicate information. This step is quite important to do because duplicate data can affect the result later in rules that look frequently appear, even though only repeated the same data. Duplicate data were removed using Employer ID and AP ID as references. Employer ID as identifier of company and AP ID as identifier of each posted job position. Next step is data transformation. Data transformation used to alter gender code stored on system. “M” is saved for Male, “F” is saved for Female, and “-“ is saved for both gender. The transformation change those saved gender code from “M” to “Male”, “F” to “Female”, and “-“ to “BothGender”. This step is taken to avoid losing important job vacancy criteria at the text processing stage later.

Job vacancies posted by companies are vary in use of language. Major languages used are English and Indonesia. Each language using different method of text processing due to language characteristic. To be able separate job vacancies based on their written language, language identification is required. Language identification can be achieve using Google Sheet’s formula namely DETECTLANGUAGE(). This formula work by detecting Position value and return the language code. Code EN for English and ID for Indonesia. This language code then stored in a new variable called Language. This variable that would be use later as a sorting parameter text processing.

After several pre-processing stages, it’s time to do data reduction. This stage focusses on reducing variables that only used for identifier, timestamp, etc. Those variables don’t affect any result for association analysis because this kind of data do not have any information related to job

vacancy. This also help to load data faster on RapidMiner Studio because of lighter dataset.

Processed job vacancy data throughout pre-processing steps can be refer on Table 3 as written in English and Table 4 as written in Indonesian.

Table 4: Example of processed job vacancy written in English

Variable	Stored Value
Employment Status	Full-Time
Jobseeker Gender	Both
Min Salary	3000000
Max Salary	4000000
Work Experience	0
Minimum Education Degree	Bachelor (S1)
City	Jakarta
Country	Indonesia
Additional Information	education DIII IV S1 Management Informatics Technical Informatics System Informatics equivalent 3 years work experience knowledge applications hospitals healthcare algorithms program Experienced Microsoft Visual Basic 6 Microsoft SQL Server 2005 2008 database Crystal Report Version 8 Microsoft Visual Studio 2008 program languages PHP Java work pressure independently within team work hard meticulous discipline Accustomed document work well cleanly good communication skills
Language	EN
Min GPA	None
Max Age	35
Number Employee Needed	1
Major of Employee Needed	,Computer Science,

Table 5: Example of processed job vacancy written in Indonesian

Variable	Stored Value
Employer Name	PT. UVW
Industry	Pharmaceutical
Title	Programmer Internship
Position	IT/MIS-Programmer
Job Type	Intern
Work Location	Office
Employment Status	Other
Jobseeker Gender	Both
Min Salary	2500000
Max Salary	3000000
Work Experience	0
Minimum Education Degree	Bachelor (S1)
City	DKI Jakarta
Country	Indonesia
Additional Information	*mahasiswa akhir skripsi internship 6 paham pemrograman joomla asp net java script html head office industri pulo gadung*
Language	ID
Min GPA	None
Max Age	None
Number Employee Needed	1
Major of Employee Needed	,Computer Engineering,Computer Science,Information Systems,Information Systems,

2.3 Text Processing

Text processing performed to stripping unwanted text from stored job vacancies criteria in a form of long sentences also known as text mining. Basically, text mining contains all features of data mining, but can handle unstructured or semi-structured data [16]. Text data unlike numerical data, that can be easily processed by computer, therefore text processing is needed to get more contextual information [17]. Text processing consist of tokenization, removing HTML element, stop word removal, stemming, lemmatizing. All process done using Python 3. Library were utilized like BeautifulSoup4, Natural Language Toolkit (NLTK), and Sastrawi.

Stemming is a prerequisite step to a text mining based, to extract the original root word [18]. Also stemming is a crucial step to do to reduce the feature vector and memory use. Words of "connected", "connection", and "connecting" are the same root word of "connect" [19]. Stemming also more suitable for Indonesian text [20].

On the other hand, lemmatizing is a modified version of stemming [18]. In general, lemmatizing based on the textual form provided by a corpus. The inflection of words is grouped and form to its basic form or lemma [21]. Word of "goes" and "went" are inflection form of "go", lemmatizing can be identifying those word and turn it to the lemma form [22]. Then job vacancy written in English instead of going through stemming process, it will be processed using lemmatizing.

Algorithm 1. Text processing

```
Open File OriginalText.txt
For each line from OriginalText.txt
    Remove HTML Element
    Tokenize word
    If TokenizedWord not in Stop Word
        Stemming TokenizedWord
        Join Words
    End If
    Write to ProcessedText.txt
```

End For

Text processing is used to process Additional Information variable. Previously created Language variable is used as filter to differentiate English and Indonesian text. Each language is processed separately because of the different method used for English and Indonesian. The different only on the stemming area, English text use lemmatizing method and Indonesian text use stemming method, the rest process is the same for both languages. Unprocessed text copied to a .txt file. Python code then open and read the unprocessed text file, process through all words, and write it to a new .txt file as an output of processed text. Output result then put back into previous Excel file as processed text.

2.4 Association Analysis

Association analysis worked by detecting one or more characters that often appear together on the dataset to gain more information between feature correlation [23]. To be able found any correlation from dataset, we're utilizing RapidMiner Studio as a tool to perform association analysis. Excel file that already goes through pre-processing and text processing stages loaded into RapidMiner Studio. Then processing data that consist of filtered variable, selected variable, document processing, and convert numerical data to binomial data.

The algorithm used is Frequent Pattern Growth (FP-Growth). FP-Growth algorithm chosen instead of Apriori algorithm because of FP-Growth algorithm is proven to be better than Apriori algorithm especially categorical data [24]. FP-Growth algorithm also has a faster execution this is because FP-Growth only need two times to read through all dataset in order to generate rules. Therefore, Apriori algorithm need to re-read all dataset for each iteration. FP-Growth also based on FP-Tree to work, each node can be shared for other, so memory usage is also considered smaller (Ünvan, 2020).

FP-Growth work by setting minimum support threshold for corelated of each word. If a word pair passed the given threshold, then those words considered as words that often appear together. On this research, the minimum support threshold is set to 0.1.

$$Support\ (A \cup B) = \frac{\sum vacancies\ word\ A \cup B}{\sum job\ vacancies} \quad (1)$$

Words combination that passed the given minimum support, then processed on association rule to generate rules. To be able generating rules, association rule needed minimum confidence threshold. The higher minimum confidence is set, then this defines the likeliness of occurrences of consequent on dataset that already has the antecedents. On this research, the minimum confidence value is set to 0.95.

$$Confidence = P(B|A) = \frac{\sum vacancies\ word\ A \cup B}{\sum vacancies\ word\ A} \quad (2)$$

Association rule also generating lift number for each rules. The bigger lift number is, it signifies the strong bond between those item.

$$Lift\ (A \rightarrow B) = \frac{Confidence\ (A \rightarrow B)}{Support\ B} \quad (3)$$

To be able get the rules that contain criteria with the most frequently used by the industry, the support value is used. To determine a low or high support value, the average support value is needed. This applies to each year of the produced rules. Low support value determined by value below the average support value and high support value determined by a value above the average support value. In the other hand, the same concept also apply for the lift value to determine the stronger and the weaker bond of itemset criteria that the rules produce.

To simplify the understanding of this concept, the support and lift distribution will be made in a form of two axis bubble chart. The Chart is divided into 4 quadrants by 2 line that represent the average value of support on the vertical line and the average value of lift on the horizontal line. Detailed chart zones of each quadrant can be seen on Figure 2. Produced rules are represent in the shape of bubble. Where is each bubble can contain more than one rules. The greater the number of rules with the same support and lift value, then the bubble will be bigger.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

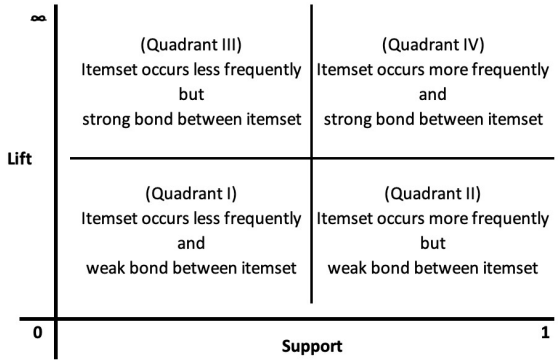

Figure 2: Zones on the chart

3. RESULT AND DISCUSSION

Produced rules in year of 2016 are 55 types of rules by support and lift value combination with total of 118 rules with the average value of support is 0,1835 symbolized by horizontal orange line and average value of lift is 3,3877 symbolized by vertical green line. Visualization of generated rules placement of corresponding quadrant can be referred to Figure 3.

In 2017, there are 51 types of rules by support and lift value combination with total of 110 rules with the average value of support is 0,2071 symbolized by horizontal orange line and average value of lift is 3,1054 symbolized by vertical green line. Visualization of generated rules placement of corresponding quadrant can be referred to Figure 4.

In 2018, there are 62 types of rules by support and lift value combination with total of 110 rules with the average value of support is 0,1344 symbolized by horizontal orange line and average value of lift is 3,4400 symbolized by vertical green line. Visualization of generated rules placement of corresponding quadrant can be referred to Figure 5.

In 2019, there are 40 types of rules by support and lift value combination with total of 106 rules with the average value of support is 0,1760 symbolized by horizontal orange line and average value of lift is 3,0166 symbolized by vertical green line. Visualization of generated rules placement of corresponding quadrant can be referred to Figure 6.

Based on the visualization of the four charts from 2016 until 2019, there is only 8 specific Positions with total of 21 rules that qualified as a into Quadrant IV. These rules are qualified for passing the support and lift threshold for each corresponded year. Refer to Table 2, job vacancy criteria of "web" and "aws" only found on Position IT/MIS-Quality Assurance, this combination often appear together with the support and lift value greater than the determined threshold. This indicate the criteria mostly needed by industry in 2016 are "web" and "aws". Whereas in 2017, criteria mostly needed and also appear together is "windows" and "server" only on Position IT/MIS-Administrator.

In 2018, several criteria were found in five different Position. First Position is Engineering-Computer, produced as many as five rules consist of a combination of "develop, swift, ui, user" and "andoid", "mobile, swift, ui, user" and "app", "applications, swift, ui, user" and "Ios", "android, standards, ui, user" and "java", and "performance, swift, ui, user" and "debug". Second Position is IT/MIS-Multimedia Designer, produced as many as four rules consist of a combination of "project, video, visual" and "market", "communication, video, visual" and "skill", "video, visual" and "edit", and "edit, visual" and "video". Third Position is IT/MIS-Product Management, produced only one rule consist of a combination of "issue" and "web". Fourth Position is IT/MIS-Project Manager, produced as many as three rules consist of a combination of "*koordinasi*, *pmbook*, *pmp*, *visio*" and "*projek*", "*koordinasi*, *pmbook*, *pmp*, *visio*" and "*komunikasi*", and "*excell*, *pmbook*, *pmp*, *visio*" and "*ipk*". Fifth Position is IT/MIS-Technical Support, produced two rules consist of a combination of "experience, support, verbal, years" and "system" and "it, system, verbal, years" and "monitor". Overall, the most criteria needed also appear together among all of the produced rules in 2018 is "issue" and "web" with the highest support value and lift value.

In 2019, only position of IT/MIS-Technical / Functional Consultancy able to produce rules up to four rules. Rules consist of a combination of "cloud" and "support, computer, technologies", "development, communication, think" and "project", "computer" and "language, technologies", and "computer, technologies" and "maintain". First, third, and fourth rules are indicated as the most criteria needed also appear together in 2019 with highest support value and lift value.

Thus, are all of the combination that often appear together which the applicant should master it. These findings also indicate criteria needed by company nowadays not focused only on hard skill but also soft skill, refer to rule "development,

communication, think" and "project" as in section IT/MIS-Technical / Functional Consultancy 2019.

4. CONCLUSION

A The workforce rate and university graduates in Indonesia keep raising every year. To ensure all university graduates are acceptable by the industries, they need to meet the industry expectancy. Often the skill mastered by the graduates doesn't suitable with the industry currently needed. This could lead to an unemployed for those university graduates.

This research proposes Association Analysis method to analyze and learn the pattern that can be found in the posted job vacancy. The existed job vacancy can be learned and explored using FP-Growth to produce interest rules. These formed rules provide an insight overview of job criteria that most frequently used by industry company also the criteria often appear together in a job opening. Thus, were indicate by rules with high Support and Lift value for each of the corresponding year. The findings reported here, therefore, gave results that in 2016, "web" and "aws" being the most frequently appear together. In 2017, "windows" and "server" being the most frequently appear together. In 2018, "issue" and "web" being the most frequently appear together. In 2019, "cloud" and "support, computer, technologies", "computer" and "language, technologies", and "computer, technologies" and "maintain". This method proves that it can deliver certain knowledge in job vacancy criteria to help related faculty and student to be able to understand the criteria mostly needed by the industry. This method also could be implemented for non-Technology Information job vacancy in the future to accommodate wider coverage of knowledge for thus majors.

This kind of research hasn't been done especially in Indonesia's job vacancy scope with the focus on finding most criteria needed and often appear together using association analysis to provide an overview skillset expected by industries. These findings also could help BINUS University to developing future learning outcomes to suit the industry needs.

There are several deficiencies on this method like the minimum existing job vacancy is three to be able generate rules. Data collection stage also could be improved in the future to be retrieve data directly from data source and simplified process to present data close to real time. Also, this could minimalize the impact of growing data rapidly in the future for lighter process. The text processing also could be improved in the future to carry out more comprehensive text processing in English and Indonesian text by applying more in-depth word filtering and form of inflected word.

ACKNOWLEDGEMENTS

We thank BINUS Career Center for providing vacancies data used in this research.

REFRENCES:

[1] Badan Pusat Statistik, "Berita Resmi Statistik", *bps.go.id*, 2018. https://www.bps.go.id/pressrelease/2018/05/07/1484/februari-2018--tingkat-pengangguran-terbuka--tpt--sebesar-5-13-persen--rata-rata-upah-buruh-per-bulan-sebesar-2-65-juta-rupiah.html (accessed Jan. 28, 2020).

[2] Badan Pusat Statistik, "Berita Resmi Statistik", *bps.go.id*, 2019. https://www.bps.go.id/website/materi_ind/materiBrsInd-20190506113732.pdf (accessed Jan. 29, 2020).

[3] Castelo-Branco, F., Reis, J. L., Vieira, J. C., and Cayolla, R., "Business intelligence and data mining to support sales in retail" in *Marketing and Smart Technologies. Smart Innovation, Systems and Technologies*, vol 167, 2020, pp. 406-419, doi: https://doi.org/10.1007/978-981-15-1564-4_38.

[4] Antomarioni, S., Pisacane, O., Potena, D., Bevilacqua, M., Ciarapica, F. E., and Diamantini, C., "A predictive association rule-based maintenance policy to minimize the probability of breakages: application to an oil refinery" in *The International Journal of Advanced Manufacturing Technology*, 2019, pp. 3661-3675, doi: https://doi.org/10.1007/s00170-019-03822-y.

[5] Solanki, S. K., and Patel, J. T., "A Survey on Association Rule Mining" in *International Conference on Advanced Computing & Communication Technologies*, 2015, pp. 212-216, doi: https://doi.org/10.1109/ACCT.2015.69.

[6] Nengsih, W., "A Comparative Study on Market Basket Analysis and Apriori Association Technique" in *International Conference on Information and Communication Technology*, 2015, pp. 461-464, doi: https://doi.org/10.1109/ICoICT.2015.7231468.

[7] Ünvan, Y. A., "Market basket analysis with association rules" in *Communications in Statistics - Theory and Methods*, 2020, pp. 1615-1628, doi: https://doi.org/10.1080/03610926.2020.1716255.

[8] Yamada, S., Funayama, T., and Yamamoto, Y., "Visualization of relations of stores by using Association Rule mining" in *International Conference on ICT and Knowledge Engineering*, 2015, pp. 11-14, doi: https://doi.org/10.1109/ICTKE.2015.7368463.

[9] Thongsatapornwatana, U., and Chuenmanus, C., "Suspect Vehicle Detection Using Vehicle Reputation with Association Analysis Concept" in *Intelligent Systems Reference Library*, 2015, pp. 151-164, doi: https://doi.ieeecomputersociety.org/10.1109/IIAI-AAI.2014.94.

[10] Chaudhuri A.K., Das A., and Addy M., "Identifying the Association Rule to Determine the Possibilities of Cardio Vascular Diseases (CVD)" in *Advanced Machine Learning Technologies and Applications (AMALTA 2020) Advances in Intelligent Systems and Computing*, vol 1141, 2021, pp. 212-219, doi: https://doi.org/10.1007/978-981-15-3383-9_20.

[11] Thiyagarajan, M., Raveendra, C., P., T., and Priya, S. K., "Role of Association Rules in Medical Examination Records of Geostational Diabetes Mellitus" in *Computing, Communication and Automation*, 2017, pp. 78-81, doi: https://doi.org/10.1109/CCAA.2017.8229775.

[12] Gangar, D., Kamdar, P., Chitre, A., and Patil, M., "Job Portal using Data Mining Techniques for Adaptive Analysis" in *National Conference on Technological Advancement and Automatization in Engineering, International Journal for Scientific Research and Development*, 2016, pp. 610-612.

[13] BINUS University, "Computer Science Program", *binus.ac.id*, 2020. https://curriculum.binus.ac.id/program/computer-science-pjj/ (accessed Feb. 02, 2020).

[14] BINUS University, "Information Systems Program", *binus.ac.id*, 2020. https://curriculum.binus.ac.id/program/information systems-pjj/ (accessed Feb. 02, 2020).

[15] BINUS University, "Computer Engineering Program", *binus.ac.id*, 2020. https://curriculum.binus.ac.id/program/computer-engineering/ (accessed Feb. 02, 2020).

[16] Salloum, S. A., Al-Emran, M., and Shaalan, K., "Mining Social Media Text: Extracting Knowledge from Facebook" in *International Journal of Computing and Digital Systems*, 2017, pp. 73-81, doi: http://dx.doi.org/10.12785/ijcds/060203.

[17] Doo, I. C., Shin, H. D., and Park, M. H., "Automated Product Review Collection and Opinion Analysis Methods for Efficient Business Analysis" in *International Journal of Computing and Digital Systems*, 2021, pp. 37-45, doi: http://dx.doi.org/10.12785/ijcds/100104.

[18] Gupta, R., and Jivani, A. G., "Analyzing the Stemming Paradigm" in *Information and Communication Technology for Intelligent Systems*, 2018, pp. 333-342, doi: https://doi.org/10.1007/978-3-319-63645-0_37.

[19] Alsaidi, S. A., Sadiq, A. T., and Abdullah, H. S., "English Poems Categorization using Text Mining and Rough Set Theory" in *Bulletin of Electrical Engineering and Informatics*, 2020, pp. 1701-1710, doi: https://doi.org/10.11591/eei.v9i4.1898.

[20] Ramdhani, M. A., Maylawati, D. S., and Mantoro, T., "Indonesian news classification using convolutional neural network" in *Indonesian Journal of Electrical Engineering and Computer Science*, 2020, pp. 1000-1009, doi: http://doi.org/10.11591/ijeecs.v19.i2.pp1000-1009.

[21] Sáenz, M. T., "The regularization of Old English weak verbs" in *Revista de Lingüística y Lenguas Aplicadas*, 2015, pp. 78-89, doi: http://dx.doi.org/10.4995/rlyla.2015.3583.

[22] Stratogiannis, G., Siolas, G., and Stafylopatis, A. "Semantic Question Answering Using Wikipedia Categories Clustering" in *International Journal on Artificial Intelligence Tools*, 2014, pp. 1460014-1 – 1460014-20, doi: https://doi.org/10.1142/S0218213014600148.

[23] Agarwal, V., Hubballi, N., Chitrakar, A. S., and Franke, K., "Identifying Anomalous HTTP Traffic with Association Rule Mining" in *IEEE International Conference on Advanced Networks and Telecommunications Systems*, 2019, doi: https://doi.org/10.1109/ANTS47819.2019.9118146.

[24] Bala, A., Shuaibu, M. Z., Lawal, Z. K., and Zakari, R. Y., "Performance Analysis of Apriori and FP-Growth Algorithms (Association Rule Mining)" in *International Journal of Computer Applications in Technology*, vol 7, 2016, pp. 279-293.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

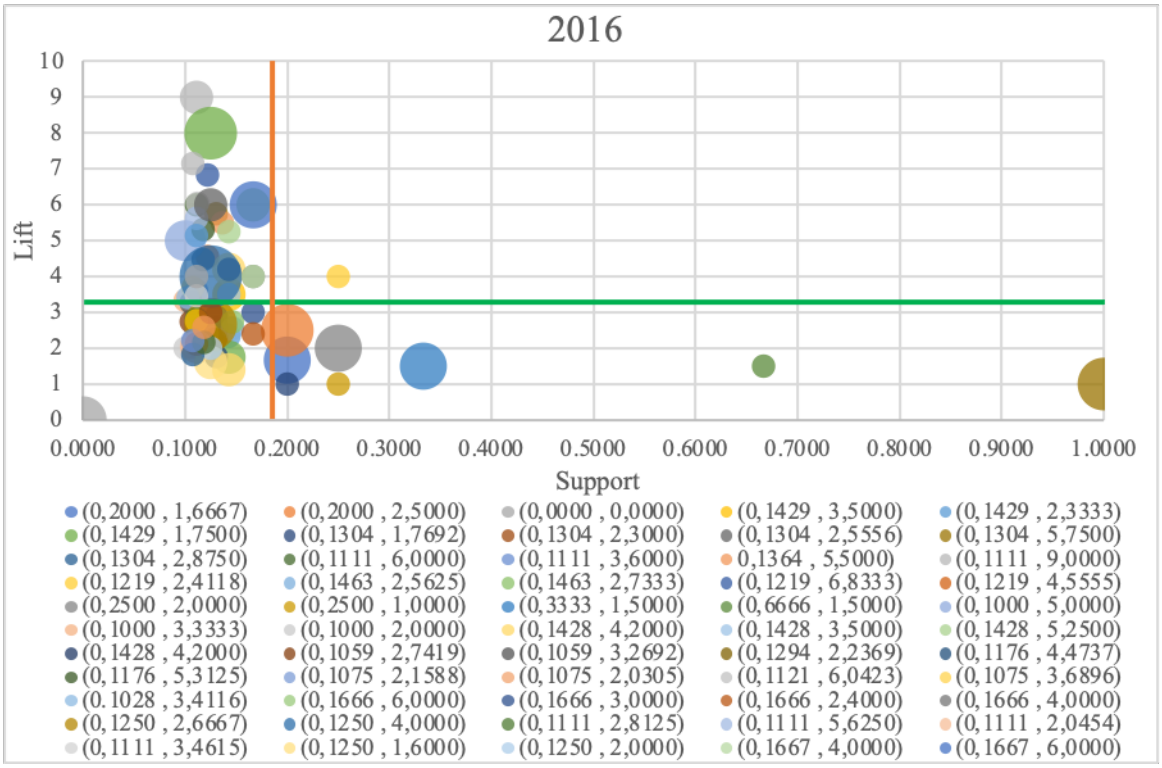

Figure 3: Visualization of generated rules based on support and lift value on vacancies in 2016 on each quadrant placement

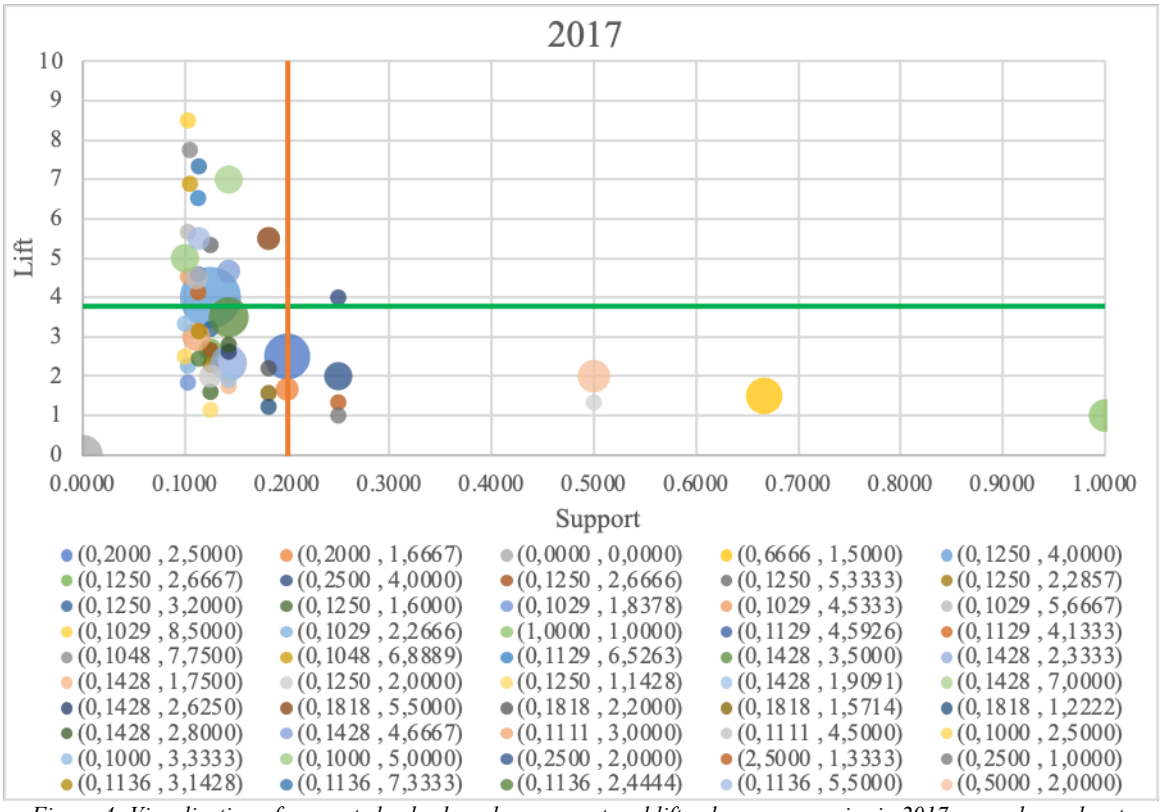

Figure 4: Visualization of generated rules based on support and lift value on vacancies in 2017 on each quadrant placement

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

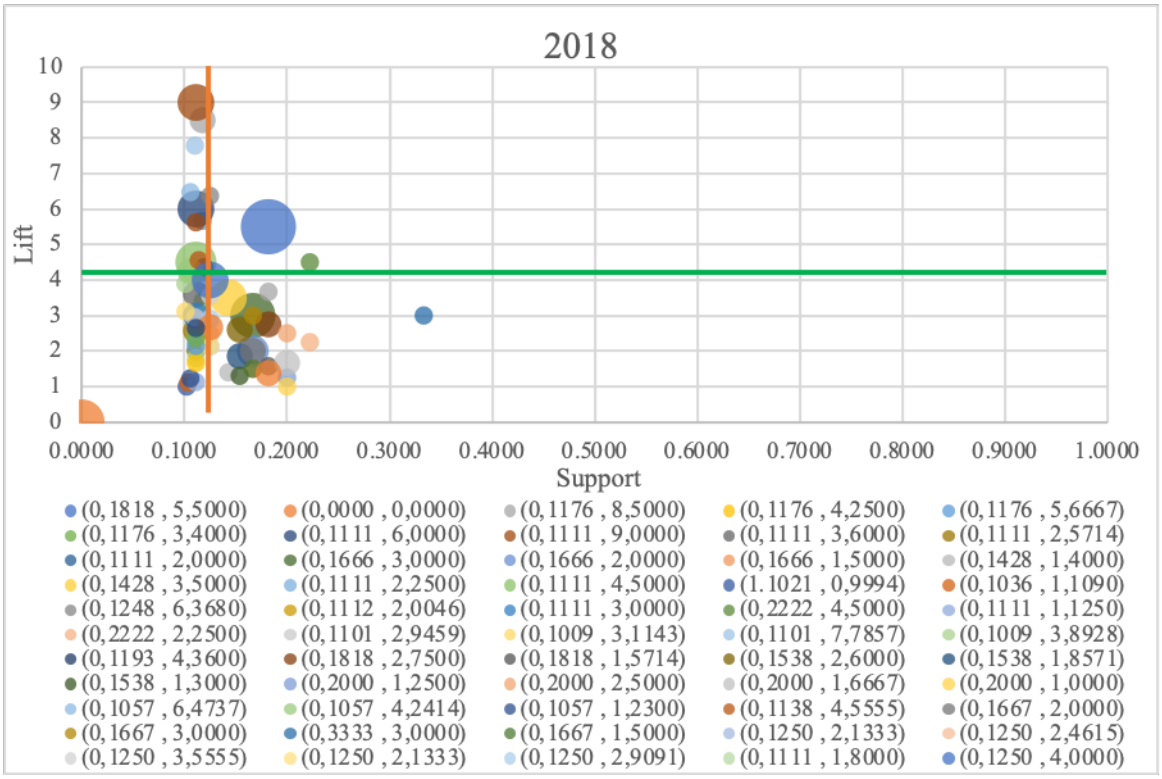

Figure 5: Visualization of generated rules based on support and lift value on vacancies in 2018 on each quadrant placement

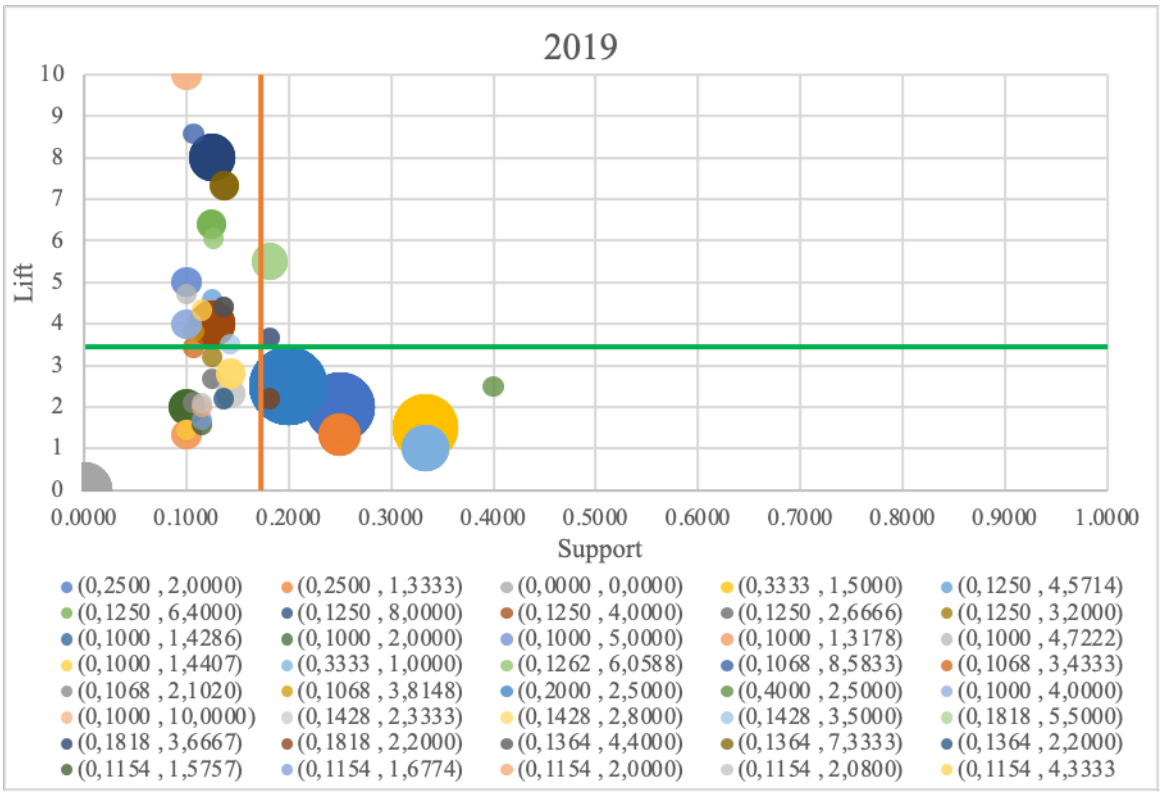

Figure 6: Visualization of generated rules based on support and lift value on vacancies in 2019 on each quadrant placement

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 6: All Categorized Rules into Quadrant IV

Premises	Conclusion	Support	Confidence	Lift
IT/MIS-Quality Assurance - 2016				
web	aws	0.2500	1	4.0000
IT/MIS-Administrator - 2017				
windows	server	0.2500	1	4.0000
Engineering-Computer - 2018				
develop, swift, ui, user	android	0.1818	1	5.5000
mobile, swift, ui, user	app	0.1818	1	5.5000
applications, swift, ui, user	Ios	0.1818	1	5.5000
android, standards, ui, user	java	0.1818	1	5.5000
performance, swift, ui, user	debug	0.1428	1	5.5000
IT/MIS-Multimedia Designer - 2018				
project, video, visual	market	0.1428	1	3.5000
communication, video, visual	skill	0.1428	1	3.5000
video, visual	edit	0.1428	1	3.5000
edit, visual	video	0.1428	1	3.5000
IT/MIS-Product Management - 2018				
issue	web	0.2222	1	4.5000
IT/MIS-Project Manager - 2018				
koordinasi, pmbook, pmp, visio	projek	0.1818	1	5.5000
koordinasi, pmbook, pmp, visio	komunikasi	0.1818	1	5.5000
excell, pmbook, pmp, visio	ipk	0.1818	1	5.5000
IT/MIS-Technical Support - 2018				
experience, support, verbal, years	system	0.1818	1	3.6667
it, system, verbal, years	monitor	0.1818	1	5.5000
IT/MIS-Technical / Functional Consultancy - 2019				
cloud	support, computer, technologies	0.1818	1	5.5000
development, communication, think	project	0.1818	1	3.6667
computer	language, technologies	0.1818	1	5.5000
computer, technologies	maintain	0.1818	1	5.5000