

Kragujevac, 22 August 2023

Prof. dr Zoran Kalinic
Faculty of Economics
University of Kragujevac
Serbia

Review of the doctoral dissertation of
Mr. Dwijendra Nath Dwivedi, M. Phil.
entitled „**Analysis of the employer's expectations towards competencies of candidates for employment in the Banking Sector in The UAE using Text Mining Approach**”
written under the scientific supervision of prof. dr hab. Paweł Lula
at the Krakow University of Economics, Poland

This review is prepared based on the decision of the Council for Academic Discipline of Management and Quality Sciences of the Krakow University of Economics from 13 July 2023, which appointed me as a reviewer of the doctoral dissertation of Mr. Dwijendra Nath Dwivedi. The doctoral dissertation entitled „Analysis of the employer's expectations towards competencies of candidates for employment in the Banking Sector in The UAE using Text Mining Approach” was prepared under the scientific supervision of prof. dr hab. Paweł Lula, at the Krakow University of Economics, Poland.

GENERAL REMARKS

The reviewed doctoral dissertation entitled „Analysis of the employer's expectations towards competencies of candidates for employment in the Banking Sector in The UAE using

Text Mining Approach" deserves generally positive assessment. The dissertation has a structure adequate for this kind of scientific work, with a proper review of relevant scientific literature, sound theoretical background and, as the most significant part of the dissertation, original research study in the area of text mining, which provides new insights and scientific contribution to the area of management and quality sciences.

However, there are some issues and weaknesses of the reviewed doctoral dissertation, that will be discussed in the rest of this review. These comments do not significantly influence positive overall assessment and should be considered as suggestions how to improve the academic skills of the author as well as how to prepare better scientific publications, based on the research presented in the dissertation.

The detailed evaluation of the doctoral dissertation is prepared taking into account following five criteria:

- Topic and general structure of the dissertation,
- Research methodology (assumptions, objectives, hypotheses, methods),
- Literature,
- Detailed assessment of the content,
- Language and editing.

DETAILED REMARKS

Topic and general structure of the dissertation

The topic of the doctoral dissertation is very interesting and important in contemporary science and practice in the academic area of management and quality sciences. We often witness the gap between competences of graduate students and competences necessary for job positions in many areas. Therefore, the dissertation addresses important problem in practice of many modern companies and organizations. The application of text mining i.e. the analysis of unstructured data such as various forms of documents and short texts is very popular in contemporary scientific research, as today it is very easy to access enormous

quantity of textual data on the web, but also modern methods, including Artificial Intelligence and Machine Learning, and availability of computing power and software solutions enable us to perform the analysis of large datasets, which was not possible in prior decades. The originality and the main contribution of the doctoral dissertation lies in its context i.e. the application of text mining to the analysis of job offers and required competences in the banking sector in the United Arab Emirates (UAE) – the area which is not, to the best of my knowledge, properly investigated in contemporary scientific literature.

The reviewed dissertation consists of 198 pages, where the main body i.e. text of the dissertation is presented on 160 pages (pp. 10-170), which is in range of an average volume of doctoral dissertations in this academic field. Generally, the dissertation is traditionally structured and it consists of the introduction, six main chapters, followed by the conclusions, recommendations and further research (Chapter 7) and list of bibliography.

Chapter 1 introduces UAE economy and labor market analysis, with focus on banking sector. In my personal opinion, maybe the dissertation should have started with theoretical, more general chapters, and the PhD candidate could position this chapter as an introduction to empirical part of the dissertation i.e. to use this chapter to introduce specific context of the banking sector in the UAE as a case study on which the theoretical model will be tested. Chapters 2 and 3 present theoretical and methodological grounds for the dissertation: Chapter 2 focuses on competences and various competency definitions and models, while Chapter 3 introduces text mining as a valuable analytical method and discusses various concepts of text mining which will be used in the empirical part (more detailed analysis of the content of all chapters will be presented later in this review). Empirical part of the dissertation starts with Chapter 4, which presents the analysis of competency demand in banking sector in the UAE. Chapter 5 is again theoretical, and it presents the proposal of the network model of selected aspects of the labor market generally, with the analysis of the main indicators related to the importance and relationships between competences. Chapter 6 presents original empirical analysis and model of employers' expectations towards candidates in banking sector in the UAE, followed with Chapter 7, which presents main conclusions of the doctoral dissertation, recommendations and finally, avenues for further research.

Research methodology (assumptions, objectives, hypotheses, methods)

The main and intermediate objectives/goals, as well as hypotheses are presented in the Introduction section of the dissertation (p. 10-11). Generally, the presentations of objectives and hypotheses are proper and meet scientific standards. The main objective of the reviewed dissertation is "*to develop a software to perform the competency gap assessment in the UAE market mainly focusing on the banking sector*" (p. 10). In order to assess the competency gap, it would be needed to analyze both sides, the demand and the offer, i.e. the competences requested by the employers and the present competences of the potential employees. Since the dissertation was focused only on the analysis of the job offers i.e. on the competences requested by the employers, it would be better not to use the word "*gap*" in the objective definition. In addition, maybe a part "*to develop a software*" is not necessary in the objective definition, because software development is just a tool to achieve the main objective – the analysis of the employers' expectations towards competencies of candidates.

To achieve this main objective, the Author also sets five intermediate objectives/goals (p. 11):

IG1: *Understanding of the UAE banking sector and the labor market (Chapter 1).*

IG2: *Identify competency schemas for banking industry (Chapter 2).*

IG3: *Application of corpus-based approach in analysis of published job description in banking sector (Chapter 3).*

IG4: *Application of ontology-based approach in the analysis of competencies expected in the banking sector in the UAE (Chapter 4).*

IG5: *Importance of key competencies (Chapters 5 and 6).*

Generally, these objectives fit with the main objective, and all are, in the most part, achieved by the PhD candidate.

The main contribution of the dissertation is its empirical research, in which three research hypotheses were formulated (p. 10):

H1: Exploratory text analysis of job descriptions can be used for identification of crucial competencies expected in the banking sector in the UAE.

H2: Competency schemas allow for an in-depth description and examination of relationships between expected competencies.

H3: Network models are useful for description of various aspects of labour market in banking sector in the UAE.

Overall, the hypotheses are well formulated. To answer to these hypotheses, the PhD Candidate used a range of analytical methods, typical and appropriate for text mining studies.

Literature

The bibliography list includes 240 sources presented in alphabetical order, which is sufficient for a doctoral dissertation. Most of them are articles from scientific journals, but the list also includes various books, conference proceedings, analytical reports, etc. Overall, the references are well selected and relevant for the dissertation. All references are in English, and the publications published in the last 10 years (2013 onwards) represent significant portion of the list. Therefore, the selection of literature can be evaluated as appropriate for a doctoral dissertation.

Regarding reference list, only minor issues could be mentioned: some references mentioned in the text are missing in the bibliography list (e.g. Lula et al. (2017), mentioned at page 113, couldn't be found in the reference list); several references in the list are not formatted using the same style as the others (e.g. Atkociuniene (2010); IDC (2021); Jackson and Schuler (2003); Ra. Baeza-Yates et al. (1999); etc.); some of the references are missing publisher and/or journal volume/number (e.g. Askar & Altun (2009); Bandura et al. (1999); Cao & Cui (2016); Chang & Blei (2010); Oussii et al. (2017); Petrović (2007); etc.); and some of the are simply not presented in alphabetical order (e.g. Durcker (2014); Musen (1992); Yuvaraj (2011) page – 193; Jones et al. (2000) – page 195; etc.).

Detailed assessment of the content

In terms of the content, the reviewed doctoral dissertation ultimately receives a positive assessment, with minor weaknesses discussed below. The dissertation starts with an introductory chapter, which brings very short description of the research gap discovered in the literature review, followed by hypotheses and objectives of the dissertation. The introduction is very short, but sufficient to understand the problem to be analyzed in the dissertation.

Since the dissertation is focused on the analysis of the competences required by banking sector in the UAE, **Chapter 1** presents basic statistical data on the UAE economy, and particularly its banking sector. In this chapter, the various general economic indicators were discussed, as well as some indicators related population and unemployment. Finally, Section 1.5 presents detailed analysis of the banking sector in the UAE, including details about the public and the private bank structure. Overall, the chapter is well structured and presents information important to one to understand the context in which the research will be performed. Some minor issues can be observed in this chapter. As previously mentioned, this chapter could also be positioned later, after more general, theory-based chapters. The discussion on human development indications (p. 12) is based on UNDP report from 2014, which is almost 10 years old. It would be better to use more recent statistics to support the ideas and concepts. The same stands for the analysis of all indicators in this Chapter, which is usually based on the period 2010-2018 – again, one would expect more recent data to be presented and analyzed in the dissertation. I guess that because of the COVID-19 pandemic maybe some data was not even collected or it's hard to find, but I'm sure that some of them it wouldn't be too hard to get. Some of presented statistics is hard to relate to the topic of the dissertation – for example, percentage distribution of unemployed by marital status (p. 34).

Chapter 2 focuses on the competencies and their analysis, which is very important for the dissertation. It starts with the definition of the competency, followed by discussion on behavioral and professional competences. The definitions of competency, presented in Table 1 (p. 42) are generally based on rather old references – it would be better if some more recent

sources were also mentioned. In addition, I don't see a reason why, among several definitions of competency in general, the author here introduces a definition of company competences (Foss & Knudsen, 1996) – in my opinion, this definition should be excluded from the Table. After discussing how to match market demand with competences and competences in different industry sectors, the PhD candidate focuses on the competences in banking sector and presents various competency models. In my opinion, section 2.4. Match market demand with competency is rather short and it would be better if this important topic was better elaborated. General remark for the dissertation is that in several cases the author presents some figure of table, without discussing it. For example, he presents Table 2 (p. 48) with employee competences in banking sector, without any previous or further elaboration. In section 2.5, while discussing competences in banking sector, the PhD candidate refers to the study in pharmaceutical industry (p. 49) – this should be shifted to the previous section. Finally, section 2.7 discusses various competency models, which is very important aspect of the dissertation, and although it brings important concepts and methods, maybe it could be a bit more detailed and based on more recent references.

Chapter 3 provides theoretical background and key methodological aspects of text mining, which will be used to analyze job postings, published on selected web platforms in the UAE. Text mining is a technology that uses natural language processing (NLP) to extract meaningful information from large amounts of textual information. The chapter presents brief but very good introduction to text mining and classification of the problems solved by text mining. Next, representation of the documents in form of frequency matrix was discussed. Very important section 3.4 discusses Latent Semantic Analytics (LSA) approach, which will be used in empirical part of the dissertation. Unfortunately, this was done mainly based on rather old references (1990-1998, p. 58). In addition, the explanation of the singular value decomposition of A (p. 59-60) is a bit confusing. On the other hand, I find it very useful the introduction of an example of LSA application in section 3.4.3 (p. 61-64). Chapter 3 also introduces Latent Dirichlet Allocation (LDA) - a statistical model which can be used to obtain relationships between multiple documents. Good literature review of the LDA applications is presented in section 3.6.2, followed by detailed analysis of LDA theory and implementation. Section 3.6.7 discusses

model visualization, on a practical example. What I was missing here is a brief introduction and better explanation of the example itself, i.e. it is not clear which data was used in the analysis, modeling and visualization.

The analysis of competences requested by banking sector in the UAE using text mining methods is presented in **Chapter 4**. Here, the PhD candidate starts with short description of the sampling procedure (section 4.1). The procedure itself is well explained, although the graphical quality of Figure 1 is quite low and it is hardly readable. It is very good that the author used not one but three sources of job posts i.e. three most popular web platforms for job posting in the UAE. What is missing is better explanation of the sample itself: I couldn't find the sample size i.e. the number of job posts used in the analysis. In addition, some basic statistics of the sample should be provided: the number of posts per source website, per city, sampling period, etc. Next, the candidate presents empirical results of LDA analysis of topics in job postings, starting with very good explanation of data preprocessing (section 4.2.1), followed by the results of topic modeling (section 4.2.2). Here, the candidate introduces the analysis of pre-COVID and post-COVID period, without proper explanation why these two periods were compared. The sample(s) should be better explained, at least the sample size, period of collecting data, etc. Candidate also compares topic distribution in pre- and post-COVID period (section 4.2.3), but without presenting detailed results from post-COVID period. In addition, it is good that the candidate tried to compare his results with other studies, but the choice for comparison could be better, as presented studies were mainly dealing with income decline during COVID, which was not explored by the author in this chapter. Finally, section 4.2.4 presents concluding remarks and discussion, which should be done with more details and by deeper discussion on obtained results.

Chapter 5 focuses on theoretical background of graph-based models, which can be used to model co-occurrence of words in text segments or in documents. The author starts with short, but interesting literature review of studies related to various aspects of connection between social networks and job search and placement. The role of social networks in labor market outcomes (section 5.1.4) is explained in just one sentence and with just one reference, which is not enough for a section. Next, the author presents original concept of the model of

employers' expectations towards candidates for employment in banking sector in the UAE (section 5.2). The proposed competency model is hierarchical, with 5 main groups of competences (academic, industry-wide, personal effectiveness, technical and workplace), which I find very interesting and useful. Although it was used in specific context (banking sector in the UAE), the model itself is general enough to be applied in other counties and sectors. The only thing which I was missing here is an explanation how the author came to these five groups of competences, as I couldn't find any referencing or explanation. Next, a set of competences for each group is proposed by the author, which was also very interesting. But, here I two observations: maybe the author shouldn't mix real competences (such as critical thinking, customer service, sales, time management, etc.) with jobs (Business Analyst, Procurement manager, Transportation manager, etc.) – it would be better if these jobs were translated to most important competences. Also, it is not clear do tables 2-6 present finite lists of competences in each group. In pages 141 and 150, the candidate states that the model is based on 84 competences, so it is not clear why all 84 were not presented in these tables. Section 5.3 presents theoretical basis of the analysis of competences and relationships among them i.e. the process of building a matrix of competency co-occurrence, evaluation of competency importance based on the number of occurrences, key centrality measures and Gini impurity index. Here, Weighted Degree Centrality (section 5.3.4) need better explanation, particularly what are its advantages and how the weights are selected. Section 5.3.6 introduces identification of groups of competences, presented on an example of Polish labor market (but here the example i.e. the sample needs at least a short description). Section 5.4 introduces a comprehensive set of indices used to evaluate relationships between competences. Overall, this part is very good, but in some cases, a proper description of the presented equations and its elements is missing (for example, in section 5.4.2.3 it is not clear what are M and N and why N_C is mentioned). Finally, Summary and conclusion section should be longer than one sentence, since this was very important chapter of the dissertation, with a lot of important concepts and metrics presented.

Empirical results of the analysis of competencies requested by banking sector in the UAE is presented in **Chapter 6**. Here, the theoretical concepts, models and indices, discussed in

Chapter 5, were implemented on research sample of job postings in banking sector in the UAE. The chapter starts with an explanation of data retrieval (partially already explained in Chapter 4) and modelling processes. Next, an interesting and detailed analysis of competences was presented, including sample of detailed competency co-occurrence matrix and graph, and key centrality measures. Here, the author introduces Heat map (section 6.2.2.2), which is good, but this section needs further analysis and discussion of the map shown in Figure 4 (p. 148). Final sentence in section 6.2.4 is missing Human Resource as one of the most significant competences. In addition, it would be beneficial if the results presented in a table in section 6.2.5.2 would be analyzed and further elaborated. Section 6.3 presents key findings related to bipartite network indices i.e. comparison of competencies network with and without cities of posting. Test of independence of competences across cities is presented in section 6.4, followed by segmentation of job descriptions based on key competences at a high level (five groups of competences). In section 6.4.2, Table 10 has minor errors. In addition, it would be very useful if threshold for statistical significance used in this research is presented (it is usually 0.05, but here Probe for Legal and Risk management is higher, yet presented in the Table as significant). Very interesting and useful results of segmentation are presented in section 6.5. The chapter ends with short but effective summary and conclusions.

Finally, **Chapter 7** introduces conclusions, recommendations, and avenues for further research. It starts with conclusions, followed by recommendations for practitioners, further research and policy makers. Regarding structure of the chapter, maybe it would be better to have Recommendation for further research as a final section of the chapter. The Conclusion section is well written, elaborating and confirming all three research hypotheses. There is no comparison with previous studies, but taking in mind specific context, it was hard to find comparable research in the literature. What is maybe missing here, is a deeper discussion of the impact of results, particularly which results could be generalized for other industries/countries, and which are really specific only to examined banking sector in the UAE. In addition, although limitations of the study were briefly mentioned in the chapter, in my opinion it would be better to present them in a separate section. Finally, this chapter should also include some theoretical implications of the presented research. The section 7.2.1 with

recommendations for practitioners discusses the results but doesn't really give any recommendations to practitioners, how they could use the presented results of the study to improve their operations (like it was done in section 7.2.3). I liked the idea of introducing recommendations for policy makers, which are surely significant. Although recommendations for further research are well written and cover several important avenues for extended study, I would surely add the exploration of the offer side i.e. analysis of the profiles of potential employees and their competences, in order to assess the gap between what employers would like to get and what potential candidates really offer. However, despite the aforementioned weaknesses, the concluding chapter confirms that the PhD candidate understands the research process and is able to perform scientific research and present its results. Therefore, overall impression and assessment is positive.

Language and editing

Overall, the quality of language and editing is satisfactory. It is clear that the author is not a native speaker, but minor language issues and shortcomings do not significantly influence readability and understandability of the dissertation (although, when preparing publications for eminent journals, proof-reading by a native speaking person is strongly advised to the PhD candidate). The terminology and vocabulary used in English are correct and appropriate for the scientific discipline of the dissertation.

As far as editing issues are concerned, some figures are of low graphical quality (low resolution) and therefore hard to read (e.g. Figure 1, p. 96; Figure 8, p. 137; etc.). There are some typing mistakes (e.g. p14. – "... for most, if the years, ..."; p. 25 – exports instead of exparts; p. 33 – 19991 instead of 1991; p. 38 – bakery instead of bank; p. 138 – Figure not enumerated; p. 151 – "Figgure3"; p. 162 – subtitle *4.5 Conclusion and discussions* misplaced; etc.). The numeration of the equations is strongly advised, as it would be much easier to reference to them. Some sentences are repeated in the text (e.g. the sentence above and below Table 12, p. 28). Finally, numeration of the figures in some chapters is wrong (Chapter 2: after Figure 3 comes Figure 6, followed by Figure 8; Figure in p. 73 should be 12, not 10 again;

Figure in p. 93 should be 19, not 20; Chapter 3: Figure 3 twice; Figure 7 twice; Figure 12 after Figure 8 - p. 109; Figure 12 twice, followed by Figure 4 twice; Chapter 5: Figure 2 is missing; Chapter 6: Figure 4 in p. 148 and Figure 8 in p. 158 are wrongly enumerated), and in some cases the text is referenced to the wrong figure (e.g. in p. 150, the author is referencing to Figure 1, but it is obvious that it should be Figure 2, presented in p. 149). But, these are minor issues, and do not significantly influence the quality of the dissertation itself i.e. in general, the formal side of the dissertation well prepared.

CONCLUSION

Overall, my evaluation of the reviewed doctoral dissertation is positive, despite certain weak points elaborated in previous sections. Strong points of the dissertation are its multidisciplinary and empirical approach, with original and valuable empirical results.

In my opinion, the reviewed doctoral dissertation meets the legal requirements set for doctoral dissertations, as it presents an original solution to the scientific problem which contributes to the body of knowledge with its empirical findings. The PhD Candidate demonstrates sufficient level of theoretical knowledge of the scientific discipline of the dissertation. In addition, the reviewed doctoral dissertation confirms that PhD Candidate is able to independently conduct similar scientific research in the future period.

To conclude, I believe that the reviewed doctoral dissertation entitled "Analysis of the employer's expectations towards competencies of candidates for employment in the Banking Sector in The UAE using Text Mining Approach", prepared by **Mr. Dwijendra Nath Dwivedi**, meets all legal requirements for a doctoral dissertation and therefore **I recommend** to the Council for Academic Discipline of Management and Quality Sciences of the Krakow University of Economics **the admission of the PhD candidate to public defense of this dissertation.**



Prof. dr Zoran Kalinic