JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Awareness of Artificial Intelligence amongst Management Students of Indore Region

Vinod Mishra*
Tapas Jain, Gautam Ghode**, Pankhuri Mishra*****
***Associate Professor, IPS Academy, IBMR, Indore**
**** Assistant Professor, IPS Academy, IBMR, Indore**
****** Asst. Professor, Little Angles Institute of Professional Studies, Mhow.**

Abstract

In the study, researchers have identified the scope of big data analytics concepts, programs, and techniques, by exploring the awareness among scholars, industrial professionals, and academicians. The impact of demographic profiling has been measured using ANOVA, and the chi-square test on awareness of big data analytics concept. Descriptive statistics have been studied and hypothesis testing was performed in the study. In the study, it was found that the level of awareness is almost similar for different cadres of the society in Indore region and there is much need to create Big Data Awareness programs for students, teachers, and industry for big data paradigm. In the study techniques like Predictive, prescriptive, and descriptive analytics were enquired for their awareness. and software applications and programming have been also explored for SAS, R, Tableau, etc…

Keywords: Big data analytics, Awareness, Hadoop, Python, prescriptive analytics, predictive analytics.

Introduction

Many organizations are currently using data to make better decisions about their strategic and operational directions. Using data to make decisions is not new; business organizations have been storing and analyzing large volumes of data since the advent of data warehouse systems in the early 1990s. However, the nature of data available to most organizations is changing, and the changes bring with them complexity in managing the volumes and analysis of these data. Basu (2013) observed that most businesses run on structured data (numbers and categories) today. However, this does not reflect the complexity of the nature of available corporate data and their untapped hidden business value. According to IBM, 80% of the data organizations currently generated are unstructured, and they come in a variety of formats such as text, video, audio, diagrams, images and combinations of any two or more formats. Most of these unstructured data make their way to corporate data warehouse. The term 'Data warehouse' refers to a central repository of data or a centralised database. It represents an ideal vision of maintaining a central data repository and a living memory of data that can be leveraged for better decision making. Recent developments in database technologies made it possible to collect and maintain large and complex amounts of data in many forms and from multiple sources. In addition, there are analytical tools available that can turn this complex data into meaningful patterns and value, a phenomenon referred to as Big Data. Fisher et al (2012), asserted that big data is data that cannot be handled and processed straightforwardly. Yaseen and Obaid (2020), said that big data is a term for massive data sets having large, more varied, and complex structures with the difficulties of storing, analyzing, and visualizing for further processes or results. Bakshi (2013) asserted that unstructured data is the fastest-growing data and needs to be organized, structured, and analyzed to overcome the problem of big and growing data. Manyika et al, (2011) described data that is fundamentally too big and moves too fast, thus exceeding the processing capacity of conventional database systems. It also covers innovative techniques and technologies to capture, store, distribute, manage, and analyze larger-sized data sets with diverse structures.

With new concepts, critiques emerge. Some critics contested that the notion of Big in the term itself is misleading and that it does not reflect only data size but complexity. Yang (2013) pointed out the definition of Big Data has little to do with the data itself, as the analysis of large quantities of data is not new, rather Big Data refers to an emergent suite of technologies that can process mass volumes of data of various types at faster speeds than ever before. This conceptualization of Big Data was echoed by Forrester defining Big Data as "technologies and techniques that make capturing value from data at an extreme scale economical." The term economical suggests that if the costs of extracting, processing and making use of data exceed the advantages to be collected, then it is not worth indulging in the process. Dijicks. (2012), identified volume, velocity, variety and value as characteristics of big data, whereas Beyer and Laney (2012), identified 'High volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making as characteristics of Big Data. Described as under:

• Volume—a large amount of information is often challenging to store, process, transfer, analyze, and present.

• Velocity—relating to the increasing rate at which information flows within an organization—(eg, organisations dealing with financial information have ability to deal with this).

• Veracity refers to the biases, noise and abnormality in data. It also looks at how data is being stored, and meaningfully mined to the problem being analysed. Veracity also covers questions of trust and uncertainty.

• Variety—referring to data in diverse format both structured and unstructured.

• Verification—refers to data verification and security.

• Value—most importantly, has the data been utilised to generate value of the insights, benefits, and business processes, etc. within an organisation?

Yohanna et al (2017) in Forrester lamented ‘Big data warehouse (BDW) is a modern architecture that combines both of the best worlds — leveraging existing data warehouse and new big data technologies that enable organizations to support their growing analytical requirements. Bayer and Laney (2012) proposed three of the most common properties of Big Data in Gartner's report. The report made three fundamental observations: the increasing size of data, the growing rate at which it is produced, and the cumulative range of formats and representations employed. Douglas proposed a threefold definition encompassing the “three Vs” (Volume, Velocity, and Variety).

There are other properties of Big Data also, such as data validity, which refers to the accuracy of data, and volatility, a concept associated with the longevity of data and their relevance to analysis outcomes, as well as the length required to store data in a useful form for appropriate value-added analysis. In addition to these properties, there are three stages required to unlock the value of Big Data in any organization. These include data collection, data analysis, visualization, and application.

Research Objectives

1. To study the awareness level of big data analytics techniques.
2. To study the awareness level of big data analytics software.
3. To study the awareness level of big data analytics solutions.
4. To study the impact of demographic profile on Awareness level.

Research Methodology

Data Collection

The sample: The sample of 111 respondents was analysed in the study using survey method, in which a questionnaire as constructed and as sent via google forms online to various social networks. The non probability sampling technique is applied to collect the data.

Tools & Techniques

In the study impact of demographic profiling has been measured using ANOVA, and chi-square test on awareness of big data analytics concept. The descriptive statistics have been studies and hypothesis testing as performed in the study.

Results and Discussions

Demographic Profile

Age (in years)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15 - 25	102	91.9	92.7	92.7
	25 – 30	7	6.3	6.4	99.1
	30 – 50	1	.9	.9	100.0
	Total	110	99.1	100.0	
Missing	System	1	.9		
Total		111	100.0		

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	MALE	53	47.7	47.7	47.7
	FEMALE	58	52.3	52.3	100.0
	Total	111	100.0	100.0	

Qualification

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	XII	7	6.3	6.4	6.4
	UG	43	38.7	39.1	45.5
	PG	56	50.5	50.9	96.4
	Ph.D.	4	3.6	3.6	100.0
	Total	110	99.1	100.0	
Missing	System	1	.9		
Total		111	100.0		

Location Background

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	RURAL	9	8.1	8.2	8.2
	SEMI-URBAN	35	31.5	31.8	40.0
	URBAN	66	59.5	60.0	100.0
	Total	110	99.1	100.0	
Missing	System	1	.9		
Total		111	100.0		

Category you belong to

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	STUDENT	99	89.2	90.0	90.0
	TECAHER	3	2.7	2.7	92.7
	INDUSTRY PROFESSIONAL	8	7.2	7.3	100.0
	Total	110	99.1	100.0	
Missing	System	1	.9		
Total		111	100.0		

Work experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0-1 YEAR	82	73.9	77.4	77.4
	1-3 YEARS	16	14.4	15.1	92.5
	3-5 YEARS	2	1.8	1.9	94.3
	5 & ABOVE	6	5.4	5.7	100.0
	Total	106	95.5	100.0	
Missing	System	5	4.5		
Total		111	100.0		

Industry

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	IT	6	5.4	5.7	5.7
	BANKING & FINANCE	20	18.0	19.0	24.8
	ACADEMICS	37	33.3	35.2	60.0
	MANUFACTURING	7	6.3	6.7	66.7
	OTHERS	30	27.0	28.6	95.2
	RETAIL	5	4.5	4.8	100.0
	Total	105	94.6	100.0	
Missing	System	6	5.4		
Total		111	100.0		

Hypothesis Testing

S.No.	**Null Hypothesis**	**F Value/Chi-Square Value**	**P-Value**	**Result**
1	**There is no significant impact of Age on Awareness level of Big Data Streams**	.492	.613	**Accepted**
2	**There is no significant impact of Age on Awareness level of Big Data Analytics' softwares.**	**7.94**	**0.001**	**Rejected**
3	**There is no significant impact of Age on Awareness level of Big Data Techniques**	**1.231**	**.296**	**Accepted**

4	**There is no significant impact of Gender on Awareness level of Big Data Streams**	**27**	**.017**	**Rejected**
5	**There is no significant impact of Gender on Awareness level of Big Data Analytics' softwares.**	**7.8**	**.342**	**Accepted**
6	**There is no significant impact of Gender on Awareness level of Big Data Techniques**	**4.5**	**.339**	**Accepted**
7	**There is no significant impact of Qualification on Awareness level of Big Data Streams**	**.453**	**.716**	**Accepted**
8	**There is no significant impact of Qualification on Awareness level of Big Data Analytics' software.**	**5.714**	**.001**	**Rejected**
9	**There is no significant impact of Qualification on Awareness level of Big Data Techniques**	**.768**	**.515**	**Accepted**
10	**There is no significant impact of Location on**	**.749**	**.475**	**Accepted**

	Awareness level of Big Data Streams			
11	There is no significant impact of Location on Awareness level of Big Data Analytics' softwares.	.501	.608	Accepted
12	There is no significant impact of Location on Awareness level of Big Data Techniques	.010	.990	Accepted
13	There is no significant impact of Category on Awareness level of Big Data Streams	1.287	.280	Accepted
14	There is no significant impact of Category on Awareness level of Big Data Analytics' softwares.	7.880	.001	Rejected
15	There is no significant impact of Category on Awareness level of Big Data Techniques	2.819	.064	Accepted
16	There is no significant impact of Work Experience on Awareness level of Big Data Streams	1.212	.309	Accepted
17	There is no significant impact of			

	Work Experience on Awareness level of Big Data Analytics' softwares.			
18	**There is no significant impact of Work Experience on Awareness level of Big Data Techniques**	**7.658**	**.000**	**Rejected**
19	**There is no significant impact of Industry on Awareness level of Big Data Streams**	**2.603**	**.056**	**Accepted**
20	**There is no significant impact of Industry on Awareness level of Big Data Analytics' softwares.**	**1.862**	**.108**	**Rejected**
21	**There is no significant impact of Industry on Awareness level of Big Data Techniques**	**2.181**	**.062**	**Accepted**

Conclusion

In the study impact of demographic profiling has been measured using ANOVA, and chi-square test on awareness of big data analytics concept. The descriptive statistics have been studies and hypothesis testing as performed in the study. In the study it was found the level of awareness is almost similar for different cadres of the society in Indore region and there is much need of creating Big Data Awareness programs to make students, teachers and industry ready for big data paradigm.

Links

[Questionnaire]

References

1. Adams, M.N.: Perspectives on Data Mining. (2010). International Journal of Market Research 52(1), 11–19.
2. Asur, S., Huberman, B.A.(2010). Predicting the Future with Social Media. In: ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, pp. 492–499.
3. Bakshi, Kapil (2012). Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7.
4. Basu, A. (2013). Five pillars of prescriptive analytics success. Analytics, pp. 8–12, March/April Issue.[ANALYTICS-MAGAZINE.ORG].
5. Cebr (2012). Data equity, Unlocking the value of big data. in: SAS Reports, pp. 1–44
6. Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J.M., Welton, C. (2012). MAD Skills: New Analysis Practices for Big Data. Proceedings of the ACM VLDB Endowment 2(2), 1481–1492.
7. Cuzzocrea, A., Song, I., Davis, K.C.(2011). Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104 (2011).
8. J. Dijcks, (2012). " Big Data for the Enterprise," Oracle Report, 2012.
9. Beyer, M. A. and D. Laney (2012). "The Importance of “Big Data”: A Definition," Gartner report, 2012.
10. Economist Intelligence Unit (2012). The Deciding Factor: Big Data & Decision Making. Capgemini Reports, pp. 1–24.
11. Elgendy, N. (2013). Big Data Analytics in Support of the Decision-Making Process. MSc Thesis, German University in Cairo, p. 164.
12. EMC (2012). Data Science and Big Data Analytics. In: EMC Education Services, pp. 1–508.
13. Fisher, Danyel, Rob DeLine, Mary Czerwinski and Steven M. Drucker, (2012) Interaction with Big Data Analytics, *Interactions, Vol. 9 (3), pp 50-59*.
14. Gartner, (2014). Available: http://www.gartner.com/newsroom/id/2684616.
15. He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., Xu, Z. (2011). R C File: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems. In: IEEE International Conference on Data Engineering (ICDE), pp. 1199–1208.
16. Khalid Yaseen, Ahmed Mahdi Obaid (2020). Big Data: Definition, Architecture & Applications Humam, *International Journal of Informatics Visualisation*. Vol. 4 (1). pp 45-51.
17. Yohanna, Noel, Gene Leganza, and Jun Lee (2017). The Forrester Wave: Big Data Warehouse, Q2 2017, Forrester Report, 2017. https://b2bsalescafe.wordpress.com/wp-content/uploads/2017/07/the-forrester-wave-big-data-warehouse-q2-2017-june-2017.pdf. (referred on 13 December 2024)

Appendix

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Big_Data_Streams	Between Groups	15.155	2	7.578	.492	.613
	Within Groups	1647.063	107	15.393		
	Total	1662.218	109			
Software	Between Groups	38.522	2	19.261	7.974	.001
	Within Groups	258.469	107	2.416		
	Total	296.991	109			
Techniques	Between Groups	5.437	2	2.719	1.231	.296
	Within Groups	236.381	107	2.209		
	Total	241.818	109			

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Big_Data_Streams	Between Groups	21.041	3	7.014	.453	.716
	Within Groups	1641.177	106	15.483		
	Total	1662.218	109			
Software	Between Groups	41.345	3	13.782	5.714	.001
	Within Groups	255.646	106	2.412		
	Total	296.991	109			
Techniques	Between Groups	5.141	3	1.714	.768	.515
	Within Groups	236.677	106	2.233		
	Total	241.818	109			

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Big_Data_Streams	Between Groups	22.944	2	11.472	.749	.475
	Within Groups	1639.275	107	15.320		
	Total	1662.218	109			
Software	Between Groups	2.753	2	1.377	.501	.608
	Within Groups	294.238	107	2.750		
	Total	296.991	109			
Techniques	Between Groups	.047	2	.023	.010	.990
	Within Groups	241.772	107	2.260		
	Total	241.818	109			

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Big_Data_Streams	Between Groups	39.046	2	19.523	1.287	.280
	Within Groups	1623.172	107	15.170		
	Total	1662.218	109			
Software	Between Groups	38.126	2	19.063	7.880	.001
	Within Groups	258.865	107	2.419		
	Total	296.991	109			
Techniques	Between Groups	12.105	2	6.052	2.819	.064
	Within Groups	229.713	107	2.147		
	Total	241.818	109			

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Big_Data_Streams	Between Groups	54.494	3	18.165	1.212	.309
	Within Groups	1528.346	102	14.984		
	Total	1582.840	105			
Software	Between Groups	51.648	3	17.216	7.658	.000
	Within Groups	229.295	102	2.248		
	Total	280.943	105			
Techniques	Between Groups	16.288	3	5.429	2.603	.056
	Within Groups	212.778	102	2.086		
	Total	229.066	105			

Gender * Big_Data_Streams

Crosstab

Count

		Big_Data_Streams															Total
		2.00	6.00	7.00	9.00	10.00	11.00	12.00	13.00	14.00	15.00	16.00	17.00	18.00	19.00	20.00	
Gender	MALE	0	0	1	2	18	1	5	4	1	1	3	5	1	4	7	53
	FEMALE	1	1	0	0	7	9	7	4	9	1	2	3	4	1	9	58
Total		1	1	1	2	25	10	12	8	10	2	5	8	5	5	16	111

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.354[a]	14	.017
Likelihood Ratio	31.586	14	.005
Linear-by-Linear Association	.129	1	.719
N of Valid Cases	111		

a. 22 cells (73.3%) have expected count less than 5. The minimum expected count is .48.

Gender * Software

Crosstab

Count

		Software								Total
		7.00	8.00	9.00	10.00	11.00	12.00	13.00	14.00	
Gender	MALE	1	1	4	4	3	5	14	21	53
	FEMALE	0	1	0	3	4	5	12	33	58
Total		1	2	4	7	7	10	26	54	111

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.897[a]	7	.342
Likelihood Ratio	9.836	7	.198
Linear-by-Linear Association	4.295	1	.038
N of Valid Cases	111		

a. 11 cells (68.8%) have expected count less than 5. The minimum expected count is .48.

Gender * Techniques

Crosstab

Count

		Techniques					Total
		4.00	5.00	6.00	7.00	8.00	
Gender	MALE	12	8	10	6	17	53
	FEMALE	8	9	10	15	16	58
Total		20	17	20	21	33	111

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.530[a]	4	.339
Likelihood Ratio	4.654	4	.325
Linear-by-Linear Association	.657	1	.418
N of Valid Cases	111		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 8.12.