IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

The Role of Data Analysis & Research Methods for Research Project

Dr. Varun Tiwari[1], Dr. Santosh Kumar Singh[2], Dr. Ruchi Sawhney[3], Dr. Vikas Rao Vadi[4]

Associate Professor, Don Bosco Institute of Technology, Okhla Road, New Delhi, India[1]

Assistant Professor, Bosco Technical Training Society, Okhla Road, New Delhi, India[2]

Assistant Professor, Bosco Technical Training Society, Okhla Road, New Delhi, India[3]

Professor & Director, Don Bosco Institute of Technology, Okhla Road, New Delhi, India[4]

ABSTRACT: This paper is helpful to find better data analysis & research types for research projects. Here we have discussed data analysis & research types and provide one case study with the help of descriptive analysis which is useful for research projects. These different types of methods are helpful for descriptive, explanatory and inferential analyses. In this research paper we have a main focus on descriptive data analysis. We have done one case study related to descriptive data analysis.

KEYWORDS: Big Data Analysis, Preparation of data, Analysis Methods, Analysis Types, Covid-19 dataset, Research types, Data Coding

I. INTRODUCTION

Data analysis is the process of converting your data to expressive information. Preparation of data is a big task. In this part we have a process of data to convert to the numerical format. Which is readable by the machine. For processing we have used some tools like SPSS, R and Python etc. The specifics of the data preparation process vary by industry, organization, and need, but the workflow remains largely the same. First most important part of data preparation is gathering data and discovering, assessing data, Cleanse and validate data, Transform and enrich data and Store data.
There are many steps for preparation of data like cleaning of data, data coding, collecting data, finding missing value in the data and data transformation etc. Here we have described briefly about these steps in given below:
In data coding we process and transform your raw data into a more structured and manageable format. Data coding is used to assigning labels or numerical codes to different pieces of information. It is based on predefined criteria or categories. This code is work as a bridge between the raw data and the analytical phase of research. It is facilitating the organization and interpretation of data.

Data coding is playing an important role for preparing your research project. There are lots of reasons to use data coding for research project preparation.

1. It is helpful for data reduction. When the researcher collects a huge amount of data after that data coding is help to condense and summarize your data very easily. This reduction makes it possible to analyze large datasets effectively.
2. With the help of data coding we have organize, managed your data easily.
3. It is helpful to allow researchers to identify patterns, relationships and trends with raw data.
4. Researchers can run statistical tests on coded data to draw meaningful conclusions.
5. With the help of data coding researchers can compare and contrast information across different cases or groups, aiding in the generation of insights and theories.

II. TYPES OF RESEARCH

There are different types of examples for data coding in research

1. Qualitative Research: In these types of research, we have categorized and analyzed textual or narrative data. for example, imagine a study on customer feedback about a new product. Here Researchers can be defining some categories for comments of customer like “product quality”, “customer service”, “pricing” and “delivery”.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

2. Survey Research: In this type of research, we have assigned numerical values to responses on a Likert scale. Here the researcher assigns a numerical value according to the responses. Suppose the response is "strongly agree" then code will be 5, "agree" then 4, "neutral" then 3, "disagree" then 2, and "strongly disagree" then 1. These codes enable quantitative analysis of survey data.

3. Content Analysis: In this type of analysis, we have used textual or visual content, such as news articles or social media posts, into predefined categories. For instance, in a content analysis of news articles about climate change, researchers could code articles as "supportive of climate action," "neutral," or "skeptical of climate change." This coding allows researchers to assess the prevalence of different perspectives in the media.

4. Medical Research: In this type of research, we have categorized patient data into different diagnostic groups based on symptoms, test results, or medical histories.

5. Historical Research: In this type of research, data coding can be useful. This research is based on themes, time periods, or key events. This historical research to identify patterns and trends across historical records and gain new insights into the past.

Data Entry: In the data entry part, we have entered raw data into the computer. This type of data is available in different formats like text, number etc. If you collected your data manually or automated the accuracy is the most important part to ensure that data is correct and reliable.

Missing Data: At the time of data cleaning the most important problems encountered is missing data. There are lots of reasons for this error like measurement, non-response, or data loss. There are different types of ways to handle missing data at the time of cleaning.
1. Deletion: The first way to handle missing data is to delete the records. But in this method the result is a loss of information and decrease in size.

2. Imputation: In this part we have replaced missing values to estimated values. There are many methods available for imputation like median, mean and multiple imputation. For example, we have use python code for replace missing values with the mean of the remaining values in a panda Dataframe:

Example:
```
import pandas as pd
import numpy as np
data = pd.read_csv('one.csv')
data.fillna(one.mean(),inplace=True)
```

3. Interpolation: It is a statistical method that involves using the values of other variables to estimate the missing values. For example, we have used panda dataframe for replace missing values using Linear Interpolation:

Example:
```
import pandas as pd
import numpy as np
data = pd.read_csv('one.csv')
data.interpolate(method='linear', inplace=True)
```

4. Regression: It is also a statistical method. It involves using the values of other variables to predict the missing values. It can be used to relate between the variables in the data. For example, we have used panda data frame in python for replace missing values using Linear Regression:

Example:
```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
data = pd.read_csv('one.csv')
lr = LinearRegression()
lr.fit(data.dropna().drop('missing_column', axis=1), data.dropna()['missing_column'])
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

```
pred = lr.predict(data[data['missing_column'].isnull()].drop('missing_column', axis=1))
data.loc[data['missing_column'].isnull(), 'missing_column'] = pred
```

Data transformation: It is the process of data converting, data cleaning and structuring of data into usable format. Transformation has contained many processes like data integration, migration of data and warehousing. Data Transformation process is given below:

1. Constructive
2. Destructive
3. Aesthetic
4. Structural

III. TYPES OF DATA ANALYSIS WITH EXAMPLE

There are different types of Data Analysis methods available. There are eight types of methods available.

- **Descriptive analysis:** This method is also known as the least amount of effort. This method is used for large amounts of data.

A. First Analysis
B. Generates simple summaries for samples & measurements
C. Involves common, descriptive statistics for measurement of central tendency, variability, frequency and position.

Example
Best example of this analysis is COVID-19 dataset for finding summary of total cases/death according to the country which is infected by the virus using a line graph.

- **Diagnostic analysis:**

A. Findings and Investigating using certain patterns in data.
B. Past data, to reveal more insights into current data trends.
C. Used exploring patterns in data to explain anomalies.

Example
If you want to find the website traffic levels over the previous 6 months on Footwear store now you have compiling & assessing data by using Diagnostic Analysis.

- **Exploratory analysis:** It is used to describe unknown relationships and discover new connections, define future studies or questions.

A. It is used to discover relationships between measures in your data, which are not evidence for the existence of the correlation, as denoted by the phrase, "Correlation doesn't imply causation."
B. It's useful for discovering new connections and forming hypotheses. It drives design planning and data collection.

Example:
The best example of an exploratory data analysis on climate change involves taking the rise in temperature over the years from 1980 to 2024 and the increase of human activities and industrialization to find relationships from the data. For example, you may increase the number of factories, cars on the road and airplane flights to see how that correlates with the rise in temperature.

- **Inferential analysis:** In this method we have used a small sample to conclude a bigger population. It is used for observational, retrospective data sets and cross-sectional time study.

A. Involves using estimated data that is representative of a population and gives a measure of uncertainty or standard deviation to your estimation.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

B. It is based on your sampling scheme. If the sample isn't representative of the population, the generalization will be inaccurate.

Example:
A psychological study on the benefits of sleep might have a total of 500 people involved. When they followed up with the candidates, the candidates reported to have better overall attention spans and well-being with seven-to-nine hours of sleep, while those with less sleep and more sleep than the given range suffered from reduced attention spans and energy. This study drawn from 500 people was just a tiny portion of the 7 billion people in the world, and is thus an inference of the larger population.

- **Predictive analysis:** It is utilizing historical & current facts to reach future predictions.

A. Accuracy of prediction is completely dependent on input variables.
B. Models are used for its working. Linear Model is used for Predictive Analysis.
C. Using a variable to predict another one doesn't denote a causal relationship.

The 2024 India Election is a popular topic and many prediction models are built to predict the winning candidate. 303 did this to forecast the 2019 elections. Prediction analysis for an election would require input variables such as historical polling data, trends and current polling data in order to return a good prediction. Predictive analysis takes data from the past and present to make predictions about the future.

- **Explanatory or Causal analysis:** It uses randomized trial data sets.

A. It is used to discover the hidden mechanisms underlying the correlations.
B. Applied in Randomized studies focused on identifying causation.
C. Causal analysis is the gold standard in data analysis and scientific studies where the cause of a phenomenon is to be extracted and singled out, like separating wheat from chaff.
D. Analyzed in aggregate (multiple groups), and the observed relationships are just average effects (mean) of the whole population.

Say you want to test out whether a new drug improves human strength and focus. To do that, you perform randomized control trials for the drug to test its effect. You compare the sample of candidates for your new drug against the candidates receiving a mock control drug through a few tests focused on strength and overall focus and attention. This will allow you to observe how the drug affects the outcome.

- **Mechanistic analysis:**
  A. Applied in physical or engineering sciences, situations that require high precision and little room for error, only noise in data is measurement error.
  B. Designed to understand a biological or behavioral process, the pathophysiology of a disease or the mechanism of action of an intervention.

Example:
Many graduate-level research and complex topics are suitable examples, but to put it in simple terms, let's say an experiment is done to simulate safe and effective nuclear fusion to power the world. A mechanistic analysis of the study would entail a precise balance of controlling and manipulating variables with highly accurate measures of both variables and the desired outcomes.

- **Prescriptive analysis:**

A. Prescriptive analysis may come right after predictive analysis, but it may involve combining many different data analyses.
B. Companies need advanced technology and plenty of resources to conduct prescriptive analysis. AI systems that process data and adjust automated tasks are an example of the technology required to perform prescriptive analysis.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

Example

This analysis is pervasive in everyday life, driving the curated content users consume on social media. On platforms like TikTok and Instagram, algorithms can apply prescriptive analysis to review past content a user has engaged with and the kinds of behaviors they exhibited with specific posts.

IV. OBJECTIVE OF THE STUDY

1. Descriptive Analysis of Covid-19 dataset according to parameter.
2. To visualize a dataset using different statistical tools.
3. To study different types of Data Analysis methods.
4. To study different types of Research.

The paper is organized in sections, First Section Introduction, Second Section types of research, Third Section types of data analysis, related research work, proposed work, in proposed work we have done one case study with descriptive analysis, for example Covid-19 dataset, and the last section is conclusion.

V. RELATED WORK

There are lots of research papers available on descriptive analysis for covid-19. Some studies have been published on types of research, data analysis types. We have shown related work according to different published research paper which are as follows:

- Joel Hellewell et al. **[6],** conducted a feasibility study of controlling COVID-19 by isolation and quarantine. They developed a transmission model to parametrize the outbreak of COVID-19. This model is found that in some plausible scenarios, case isolation alone would be unlikely DESCRIPTIVE ANALYSIS OF COVID-19 PATIENTS 5 to control transmission within 3 months. Case isolation was more effective when there was little transmission before symptom onset and when the delay from symptom onset to isolation was short.

- Ensheng Dong et al**. [7],** developed a dashboard to visualize and to track reported cases of China, USA, Canada and Australia. To collect the data they monitor, online news Twitter feeds and direct communication through the dashboard. The proposed dashboard is updated every 15 minutes from DXY of all countries. The proposed dashboard is hosted by Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, Baltimore, MD, USA.

- Rajesh Singh and R. Adhikari **[8]** studied the impact of age structured on social distance. They compared the social contact structure and age of Chinese, Indian and Italian populations. They developed a mathematical model to assess the effect of social distancing measures. They conclude that three-week lockdown is not enough at all, but sustained periods of lockdown will reduce the cases of COVID-19 in India.

- Surabhi Shinde,Sayali Shetye,Taruna Rawat, 4 Prof. Prashant NitnawareIn **[18]** In this paper the author will review the relevant techniques in the literature. Describes the different techniques used in the work. The Literature Review is an objective and critical summary of published research literature related to a research topic. Its purpose is to become familiar with current thinking and research on a particular topic and may guarantee future research in previously overlooked or unexplored territory.

VI. PROPOSED WORK

We have done one case study related to Descriptive Analysis. This descriptive analysis is also known as the least amount of effort. This method is used for large amounts of data.

A. First Analysis
B. Generates simple summaries for samples & measurements
C. Involves common, descriptive statistics for measurement of central tendency, variability, frequency and position.

Descriptive Analysis Example with Case Study:

Best example of this analysis is COVID-19 dataset for finding summary of total cases/death according to the country which is infected by the virus using line graphs. Firstly, find the data of covid-19 in csv file format on

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

https://api.covid19india.org/csv/latest/state_wise.csv site and load csv data on python then analyzed these data and find the result:

1. Preparation or Collection and load of Dataset: Now we collect data from API and upload covid-19 latest data on python through cloud-based technology Google Colab and now import a dataset from file:

```
stat='https://api.covid19india.org/csv/latest/state_wise.csv'
rs=pd.read_csv(stat)
```

```
rs.head()
```

	State	Confirmed	Recovered	Deaths	Active	Last_Updated_Time	Migrated_Other	State_code	Delta_Confirmed	Delta_Recovered	Delta_Deaths	State_Notes
0	Total	22992039	19021313	250027	3710900	11/5/2021 08:06:21	9799	TT	0	0	0	NaN
1	Maharashtra	5138973	4469425	76398	590818	10/05/2021 20:33:18	2332	MH	0	0	0	[Dec 16]:10,218 duplicate cases & other state ...
2	Kerala	1930116	1504160	5880	419725	10/05/2021 19:30:13	351	KL	0	0	0	Mahe native who expired in Kannur included in ...
3	Karnataka	1973683	1383285	19372	571006	10/05/2021 19:34:28	20	KA	0	0	0	NaN
4	Andhra Pradesh	1302589	1104431	8791	189367	10/05/2021 19:30:15	0	AP	0	0	0	NaN

Fig 1.1: Head of State_wise.csv file

```
[ ] rs.shape

    (38, 12)

[ ] rs.columns

    Index(['State', 'Confirmed', 'Recovered', 'Deaths', 'Active',
           'Last_Updated_Time', 'Migrated_Other', 'State_code', 'Delta_Confirmed',
           'Delta_Recovered', 'Delta_Deaths', 'State_Notes'],
          dtype='object')
```

Fig 1.2: Show shape and columns of dataset

1. Data filtration and processing: Suppose if you want to show the all-state data then you use some command for showing this dataset.

```
rs.iloc[::-1]['State_code']
```

Suppose you want to find the latest current cases in Delhi. One big question is how to perform data processing.

```
bd=rs.loc[rs['State'] =='Delhi']
```

	State	Confirmed	Recovered	Deaths	Active	Last_Updated_Time	Migrated_Other	State_code	Delta_Confirmed	Delta_Recovered	Delta_Deaths	State_Notes
6	Delhi	1336218	1231297	19663	85258	10/05/2021 19:30:17	0	DL	0	0	0	[July 14]: Value for the total tests conducted...

Fig 1.3: Data Filtration and process

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

1. Visualization of Data:

A. Analysis Phase 1: Find the Confirmed Cases according to 6 State:

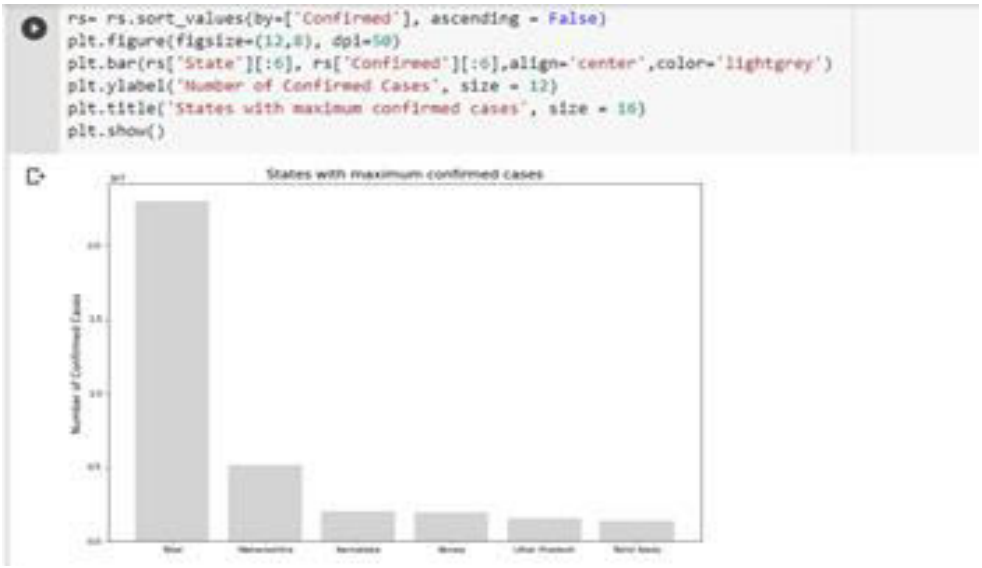

Fig 1.4: Plotting of Confirmed Cases of Six State

B. Analysis Phase 2: Find the no. of death of 5 state:

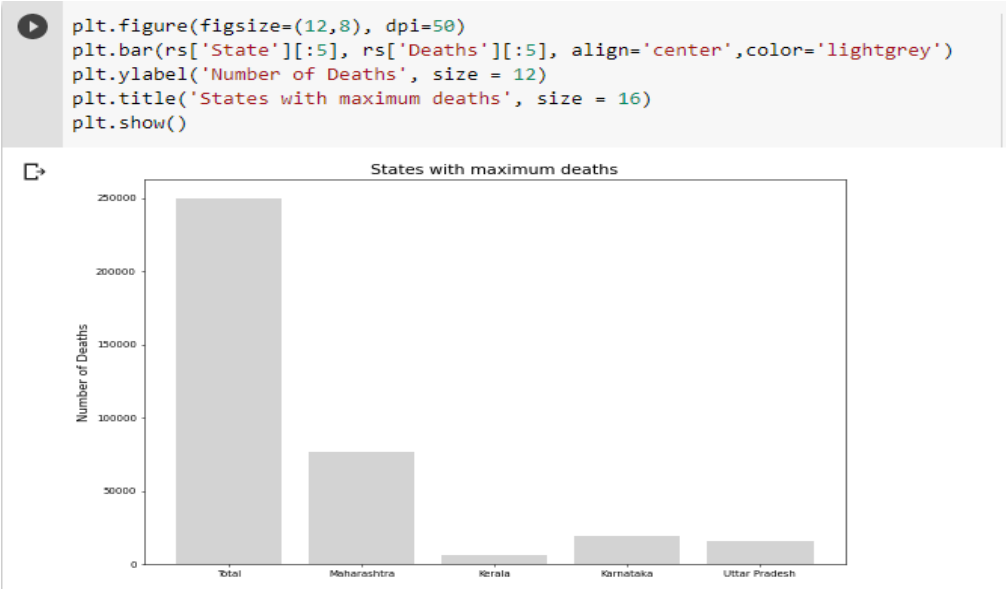

Fig 1.5: Plotting of Death Cases of Five State

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

C. Analysis phase 3: Find the Recovered Cases according to state:

```
rs= rs.sort_values(by=['Recovered'], ascending = False)
plt.figure(figsize=(12,8), dpi=50)
plt.bar(rs['State'][:6], rs['Recovered'][:6],align='center',color='lightgrey')
plt.ylabel('Number of Recovered Cases', size = 12)
plt.title('States with Recovered cases', size = 16)
plt.show()
```

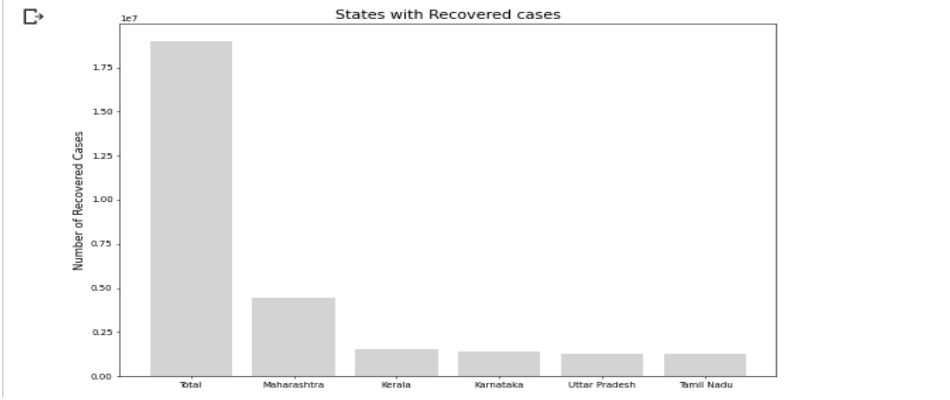

Fig 1.6: Plotting of Recovered Cases of Six State

VII. CONCLUSION

In this paper we have provided a description of the most common data analysis techniques. In this paper we have described data analysis types and research types which are an essential process in analyzing data. Then, we have defined descriptive analysis with examples and case study, and we have discussed the data analysis types & research types which are used for data analysis. We have provided a case study of COVID-19 dataset which is based on descriptive analysis specifically in the final section.

REFERENCES

[1] Barkur G. Vibhab, Kamath G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry, 51*(June), 102089.
[2] Cohut M. (2020, April 24). COVID-19 global impact: How the coronavirus is affecting the world. *Medical News Today*.
[3] World Health Organization, https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf, accessed: 2020-04-06.
[4] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, C. Zheng, Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study, The Lancet Infectious Diseases (2020).
[5] Bhatnagar, V., Poonia, R. C., Nagar, P., Kumar, S., Singh, V., Raja, L., & Dass, P. (2021). Descriptive analysis of COVID-19 patients in the context of India. Journal of Interdisciplinary Mathematics, 24(3), 489–504. https://doi.org/10.1080/09720502.2020.1761635.
[6] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al., Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts, The Lancet Global Health (2020)
[7] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track covid-19 in real time, The Lancet infectious diseases (2020).
[8] R. Singh, R. Adhikari, Age-structured impact of social distancing on the covid-19 epidemic in india, arXiv preprint arXiv:2003.12055 (2020).
[9] Varun Tiwari, Atul D. Newase "A Study on Big Data Analytical Tools" PNo. 2045 - 2060, vol 21,No 1 (2022) , Telematique (WOS group).
[10] Tiwari Varun, Newase Atul D. "BCA STUDENTS RESULT ANALYSIS USING R", NeuroQuantology,Vol. 20, Iss. 8,(2022): 9652 - 9661. DOI:10.14704/nq.2022.20.8.NQ22983.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204328 |

[11] KL Bharti, TA Abbasi, V Tiwari "A Brief Survey on Data Mining using Latest Technology (Big Data)" IJRASET (July 2018) (ISSN: 2321-9653)

[12] Vadi, V.R., Tiwari, V., Abidin, S. (2021). Big Data Analysis Using R and Hadoop. In: Tavares, J.M.R.S., Chakrabarti, S., Bhattacharya, A., Ghatak, S. (eds) Emerging Technologies in Data Mining and Information Security. Lecture Notes in Networks and Systems, vol 164. Springer, Singapore. https://doi.org/10.1007/978-981-15-9774-9_76

[13] V Tiwari, G TIPS "Data Integrity and Authentication in WSNs" HMR Interdisciplinary Journal of Science, Technology 2018

[14] V Tiwari, M Sharma, VR Vadi "Big Data Analysis using R integration with Hadoop" Journal of Systems Engineering and Electronics (ISSN NO: 1671-1793) Volume 33 ISSUE 9 2023

[15] Varun Tiwari and Dr. Atul Dattatray Newase "Trends And Services For Big Data Analysis On Cloud Data Centers" Webology (ISSN: 1735-188X) Volume 18, Number 5, 2021

[16] Varun Tiwari, Vikas Rao Vadi, Mukta Sharma "WOT Enabled Smart Cities Integrated Big Data Analytics" International Journal of Computer Applications Foundation of Computer Science (FCS), NY, USA Volume 177 - Number 112019 DOI: 10.5120/ijca2019919540

[17] K.Bharti,T.A.Abbasi,V Tiwari, A Brief Survey on Data Mining using Latest Technology (Big Data) , International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue VII, July 2018.

[18] Surabhi Shinde,Sayali Shetye,Taruna Rawat, 4 Prof. Prashant Nitnaware "DATA ANALYSIS AND PREDICTION OF COVID-19 CASES IN INDIA" International Journal of Creative Research Thoughts (IJCRT).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details