

# Systematic Literature Review on Big Data and Data Analytics for Employment of Youth People: Challenges and Opportunities

Aniss Qostal<sup>1</sup><sup>a</sup>, Aniss Moumen<sup>2</sup><sup>b</sup> and Younes Lakhri<sup>1</sup><sup>c</sup>

<sup>1</sup>Intelligent Systems, Georesources and Renewable Energies Laboratory (SIGER IN FRENCH),  
Sidi Mohamed Ben Abdellah University, FST Fez, Morocco

<sup>2</sup>System Engineering Laboratory, National School of Applied Sciences,  
Ibn Tofail University Kenitra, Morocco

**Keywords:** Big Data, Data Analytics, Systematic Literature Review, Labour Market.

**Abstract:** The emergence of ICT and the digitization wave that has taken place over the past decades has led to an exponential growth in the data generated - both structured and unstructured -. Opening up a promising path and a new area of research known as the Big Data (BD) paradigm. Therefore, one of many ways to highlight the capabilities and potential opportunities offered by the BD is to conduct an exploratory and empirical study on this subject.


In this paper, we will perform a systematic literature review (SLR) on the potential role of big data and data analytics as a way to develop the labour market and explore the opportunities for the employment development. First, we will identify BD's capabilities, areas of influence, and workplace applications. Second, explore gaps to be filled in order to meet the aspirations of youth job seekers. This layer of society is the most concerned with the employment issues, moreover, youth people are extremely active on social networks, generating a huge volume of data which make them a great group for a qualitative and quantitative analysis in order to define patterns and create opportunities for them.


## 1 INTRODUCTION


The employability concept, according to the most definitions and as described by many literatures, consists of understand the personality traits, attributes, skills and qualifications that make a person more employable for a specific position and perform specific tasks that meet the needs of the potential employers (Hillage et al., 1999). For The Confederation of British Industry (CBI), the employability is the possession by an individual of the qualities and competencies required to meet the changing needs of employers and customers and thereby help to realize his or her aspirations in the workplace (McQuaid and Lindsay, 2005). Furthermore, Groot and Maassen van den Brink considered the employability as the individual ability to fulfil a variety of functions in a given labor market. Another example of this core definition is provided

by Feyter et al who define employability as “the ability of employees to carry out various tasks and functions properly” (McQuaid and Lindsay, 2005), In addition, the International Labour Organization (ILO) believes that the employability exceeds the ability to obtain a first job. Rather, it is the development of capacities and skills that give jobseekers more flexibility to adapt to the labour market variables, the ability to ask questions, learn new skills, identify and assess options and understand rights at work, including the right to a safe and healthy working environment (Brewer et al., 2013).

Today, many disciplines, each in its field of competence, are trying to master the problem of employability and identify the main causes and obstacles that prevent youth people and fresh graduates to easily integrate the job market or finding jobs corresponding to their skills, as well as help them to improve their soft and hard skills. Therefore, many

<sup>a</sup> <https://orcid.org/0000-0003-1016-8467>

<sup>b</sup> <https://orcid.org/0000-0001-5330-0136>

<sup>c</sup> <https://orcid.org/0000-0002-2537-8283>

researches(Beaumont et al., 2016), (De Mauro et al., 2018) including(ILO) reports have indicated the main cause of the unemployment among recent graduates is the lack of employability skills such as people management, communication, teamwork, professionalism, knowledge development, problem-solving and decision-making. Consequently, the unemployment rate among Moroccan youth increased to 21.9%, distributed between youth men and youth female, respectively with 22.05% and 22.8% according to the ILOSTAT database of June 21, 2020.

Nowadays, the application of approaches that repose and focus on the analytics, recommendations and orientation can be a solution not to demolish the employment issues, rather to mitigate it and give a roadmap to policymakers. Accordingly, we will conduct in this paper a systematic literature review on the employability and what big data can offer through its capabilities and analytics to discover opportunities hidden in the data shared, generated or even consumed on electronic platforms related to youth interest specially related the employment context. With the objective to understand more on youth behaviours, skills and even responds to these major questions: 1) what is the interference and area of application of the big data on the employability subject? 2) Can big data analytics be a booster to the employability and facilitator to the integration in labour market? 3) What are the architectures developed on this subject and what gap need to be more developed?

## **2 CONTEXT OF RESEARCH AND RELATED WORKS**

With the growth in the use of information and communication technologies, accompanied by the digitization of services and the dissemination of professional social media platforms such as e-learning, e-recruitment and e-commerce. Flooding the virtual world with a large amount of data (structured, semi-structured and unstructured) that can be useful on many ways and can be a base for research on the employability through extraction of indicators and valuable information that need to be processed and analyzed using big data approaches(Elgendy and Elragal, 2014)(Günther et al., 2017).

However, the application of big data on the employability context begins from the identification of data sources layer that can be a way to build an image of skills, preferences and competences of

candidates. This layer can be alimented with three types of data sources:

- Education platforms: Refer to E-learning systems and as defined by Horton (2001) as “the use of Internet and digital technologies to create experience that educate fellow human beings”. Providing a rapid access to specific knowledge and information and encompasses a wide set of applications and processes such as computer-assisted learning, web-based training, virtual classrooms, and digital collaboration. The widespread use of these platforms by students and even recent graduates generate a large amount of data, which includes indicators of the knowledge learned and skills acquired and it can be used as a means of ranking and classify applicants profiles to match them with the skills required in the labour market and identify their weaknesses and strengths (Dascalu et al., 2016)(Dascalu et al., 2017).
- Recruitment platforms: Refer to E-recruitment or E-job applications (Benabderrahmane et al., 2017), there are many platforms used around the world by job seekers, job recruiters and even human resources managers. The interactions between these actors through the ranking and recommendation on the profiles needed on the labour market can be a way to improve the existing human capital, to know the existing skills and what needs to be done to increase the chances of youth people in the labour market (Zang and Ye, 2015).
- Social Networking Site (SNS): Can be divided into two types, Social Media such as Facebook and Twitter and Professional Social Networking (PSN) such as LinkedIn, Xing, and Sumry. Both types are important in the analysis of shared data on professional projects, training and expertises. The data collected can be a source to analysis the capacities and skills of candidates (Landers and Schmidt, 2016), (Batagan and Boja, 2015).

The data generated on these potential sources are important but not sufficient without an adequate analysis to complete and establish an approach and architecture intended for the objective of employability. For that, in the following sections a qualitative and quantitative analysis of existing studies will be presented to establish a roadmap and a guide for the future studies.

### 3 METHODOLOGY

In order to conduct a systematic review on the correlation and the interference between big data and employability especially for youth people, we have designed a 4 steps methodology to create our corpus. First, we started our review by identifying the words related to our topic of research and possible queries for more precision on term results by combining words related to "Big Data" and "employability" as indicated in table 1 and existing on the title, the abstract or the keywords of the articles. Second we have selected the main indexed journals and digital databases represented on: IEEE Xplore Digital Library, Scopus, ScienceDirect and Web Of Science. Third, the results obtained represented in RIS format were imported on the ZOTERO platform to do the cleaning phase by deleting documents not related to our research, filling the missing fields for the meta-analysis, merge double documents and accept the document in the time range over the current decade[2010-2020] as shown in the Figure 1 Fourth, the last phase oriented to exploit the output of the third phase(cleaned RIS database) using NVIVO for qualitative and quantitative processing and outputs will be explored on the next sections.

Table 1: The queries on each digital database

Digital Library	Advanced Queries
Scopus	TITLE-ABS-KEY(*employment *employability OR "job seeker" OR job-seeker ) AND TITLE-ABS-KEY ( "big data" OR bigdata ) AND PUBYEAR > 2009 AND PUBYEAR < 2021
ScienceDirect	(*employability OR *employment OR "job seeker" AND ("big data" OR bigdata OR BigData))
IEEE Xplorer	((("Abstract":*employability OR *employment OR 'job seeker' OR job-seeker) AND ("Abstract": "big data" OR bigdata OR BigData)))
Web of Science	Using assistance module to create the queries

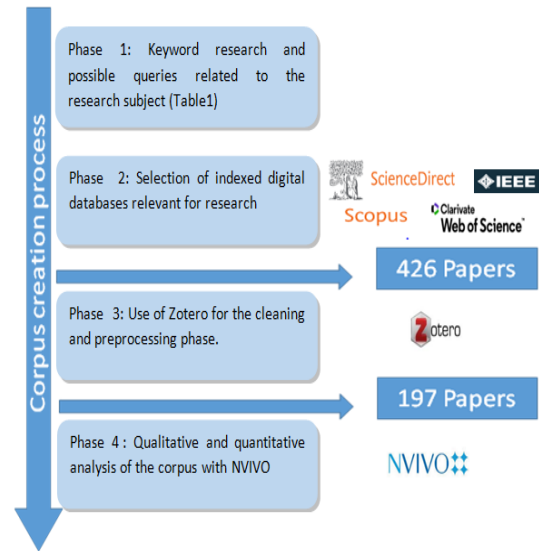


Figure 1: SLR methodology steps.

### 4 RESULTS

#### 4.1 Meta-analysis

The purpose of the meta-analysis is to discover the typology of the documents represented in the corpus (table 2) and their proportion between scientific journals (figure.2). From table 2, we carry out the dominant Conference Papers with 51% of all the documents.

Table 2: the ratio of documents in the corpus

Type	Number of docs	Ratio (%)
Conference Papers	101	51
Review Article	74	38
Others	22	11

The detail of this table is shown in the figure 2, we can notice that Scopus occupies the first place with 56% of the references, Web Of Science in second place with 27% of the references followed by IEEE Xplore (11%) and finally ScienceDirect in the last place with 6%.

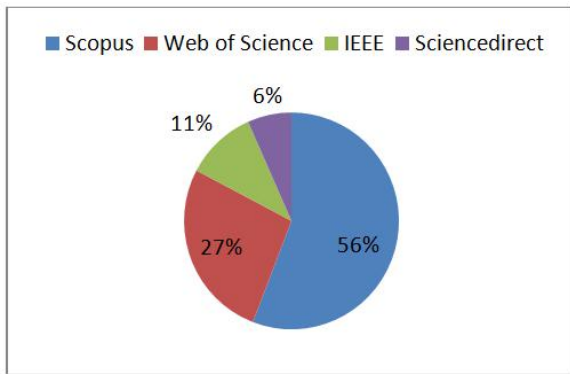


Figure 2: Number of references per Scientific Database

## 4.2 Word Analysis

The word Analysis or the word study is a way to highlight the dominants words in the corpus through the visual representation of words. Using the word cloud (Figure 3), we can easily see the most common words on the abstract, title and keywords of documents in the corpus. The important terms can be easily identified by the font and the size and how much is near to the centre of the cloud. Obviously, the big data word is the most present followed with employment and we can find words related to the analytics approaches and potential sources related to the big data.



Figure 3: Word Cloud obtained from the corpus

Another way for the word analysis, can be through the statistical analysis of the words (Figure 4), we can see the term of "big data" at first place with 110 occurrences followed by the employment, the rest of words have similar values and can be grouped on sub-subjects such as analytics approaches, stakeholders (student, government) and platforms

related to recruitment process. This analysis gives a view on how much to corpus is pertinent and related to the subject of research. For the collected corpus, it is clear that our database has the orientation "big data" and "employability".

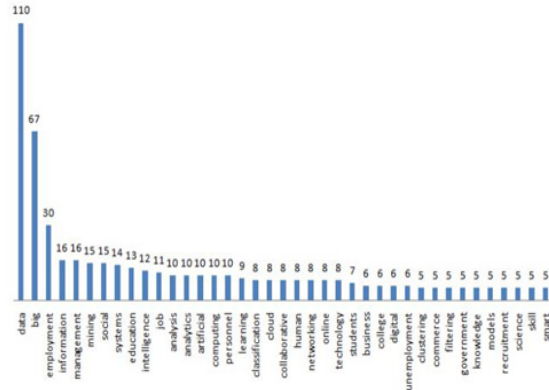


Figure 4: Word frequency

## 5 BIG DATA ARCHITECTURES FOR THE EMPLOYMENT

From the analysis carried out on the corpus, we found many researchers have highlighted the role that big data can play on the subject of employability, the majority of the proposed frameworks aim to give recommendations and even predict the job adequate to the job-seeker and profiles proposals needed by potential employers. This treatment is assured through using analysis techniques and possible ecosystems of implementation. Most of the proposed frameworks have divided their architectures on many layers for more segmentation and control of the outputs :

- Data source layer: contain all possible data sources that can feed the 3V characteristics of big data, as described in the previous section.
- Analytics layer: this layer is the main core of each framework and the innovative piece, this layer aims to build an intelligent algorithm that has the capability to extract useful information from the input data and give prediction and recommendation to stakeholder related to the employability subject.
- Visualization layer: this layer represents dashboard interfaces, in order to give the system the possibility of interactivity with the actors.

## 5.1 Employability Prediction System (EPS)

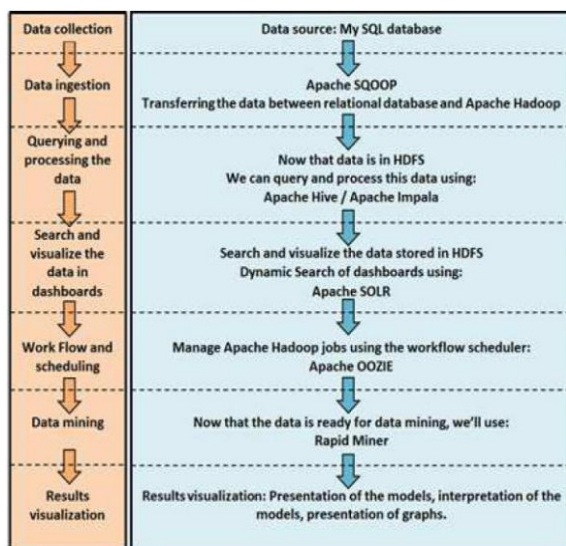


Figure 5: Architecture for Employability Prediction System (EPS)(Saouabi and Ezzati, 2019)

The architecture proposed by the author in the figure 5 has been fragmented over seven phases and shows an abstraction view on each phase with its correspondence on the ecosystem of the implementation (Hadoop ecosystem). However, the weaknesses of this architecture is its generality that can be implemented in many fields beside the employability context (health and transportation etc), secondly the layer of stakeholders has not mentioned and finally the data source layer have only one type of source represented with MySQL database and as known one of the major characteristic of big data is the heterogeneity of data as input.

## 5.2 The Hybrid Recommendation System for Job Recommendation

The architecture proposed on the figure.6, the author have separated each layer with the introducing of the data source layer including a variety of sources of information giving the possibility and flexibility on terms characteristics of data supported also adding more detail on the phase of analytics with the adoption of the hybrid recommendation analytics combining the collaborative recommendation (Filtering algorithm for calculating the intersection between User-User and Item-Item based recommended list) and Content-based filtering selects items based on the similarities between the

content description of an item and the users preferences which are used by LinkedIn(S. Ahmed et al., 2016).

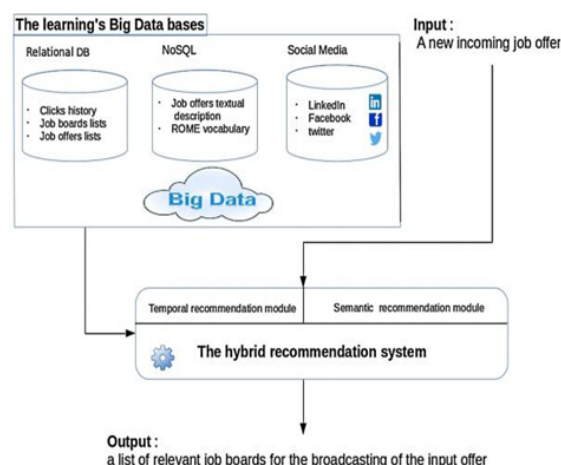


Figure 6: Architecture of the hybrid recommendation system for job recommendation (Benabderrahmane et al., 2017)

## 5.3 Expandingly Tree-based and Dynamically Context-Aware Online Learning Algorithm (ETDC)

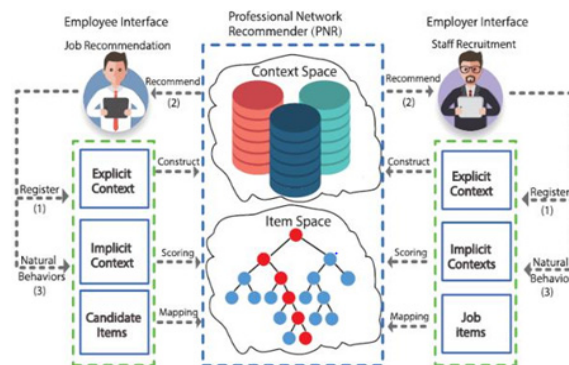


Figure 7: Professional network recommenders based on ETDC algorithm Chen et al., 2018)

The architecture of Figure 7 clearly separated each layer according to the categories of events. Therefore, the system is more interactive and built on explicit information of the context of use depending on their acceptance or rejection of the proposed items.

An Expandingly Tree-based and Dynamically Context-aware Online Learning algorithm (ETDC) was developed to observe the context and recommend an item to the user based on the current context, the historical information about users, items, contexts and rewards after the recommendation. The system extracts reward from the user's click behaviours. Then



adds the interaction log to the database which will be used in future recommendation (Chen et al., 2018).

## 6 DISCUSSION

From the three architectures introduced on the previous section, we can observe the adoption of architecture segmented on many layers, this approach can be the key of success of any proposed framework, more dynamics, adaptive as well as flexible to variables that control the employment process of youth people. Consequently, the proposed approach will be based on three main axes. First, the identification of stakeholders and actor, second, the elaboration on Data source Analysis to determine the potential sources of data related to the context and finally, the development of an architecture capable to combines ecosystems with analytics techniques to meet the ambitions of stakeholders.

- Stakeholders analysis: Stakeholders are “any group or individual who can affect or is affected by the achievement of the organization’s objectives” (Freeman 1984) Freeman (2004). On big data subject, stakeholders are represented by all entities or groups that interact directly or indirectly with the generation or exploitation of data.
- Data source analysis: The big data concept is based on the considerable volume of data produced in an accelerated manner with different formats. Today, the generation of data become easiest task with the digitalization wave, emergence of Internet Of Things (IoT), the proliferation of hyper connected devices.
- Approach to extract values: The model should be able to collect, clean and store the tremendous and heterogeneous datasets generated over distributed sources.

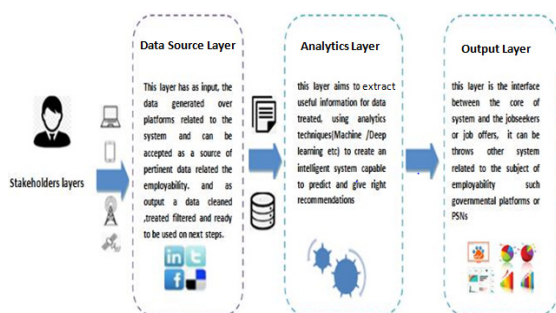


Figure 8: General view on the proposed architecture on the big data and employability framework

## 7 CONCLUSIONS

As a conclusion and as a response to the three majors questions declared on the introduction, the big data is present on every activity we do, today we generate more of data than before and the collection of this data and its treatment can be used to give solutions to very complicated problems. Mainly, the employability of youth people is not an exception; the digitalization of various services related to youth people can be a source of data including indicators, their behaviours, competences and skills. Using the intelligence artificial including Machine learning and other approaches can easily match the profile of each youth with the opportunities in labour market. Despite this, this field of research is still in its infancy and must be developed on a system adaptable to each case to respond to the specificity of each case.

## REFERENCES

- Benabderrahmane, S., Mellouli, N., Lamolle, M., Paroubek, P., 2017. Smart4Job: A Big Data Framework for Intelligent Job Offers Broadcasting Using Time Series Forecasting and Semantic Classification. *Big Data Res.* 7, 16–30. <https://doi.org/10.1016/j.bdr.2016.11.001>
- Brewer, L., International Labour Office, Skills and Employability Department, 2013. Enhancing youth employability: What? Why? and How? Guide to core work skills. ILO, Geneva.
- Chen, W., Zhou, P., Dong, S., Gong, S., Hu, M., Wang, K., Wu, D., 2018. Tree-Based Contextual Learning for Online Job or Candidate Recommendation With Big Data Support in Professional Social Networks. *IEEE Access* 6, 77725–77739. <https://doi.org/10.1109/ACCESS.2018.2883953>
- Dascalu, M.-I., Bodea, C.N., Moldoveanu, A., Dragoi, G., 2017. Towards a Smart University through the Adoption of a Social e-Learning Platform to Increase Graduates’ Employability, in: Popescu, E., Kinshuk, Khribi, M.K., Huang, R., Jemni, M., Chen, N.-S., Sampson, D.G. (Eds.), *Innovations in Smart Learning*. Springer Singapore, Singapore, pp. 23–28.
- Dascalu, M.-I., Tesila, B., Nedelcu, R.A., 2016. Enhancing Employability Through e-Learning Communities: From Myth to Reality, in: Li, Y., Chang, M., Kravcik, M., Popescu, E., Huang, R., Kinshuk, Chen, N.-S. (Eds.), *State-of-the-Art and Future Directions of Smart Learning*. Springer Singapore, Singapore, pp. 309–313.
- De Mauro, A., Greco, M., Grimaldi, M., Ritala, P., 2018. Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Inf. Process. Manag.* 54, 807–817. <https://doi.org/10.1016/j.ipm.2017.05.004>
- Elgendy, N., Elragal, A., 2014. Big Data Analytics: A Literature Review Paper, in: Perner, P. (Ed.), *Advances*

- in Data Mining. Applications and Theoretical Aspects, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 214–227. [https://doi.org/10.1007/978-3-319-08976-8\\_16](https://doi.org/10.1007/978-3-319-08976-8_16)
- Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M., Feldberg, F., 2017. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* 26, 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- Hillage, J., Pollard, E., Great Britain, Department for Education and Employment, 1999. *Employability: developing a framework for policy analysis*. Dept. for Education and Employment, London.
- Landers, R.N., Schmidt, G.B., 2016. Social media in employee selection and recruitment: Theory, practice, and current challenges, *Social Media in Employee Selection and Recruitment: Theory, Practice, and Current Challenges*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-29989-1>
- S. Ahmed, M. Hasan, M. N. Hoq, M. A. Adnan, 2016. User interaction analysis to recommend suitable jobs in career-oriented social networking sites, in: 2016 International Conference on Data and Software Engineering (ICoDSE). Presented at the 2016 International Conference on Data and Software Engineering (ICoDSE), pp. 1–6. <https://doi.org/10.1109/ICODSE.2016.7936143>
- Saouabi, M., Ezzati, A., 2019. Proposition of an employability prediction system using data mining techniques in a big data environment. *Int. J. Math. Comput. Sci.* 14, 411–424.
- Zang, S., Ye, M., 2015. Human Resource Management in the Era of Big Data. *J. Hum. Resour. Sustain. Stud.* 03, 41–45. <https://doi.org/10.4236/jhrss.2015.31006>