# Awareness of Big Data in Select Sectors in Hyderabad; India

**A.Rajini[a,b], Santi Rohit Rao[c,b] , and V V V Achutamba[b]**

[a]Research Scholar-Osmania University, India
[b]Assistant Professor, Bhavan's Vivekananda College of Science, Humanities and Commerce, Sainikpuri, Secunderabad, TS, India.
[c]Research Scholar-KLU, Vijayawada, India

**Article History**: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

**Abstract:** Big data produces massive content of accelerated information with variety and in less time by using advanced technology and various meticulous methods for converting the information produced by big data into value. The study aims to analyze the big data awareness amongst the employees in various sectors, investigate different management practices on developing a better strategy in the implementation of big data and study the impact of factors on organization culture in the implementation of big data. Probability random sampling adopted and data analyzed using Chi-square, ANOVA, and Stepwise regression analysis. The survey shows the impact of different select management practices on developing a better strategy in the execution of massive amount of data the Management Practices and Company Strategy is mostly influenced with Support by top level management. Thus, the research hypothesis proves that there is a significant relationship Management Practices and Company Strategies. Thus, the organization culture is mostly influenced by the Cost of data and the Difficulties of data. Therefore, the research hypothesis states that there is a significant relationship between various factors and culture of an organization.

**Keywords**: Big data, Sectors, Management practices, Organization culture, Company strategy.

## 1. Introduction

Many Eminent processes have been introduced in leading software companies and other firms in the last two years. When the big data has come into the business scene it has become a sensation in the last decade and every organization like Google, eBay, yahoo, are concerned about this big data. It produces the data which is very unique in nature and the world's richest data which shows details about behavior patterns, activities and events that happen around the world. Access will be given from enormous resources to various types of data-by-data analytics in a less span of time so that with the help of this new data companies may find out new and innovative ways of earning income. Organizations need a reason for applying big data analytics to determine shape out about the data collection and how it will be sorted and process into final data by using the big data analytics. Majority of the organizations are not aware with the sources of information or conventional forms of collecting data, due to this there is no absorption or acculturation of old technology with big data analytics. For example, A software application, Hadoop works besides the mainframe of IBM classifies enormous types of data. In telecommunication sector few companies implemented Big data through which large volume of subscriber's data collected, sorted and process into final data which helped them to increase their sales. Another example, Walt Disney co. also implemented big data analytics with this there was a rise in their revenue by 20% by introducing "magic wristbands" in their parks. So that they could easily pay attention and manage more people at a time Amazon also get benefited by using the big data to understand the customer in a $360^{O}$ view to increase their business. There are a lot of opportunities in the organizations for big data which leads to business intelligence due to this new technology is used to understand the market competence and take accurate decisions on time.

to understand the market competence and take accurate decisions on time.

### 1.1 Need of the Study

Big data has come into a lot of pre - eminence in recent times. Hence, there is a need to understand the awareness levels of the employees regarding the same.

### 1.3 Scope of the Study

Confined to different sectors and all levels of employees in Hyderabad.

### 2.1 Research Objectives

- To study the big data awareness amongst the employees in various Industries.
- To know the impact of demographic factors on the awareness levels of the big data.
- To understand the impact of different management practices on developing a better strategy in the implementation of big data
- To identify the impact of factors on organization culture in the implementation of big data.

- To provide suggestions to improve the awareness levels.

**2.2 Prior Research**

Akoka et al. (2017) for the last 5 years there was a huge increase in research articles on big data. They observed the interest among the researchers in big data is diverse such as framing the objectives of the research gap , using and application of big data in research articles and also identified various techniques of, how the data is collected, segregated and process to final data. Based on four axis technology, technique, information, and impact, De Mauro, Greco, and Grimaldi (2015) suggested the following definition: Big data produces massive content of accelerated information with variety and in less time by using advanced technology and various meticulous methods for converting the information produced by big data into value Izhar et al., (2013) discovered Big data analytics provided eminent opportunities for various firms where it has an impact on different types of processes in companies. Davenport & Dyche, (2013) noticed in their daily business activities many organizations create a huge volume of data but the problem here is, data is conceived and apprehended in many different formats, which is very difficult to identify the existing relationship of various data , due to different formats of data large content of data is becoming unessential which is unable to correlate with the aims and objectives of the companies. Big Data can be transformed into a unit of information equal to one thousand million which cannot be combined with other technology easily. Cuzzocrea, Song, and Davis (2011) explained that big data provides different characteristics of processed information which consists of proper amount of volume and composition. At the same time it is evident from Bizer, Boncz, Brodie, and Erling (2012) definition of big data not only includes fact traits but also additional attributes such as horizon, aim and composition of the data. At the same time Jacobs (2009) identified the approachability and also the statistics about the big data usage. However, Chen, Chiang, and Storey (2012) mantled the infrastructure of information technology and various technologies used for big data. Madden (2012) incorporated various characteristics of data and infrastructure in information technology. According to Rodríguez-Mazahua et al. (2016), Big Data is a organized knowledge sometimes it may be in the form semi organized and unorganized in various fields such as chemistry, bio chemistry, physics and various business which needs huge applications and technology. Thus, collection of research studies engrossed on the impact of big data in various zones and also noticed how big data will do the contributions in versatile fields. These contributions in various fields are classified into three they are climate science, genetic and the second one is big data paradigm thus the last one is cluster of all fields which uses big data like hospitals, government institutions etc. Akoka et al. (2017): Goes ,(2014)

**2.3 Research Gap**

1. lack of skill and technology in organizations for using big data
2. lack of experience and knowledge in implementing the big data in various business process for decision making
3. lack of encouragement for managers in using the big data for processing the data into fine valuable information.

There is a need to fill the gap in knowledge of big data and the technology used in the organizations for implementation of big data

**2.4 Data Collection & Methodology**

The present research has been conducted on different sectors of employees on the awareness of big data in Hyderabad area. The main aim of this research is to highlight the various factors on developing a better strategy and organization culture by adopting an empirical research method. The research is conducted on a sample of 64 respondents which is collected through a structured questionnaire using Google forms. Statistical tools- ANOVA, Chi-square test, Regression Analysis using IBM SPSS statistics version 25 were used to analyze the data and interpret the results.

**2.5 Demographic Factors**

(Table 1) reveals about the demographic factors that is out of the total sample of 64 employees, 63(62.5%) constitute male employees while the rest 38(37.5%) comprise female. Thus, the majority of the sample represented male employees.

Considering the age , the distribution of the employees is relatively normal, that is 9(9.375%) in the age group of 26-30 years, 19(18.75%) in the age group of 26-30 years, 19(18.75%) in the age group of 31-35 years, 44(43.75%) in the age group of 36-40 years, 19(18.75%) in the age group of 41-45 years, 6(6.25%) in the age group of 46-50, years and 3(3.125%) in the age group of more than 50 years.

The data on Organization, 22(21.875%) belongs to Educational Sector, 17(17.18%) belongs to Manufacturing Sector, 31(31.25%) belongs to IT Sector, 22(21.875%) belongs to Pharma Industry and 8(7.81%) belong to Other Sectors.

Table 1: Frequency distribution of Demographic Factors

| S.No | Variables | Category | Frequency | Percentage |
|---|---|---|---|---|

| . | | | | |
|---|---|---|---|---|
| 1 | Sex | Male | 40 | 62.5 |
| | | Female | 24 | 37.5 |
| | | Total | 64 | 100 |
| 2 | Age | 20-25 | 0 | 0 |
| | | 26-30 | 6 | 9.375 |
| | | 31-35 | 12 | 18.75 |
| | | 36-40 | 28 | 43.75 |
| | | 41-45 | 12 | 18.75 |
| | | 46-50 | 4 | 6.25 |
| | | >50 | 2 | 3.125 |
| | | Total | 64 | 100 |
| 3 | Employee organization | Education | 14 | 21.875 |
| | | Manufacturing | 11 | 17.1875 |
| | | IT | 20 | 31.25 |
| | | Pharma Industry | 14 | 21.875 |
| | | Other | 5 | 7.8125 |
| | | Total | 64 | 100 |

Figure 1 shows that Most of the organizations are using 53.1% of Numerical data and 37.5% of Text data for analyzing the context of Big data. Here observe that 31.3% of the employees don't know which data they are using.

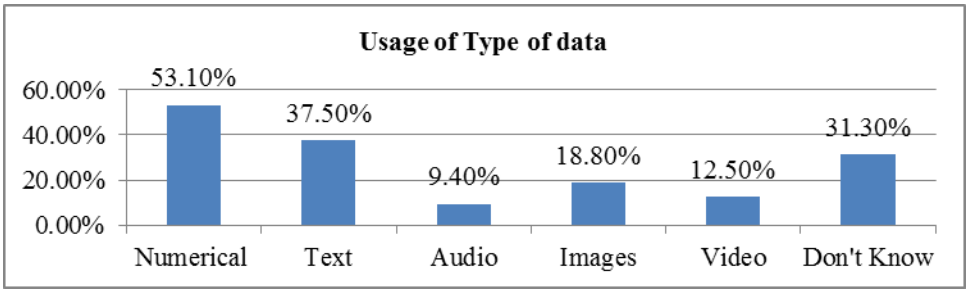


Figure 1: Usage of type of data

From Figure 2 we observe that understanding the applications of big data is very high in Information Technology (31.3%) and low in Customer service, Direct and Online Marketing, Supply chain management and Logistics, Finance and Administration, Risk management (3.1%). 25% of the employees don't know the application of big data in their organization.

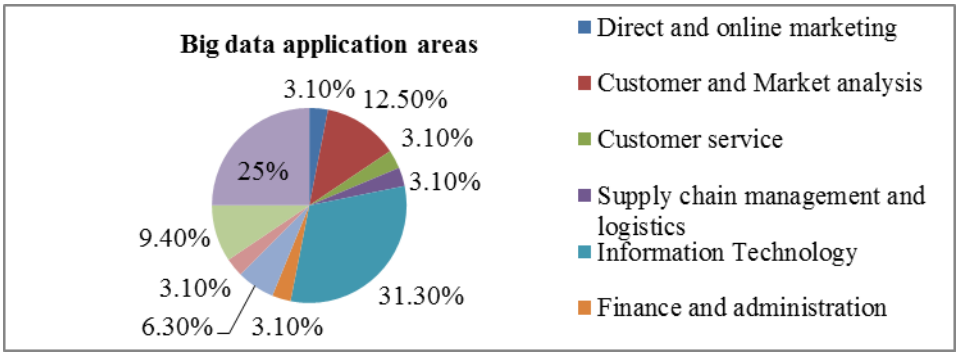


Figure 2: Big data Application areas

Here, in Figure 3 observe that employees' believe that 59.4% of a clear company strategy, 43.8% of training, 34.4% of an organization structure and 21.9% of support by higher management are important factors for successful big data implementation.

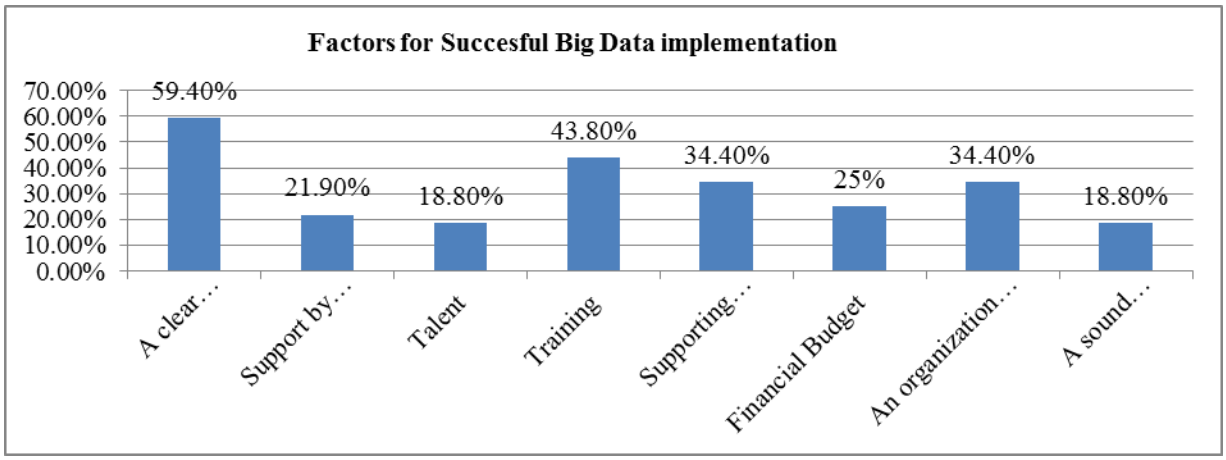


Figure 3: Factors for successful big data implementation

Figure 4 reflects that 31.3% of employees noticed that they have right analytical tools to handle big data in their organization, 18.8% employees are expecting to have them in 5 years, 15.6% employees are having no plans for analytical tools and 34.4% employees don't know about the usage of analytical tools in their organizations.

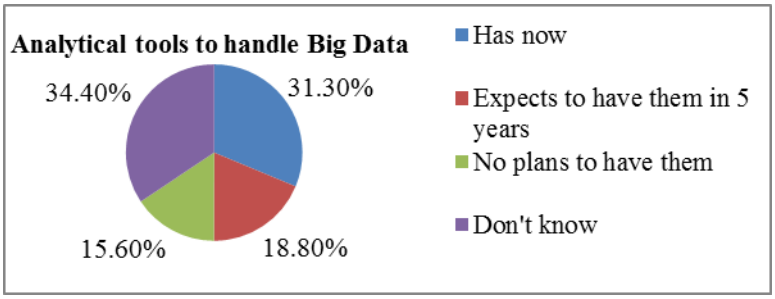


Figure 4: Analytical tools to handle big data

**2.6 Hypotheses of the Study**

Based on the objectives of the study the research hypothesis has been framed:

$H_{11}$: There is an association between Sex and awareness of the big data amongst the various employees.

$H_{12}$: There is an association between Age and awareness of the big data amongst the various employees.

$H_{13}$: There is a significant difference between the various management factors on developing a better strategy in the implementation of big data.

$H_{14}$: There is a significant difference between organization cultures in the implementation of big data.

Chi-Square test is used to test the Hypotheses $H_{11}$ and $H_{12}$ which has taken for the study. From table 2 we can see that p-value is greater than 0.05. Therefore the alternative hypothesis $H_{11}$ is rejected. It is also observed from table 3 that p-value is less than 0.05. Therefore, the hypothesis $H_{12}$ is accepted. Hence, we conclude that there is no relation between Sex and awareness of the big data but, there is a relation between Age and awareness of the big data. That is males and females are having equal knowledge about big data but coming to various age groups this is different.

Table 2: Chi-Square Test for Sex Vs Awareness of big data

| | Value | Df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 6.552[a] | 4 | .162 |
| Likelihood Ratio | 7.665 | 4 | .105 |
| Linear-by-Linear Association | 4.714 | 1 | .030 |
| N of Valid Cases | 64 | | |

Table 3: Chi-Square Test for Age Vs Awareness of big data

| | Value | Df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 33.725[a] | 20 | .028 |
| Likelihood Ratio | 37.878 | 20 | .009 |
| Linear-by-Linear Association | 11.270 | 1 | .001 |
| N of Valid Cases | 64 | | |

**2.7 Impact of Different Select Management Practices on Developing A Better Strategy in Implementation of Big Data**

Here the seven practices which is important for developing a better strategy for implementing big data. Those seven practices has been considered and data has been obtained from 64 respondents of the sample. The regression model has been applied in that company strategy has taken as the dependent variable and the seven practices of company strategy as independent variables viz., support by higher management, talent, training, supporting systems and procedures, financial budget, an organizational structure that supports multi-disciplinary projects and a sound procedure for legal, ethical and reputational issues. The adjusted R-square value, in Table-4, tells that 71.5 percent of the variation in dependent Variable, Company strategy is explained by the independent variable and the Durbin-Watson value is 1.591 which is less than 2 produce the model very explanatory. From Table -5 it can observe that the model is good fit as p value is less than 5% level.

Table-4: Regression Summary of Management Practices and Company Strategy

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | .848 | .719 | .715 | .551 | 1.591 |

Table 5: ANOVA Results for Management Practices and Company Strategy

| Model | | Sum of Squares | Df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 48.198 | 1 | 48.198 | 158.930 | .000 |
| | Residual | 18.802 | 62 | .303 | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Total | 67.000 | 63 | | | |

The Table-6 and Table-7 disclose that the p-values are significant at a 5 percent level and hence management practice that is supported by higher management is explicable. The p-value for other practices coefficients are more than 5% level of significance and hence are excluded in the regression expression.

Table-6: Regression statistics: Coefficients and Test Results for Management Practices and Company Strategy

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | .113 | .172 | | .659 | .012 | -.457 | .231 |
| | Support by higher management | .950 | .075 | .848 | 12.607 | .000 | .799 | 1.100 |

Table-7: Excluded Variables in Regression model of Company Strategy

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics |
|---|---|---|---|---|---|---|
| | | | | | | Tolerance |
| 1 | Talent | .055 | .601 | .550 | .077 | .550 |
| | Training | -.072 | -.978 | .332 | -.124 | .839 |
| | Supporting systems and procedures | .061 | .669 | .506 | .085 | .545 |
| | Financial budget | -.028 | -.337 | .737 | -.043 | .654 |
| | An organizational structure that supports multidisciplinary projects | .032 | .285 | .776 | .037 | .372 |
| | A sound procedure for legal ethical and reputational issues | .021 | .192 | .848 | .025 | .394 |

Therefore, the regression equation for Management Practices and Company Strategy is:

Company Strategy = 0.113 + 0.950*Support by higher management

Thus, the Management Practices and Company Strategy is mostly influenced by Support by higher management. Hence, the research hypothesis tells that there is a significant relationship between Management Practices and Company Strategies.

**2.8 Impact of Factors on Organization Culture in Implementation of Big Data**

Here the seven practices which is important for developing a better strategy for implementing big data. Those seven practices has been considered and data has been obtained from 64 respondents of the sample. The regression model has been applied in that company strategy has taken as the dependent variable and the seven practices of company as independent variable. The adjusted R-square value, in Table-8, shows that 85.6 percent variation in dependent variable that is organization culture is explained by independent variable and Durbin-Watson value is

1.881 which is less than 2 produce the model very explanatory. From Table -9 it can observe that the model is good fit as p value is less than 5% level.

Table-8: Regression Summary of Various Factors on Organization Culture

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|---|---|---|---|---|---|
| 1 | .928 | .861 | .856 | .410 | 1.881 |

Table-9: ANOVA Results for Various Factors on Organization Culture

| Model | | Sum of Squares | Df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 63.487 | 2 | 31.743 | 188.670 | .000 |
| | Residual | 10.263 | 61 | .168 | | |
| | Total | 73.750 | 63 | | | |

The regression statistic Table-10 and Table-11 reveals that the p-values are significant at 5 percent level of significance and hence, the cost of data and difficulties of data are explicable. The p-value for other factors coefficient is more than 5% level are excluded in the regression expression.

Table-10: Regression statistics: Coefficients and Test Results for Various Factors on Organization Culture

| Model | | Unstandardized Coefficients | | Standardized Coefficients | T | Sig. | 95.0% Confidence Interval for B | |
|---|---|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | .229 | .135 | | 1.695 | .045 | .041 | .499 |
| | Cost of data | .750 | .085 | .659 | 8.864 | .000 | .581 | .919 |
| | Difficulties of data | .318 | .074 | .320 | 4.309 | .000 | .170 | .465 |

Table-11: Excluded Variables in Regression model of Organization Culture

| Model | | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics |
|---|---|---|---|---|---|---|
| | | | | | | Tolerance |
| 1 | Timeliness | -.037 | -.466 | .643 | -.060 | .369 |
| | Overwhelming volume | -.108 | -1.436 | .156 | -.182 | .396 |

| | | | | | |
|---|---|---|---|---|---|
| Managing unstructured data | -.134 | -1.756 | .084 | -.221 | .379 |
| Data quality | -.090 | -.962 | .340 | -.123 | .261 |
| Availability of data | .026 | .420 | .676 | .054 | .585 |
| Access rights to data | -.080 | -1.189 | .239 | -.152 | .496 |
| Data ownership issues | .017 | .219 | .828 | .028 | .366 |
| Lack of facilities, infrastructure | -.094 | -1.241 | .220 | -.158 | .395 |
| Lack of pre-processing facilities | -.065 | -.767 | .446 | -.099 | .321 |
| Lack of technology | -.024 | -.269 | .789 | -.035 | .281 |
| Shortage of talent/skills | -.014 | -.195 | .846 | -.025 | .438 |
| Privacy concerns and | .028 | .254 | .800 | .033 | .196 |
| regulatory risks | .062 | .557 | .580 | .072 | .189 |
| Security | .037 | .393 | .696 | .051 | .267 |
| Portability | .098 | .861 | .393 | .111 | .177 |

Therefore, the regression equation for Various Factors on Organization Culture:

Organization culture = 0.229 + 0.750* Cost of data + 0.318* Difficulties of data

**3. Interpretation**

Every area of the organization has an impact of big data analytics. When compared with cloud computing and latest technologies for decision making Big data plays a vital role .Still there is no coherence to assimilate big data and knowledge management to improve in-time decision making and business analytics in organizations. Though research contributions in big data analytics trying to illuminate the industry in what best way it will understand the technical opportunities and accept the challenges in developing and implementing big data, knowledge management and analytics. Organizations has to come up with productive approaches simultaneously has to rationalize the knowledge of the organization to support in order to create and deliver the knowledge in the big data era.

**4. Conclusion**

Because of the exploitation of big data analytics in this industrial revolution is making a way for agility and productive industrial performance leads to growth of the industries in the future. Where there is a metamorphosis in the industrial revolution towards big data analytics, the decision makers in the organizations will get an opportunity to employ more data by taking into account many actions by improvising the goals and objectives of the organization by only developing and implementing big data analytics. Because of implementation of big data goals will be set to higher standard and performance of the organization will be maximized and also they can preferably predict already unpredictable things and upgrade the process performance.

In this study we came to know that there is no relationship between Sex and awareness of the big data but, there is an association between Age and awareness of the big data. It means number of male and female is having an equal knowledge but different age groups of male and female knowledge about big data is different. Because of the impact of the different select management practices there was a development of better strategy in the implementation of big data and also there was a constant and continuous support by the higher management in this regards. So this shows that there is an association between various practices of management and the strategies of the organizations . Thus, the organization culture is mostly influenced by the Cost of data and the Difficulties of data. As well as there is an high amount of association between various factors and culture of the companies. There are more chances of manipulation of big data analytics in the industries of different processes and operations because of various competencies and organizational factors and strategies. Big data may be useful in changing the decision making of every successful organization which are implementing but big data is creating an awareness to protect the business process from various risks with which it is associated.

**5. Limitations And Future Work**

The research study is limited to Hyderabad area . The study shows that only a few factors showing the impact on the strategy of a company and the culture of an organization in select sectors. The other factors may be showing an impact on any other dependent variable. This can be studied in further research and more sample size can be considered for further study.

**6. Suggestions**

- There is a need for employees to understand the data that they are using.
- Irrespective of age group the awareness of big data should be in practice.
- The requirement of advanced analytical methods in Big Data applications in the organizations is inevitable.
- There is not much awareness on the usage of analytical tools. So, there is a need for creating awareness of Big Data.

**References**

1. Akoka, J., Comyn-Wattiau, I., & Laoufi, N. (2017). Research on Big Data–A systematic mapping study. *Computer Standards & Interfaces*, *54*, 105-115.
2. Arunkarthikeyan, K. and Balamurugan, K., 2021. Experimental Studies on Deep Cryo Treated Plus Tempered Tungsten Carbide Inserts in Turning Operation. In Advances in Industrial Automation and Smart Manufacturing (pp. 313-323). Springer, Singapore.
3. Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U.S. and Warrier, K.G.K., 2019. Predicting correlations in abrasive waterjet cutting parameters of Lanthanum phosphate/Yttria composite by response surface methodology. Measurement, 131, pp.309-318.
4. Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U.S. and Warrier, K.G.K., 2017. Mathematical modelling on multiple variables in machining LaPO4/Y2O3 composite by abrasive waterjet. International Journal of Machining and Machinability of Materials, 19(5), pp.426-439.
5. Balamurugan, K., Uthayakumar, M., Sankar, S., Hareesh, U.S. and Warrier, K.G.K., 2018. Preparation, characterisation and machining of LaPO4-Y2O3 composite by abrasive water jet machine. International Journal of Computer Aided Engineering and Technology, 10(6), pp.684-697.
6. Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. *ACM Sigmod Record*, *40*(4), 56-60.
7. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.
8. Cuzzocrea, A., Song, I. Y., & Davis, K. C. (2011, October). Analytics over large-scale multidimensional data: the big data revolution! In *Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP* (pp. 101-104).
9. Davenport, T., & Dyché, J. Big Data in big companies, International institute for analytics, 2013.
10. De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.
11. Goes, P. B. (2014). Editor's comments: big data and IS research.
12. Izhar, T. A. T., Torabi, T., Bhatti, M. I., & Liu, F. (2013). Recent developments in the organization goals conformance using ontology. *Expert Systems with Applications*, *40*(10), 4252-4267.
13. Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM*, *52*(8), 36-44.
14. Latchoumi, T.P., Loganathan, J., Parthiban, L. and Janakiraman, S., 2016, August. OFS method for selecting active features using clustering techniques. In *Proceedings of the International Conference on Informatics and Analytics* (pp. 1-4). https://doi.org/10.1145/2980258.2982108
15. Latchoumi, T.P., Reddy, M.S. and Balamurugan, K., 2020. Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention. European Journal of Molecular & Clinical Medicine, 7(02), p.2020.
16. Latchoumi, T.P. and Parthiban, L., 2016. Secure Data Storage in Cloud Environment using MAS. Indian Journal of Science and Technology, 9, pp.24-29.
17. Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, *16*(3), 4-6.
18. Ranjeeth, S., Latchoumi, T.P. and Paul, P.V., 2020. Role of gender on academic performance based on different parameters: Data from secondary school education. Data in brief, 29, p.105257.
19. Rodríguez-Mazahua, L., Rodríguez-Enríquez, C. A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández, G. (2016). A general perspective of Big Data: applications, tools, challenges and trends. *The Journal of Supercomputing*, *72*(8), 3073-3113.