





Universities students' knowledge about big data analysis

Elham Nazari¹ , Marjan Rasoulian¹, Samane Sistani¹ , Maryam Edalati Khodabandeh¹ ,
Hamed Tabesh^{1*} 

¹Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

Article Info

Article type:
Research

Article History:

Received: 2020-09-12
Accepted: 2020-09-30
Published: 2020-12-01

* Corresponding author:

Hamed Tabesh

Department of Medical Informatics,
Faculty of Medicine, Mashhad
University of Medical Sciences,
Mashhad, Iran

Email: Tabeshh@mums.ac.ir

Keywords:

Big Data
Benefits
Challenges
Analyses
Universities Students

ABSTRACT

Introduction: Big data analysis has raised controversies today and attracted many students and academics for its dramatic advantages. The present research aims to investigate the extent to which students in different universities of Mashhad are familiar with this type of analysis.

Material and Methods: The present cross-sectional research was conducted on university students of different fields of study in Mashhad, Iran. A questionnaire was developed based on a review of the related literature in PubMed, Google Scholar, Science Direct and EMBAS. The target questionnaire explored students' knowledge of big data analysis. To this aim, 142 students participated in this research and completed the target questionnaire. Their responses were analyzed descriptively.

Results: The majority age of participants ranged between 21 and 28 years. 59% of these participants were female; 27% had less than a year of work experience; the academic grade of the majority of participants was Master's or Ph.D. 42% enjoyed a desirable knowledge of big data analysis. The largest number of hours of scientific and non-scientific studies belonged to basic science students and more specifically that of pharmacology.

Conclusion: Despite the significance and benefits of big data analysis, students' unfamiliarity with the essentiality of these analyses in industries and research is considerable. It seems that the field or grade of studies has no effect on one's knowledge of big data analysis. Probably, the design of specialized educational courses with this concern can help to promote individuals' knowledge of big data analysis.

Cite this paper as:

Nazari E, Rasoulian M, Sistani S, Edalati Khodabandeh M, Tabesh H. Universities Students' Knowledge About Big Data Analysis. *Front Health Inform.* 2020; 9: 50. DOI: [10.30699/fhi.v9i1.242](https://doi.org/10.30699/fhi.v9i1.242)

INTRODUCTION

Today, the world is faced with loads of data produced known as big data which has turned into a global controversy especially among academics. It has attracted many researchers too. Such data are marked by such features as large volume, variety, scalability, fast production and so on. Due to the same features, they are hard to manage and analyze via ordinary software and hardware [1, 2]. The types of analysis often used for such voluminous data are known as big data analysis. They enjoy certain benefits such as the discovery of useful data-driven patterns, extraction of key features, information abstraction and reduced cost [3]. These analyses are of much application in different industries, banking, medicine, transportation, insurance and so on [4-8]. Besides the great many benefits of these analyses, there are certain challenges too. If disregarded, the

consequences might be multifarious. These can include deficient expertise, unfamiliarity with the required instruments and methods, type of data, security issues and budgeting [9, 10]. Big data analysis is of a great significance in different industries. Besides, university students and their research in practice relate industry and applied research. Moreover, this domain of research is still at its preliminary stages in Iran and its relevant concepts are not yet well understood. Thus, the present research aims to investigate the extent to which university students from different fields of study are familiar with big data analysis.

MATERIAL AND METHODS

The present cross-sectional research was conducted on 142 students at Ferdowsi University of Mashhad as well as the University of Medical Sciences. The aim

was to explore the extent to which students from different fields of study were familiar with the target concept. Mashhad, as the main metropolis in the east of Iran has a population of 3 million people. It lies at the border of Afghanistan and Turkmenistan located on the well-known Silk Road. There are two major state universities in Mashhad, Ferdowsi University and University of Medical Sciences. The former hosts students from different fields of study including engineering and basic sciences. The latter hosts students from medical fields of study such as medicine and biology. To evaluate students' knowledge and awareness of big data analysis in different fields of study in Mashhad universities, a questionnaire was developed.

The target questionnaire was closed-ended. The primary version of the questionnaire was derived from a review of related content in Google Scholar, Science Direct and EMBASE via a Delphi method and the help of a panel of 10 experts from different fields (medical informatics, biostatistics, TIH and computer sciences).

This questionnaire contains 5 items concerned with one's knowledge of how to analyze big data. The relevant items can be seen in Table 1:

Table 1: Questionnaire content

Item	Descriptor
QK1	What is the definition of big data?
QK2	What is the hardware prerequisite for data analysis?
QK3	What is big data analysis based on?
QK4	What are the benefits of big data analysis?
QK5	What are the challenges of big data analysis?

The reliability and validity of the questionnaire was confirmed as a panel of 10 experts confirmed the validity and Cronbach's alpha was estimated to test reliability and was estimated at.73. Then, the questionnaires were submitted to 150 students. The present research attempted to include students of different fields of study. These included the following within medical sciences: medicine and dentistry, biotechnology, toxicology, nano-medicine,

biotechnology, nutrition, medical imaging, radiology, microbiology, physiology, genetics, medical informatics, biochemistry, immunology, HIT, Molecular/cellular sciences and medical physics. As for engineering, the fields of study included mechanical engineering, natural resources, aquatic sciences, industries, aerospace, metallurgy, computer and civil engineering. Concerning basic sciences, the majors included mathematics, physics and chemistry. The required data were collected and it was made sure that all questionnaires were completed. From among the initial 150 questionnaires, 142 were completed and returned. Data entry and analysis were done using SPSS21 and Excell-2007.

RESULTS

To conduct the present study, 150 university students participated from different fields of study and grades.

As it can be observed in Table 2, the majority of students belonged to the 21-24 and 25-58 age groups. They were mostly female and only a minority had work experience. Most of the participants studied medical sciences, professional Ph.D. and Ph.D. of Pure basic sciences. The majority were M.S. or professional Ph.D. candidates and their previous academic field of study was experimental sciences.

Considering the probability of correct answers by chance, any single participant's score was categorized in two. Those answering fewer than 3 items correctly were categorized as of low knowledge, and those answering 3 or more items correctly were taken as having desirable knowledge. Accordingly, 82 participants (58%) found to belong to the former and 60 (42%) with the latter category.

The mean hours of participants' scientific and non-scientific hours of studies are shown in Table 3.

As it can be observed, basic sciences and pharmacological basic sciences enjoyed the longest hours of scientific studies while pharmacological basic sciences and dentistry enjoyed the longest hours of nonscientific studies.

Table 2: Research participants' demographic information (n=142)

Variables	Groups	Frequency	Percentage
Age	<=20 year	17	12
	21-24 year	45	32
	25-28 year	39	27
	>=29 year	34	24
	Missing	7	5
Sex	male	58	41
	female	84	59
Work experience	None or less than 1 year	103	73
	Yes	39	27
Field of study	Professional Ph.D. candidate of medicine	29	21
	Medical student	44	31
	Professional Ph.D. of pharmacology	3	2
	Pharmacology basic sciences	18	12
	Dentistry	8	6
	Engineering	19	13
	Pure sciences	21	15
Grade	B.S.	29	20
	Professional Ph.D.	39	28
	M.S.	45	32
	Specialized Ph.D.	29	20
Previous academic field of study	None	66	46
	Medical basic sciences	40	28
	Pharmacology	3	2
	Pharmacology basic sciences	7	5
	Dentistry	3	2
	Engineering	13	9
	Basic sciences	11	8

Table 3: Mean and standard deviation of participants' hours of scientific and non-scientific studies across fields of study

Field of study	Scientific (Mean±SD)	Non-Scientific (Mean±SD)
Dentistry	12.25±8.43	8.29±4.96
Engineering	17.76±13.38	5.7 ±5.09
Basic sciences	20.85±16.51	6.00±3.13
Professional Ph.D. of medicine	17.72±10.77	5.43±5.22
Medicine	18.70±17.96	7.4 ±7.44
Professional Ph.D. of pharmacology	15.00±.00	4.50±.71
Basic sciences of pharmacology	21.60±19.08	10.15±13.51
Total	18.55±15.19	7.01±7.35

Participants' knowledge was as the following (Table 4). The frequency of correct answers to questions (n=5).

Table 4: Research participants' level of knowledge

#	f.	%
0	6	4.2
1	37	26.1
2	39	27.5
3	34	23.9
4	22	15.5
5	4	2.8
Total	142	100

Table 5: Distribution of participants' age and level of knowledge (frequency and percentage in each age group)

Age groups			18-24	25-34	35-44	45-54	Total
score	poor	F (%)	13 (54.2)	28 (62.2)	22 (56.4)	19 (55.9)	82 (57.7)
	good	F (%)	11 (45.8)	17 (37.8)	17 (43.6)	15 (44.1)	60 (42.3)
Total		F (%)	24 (100)	45 (100)	39 (100)	34 (100)	142 (100)

As it can be observed in Table 5, the age range 25-44

enjoyed the highest score and the 18-24 and 45-54 received the lowest score.

Table 6: Distribution of participants' sex and level of knowledge (frequency and percentage in each sex)

Sex		Male	Female	Total	
score	poor	F (%)	35 (60.3)	47 (56.0)	82 (57.7)
	good	F (%)	23 (39.7)	37 (44.0)	60 (42.3)
Total		F (%)	58 (100)	84 (100)	142 (100)

As it can be observed in Table 6, the majority of participants were female, yet the level of knowledge diverged among them from low to high.

Table 7: Distribution of participants' background experience and level of knowledge (frequency and percentage in each category)

Experience			<1 year	>1 year	Total
Score	poor	F (%)	59 (57.3)	23 (59.0)	82 (57.7)
	good	F (%)	44 (42.7)	16 (41.0)	60 (42.3)
Total		F (%)	103 (100.0)	39 (100.0)	142 (100.0)

Table 8: Distribution of participants' field of study and knowledge score (frequency and percentage in each category)

Field of study			Dentistry	Engineering	Basic sciences	Professional Ph.D.	Medical sciences	Pharmacology professional Ph.D.	Pharmacology basic sciences	Total
score	poor	F (%)	5 (62.5)	10 (52.6)	12 (57.1)	19 (65.5)	24 (54.5)	1 (33.3)	11 (61.1)	82 (57.7)
	good	F (%)	3 (37.5)	9 (47.4)	9 (42.9)	10 (34.5)	20 (45.5)	2 (66.7)	7 (38.9)	60 (42.3)
Total		F (%)	8 (100)	19 (100)	21 (100)	29(100)	44 (100)	3 (100)	18 (100)	142 (100)

As it shown in Table 7, the majority of participants had less than a year experience, they had both poor and good scores.

According to the Table 8, good scores belong to medical sciences and professional Ph.D. and poor scores were those of dentistry and pharmacology professional Ph.D.

As the results showed (Table 9), good knowledge scores belonged to M.S. and professional Ph.D. grades while low scores were obtained by B.S. and Specialized Ph.D.

As indicated in Table 10, good scores were obtained by students whose previous field of study was either

medical basic sciences or none whereas the poor scores were those of dentistry and pharmacology.

According to the results presented in Table 11, the hours of scientific and non-scientific studies in both good and low categories were about equal. Yet, it can be concluded that those with longer hours of scientific and non-scientific studies enjoyed a good knowledge score.

Table 9: Distribution of participants' grade of education and knowledge score (frequency and percentage in each category)

Grade			B.S.	Professional Ph.D.	M.S.	Specialized Ph.D.	Total
Score	Poor	F (%)	17 (58.6)	25 (64.1)	23 (51.1)	17 (58.6)	82 (57.7)
	Good	F (%)	12 (41.4)	14 (35.9)	22 (48.9)	12 (41.4)	60 (42.3)
Total		F (%)	29 (100)	39 (100)	45 (100)	29 (100)	142 (100)

Table 10: Distribution of participants' previous field of study and knowledge (frequency and percentage in each category)

Previous field of study			None	Dentistry	Engineering	Basic sciences	Medical basic sciences	Dentistry	Pharmacology basic sciences	Total
Score	Poor	F (%)	41 (62.1)	1 (50)	5 (38.5)	5 (45.5)	24 (60)	2 (66.7)	4 (57.1)	82 (57.7)
	Good	F (%)	25 (37.9)	1 (50)	8 (61.5)	6 (54.5)	16 (40)	1 (33.3)	3 (42.9)	60 (42.3)
Total		F (%)	66 (100)	2 (100)	13 (100)	11 (100)	40 (100)	3 (100)	79 (100)	142 (100)

Table 11: Participants' mean and standard deviation of scientific and non-scientific hours of studies and level of knowledge

Score	Scientific hours studying Mean \pm SD (n)	Nonscientific hours studying Mean \pm SD (n)
Poor	18.75 \pm 15.54 (77)	7.10 \pm 7.98 (37)
Good	18.26 \pm 14.81 (54)	6.86 \pm 6.28 (37)
Total	18.55 \pm 15.19 (131)	7.01 \pm 7.35 (98)

DISCUSSION

It is essential to explore university students' knowledge of big data analysis. In the present research, students' knowledge was explored across different fields of study in Mashhad. The majority of students belonged to the 21-58 age group; they were female; a minority had work experience. Most of the participants studied medical sciences, professional Ph.D. and pure basic sciences.

Those belonging to the 25-44 year age group received the best scores and those in the 18-24 and 45-54 year age group received the lowest scores. It seems that youth is the best time for knowledge acquisition.

CONCLUSION

Considering the fact that the present research participants were affiliated with the best reputed universities in Iran, they were expected to have a better knowledge of big data analysis. Moreover, the lacking correlation between hours of scientific or non-scientific studies and knowledge of big data can show the fact that well-planned education is severely lacking with this respect. Thus, preliminary

education on this is essential in many fields of study. Holding conferences and seminars can also be effective.

ACKNOWLEDGEMENTS

The present study is the result of research project approved by the vice chancellor for research of Mashhad University of Medical Sciences (grant number 961731).

AUTHOR'S CONTRIBUTION

The authors agree on this final form of the manuscript, and attested that all authors contributed in the final draft of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

- Manogaran G, Lopez D, Thota C, Abbas KM, Pyne S, Sundarasekar R. Big data analytics in healthcare Internet of Things. In: Qudrat-Ullah H, Tsasis P [Eds].
- Innovative healthcare systems for the 21st century. Springer, Cham; 2017.
- Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare.

-
- Biomed Res Int. 2015; 2015: 370194. PMID: 26229957 DOI: 10.1155/2015/370194 [[PubMed](#)]
3. Srinivasan U, Arunasalam B. Leveraging big data analytics to reduce healthcare costs. *IT Professional*. 2013; 15(6): 21-8.
 4. Kim GH, Trimi S, Chung JH. Big-data applications in the government sector. *Communications of the ACM*. 2014; 57(3): 78-85.
 5. Hansen MM, Miron-Shatz T, Lau AY, Paton C. Big data in science and healthcare: A review of recent literature and perspectives. *Yearb Med Inform*. 2014; 9(1): 21-6. PMID: 25123717 DOI: 10.15265/IY-2014-0004 [[PubMed](#)]
 6. Archana J, Anita EM. A survey of big data analytics in healthcare and government. *Procedia Computer Science*. 2015; 50: 408-13.
 7. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inform*. 2008; 77(5): 291-304. PMID: 17951106 DOI: 10.1016/j.ijmedinf.2007.09.001 [[PubMed](#)]
 8. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph*. 2007; 31(4-5): 198-211. PMID: 17349778 DOI: 10.1016/j.compmedimag.2007.02.002 [[PubMed](#)]
 9. Acharjya DP, Ahmed K. A survey on big data analytics: Challenges, open research issues and tools. *International Journal of Advanced Computer Science and Applications*. 2016; 7(2): 511-8.
 10. Alharthi A, Krotov V, Bowman M. Addressing barriers to big data. *Business Horizons*. 2017; 60(3): 285-92.