International journal of innovation in Engineering

journal homepage: www.ijie.ir

Research Paper

A Survey of Students' Attitudes to Big Data Analysis in Iranian Universities

Elham Nazari[a], Somayeh Norouzi[b], Tahmineh Aldaghi[c], Marjan Rasoulian[a], Mohammad Hasan Shahriari[d], Azam Kheirdoust[a], Hamed Tabesh[a1]

[a]*Department of Medical Informatics, faculty of medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

[b]*PhD Student of medical informatics, Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran;*

[c]*Department of Industrial Engineering, Tarbiat Modares University, Tehran, Iran*

[d]*Department of Electrical Engineering, Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran*

ARTICLE INFO

Received: 02 September 2021

Reviewed: 16 September 2021

Revised: 18 October 2021

Accepted: 22 October 2021

Keywords:

Big Data, Benefits, Challenges, Analysis

ABSTRACT

Today, with the emergence of new technologies and massive data, big data analysis has attracted the attention of researchers, industries and universities on a global scale. The present research aims to explore students' attitude to big data analysis in different fields of study. The present cross-sectional study was conducted with students at different universities and fields of study in Iran. A questionnaire was developed. This questionnaire explored students' attitude toward big data analysis. To this aim, 359 university students participated in the research. The data were analyzed using descriptive and inferential statistics. The age of the students ranged between 25 and 34 years. 55.2% were female and 54% were economically active. 40.9% had a work experience of less than a year. The academic degree of the majority of participants was master's degree. 93.9% of the participants believed that big data analysis was essential for the country. 43.2% maintained that big data mostly belonged to the communication industry. 28.1% perceived MATLAB useful software for analysis. 40.9% were familiar with the benefits of analysis. Engage in economic activities, less than 1 year of experience and studies for a Master's degree showed to be significantly correlated with familiarity with the benefits of big data ($p \leq 0.01$). Such issues as high costs, managers' unfamiliarity and lack of expertise and complexity were raised by the respondents. Considering the undeniable benefits of big data analysis, it seems essential to familiarize university students with these analyses through particular training courses, conferences and so on.

[1] **Corresponding Author**
rp666600@gmail.com

1. Introduction

Along with the emergence of new technologies such as smartphones and PCs and the ever increasing use of the internet in society and all industries, bog loads of data are produced (Ahmed et al., 2017). The total amount of data produced by humans until 2003 was 5 Exabyte (1,018 bytes). Yet, today this load of data is created only in two days. In 2012, a total amount of 2.72 Zettabyte (1,021 bytes) was developed and predicted. This amount is doubled every two years. There are 6 billions of mobile phone subscribers worldwide, which means 10 billion text messages are sent a day. Until 2020, 50 billion devices will be connected to the internet and networks. 571 new websites are created per minutes. In the next decade, the size of information will be 50 fold. However, the number of information technology experts to store that information will be 1.5 fold (Sagiroglu & Sinanc, 2013). It is believed that this information fuels future economic information (Jin et al., 2015). These types of data result from log files, texts, images, etc., known as big data. They are characterized by high load, variety, complexity and uncertainty. Therefore, they are not manageable by simple software and hardware. Correct analysis of these data can contribute to making important decisions. There are quite many benefits of big data analysis, such as finding latent patterns of data, finding and prioritizing the most important information, data accumulation and sharing and reduced costs (Kaisler et al., 2013; Nozari et al., 2021). These analyses are used in different industries including agriculture, mining, information, communication, education, transportation (Zhang et al., 2018). Besides the many applications and benefits, there are certain challenges too, such as low quality data, insufficient experts, and security issues and so on. These analyses would require familiarity with relevant methods, instruments, platforms and concepts (Bahrmani et al., 2021; Bossé & Solaiman, 2016; Ristevski & Chen, 2018; Saheb & Saheb, 2020). Considering the significance of big data analysis in different industries and as university students and their research pave the way for communications with industry and applied research. Moreover, this field of study is still at a novice stage in Iran and there is a lack of familiarity with the relevant concepts. The present research aims to explore university students' attitude toward big data analysis in different universities.

2. Material and Methods

2.1. Study Location

The present survey was conducted on 359 students in different universities of Iran to explore students' attitude to big data analysis. These students studied different majors and they expressed their attitude through a questionnaire.

2.2. Instrument

First, a self-administrated semi-structured questionnaire was formed based on related literature in Google Scholar, Science Direct and EMBASE. The questionnaire consisted of closed-ended items and one open-ended item. Then, a Delphi method was used in two rounds facilitated with 10 experts from different majors (Health Informatics, Biostatistics and Computer Sciences).

There were 5 items overall, 4 closed-ended and 1 open-ended to explore university students' attitude to big data analysis. The items can be observed in Table (1).

Table 1. Items exploring university students' attitude towards big data analysis

Item	Descriptor
1	Do you think big data management should be implemented in Iran?
2	In which industry do we need more big data?
3	Which software can best fit big data analysis?
4	What is the best benefit of big data analysis?
5	Why isn't the big data field considered seriously in Iran?

2.3. Questionnaire Validation

The reliability and validity of the questionnaire were confirmed. A panel of 7 experts confirmed the validity of the questionnaire and its reliability was confirmed by the test-retest method on 30 university students selected through convenience sampling method. To test the reliability, Cronbach's alpha was estimated which was 0.73.

2.4. Questionnaire Validation

The questionnaires were submitted to 359 university students. It was attempted to include students from different majors such as medicine, dentistry, pharmacology, biotechnology, toxicology, nanomedicine, bioinformatics, nutrition, medical imaging, microbiology, physiology, genetics, medical informatics, biochemistry, immunology, health IT, molecular cellular, medical physics, mechanics, natural sciences, water, industries, aerospace, metallurgy, computer sciences, civil engineering, engineering sciences, statistics, math, physics, chemistry and basic sciences. The data were collected and it was ensured that all the questionnaires were completed. Incomplete or defective questionnaires were excluded from the study.

2.5. Data Analysis

The IBM SPSS version 21 and Excel 2007 were used to analyze data. Statistical significance for all of the analysis was defined as $p \leq 0.05$. Descriptive statistics were used to summarize the demographic information about the participants, knowledge level, attitude and, challenges. Chi-square test and t-test were used to compare the differences in knowledge and attitude within the demographic characteristics. As well, the relationship between the knowledge levels and the attitude were assessed by Chi-square test.

3. Results

In the present research, 359 university students of different fields of study and degrees participated. The results are presented below.

As it can be observed, the majority of participants are between 24 and 25 years of age. They are mainly female. Few have the experience of economic activity and their work experience is under 1 year. The majority of student's study for a Master's degree in computer engineering, medical informatics, HIT and so on. The average hours of academic studies per week exceed 5 hours and the hours of non-academic studies per week is less than an hour.

As for the necessity of big data management, 94% of the participants agreed with the essentiality of the issue. Concerning the industry, which can mostly be associated with big data, each participant could choose more than one answer. As for familiarity with the benefits of big data, 40.9% of the participants answered correctly.

Table 2. Participants' demographic information (n=359)

Variables	Sub-variables	Frequency	Percentage
Age	18-24 year	115	32.0
	25-34 year	187	52.1
	35-44 year	52	14.5
	45-54 year	۳	.8
	no response	۲	.6
sex	male	155	43.2
	female	198	55.2
	no response	6	1.7
Are you currently involved in any economic activity?	yes	160	44.6
	no	194	54.0
	no response	5	1.4
previous experience	<1 year	147	40.9
	1-5 year	88	24.5
	5-10 year	51	14.2
	>10 year	39	10.9
	no response	34	9.5
Field of study	Professional medical Ph.D.	51	14.2
	Basic medical sciences (MS,Ph. D)	66	18.4
	Professional Ph.D. of pharmacology	24	6.7
	Higher education of pharmacology(MS, Ph. D)	3	.8
	Professional Ph.D. of dentistry	11	3.1
	Pure basic sciences(BS, MS, Ph. D)	8	2.23
	Specific majors*	۱۱۶	32.31
	Engineering	73	20.3
	no response	8	2.2
Degree	B.S.	87	24.2
	M.S.	138	38.4
	Professional Ph.D.	79	22.0
	Specialized Ph.D.	55	15.3
Previous academic field of study	None	33	9.2
	Medical basic sciences	38	10.6
	Pharmacology	5	1.4
	Pharmacology basic sciences	1	.3
	Dentistry	31	8.6
	Engineering	46	12.8
	Basic sciences	48	13.4
	(Computer, Informatics, HIT, Medical records)	48	13.3
	no response	142	39.6
Average hours of field-related studies per week	<1 hour	56	15.6
	1-3 hour	75	20.9
	3-4 hour	54	15.0
	>5 hour	171	47.6
	no response	3	.8
Average hours of non-academic studies per week	<1 hour	156	43.5
	1-3 hour	116	32.3
	3-4 hour	37	10.3
	>5 hour	48	13.4
	no response	2	.6

* Specific majors are those expected to be more relevant to big data than others. Instances are computer sciences, medical records, medical informatics, IT[2] and HIT[3].

[2] Information Technology
[3] Health Information Technology

Industries that according to participant's opinion can benefit from big data analytics are listed below.

Table 3. Distribution of participants' attitude toward the industry, which can most benefit from the big data analysis

Industry	Frequency	Percentage
Agriculture, forestry	60	16.7
Industry and mining	74	20.6
Energy provision	109	30.4
Transportation and storage	98	27.3
Information communication	155	43.2
Financial affairs and insurance	113	31.5
Professional affairs	93	25.9
Official affairs	61	17.0
Education	111	
Health-related affairs	117	32.6
Others	16	4.5

Over 30% of the participants believed that big data analysis could be beneficially used in energy provision, financial affairs and insurance, education and health-related activities. The range of software that can be used for big data analysis by the participants is presented in the following table.

Table 4. Distribution of participants' attitude toward the software that can be best used in the big data analysis

Software	Frequency	Percentage
R	94	26.2
Matlab	101	28.1
Python	79	22.0
Hadoop	64	17.8
Spark	57	15.9
Flink	42	11.7
SPSS	75	20.9
Weka	26	7.2
Rapidminder	24	6.7
SQL	72	20.1

Among the range of software introduced, MATLAB was on the top in terms of frequency. The next ranks belonged, respectively, to R, Python, SPSS and SQL. In the following, a table is presented to indicate the necessity of big data management as perceived by age, sex and economic activity.

Table 5. Distribution of participants' attitude toward the necessity of big data management and three demographic variables, age, sex and economic activity

variable	Sub-variable	Big data management		Total
		Agree	Disagree	
Age	18-24 years	104	8	112
	25-34 years	180	5	185
	35-44 years	48	3	51
	45-54 years	3	0	3
Sex	male	146	8	154
	female	185	8	193
Economic activity	yes	154	5	159
	no	178	11	189

Among 351 participants responding the necessity of big data management item, 16 expressed disagreements. These were distributed in different age groups. Chi-squared test results showed no statistically significant correlation between the perceived necessity of big data management and sex ($p=0.643$). Among 348 participants responding to the necessity of big data management item, 16 expressed disagreements. These respondents were among those with little or no economic activity. Chi-squared test results revealed no

statistically significant correlation between the necessity of big data management and economic activity (p=0.235). The following table indicates the perceived benefits of big data and two demographic variables of sex and age along with other variables including economic activity, work experience, academic grade, and average hours of academic and non-academic studies.

Table 6. Distribution of participants' attitude towards the benefits of big data and sex, age, economic activity, work experience, academic grade, average hours of academic and non-academic studies

Variable	Sub-variable	Benefit of big data		p-value*
		correct	incorrect	
Age	18-24 years	85 (77.3%)	25 (22.7%)	<.001
	25-34 years	84 (47.2%)	94 (52.8%)	
	35-44 years	25 (51.0%)	24 (49.0%)	
	45-54 years	1 (33.3%)	2 (66.7%)	
	n.	195 (57.4%)	145 (42.6%)	
Sex	male	77 (51.3%)	73 (48.7%)	.056
	female	116 (61.7%)	72 (38.3%)	
	n	193 (57.1%)	145 (42.9%)	
Economic activity	Yes	69 (45.7%)	82 (54.3%)	<.001
	No	124 (66.3%)	63 (33.7%)	
	n	193 (57.1%)	145 (42.9%)	
Experience	<1 year	89 (62.7%)	53 (37.3%)	.028
	1-5 year	44 (51.8%)	41(48.2%)	
	5-10 year	19 (38.8%)	30 (61.2%)	
	>10 year	18 (51.4%)	17 (48.6%)	
	n	170 (54.7%)	141 (45.3%)	
Degree	B.S	53 (64.6%)	29 (35.4%)	.001
	M.S	56 (43.1%)	74 (56.9%)	
	Professional Ph.D.	53 (69.7%)	23 (30.3%)	
	Specialized Ph.D.	33 (61.1%)	21 (38.9%)	
	n	195 (57.0%)	147 (43.0%)	
Hours of academic studies	<1 hour	27 (50.9%)	26 (49.1%)	.206
	1-3 hour	46 (62.2%)	28 (37.8%)	
	3-4 hour	23 (46.0%)	27 (54.0%)	
	>5 hour	98 (59.8%)	66 (40.2%)	
	n	194 (56.9%)	147 (43.1%)	
Hours of non-academic studies	<1 hour	85 (57.0%)	64 (43.0%)	.794
	1-3 hour	65 (57.0%)	49 (63.0%)	
	3-4hour	21 (63.6%)	12 (36.4%)	
	>5 hour	24 (52.2%)	22 (47.8%)	
	n	195 (57.0%)	147 (43.0%)	

It appears that the proportion of those in the 25-34 years' age group who agree with the necessity of big data management is higher than other age groups. Though the test of significance does not confirm any statistically significant correlation, it seems that a higher proportion of male participants agreed with the benefits of big data. Chi-squared test results showed that a higher proportion of those economically active agreed with the benefits of big data than those who are less economically active. It can be observed that the proportion of those with less than a year of work experience, but familiar with the benefits of big data is higher than other groups. Moreover, a higher proportion of students in a Master's degree are familiar with the benefits of big data. Chi-squared test results showed no statistically significant correlation between the hours of academic studies and belief in the benefits of big data (p=0.206). Similarly, no significant correlation was found between the hours of non-academic studies and belief in the benefits of big data analysis (p=0.794). In the following, Frequency distribution of useful software for analyzing Big data by study participants based on age and economic activity, experience and necessity of big data analysis, field of study, degree and hour of academic study is shown.

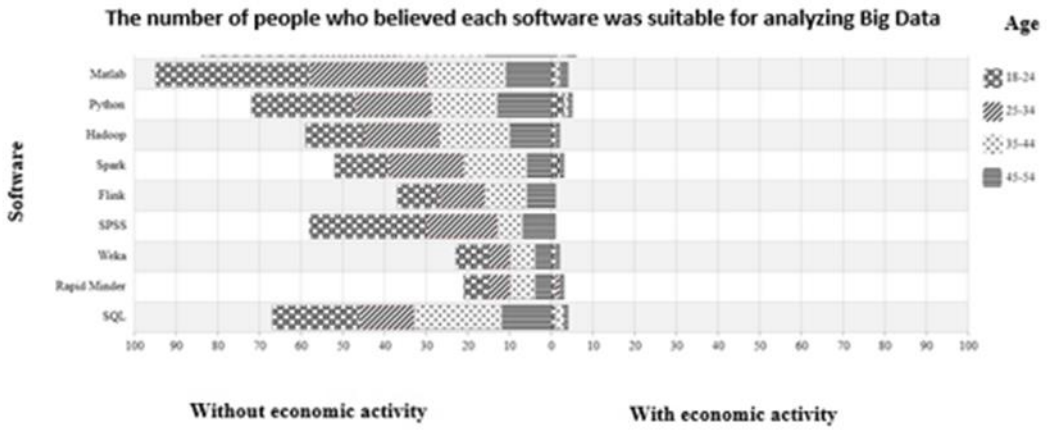

Fig. 1. Frequency distribution of useful software for analyzing Big data by study participants by age group or lack/have on economic activity

As can be seen, most of the people who chose MATLAB were between the ages of 24-18 year and were economically active.

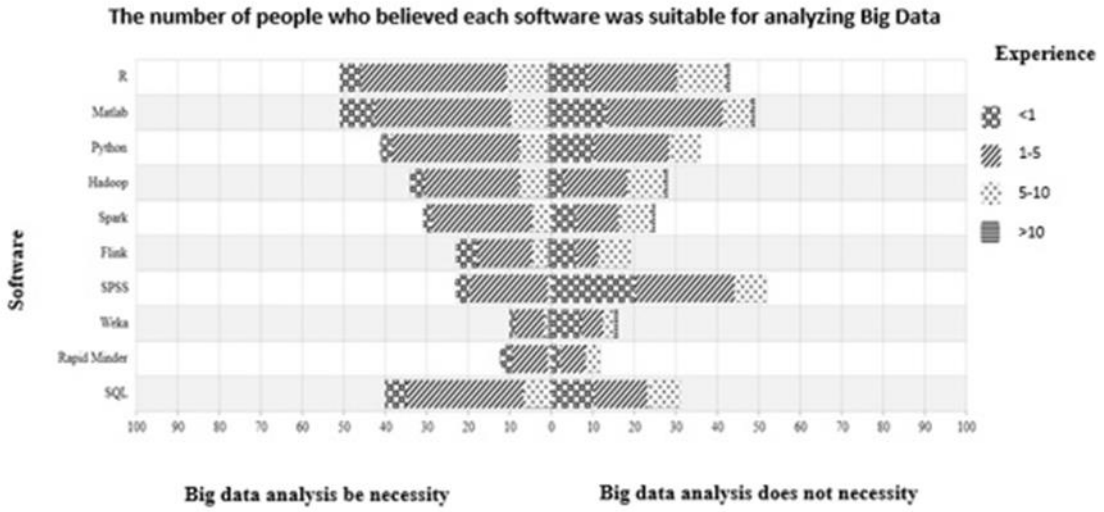

Fig. 2. Frequency distribution of useful software for analyzing big data by study participants in terms of work analysis experience, with or without knowledge of big data analysis

As can be seen, most people with work experience found R & MATLAB to be a useful software. Label for field of study on Figure (3) is displayed in the Table (7).

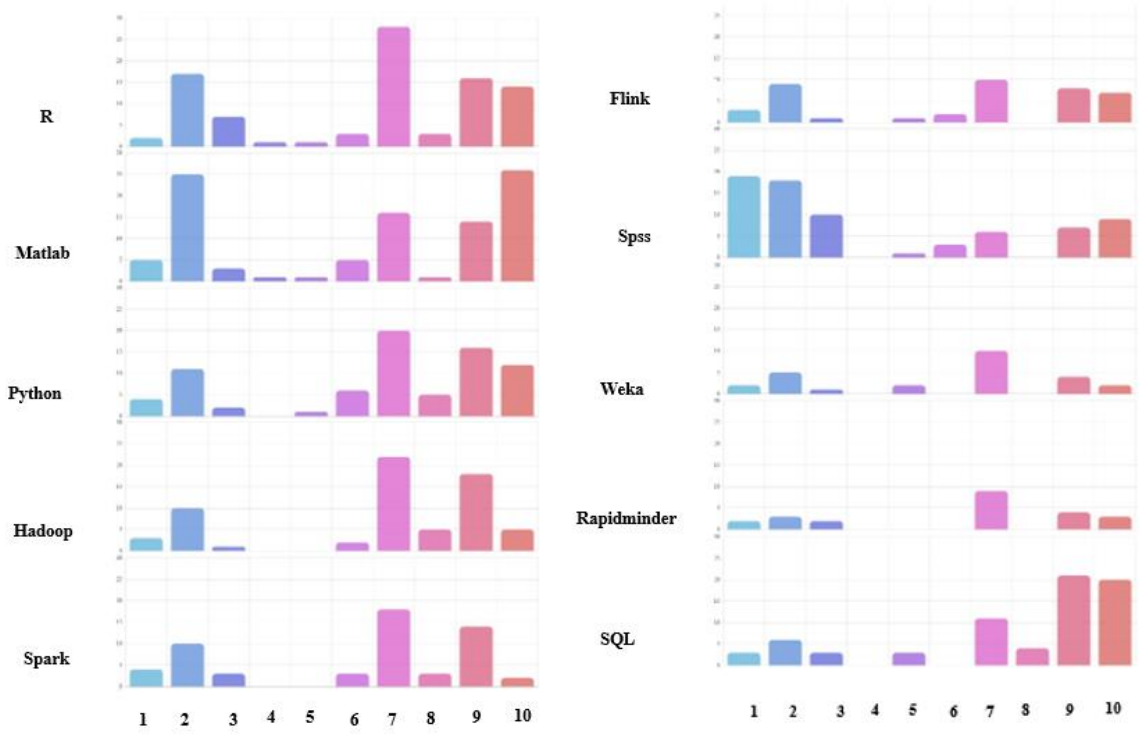

Fig. 3. Frequency distribution of student who believed each software's were suitable for big data analysis by field of study

Table 7. Label for field of study

Label	Field of Study
1	Professional medical Ph.D.
2	Basic medical sciences/ (Ms,phD)
3	Professional Ph.D. of pharmacology
4	Higher education of pharmacology(Ms,phD)
5	Professional Ph.D. of dentistry
6	Pure basic sciences/(BS,MS,phD)
7	Specific majors
8	IT
9	Computer engineering
10	Other engineering

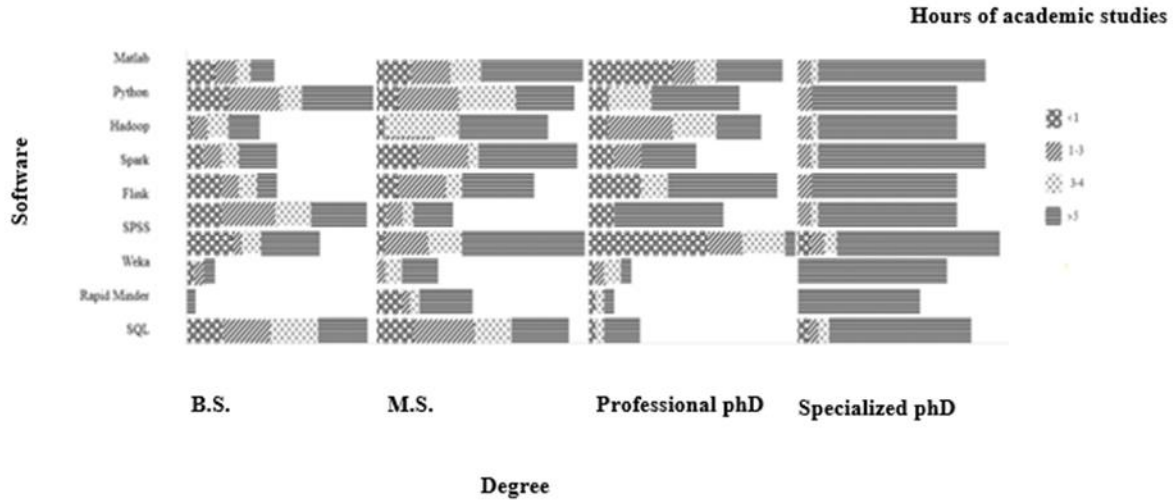

Fig. 4. Frequency distribution of appropriate software for analyzing big data by study participants according to the degree and number of hours' study.

As can be seen, people who had more hours of scientific study and they were at the specialized PhD degree were more familiar with the software

4. Students' perceived challenges of big data

Students were asked to comment on why big data analysis has not been taken seriously in Iran. The reasons explored in an open-ended item within the questionnaire were: high costs, unawareness of the significance of big data analysis, managers' unfamiliarity and lacking expertise, absence of a systematic procedure for data recording in an organization, low quality of data, complexity and interest in conventional methods.

5. Conclusion

The emerging high loads of big data and necessity of big data analysis in industry have made it essential to familiarize students and researchers with the topic at hand. Thus, the present research aimed to explore university students' attitude toward big data analysis in Iran.

The majority of student participants in this research were young; they were mostly female and also economically active. The majority of participants had less than a year's work experience and their academic grade was M.S; the average hours of academic studies, mostly exceeded 5 hours and that of non-academic studies were less than an hour. A vast majority of participants perceived big data analysis essential for the country and believed that the largest amount of big data belonged to the information communication field. A lower proportion of participants were familiar with the benefits of big data analysis. This would require certain plans to raise their awareness of educational and industrial centers. Involvement in economic activity, less than one year of work experience and studies for a Master's degree showed to be significantly correlated with familiarity with the benefits of big data analysis. It seems that because M.S. students are at the outset of academic research, they spend more time getting to know the relevant concepts. That can be why these students are better familiar with the benefits of big data analysis. Similarly, those who are more economically active and have work experience probably enjoy more experience gained from work and, thus, are more aware of the benefits.

Concerning why in Iran big data analysis has not been taken seriously, a variety of comments were made, including high cost, unawareness of the significance of big data analysis, managers' unawareness and lacking expertise, low quality of the collected data, complexity of data, interest in conventional methods and so on. It seems that interest in conventional methods results from a fear of change and the probable more work load

imposed and threat to job security. Awareness of the benefits of big data analysis is another factor that prevents the use of such analyses. If raising the awareness of organizations, managers and personnel are prioritized and the benefits and applications of these analyses are clarified to organizations, people and government, effective measures can be taken to further motivate people to use big data analysis. Managers can use big data analysis to collect the required data. Moreover, experts can be trained to cut down on the existing complexities. Preliminary education can be provided on this topic in many fields of study. Holding conferences and conventions can be helpful in this regard.

References

- Ahmed, E., Yaqoob, I., Hashem, I. A. T., Khan, I., Ahmed, A. I. A., Imran, M., & Vasilakos, A. V. (2017). The role of big data analytics in Internet of Things. Computer Networks, 129, 459-471.
- Bahrami, F., Kanaani, F., Turkina, E., Moin, M. S., & Shahbazi, M. (2021). Key Challenges in Big Data Startups: An Exploratory Study in Iran. Iranian Journal of Management Studies, 14(2), 273-289.
- Bossé, É., & Solaiman, B. (2016). Information fusion and analytics for big data and IoT. Artech House.
- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National science review, 1(2), 293-314.
- Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. Big data research, 2(2), 59-64.
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In 2013 46th Hawaii international conference on system sciences (pp. 995-1004). IEEE.
- Nozari, H., Fallah, M., Kazemipoor, H., & Najafi, S. E. (2021). Big data analysis of IoT-based supply chain management considering FMCG industries. Бизнес-информатика, 15(1 (eng)).
- Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. Journal of integrative bioinformatics, 15(3).
- Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE.
- Saheb, T., & Saheb, T. (2020). Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works. Journal of Big Data, 7(1), 1-26.
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. Information Fusion, 42, 146-157.

This work is licensed under a Creative Commons Attribution 4.0 International License.