Development of Talent Model based on Publication Performance using Apriori Technique

Zulaiha Ali Othman[1], Noraini Ismail[2], Mohd Zakree Ahmad Nazri[3], Hamidah Jantan[4]

Center for Artificial Intelligence Technology, Faculty of Information Science and Technology[1, 2, 3]
Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Dungun, Terengganu, Malaysia[4]

Abstract—The main problem or challenge faced by Human Resource Management (HRM) is to recognize, develop and manage talent efficiently and effectively. This is because HRM is responsible for selecting the correct talent for suitable positions at the right time, aligned with their existing qualifications, talents and achievements. Furthermore, the decision in identifying talent for a position must be fair, truthful and appropriate. In the academic field, publication is a core component in the evaluation of academic talent that is affected by research, supervision and conference. Therefore, this study proposed an academic talent model based on publication factor using the Apriori technique for the purpose of promotion. This study applies the Apriori based Association Rules algorithm to identify a set of meaningful rules for the assessment of significant relevant talents for the promotion of academic staff in local universities. The findings have successfully developed a model based on talent acquisition of knowledge related to the issuance and have been evaluated by comparing the guidelines for the promotion of academic experts. This knowledge helps to improve the quality of the evaluation process of academic talent management and future planning in HRM.

Keywords—*Human resource management; apriori based association rules; promotion's guideline*

I. INTRODUCTION

The presence of 21st century has demanded the need for creation of new knowledge through a solid synergy among intellectuals. Hence, the strategy to mobilize all potential intellectuals, especially among academicians in local universities, is crucial. In addition, strengthening the field of education and driving excellence at local and global levels are in line with Malaysia's aspiration to achieve a fully developed nation. Therefore, HRM in any organization including higher educational institutions plays an important role too.

According to [1], human resources are the essential drivers that will determine the success of an organization. Besides that, having a talented and competitive workforce is very important in an organization to address any challenges [2]. Thus, HRM should set a goal in selecting the right and suitable academicians for proper positions. This is to ensure that the positions given are aligned with their existing qualifications, talents and achievements. However, the main problem or challenge faced by HRM in local universities is in recognizing, developing and managing talent efficiently and effectively. This is because human judgment and decision making have limits and can affect their fairness, truthfulness and appropriateness in identifying talent for a position [3].

Therefore, the use of data mining approach should be applied in the field of HRM through the diversification of techniques. Data Mining is a popular method that is able to explain the acquisition of useful knowledge and patterns from the data. Then, the knowledge can be used to help in decision-making processes [4]. According to [5], the researchers point out that rule mining has proven to be an effective technique to extract useful information from a large database. Also, reasoning with logical rules is more acceptable and understandable to users than a black box system.

The knowledge acquisition through data exploration is very useful for a variety of purposes. Based on the study in [6], the knowledge gained from the data collections using association rules helps in improving the professionalism level of academicians in the studied departments. In that regard, this data analysis gives many advantages to the organization and helps them in decision making to formulate more accurate planning [7]. The studies in [6] and [7] discussed the development of an overall academic model, but this study will focus on specific knowledge for the academician publication model which is a key element in the promotion process in local universities.

The main objective of this study is to determine meaningful rules that can be used as a guidance in the promotion of academicians using a data set from HRM in a local university. This study includes four main sections: Section II describes the proposed framework and technique used. Then, the results are presented and analyzed in Section III. Finally, Section IV concludes the major finding of this study.

II. DATA PREPARATION AND METHOD

In data mining, the preparation of data plays an important role, especially in choosing meaningful and relevant attributes. This process greatly can influence the finding results obtained at the end of the study in which the knowledge produced has precise, understandable, reliable and interesting features [7]. In this study, the data preparation process involves five steps; i) data collection, ii) data cleansing, iii) data integration, iv) data reduction, and v) data transformation. The method or technique used in extracting a meaningful pattern in the data set is the Apriori based Association Rules Technique. All the steps taken in data preparation and the technique used in this study are explained as follows:

A. Data Collection

This research is a continuation of the study in [8] that has succeeded in producing an academic talent prediction model for selected Higher Education Institutions involving experiments on raw data from HRM academic section databases from 1994 to 2010. The study in [8] was carried out using 15 raw data sets comprising 3220 data records, with a total number of attributes at 1140. The raw data sets include Demographics, Publications, Supervision, Conference, Research, Personnel History, Awards, Administrative and Performance Tasks.

However, this study does not take all the talent attributes from the study in [8], but focuses only on the publishing talent attributes. This is because the publishing attribute is a key component of academic achievement measurement for the process of knowledge development in Higher Education Institutions that leads to increased supervisory, research and conferencing. Therefore, this study only selected 1938 data records and 35 attributes which involved the collection of all academic staff data including Lecturer, Senior Lecturer, Associate Professor and Professor.

B. Data Cleansing

Data cleansing is the first step in the pre-processing phase that involves the task of filling in the missing values, controlling noise data, identifying outliers and correcting inconsistent data which can help in improving the quality of the data [9], [10].

In the data cleansing phase of this study, HRM data patterns from past researches have to be identified by reviewing the entire contents of items or values by columns (attributes of academicians such as Gender, Race, Job Status, Current Status, Education, Length of Service, Year Promoted to Lecturer, Year Promoted to Senior Lecturer, Year Promoted to Associate Prof., Year Promoted to Prof., Actual Position, Number of Publications, Number of Supervision, Number of Research and Number of Conferences) and rows (Records of Academic ID starting from K01 up to K3220). Based on the detailed review, there were some error problems. Among the forms of error in the survey data are non-uniform distributions of data, data containing unrelated or irrelevant attributes, repeating attributes, and data values in inconsistent attributes. Therefore, it is important to undergo data cleaning steps first to address the problem.

This study focuses on publishing factors that have these five main criteria which are Demography, Publication, Supervision, Research and Conference, as shown in Table I. Therefore, attributes that are irrelevant such as Total Performance Scores, Number of Awards and Number of Holding Administrative Positions are removed from the data set. In addition, the data set is also inconsistent because it contains too many values of 0. The attributes containing the value of 0 for a record of 80 percent and above is also eliminated from the data set.

C. Data Integration

Integration data involves the process of merging data to obtain a balanced set of data. The process of merging data in this study is carried out through the decision or assessment by the expert and the relevant analysis which involves the merger of the attributes and the merging of items within the attributes. According to [11], the merger phase or data integration will contribute to the three aspects of data quality that is complete (completeness), sufficient (minimality) and understandability.

A total of 6 attributes are selected for the data integration process involving the attributes of Races, Job Status, Current Status, Education, Publication and Role of Research. In past research, the Races attribute consisted of 27 types of races but this study combines the attribute into 2 categories only, which are Malay and non-Malay. For the attribute of Job Status, this study combines 8 categories into 2 categories that are permanent and non-permanent. Next, the Current Status attribute that contains up to 8 current status types is combined into 2 active and inactive categories. The attribute for qualification in education is consolidated into three categories, namely doctors, masters and others, which previously had 31 categories altogether. The publication attribute comprising 20 different types of publications were divided into four major publications, namely book publications, proceedings, journals, and other publications. For the research related attribute, the role of academics in research with 9 types of roles in previous research were combined into 2 main roles, namely the role of research head and researcher. Table II shows the summarization of the selected attributes that underwent the integration process.

TABLE I.   MAIN CRITERIA OF ACADEMIC TALENT

No.	Criteria	Attributes
1.	Demographic	Gender, Race, Job Status, Current Status, Education, Length of service, Year promoted to Lecturer, Year promoted to Senior Lecturer, Year promoted to Associate Prof., Year promoted to Prof. and Actual Post.
2.	Publication	Book publication, Proceeding publication, Journal publication, Other publication, and Total publications.
3.	Supervision	Total supervision
4.	Research	Research grant 2006, Research grant 2007, Research grant 2008, Research grant 2009, Research grant 2010, and Total of research grants.
5.	Conferences	Presenters, Chairman, Evaluator, Participant, Others, Department Conference, University Conference, National Conference, International Conference, and Total of Conference.

TABLE II. MERGING ATTRIBUTES

No.	Attributes	No. Attributes Merger	
		Before	*After*
1.	Races	27	2
2.	Job Status	8	2
3.	Current Status	8	2
4.	Education	31	3
5.	Publication	20	4
6.	Role in Research	9	2

D. Data Reduction

Data reduction is defined as a process of data compression to reduce the effort in analysis [10]. Prior to the data reduction process, the statistical analysis of each academic record item (row) and attribute item (column) is executed in advance to see the percentage number of item content or values in each HRM data record and attribute. The data records and attributes that have 80 per cent and above of 0 values will affect the quality of the data and consequently will affect the quality of the rules generation. Furthermore, low data quality will result in weak modeling [12]. Hence, the irrelevant task of reducing data records is important in this study to provide meaningful data for further analysis and to improve the quality of rules set production. At the same time, it can overcome constraints in data complexity and reduce data size.

Following the data reduction process, only 1938 records were selected for the experimental process. This process involves reducing the number of data records by 1282 compared to the total number of data records which was 3220. The computations for data reduction of the actual sample HRM datasets are shown in Formula 1 below:

$$\text{Percent of reduction Academician record} = \frac{Number\ of\ 0value\ data}{Academician\ record} \times 100 \qquad (1)$$

As a result, K01 id record is not removed from the data set because the value of 0 for the data record is 33.33% which fulfills the requirement that only 80 percent or above will be removed from the data record.

E. Data Transformation

Transformation data is the final process in the data pre-processing phase after data cleansing, data integration and data reduction. In data transformation, the data format were converted in order to make the mining process more efficient and the pattern of knowledge gained is easier to understand [13]. Through this study, each numerical data set was discretized using the WEKA (Waikato Environment for Knowledge Analysis) software. The data set was transformed according to its representation or value of a specified range based on an interval distribution using a binning method. The execution of discretization tasks is intended to reduce the continuous attributes by dividing the values of attribute into intervals [5]. Therefore, the data set is discriminated in the appropriate form to facilitate analysis as well as to minimize the space of the data.

In this study, the transformation technique used is the Equal Frequency Binning. This technique is applied to 29 attributes of HRM data set that has a numerical attribute value. It involves the following attributes; Length of Service, Year Promoted to Lecturer, Year Promoted to Senior Lecturer, Year Promoted to Associate Prof., Year Promoted to Prof., Actual Position, Published Books, Published Proceedings, Published Journals, Other Publications, Number of Publications, The Head of Research, The Researcher, Research Grant 2006, Research Grant 2007, Research Grant 2008, Research Grant 2009, Research Grant 2010, Total of Research Grants, Presenter, Chairman, Assessors, Participants, Others, Department Level Conferences, University Level Conferences, National Level Conferences, International Conferences and Total of Conferences. For attribute actual position, Position A represents Lecturer; Position B represent Senior Lecturer; Position C represent Associate Profesor; and Position D represent Profesor. Table III above shows the sample of discretization on HRM data set.

The next process is the development of an academic publication model using the association rules technique. The description of association rules mining is discussed in detail in the next sub-section. The development of the academic publication model will be explained more clearly in Section III.

TABLE III. ATTRIBUTE TRANSFORMATION

No.	Attribute	Discrete	Symbol
1	Published books	0	A1
		1-2	A2
		≥3	A3
2	Published proceedings	0-1	B1
		2-6	B2
		≥7	B3
3	Published journals	0-1	C1
		2-6	C2
		≥7	C3
4	Other publications	0	D1
		1-3	D2
		≥4	D3
5	Number of publications	0-3	E1
		4-10	E2
		11-26	E3
		≥27	E4
6	Student supervision	0	F1
		1-34	F2
		≥35	F3
7	Actual position	A	G1
		B	G2
		C	G3
		D	G4

F. *Apriori based Association Rules Technique*

The Association Rules technique was introduced by Agrawal and Srikant in 1994 by seeking relevant rules that exist in a database [14]. Association Rules mining refer to the findings of items which are frequently associated together in a given data set. The actual task of the Association Rules is to find a pattern or relationship between the attributes that often appear together in a database and generate a set of rules. According to [15], the goal of Association Rule is to detect the relationship or correlation between absolute variable values (attributes) in a data set. This technique also allows analysts and researchers to uncover hidden patterns in large data sets.

An important measure of mining in Association Rules is the setting of the threshold value, the minimum support level and the minimum confidence level. In order to generate meaningful rules, the experiments need to carried out by setting a high support value as well as high confidence value. However, if the outcome of the rules is not satisfactory due to the small number of rules generated or no meaningful rules are found, then the next experiment will be performed by setting a lower threshold value [10].
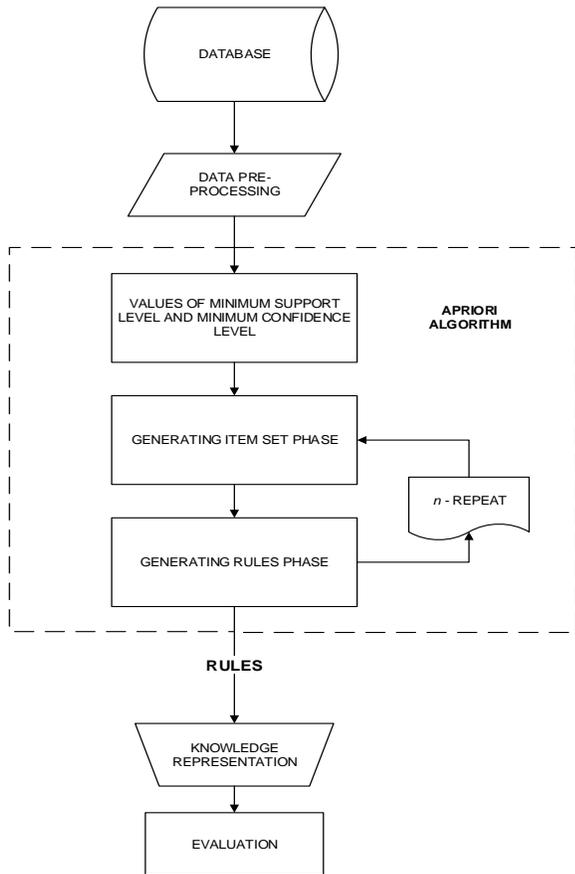

Fig. 1. Apriori based Association Rules Generation Steps.

Determining the value of minimum support and minimum confidence level is an important task as it affects the quality of the rules generated in terms of strength, legality, reliability and usability. The minimum support value measures and determines the frequency of items that often indicate the occurrence of rules in the database. Therefore, the minimum support value is an important parameter in the Association Rule because that value will control the number and form of rules generated. On the other hand, the minimum confidence value measures the strength of rules obtained from the database [16]. The minimum confidence value is calculated on the generating phase of the rules to test the strength of a formed rule. According to [17], the best rules are the rules that have a high level of confidence. Conversely, if the confidence value is low, the resulting rules are not strong and will be removed.

Apriori is the most popular algorithm in Association Rules mining. In this study, the Apriori algorithm will be used in performing experiments to produce a set of rules. The Apriori algorithm works by generating frequent items set and subsequently generating Association Rules from the regular set of items. There are two main phases in the process of implementing the Apriori algorithm which are generating frequent item sets and selection of rules from regular item sets. Fig. 1 shows the steps of generating Association Rules using the Apriori algorithm.

III. EXPERIMENT AND RESULTS

This experiment was conducted using WEKA software to find the Association Rules in HRM data sets and the Apriori algorithm was used to find data-based rules. Only 1938 data sets and 35 attributes were selected for this experiment. In this study, rule generation is governed by parameter tuning, such as minimum support and minimum confidence level. This experiment used four different minimum support values which are 0.05, 0.1, 0.15 and 0.2, and the minimum confidence level of the threshold was set from 10% to 90%. Next, each generated rules is selected based on the pre-determined selection criteria. Step selection of the rules in this study is shown in Fig. 2. The example of the generated rules is shown in Tables IV-VII.
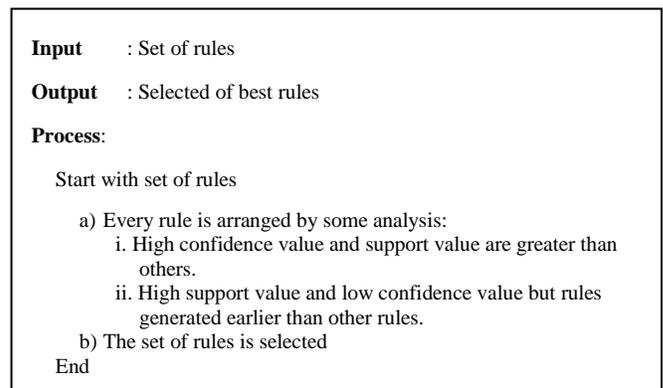

Fig. 2. Rules Selection Criteria.

A high level of minimum confidence generates good rules from a data set. However, if the experiment is set to a level of confidence that is too high, the rules generated will be less even if the resulting rules are stronger. For the purpose of meeting the needs of the study, minimum confidence is reduced to see the relevance results of the low confidence level. The experiments conducted involved the relationship between the five main criteria for publication talent factors: Demographics, Publications, Supervision, Research and

Conferences through four most important relationships in the association, namely "Publications and Demographics", "Publications and Research", "Publications and Supervision" and "Publishing and Conference". A strong relationship between the selected attributes serves as a set of rules for an experiment-based knowledge representation. The following sections illustrate the rules of the association extracted for each relationship that is arranged according to the highest minimum confidence. Section A shows the rules for "Publications and Demographics", section B indicating "Publishing and Supervision" rules, section C shows rules for "Publications and Conferences" and section D shows rules for "Publications and Research".

TABLE IV.   SAMPLE RULES FOR PUBLICATION AND DEMOGRAPHIC

No.	Rules	Min Support	Frequency	Min Confidence
1	IF otherpublication=0 ACTUALPOST=A THEN yearpromotedtobeB=0 AND yearpromotedtobeC=0 AND yearpromotedtobeD=0.	1	97	0.2
2	IF otherpublication=0 OR publishedjournal=0-1 ACTUALPOST=A THEN races=M AND yearpromotedtobeB=0 AND yearpromotedtobeC=0 AND yearpromotedtobeD=0.	0.89	163	0.15
3	IF education=Master OR otherpublication=0 OR publishedjournal=0-1 races=M ACTUALPOST=A THEN yearpromotedtobeB=0 AND yearpromotedtobeC=0 AND yearpromotedtobeD=0.	1	318	0.15
4	IF yearpromotedtobeC=0 AND otherpublication=0 ACTUALPOST=B THEN yearpromotedtobeD=0.	0.88	148	0.15
5	IF ACTUALPOST=A THEN races=M AND education=Master AND yearpromotedtobeB=0 AND yearpromotedtobeC=0 AND yearpromotedtobeD=0 AND publishedjournal=0-1 AND otherpublication=0.	0.42	1563	0.1

TABLE V.   SAMPLE RULES FOR PUBLICATION AND SUPERVISION

No.	Rules	Min Support	Frequency	Min Confidence
1	IF totalpublication=0-3 AND totalsupervision=0 ACTUALPOST=A THEN publishedjournal=0-1 OR otherpublication=0 OR publishedproceeding=0-1.	0.9	76	0.1
2	IF otherpublication=0 AND totalpublication=0-3 AND totalsupervision=0 ACTUALPOST=A THEN publishedjournal=0-1.	0.9	76	0.1
3	IF publishedbook=0 OR totalpublication=0-3 OR publishedproceeding=0-1 ACTUALPOST=A THEN totalsupervision=0.	0.8	155	0.1
4	IF publishedproceeding=0-1 OR publishedjournal=0-1 OR otherpublication=0 AND publishedbook=0 ACTUALPOST=A THEN totalsupervision=0.	0.8	155	0.1
5	IF publishedproceeding=0-1 AND totalpublication=0-3 ACTUALPOST=A THEN totalsupervision=0.	0.8	155	0.1

TABLE VI.   SAMPLE RULES FOR PUBLICATION AND CONFERENCE

No.	Rules	Min Support	Frequency	Min Confidence
1	IF publishedjournal=0-1 OR otherpublication=0 OR publishedbook=0 ACTUALPOST=A THEN conference5=0.	0.9	28	0.15
2	IF otherpublication=0 ACTUALPOST=A THEN conference2=0.	0.78	22	0.15
3	IF internationalconference=0-2 OR universityconference=0 ACTUALPOST=A THEN otherpublication=0.	0.79	302	0.1
4	IF conference1=0-3 ACTUALPOST=A THEN publishedjournal=0-1.	0.79	302	0.1
5	IF publishedproceeding=0-1 AND totalpublication=0-3 ACTUALPOST=A THEN conference5=0.	0.9	257	0.1

TABLE VII.   SAMPLE RULES FOR PUBLICATION AND RESEARCH

No.	Rules	Min Support	Frequency	Min Confidence
1	IF publishedbook=0 OR publishedjournal=0-1 OR otherpublication=0 ACTUALPOST=A THEN G2006=0-74.	0.91	154	0.15
2	IF totalpublication=0-3 AND headofresearch=0 ACTUALPOST=A THEN publishedjournal=0-1 OR otherpublication=0.	0.9	148	0.1
3	IF totalpublication=0-3 AND headofresearch=0 ACTUALPOST=A THEN publishedjournal=0-1 OR otherpublication=0 OR G2006=0-74.	0.96	803	0.1
4	IF publishedprosiding=0-1 AND headofresearch=0 ACTUALPOST=A THEN publishedjournal=0-1.	0.9	148	0.1
5	IF publishedprosiding=0-1 AND headofresearch=0 ACTUALPOST=A THEN publishedjournal=0-1 OR G2006=0-74 OR G2007=0-107.	0.9	1517	0.1

A. Experimental Study of Association between Official Position, Demographic and Publication

In producing the set of Association Rules for the acquisition of knowledge on the HRM regarding the selection of criteria in the promotion process, the study selected data at 80 percent and below for the number of items in the data record containing blank values (0). The five main criteria are Demography, Publication, Supervision, Research and Conference. This experiment is intended to see the relevance attribute of Actual Position (ACTUALPOST = n), Publication attributes (totalpublication = n, publishedjournal = n, publishedbook = n, publishedproceeding = n and otherpublications = n), and Demographics attributes (gender = n, Races = n, education = n, jobstatus = n, grade = n, performance = n and lengthofservice = n) use support weights between 0.2 to 0.05 while confidence value is set between 0.3 to 0.9.

The results of this experiment show the relevant rules for Publication and Demographics were found at the support value of 0.2 and at the confidence value of 1.0. The actual position of academic practitioners A are found to have never been in B, C and D positions. Actual position A also has a number of working years exceeding 9 years and non-permanent positions. In terms of the academic level of education, all relevant results show the level of education at Master's level. As for involvement in publications, the academic staff in position A published journals and proceedings of 0 to 1 publication and never published for other publications.

On the other hand, the resulting rules for the actual position of academic practitioners B show that academicians have never been in positions C and D. The academic level is post-Doctoral Philosophy, current status is active and the race is Malay. For engagement in publications, academic practitioners B essentially published journals of 2 to 6 and the publication of proceedings was more than 7 while never published books and other publications.

Meanwhile, actual position C has been in C position for 1 to 4 years and never been in position D. The academicians are also Malay, still in regular position, current status is active and

education level is Doctor of Philosophy. For the involvement in publications, academic practitioners C have been published in proceedings and journals of more than 7 publications and have published 3 books.

The actual position D has also never been in D position for more than 1 year and has never been in position B. The academicians are also Malay, the current status is active and education level is Doctor of Philosophy. For the involvement in publications, academic practitioners D have been published in proceedings and journals of more than 7 publications.

B. Experimental Study of Association between Publication and Supervision

The second experiment is to see the relevance of the Actual Position attributes (ACTUALPOST = n), Student Supervision (totalsupervision = n), and Number of Publications (totalpublication = n, publishedjournal = n, publishedbook = n, publishedproceeding = n dan otherpublication = n) where minimum support level is set between 0.2 to 0.05 and the minimum confidence level is set between 0.3 to 0.9.

The result shows that the rules for attributes Student Supervision and Total Publications of academicians began to be generated at the minimum support value of 0.1 and the minimum confidence value from 0.9 to 0.7 which is for the academicians at position A only. For actual position A, the relevance of the relationship is never supervised students and the derivatives produced by academicians for the publication of proceedings are 0 to 1, journal publication is 0 to 1, book publication is 0, other publications is 0, and number of total publications is 0 to 3 and 4 to 10. The publication of 4 to 10 has a low frequency value; therefore the relevance of the actual position A is appealing based on all the above relevance besides having a publication of 4 to 10.

The result for the support value of 0.05 and the confidence value of 0.7 and 0.3 has found relevance to the academic practitioners of actual position A. The rules show that position A never supervised students and has a publication number from 4 to 10. The actual position B association is found at the support value of 0.05 and the confidence value of 0.7, 0.5 and 0.4. The rules show position B never supervised students and

has no book publications at support value of 0.05 and confidence of 0.7. Subsequently, the relevance of the actual position B generated at the support value of 0.05 and the confidence value of 0.5 indicates the relevance of the actual position B, and the supervision of students is 1 to 34 and has published 2 to 6 journals or published 2 to 6 proceedings. The same minimum support value of 0.05 and the minimum confidence value of 0.4 result in association between the actual position B and supervision of students is supervised from 1 to 34 students, but never published books or other publications. This suggests that the relationship between actual position B with the supervision of students is 1 to 34, and the publication of journals is 2 to 6 or the publication of proceedings is 2 to 6.

Lastly, the association of actual position C was found at the minimum support value of 0.05 and minimum confidence values of 0.9 to 0.4. The generated rules found that position C has an association between the publication of books in excess of 3 or published proceedings in excess of 7 or the total publication is over 27 and supervision of students is in excess of 35 persons. The findings for the associations of actual position D begin with minimum support of 0.05 and minimum confidence 0.9 to 0.7. The findings indicate that job position D has an association between the publication of proceedings in excess of 7 or a total publication number exceeding 27 or journal publications exceeding 7 and supervision of students exceeding 35 persons.

C. Experimental Study of Association between Publication and Proceedings

This experiment is intended to see the relevance of the attributes Actual Position (ACTUALPOST = n), Publication (publishedjournal= n, publishedbook = n, and other publications = n), and Proceedings (conference = n) use minimum support values of 0.2 to 0.05 while minimum confidence values are set between 0.3 to 0.9.

These experimental results show the relevance rules for the Actual Position, Publication and Conference found at the support value of 0.15 and at the confidence value of 0.9. An academic practitioner of position A was never a Chairman, Registrar or Assessor and other duties during a conference. However, they have been the Presenter for 0 to 3 times and Participants for 0 to 2 times. At the conference level participated, academic practitioners of position A had never attended University or Department level conferences, but had participated in an international conference of 0 to 2 times and at a national level of 1 to 2 times. For engagement in publications, position A has never published books and other publications, while publishing journals and proceedings for 0 to 1 time. The total number of publications is 1 to 3 times.

On the other hand, academician practitioners in position B are found to be participating in conferences, whereby they have never been the Chairman, Registrar or Assessor and others. However, the academicians have been presenters for 0 to 3 times. At the conference level participated, the academic practitioner B had never attended the University or

Department conferences, but had participated in an international conference of 0 to 2 times. For the involvement in publications, the findings show that academicians in position B has never published books and other publications, but published journals and proceedings 0 to 1 time.

Furthermore, the academic practitioners in position C are also found to have been the Chairman more than 47 times, the Registrar or the Assessor more than 3 times, and the Participant more than 28 times. At the conference level participated, academic practitioners C had attended the Department level conference more than 3 times, the International level more than 13 times, the national level more than 8 times, and the University level more than 3 times. The total attendance of the conference is over 29 times. For engagement in publications, academicians in position C have published journals and proceedings more than 7 times, while the total of publications is over 27 times.

For academicians who have the actual position D, it was found that the academicians have been Spokesman more than 47 times, Chairman more than 3 times and Participants more than 28 times, and Registrars or Valuers more than 3 times. At the conference level participated, the C academic staff member had participated in the International level conference more than 13 times. The total attendance of the conference is over 29 times. For involvement in publications, academicians in position D had published journals and proceedings more than 7 times.

D. Experimental Study of Association between Publication and Research

This research is aimed to look at the relevance of the Actual Position attributes (ACTUALPOST = n), Research attributes (headofresearch = n, researcher = n, G2006 = n, G2007 = n, G2008 = n, G2009 = n, G2010 = n, and totalG = n), and Publication attributes (totalpublication = n, publishedjournal = n, publishedbook = n, publishedproceeding = n and otherpublication = n) that use minimum support of 0.2 to 0.05 while minimum confidence values are set between 0.3 to 0.9.

These experimental results show the relevance association of the Actual Position, Publication and Research was found at minimum support value of 0.15 and minimum confidence value of 0.91. Academic practitioners A are found to have never been a research head and are only a member of the researcher between 0 and 1 times. For the grants received, a researcher of position A received a research grant in 2006 from RM0.00 to RM74,000.00, in 2007 from RM0.00 to RM107,000.00, in 2008 from RM0.00 to RM830,000.00, in 2009 from RM0.00 to RM490,000.00 and in 2010 from RM0.00 to RM240,000.00. The number of research grants received from 2006 to 2010 is RM0.00 to RM13,800.00. For engagement in publications, academic practitioners of A position did not publish books and other publications, while publication of journals and proceedings were 0 to 1 time. The total number of publications for academic practitioners of A position is 0 to 3 times.
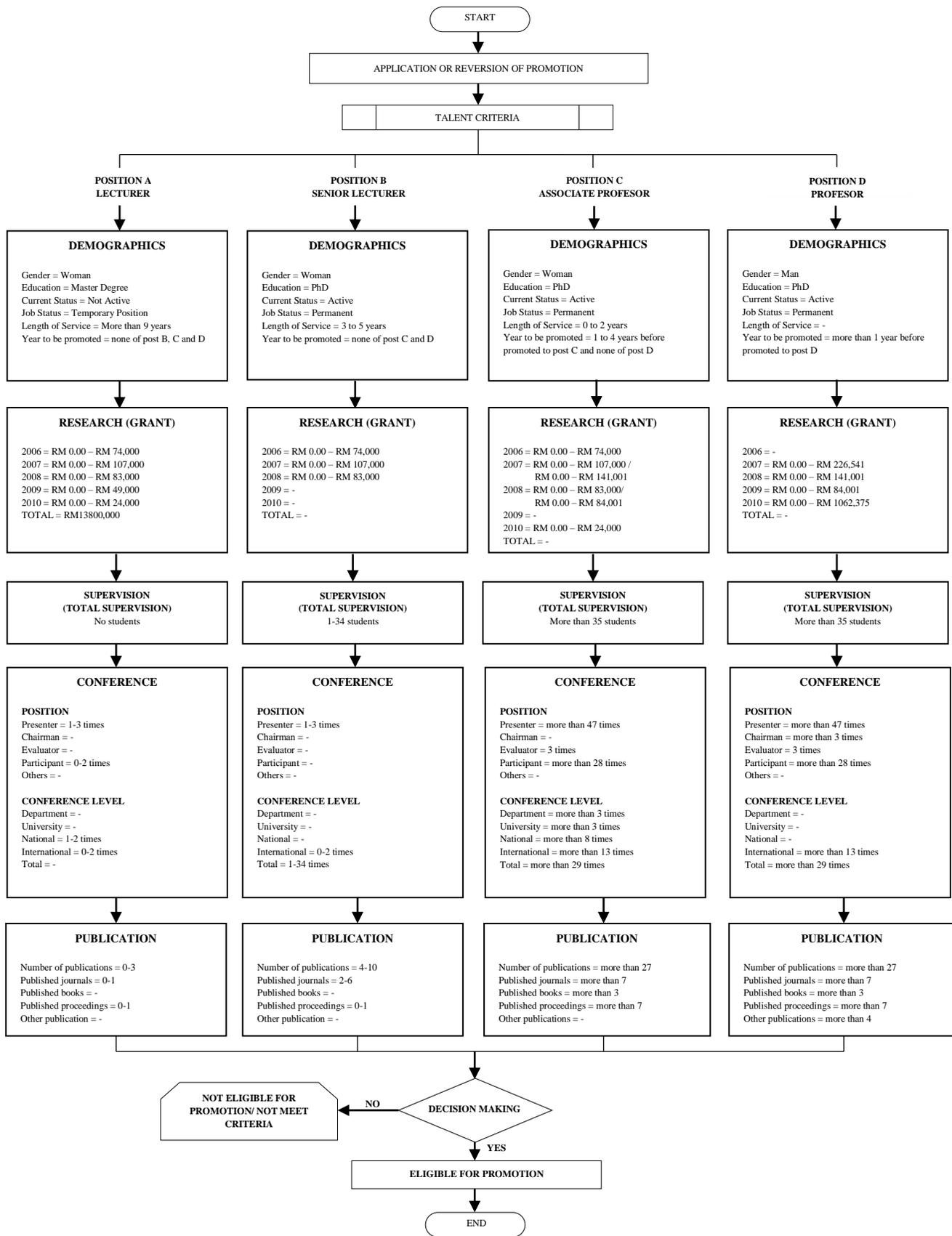
Fig. 3.   The Talent Model based on Publication Performance.

Through these experiments, the resulting rules for academic practitioners in B position show that academicians have become researchers of 1 to 2 grants and received a research grant in 2006 from RM0.00 to RM74,000.00, in 2007 from RM0.00 to RM107,000.00, and in 2008 from RM0.00 to RM83,000.00. For engagement in publications, academic practitioners in B position did not publish books and other publications, while publishing proceedings and journals were 0 to 1 time. There are also rules found that academic members of B position published journals 2 to 6 times and the total number of publications are 4 to 10 times.

This experiment also shows that academic practitioners in C position have been a research head for more than 3 times and a researcher for more than 10 times. For the grants received, they received a research grant in 2006 from RM0.00 to RM74,000.00, in 2008 more than RM141,001.00, and in 2009 more than RM 84,001.00. As for the involvement in the publication, academic practitioners of C position did not publish any books, while the publication of journals and proceedings were more than 7 times, and the total number of publications was more than 27. Other than that, there are also academic members of the C positions who obtained a research grant in 2007 from RM0.00 to RM107,000.00, in 2008 from RM0.00 to RM83,000.00 and in 2010 from RM0.00 to RM24,000.00, and did not publish any books and other publications but published proceedings and journals 0 to 1 times.

Academic practitioners of position D have also been a research head for more than 3 times and a researcher for over 10 times. For the grants, the academicians in position D received more than RM226,541.00 in 2007, in 2008 more than RM141,001.00, and in 2009 more than RM84,001.00. The number of research grants received from 2006 to 2010 is over RM106,2375.00. For the involvement in publications, academic practitioners in position D published journal and proceedings more than 7 times, other publications more than 4 times, and the total number of publication is more than 27 times.

Fig. 3 shows the academic publication model based on the decision of association rules extracted using the Apriori algorithm that will be used for promotion. The main factors that are interpreted through five main criteria are Demographics, Publications, Supervision, Research and Conferences, which have helped create interesting rules for forming links that ultimately serve as a source of reference in appraising decision-making. The development of the model is based on the results of experimental relationships between "publication with research", "publication with supervision", "publication with conference" and "publication with demographics" talent attributes. The relevance association generated through this experiment is the basis of academic rank promotion selection criteria by generating the relevancy relationship frequency between the selected attributes.

IV. CONCLUSION

This study is the continuation of [8] that uses HRM data from an IPT in talent management through Apriori based Association algorithm methods. The result of this study shows the decision of talent model based on publication performance. The difficulty in determining suitable talent for promotion is one of the major problems faced by any department or division that manages human resource-based operations. Furthermore, research in related areas are also less practiced locally or internationally, especially involving academics. Therefore, the Apriori based Association Rules technique has been applied in this study to further elaborate on other production functions in data mining, but focuses more on academic talent publication modeling. The advantage of applying the Apriori technique in the development of the talent model is that Apriori is able to generate the frequent item sets with candidate item set generation which discovers the relevant patterns in the data set. Moreover, Apriori also performs multiple scans for generating candidate sets. However, although this study has achieved its stated objectives, there are several suggestions that can be considered. One of them is the findings of relevance rules can be extended to non-academic staff for promotion. Non-academic staff also contributes in the support segment in driving excellence of academic institutions. Then the interesting attributes of the promotion have to be adjusted for the purpose of appraisal of higher rules of effectiveness. It is recommended that the assessment attributes are based on course days, number of successful and failed interviews, annual leave taken and assessment of other attributes that can be used as plus measures in HRM management and preparation of database records. Last but not least is extending the scope of the study in terms of selecting attributes from the HRM dataset.

REFERENCES

[1] M. M. Tafti, M. Mahmoudsalehi, and M. Amiri, "Critical success factors, challenges and obstacles in talent management," Ind. Commer. Train., 2017.

[2] P. F. Buller and G. M. McEvoy, "A Model for Implementing a Sustainability Strategy through HRM Practices," Bus. Soc. Rev., 2016.

[3] H. Jantan, A. Razak Hamdan, and Z. Ali Othman, "Human Talent Prediction in HRM using C4 . 5 Classification Algorithm," Int. J. Comput. Sci. Eng., 2010.

[4] M. Saron and Z. A. Othman, "Model Base On Human Resource System Using Classification Technique," IJCEM Int. J. Comput. Eng. Manag., vol. 15, pp. 1–8, 2012.

[5] S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications - A decade review from 2000 to 2011," Expert Systems with Applications. 2012.

[6] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel, and K. C. Lichtendahl Jr, Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons, 2017.

[7] J.Han, J.Pei, M.Kamber, Data Mining: Concepts and Techniques. 2012.

[8] H. R. Jantan, A. A. Hamdan, and Z. Othman, "Human Talent Forecasting using Data Mining Classification Techniques," Int. J. Technol. Diffus., 2010.

[9] S. Ramírez-Gallego, B. Krawczyk, S. García, M. Woźniak, and F. Herrera, "A survey on data preprocessing for data stream mining: Current status and future directions," Neurocomputing, vol. 239, pp. 39–57, 2017.

[10] W. A. Wan Abu Bakar, M. Y. Md. Saman, Z. Abdullah, M. M. Abd Jalil, and T. Herawan, "Mining dense data: Association rule discovery on benchmark case study," J. Teknol., 2016.

[11] S. Sarawagi and V. T. Raisinghani, "Cleaning Methods In Data Warehousing," 1999.

[12] Z. A. Othman, N. Ismail, and M. T. Latif, "Association pattern of NO 2 and NMHC towards high ozone concentration in klang," in Electrical Engineering and Informatics (ICEEI), 2017 6th International Conference on, 2017, pp. 1–6.

[13] M. A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," SIGKDD Explor., 2009.

[14] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in 94 Proceedings of the 20th International Conference on Very Large Data Bases, 1994.

[15] P. Sagar Bhise and S. Kale, "Effieient Algorithms to find Frequent Itemset Using Data Mining," 2017.

[16] S. K. Solanki and J. T. Patel, "A Survey on Association Rule Mining," in 2015 Fifth International Conference on Advanced Computing & Communication Technologies, 2015.

[17] A. A. Aziz, N. U. R. H. Ismail, and F. Ahmad, "Mining Students ' Academic Performance," Jatit Lls, 2013.