DATA-DRIVEN APPROACHES TO TACKLING PANDEMIC VIRUSES: A COMPREHENSIVE REVIEW OF ADVANCED ALGORITHMS

S. Akila[1], S. Prasanna[2]

[1]Research Scholar, VISTAS (Vels Institute of Science, Technology & Advanced Studies),
Email: Subramanian.akila@gmail.com
[2]Professor & Head Department of MCA,VISTAS (Vels Institute of Science, Technology & Advanced Studies),
Email: Prasanna.scs@velsuniv.ac.in

Article History: **Received:** 18.04.2023 **Revised:** 07.05.2023 **Accepted:** 16.06.2023

Abstract: *This survey paper presents a comprehensive overview of two crucial fields - big data analytics algorithms and advanced pandemic viruses. It explores various big data analytics algorithms, including linear regression, logistic regression, K-nearest neighbors, clustering, and classification, elucidating their applications and relevance in different industries. Moreover, it delves into significant pandemic viruses like HIV/AIDS, COVID-19, and Malaria, highlighting their origins, impact on global health, and measures to control their spread. Additionally, the paper examines the intersection between big data analytics and pandemic virus research, emphasizing how data-driven insights can enhance disease surveillance and response strategies. It also identifies future research directions and addresses challenges associated with big data analytics and pandemic virus research. Ultimately, the survey paper advocates for responsible data practices and collaboration to strengthen global health systems and combat emerging infectious diseases effectively.*

Keywords: *Disease, Review, Health, Big Data, Algorithms, Covid 19*

DOI: 10.48047/ecb/2023.12.8.509

INTRODUCTION

Big data and advanced pandemic viruses are two distinct yet interconnected fields that have significant implications for various aspects of society. Big data refers to the massive volumes of data that are generated at high speed, with incredible variety and complexity. The growth of data in today's digital age has been exponential, and it continues to shape how organizations operate and make decisions across various industries. The ability to harness and analyze big data has become a critical factor in gaining a competitive advantage and driving innovation.

The challenges associated with big data lie not only in the sheer volume of information but also in its diversity and velocity. Traditional data management systems and analytical tools are often ill-equipped to handle such vast and complex datasets effectively. Therefore, there is a pressing need to develop and implement advanced algorithms and technologies that can process, store, and analyze big data efficiently.

The application of big data analytics has transformed numerous sectors, including healthcare, finance, marketing, transportation, and more. For instance, in healthcare, big data analytics can be used to analyze patient records, medical images, and genomic data to improve diagnostic accuracy and personalize treatment plans. In finance, it enables companies to detect fraudulent activities and assess market trends to make informed investment decisions. In marketing, big data analytics can provide valuable insights into customer behavior and preferences, enabling targeted advertising and personalized recommendations.

At the same time, advanced pandemic viruses pose a persistent threat to global public health. Throughout history, humanity has faced numerous pandemics, some of which have caused widespread devastation and loss of life. Pandemic viruses, such as the HIV/AIDS, Spanish flu, COVID-19, and Malaria, have significantly impacted societies, economies, and healthcare systems. Dealing with these infectious diseases requires a coordinated global response and innovative approaches to surveillance, prevention, and treatment.

In 1898, the Dutch microbiologist Martinus Beijerinck rehashed the trials and became persuaded that the sifted arrangement contained another type of the irresistible specialist. He saw that the specialist duplicated distinctly in cells that were partitioning, yet as his examinations didn't show that it was made of particles, he considered it a contagium vivum fluidum (dissolvable living germ) and once again introduced the word infection. Beijerinck kept up that infections were fluid, a hypothesis later ruined by Wendell Stanley, who demonstrated they were particulate. In the exact year, Friedrich Loeffler and Paul Frosch passed the main creature infection, aphthovirus (the specialist of foot-and-mouth sickness), through a comparative channel.
The emergence and rapid spread of novel viruses, such as the COVID-19 virus, underscore the need for real-time data analysis and response. Big data analytics plays a crucial role in tracking the spread of pandemics, modeling infection rates, predicting hotspots, and evaluating the effectiveness of interventions. By leveraging big data, health authorities can make data-driven decisions, optimize resource allocation, and develop targeted strategies to control the spread of infectious diseases.
In recent years, the synergy between big data analytics and pandemic virus research has become increasingly evident. The analysis of vast amounts of data can provide valuable insights into the epidemiology of infectious diseases, identify risk factors, and guide public health interventions. For example, data from social media, mobile phone usage, and travel patterns can be analyzed to understand how infectious diseases spread within and between communities.
Furthermore, big data analytics can aid in the development and distribution of vaccines by predicting demand, optimizing supply chains, and identifying at-risk populations. Machine learning algorithms can also be employed to accelerate drug discovery processes and identify potential antiviral treatments.
However, harnessing the full potential of big data in pandemic virus research is not without challenges. Privacy and ethical concerns surrounding the collection and use of sensitive health data must be carefully addressed. Additionally, data quality and interoperability issues can affect the accuracy and reliability of analytical results.
In conclusion, the fields of big data analytics and advanced pandemic viruses are intertwined, representing a complex and multifaceted landscape of challenges and opportunities. The efficient analysis of big data can revolutionize how we respond to pandemics, improving disease surveillance, prevention, and treatment strategies. As technology continues to advance, and the volume of data grows exponentially, the need for sophisticated algorithms and robust data infrastructure becomes increasingly vital. The integration of big data analytics and pandemic virus research holds the potential to transform global health outcomes and strengthen our ability to tackle infectious diseases effectively. By leveraging the power of data-driven insights, we can build a more resilient and prepared world to address future challenges.
The main contribution on this review paper is mentioned below,

- Comprehensive Overview: The survey paper provides a comprehensive overview of two crucial fields - big data analytics algorithms and advanced pandemic viruses. It presents a well-researched and structured examination of the key concepts, challenges, and advancements in both domains, offering valuable insights to readers.
- In-Depth Analysis of Big Data Algorithms: The paper thoroughly explores various big data analytics algorithms, including linear regression, logistic regression, K-nearest neighbors, clustering, and classification. It not only explains the workings of each algorithm but also highlights their applications and relevance in different industries, demonstrating their significance in extracting meaningful insights from large datasets.
- Discussion of Pandemic Viruses: The survey paper delves into the world of pandemic viruses, shedding light on significant outbreaks like HIV/AIDS, COVID-19, and Malaria. It presents a concise yet informative analysis of these viruses, their origins, impact on global health, and measures taken to control their spread. This contribution provides readers with a comprehensive understanding of past and current pandemic challenges.
- Intersection of Big Data Analytics and Pandemic Virus Research: One of the notable contributions of the survey paper is its exploration of the intersection between big data analytics and pandemic virus research. By highlighting how big data analytics can enhance disease surveillance, prediction, and

response strategies, the paper showcases the potential of leveraging data-driven insights to combat infectious diseases effectively.

- Future Research Directions: The survey paper identifies future research directions for both big data analytics algorithms and pandemic viruses. It emphasizes the need for continued advancements in data analysis techniques and technologies to handle even larger and more diverse datasets. Additionally, it advocates for further research in understanding and managing emerging infectious diseases.
- Addressing Challenges: The paper addresses challenges associated with big data analytics and pandemic virus research. It emphasizes the importance of data privacy, quality, and ethical considerations in data-driven research. By acknowledging these challenges, the paper encourages researchers to adopt responsible practices and seek innovative solutions.
- Implications for Public Health: Lastly, the survey paper underscores the implications of its findings for public health and safety. By emphasizing the role of big data analytics in pandemic response and preparedness, it advocates for greater collaboration between data scientists, healthcare professionals, and policymakers to strengthen global health systems.

In summary, the survey paper makes significant contributions by providing a comprehensive and well-structured examination of big data analytics algorithms and pandemic viruses. It offers valuable insights into the applications and challenges of big data analytics and highlights the importance of data-driven strategies in managing infectious diseases. Its exploration of future research directions and implications for public health further enriches the understanding of these critical fields.

LITERATURE REVIEW

Big Data progressively benefits both examination and modern zones, for example, medical care, financial service and business proposal. The Economist says, Data are turning into another crude material of business. Monetary data is for all intents and purposes indistinguishable from capital and work. These days, the information to be investigated are dynamic and enormous in volume, additionally they are the gathering of various information types. These information come from various information sources, for example, WhatsApp, Twitter, Facebook, YouTube, Mobile telephones GPS signs and then some. Subsequently, the Big Data has the special highlights like heterogeneous, unstructured, semi organized, deficiency, high dimensional. Enormous Data has complex qualities that need amazing innovations and progressed procedures. In this way, the conventional static Business Intelligence devices can presently don't be ingenious on account of Big Data applications.A deliberate writing survey as the way toward recognizing, evaluating and deciphering all exploration results to give answers to investigate question comprises of a few exercises, specifically: determining the examination questions, choosing contemplates, removing required information, combining information and describing the outcome.

This article reviews the state of the art for the use of Big Data. At first, an outline of Big Data and its highlights are introduced. At that point, the primary parts of Big Data cycles and innovations are examined. A while later, important uses of Big Data examination are talked about. Then, Big Data Analytics are talked about when all is said in done terms and particularly for the medical care area. The article closes with a survey of the difficulties that were recognized in this investigation, trailed by the ends

Harshawardhan S. Bhosale, Devender P. Gadekar clarified briefly Big Data and 3V's.This paper clarified different issues looking with Big Data preparing like heterogeneity and deficiency of information, Scale, season of investigation and security and Privacy of information and so on They clarified Hadoop as an answer of Big Data preparing. This paper clarified Hadoop design just as HDFS and Map lessen engineering in a nutshell. Toward the finish of the paper they clarified different parts based on simultaneousness, solidness, replication techniques, data set model and simultaneousness and so on.

Cheikh Kacfah Emani, Nadine Cullot, Christopher, Nicolle et.al reviewd the possibility of Big Data. They disussed about the highlights of Big Data and furthermore clarified the means of Big Data preparing. During the administration of Big Data numerous issues can be experienced during semantic social occasion. They additionally disclosed how to handle Volume, Velocity and Variety.

S. Vikram Phannendra,E. Madhusudhana Reddy clarified that RDBMS didn't furnish total arrangement

while managing Big Data. In this paper they portrayed Big Data is not quite the same as customary information regarding five measurements this paper likewise momentarily clarified Hadoop design. Hadoop comprises of fundamentally Name Node, Data Node and Edge Node. Hadoop design can deal with enormous dataset, versatile calculation, log the executives, Extract – Transform-Load (ETL) stage. They additionally centered around different difficulties of Big Data: Data Privacy, examination and representation and so forth.

Vibha Shukla, Pawan Kumar Dubey et.al examined that information is expanding quick. With the expanded measure of information, Traditional information examination devices need to create. In this paper they clarified customary information examination versus Big Data investigation. They likewise examined different Big Data Emerging methods and innovation like NOSQL information base, Map Reduce, Hadoop, HDFS and some more. They additionally clarified different difficulties of Big Data and future exploration will focus to create answers for manage these difficulties [11].

Prity Vijay and Brigh Keshwani clarified the idea of Big Data and momentarily examined issue and difficulties looked during the handling of Big Data. Customary preparing and the board devices and systems are not reasonable to deal with Big Data since it requires the structure that can deal with unstructured information and give constant investigation adaptation to internal failure limit. This paper additionally clarified different issues when we manage Big Data with customary methodology like RDBMS. Hadoop and its connected advancements are appropriate for dealing with Big Data. This paper additionally clarified the different changes that the Hadoop requested with the time.

Big Data Analytics

Big data is a term for large data sets have a larger, more varied and complex structure with difficulty storing, analyzing and visualizing for further processing or result. Big data required technologies to capture, storage, management, and analysis the huge amount of data with more complexity and it can be characterized by its size and variety. It is a concept that describes a large volume of data along with unstructured and semi-structured data - structured data. Data are needed to isolate the hidden patterns and to find answers without over-fitting the data. The expression "Enormous Data" has as of late been applied to datasets that develop so huge that they become abnormal to work with utilizing conventional data set administration frameworks. They are informational indexes whose size is past the capacity of regularly utilized programming instruments and capacity frameworks to catch, store, oversee, just as interaction the information inside an average passed time. Enormous data sizes are increasingly expanding, currently going from a few hundred terabytes (TB) to several petabytes (PB) of data in a single set of information.

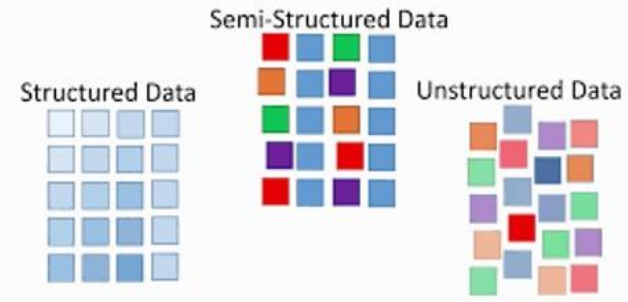

Fig 1: Types of Big Data

Thusly, some of the troubles identified with enormous information incorporate catch, stockpiling, search, sharing, investigation, furthermore, envisioning. Today, ventures are investigating enormous volumes of profoundly definite information in order to find realities they didn't know before. Thus, huge information investigation is the place where exceptional

insightful methods are applied on large informational indexes. Examination dependent on huge information tests uncovers and use business change. Nonetheless, the bigger the arrangement of information, the more

troublesome it becomes to oversee. In this part, we will begin by examining the attributes of enormous information, just as its significance.

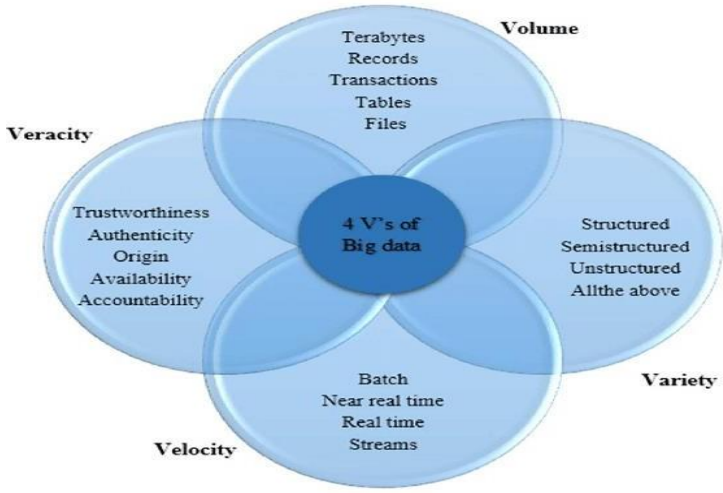

Fig 2: 4 V'S of Big Data

- **Volume:** One of the traits of big data is scale. We already know that Big Data reveals massive 'volumes' of data that are generated daily from different sources, such as social media sites. Vast quantities of data are stored in data centers. Therefore, the beginning of the Big Data function
- **Variety:** Big Data Variety applies to structured, unstructured, and semi-structured data obtained from different sources.
- **Velocity:** Velocity refers essentially to the speed at which information is produced in real-time. It involves the rate of transition, the joining of incoming data sets at different speeds and activity bursts in a broader prospect.
- **Veracity:** In general, data veracity is how reliable or truthful a data set might be. It takes on a little more importance in the sense of big data, however. Specifically, when it comes to the reliability of big data.

Algorithms

Linear Regression

Linear regression is one of the most fundamental calculations of cutting-edge investigation. This additionally makes it quite possibly the most generally utilized. Individuals can without much of a stretch envision how it is functioning and how the info information is identified with the yield information. Straight relapse utilizes the connection between two arrangements of constant quantitative measures. The previously set is known as the indicator or free factor. The other is the reaction or ward variable. The objective of straight relapse is to recognize the relationship as an equation that portrays the reliant variable regarding the free factor. When this relationship is evaluated, the needy variable can be anticipated for any case of a free factor. Perhaps the most well-known autonomous factors utilized is time. Regardless of whether your autonomous variable is income, costs, clients, use, or profitability, on the off chance that you can characterize the relationship it has with time, you can conjecture an incentive with linear regression.

Logistic Regression

Logistic regression sounds like straight relapse however is really focused around issues including order rather than quantitative forcasting. Here the yield variable qualities are discrete and limited as opposed to persistent and with boundless qualities similarly as with direct regression. The objective of calculated relapse is to order whether a case of an information variable either fits inside a class or not. The yield of

calculated relapse is an incentive somewhere in the range of 0 and 1. Results more like 1 show that the information variable all the more unmistakably fits inside the class. Results more like 0 show that the information variable probably doesn't fit inside the category. Logistic relapse is regularly used to answer plainly characterized yes or no inquiries. Will a client purchase once more? Is a purchaser credit commendable? Will the possibility become a client? Foreseeing the response to these inquiries can produce a progression of activities inside the business cycle which can help drive future income.

Classification and Regression Trees

Classification and regression trees utilize a choice to classify information. Every choice depends on an inquiry identified with one of the input variables. With each question and comparing reaction, the case of information draws drew closer to being classified with a certain goal in mind. This arrangement of inquiries and reactions and resulting divisions of information make a tree-like design. Toward the finish of each line of inquiries is a class. This is known as the leaf node of the classification tree.

These classification trees can turn out to be very huge and complex. One strategy for controlling the intricacy is through pruning the tree or deliberately eliminating levels of addressing to adjust between precise fit and reflection. A model that functions admirably with all occurrences of information esteems, both those that are known in preparing and those that are not, is principal. Forestalling overfitting of this model requires a sensitive harmony between precise fit and deliberation.

A variation of classification and regression trees is called irregular backwoods. Rather than developing a solitary tree with numerous parts of rationale, an irregular wood is a zenith of numerous little and basic trees that each assess the cases of information and decide a classification. When these straightforward trees total their information assessment, the interaction combines the individual outcomes to make a last forecast of the classification dependent on the composite of the more modest orders. This is ordinarily alluded to as an outfit technique. These arbitrary backwoods regularly excel at adjusting precise fit and reflection and have been executed effectively in numerous business cases.

Rather than calculated regression, which centres around a yes or no order, characterization and relapse trees can be utilized to foresee multivalued arrangements. They are likewise simpler to imagine and see the authoritative way that guided the calculation to a particular classification.

K-Nearest Neighbours

K-Nearest Neighbours is also a classification algorithm. It is known as a "apathetic learner" in light of the fact that the preparation period of the cycle is exceptionally restricted. The learning interaction is made out of the preparation set of information being put away. As new examples are assessed, the distance to every information point in the preparation set is assessed and there is an agreement choice concerning which class the new case of information falls into dependent on its closeness to the preparation cases.

This algorithm can be computationally costly relying upon the size and extent of the preparation set. As each new occasion must be contrasted with all cases of the preparation informational index and a distance determined, this interaction can utilize many registering assets each time it runs.

This categorization algorithm allows for multivalued orders of the information. Also, uproarious preparing information will in general slant arrangements.

K-nearest neighbour is regularly picked on the grounds that it is not difficult to utilize, simple to prepare, and simple to decipher the outcomes. It is frequently utilized in hunt applications when you are attempting to find similar items.

K-Means Clustering

K-means clustering focuses on creating gatherings of related attributes. These gatherings are cluster to as groups. When these groups are made, different occasions can be thought about in contrast to them to see where they best fit.

This method is regularly utilized as a feature of information investigation. To begin, the examiner indicates the quantity of groups. The K-implies group measure breaks the information into that number of bunches dependent on discovering information focuses with similitudes around a typical centre, called the centroid.

These groups are not equivalent to classes in light of the fact that at first, they don't have business meaning. They are simply firmly related examples of information factors. When these bunches are recognized and dissected, they can be changed over to classifications and gave a name that has business meaning.
K-means clustering is regularly utilized on the grounds that it is easy to utilize and clarify and in light of the fact that it is quick. One territory to note is that k-implies grouping is incredibly touchy to anomalies. These exceptions can altogether move the nature and meaning of these bunches and at last the consequences of examination.
These are probably the most famous calculations being used in cutting edge examination activities. Each has upsides and downsides and various manners by which it tends to be viably used to produce business esteem. The end focus with the usage of these calculations is to additionally refine the information to a point where the data that outcomes can be applied to business choices. It is this cycle of advising downstream cycles with more refined and higher worth information that is a principal to organizations turning out to be really outfitting the estimation of their information and accomplishing the outcomes that they want.

Virus
Current Pandemics
HIV/Aids
HIV, although referred to as a "global epidemic" by the WHO, is now considered a pandemic primarily concentrated in Africa. It originated in Africa and spread to the United States through Haiti between 1966 and 1972. In certain regions of southern and eastern Africa, HIV has reached pandemic proportions with infection rates as high as 25%. In 2006, the HIV prevalence among pregnant women in South Africa was 29%. Public education programs focusing on safer sexual practices and bloodborne infection precautions have helped curb infection rates in some African countries. As of 2018, HIV/AIDS has resulted in millions of infections and approximately 32-35 million deaths globally.

COVID-19
COVID-19, caused by a novel strain of coronavirus, emerged in Wuhan, China, in late December 2019. It rapidly spread and caused a severe respiratory illness, leading to the declaration of a pandemic by the World Health Organization on March 11, 2020. As of February 18, 2021, COVID-19 has affected over 200 countries and territories, with significant outbreaks in various regions, including Brazil, Russia, India, and the United States. The number of confirmed cases has reached 110,602,492 worldwide, with 85,469,351 recoveries and a death toll of 2,444,738. The actual numbers may be underestimated due to limited testing in the early stages and the presence of asymptomatic cases. Spatial-temporal analysis has been conducted to understand the initial spread of COVID-19 in China and Italy, while models have been developed to assess the likelihood of a global pandemic.

Influenza
The "Spanish influenza," which occurred during 1918-1919, is considered one of the deadliest pandemics in history. Originating from U.S. troops in Kansas, it quickly spread worldwide, infecting approximately one-third of the global population (approximately 500 million people). Within six months, around 50 million individuals died from the virus. India, the United States, and the United Kingdom were significantly affected, with millions of deaths reported. Recent efforts have led to the successful reproduction of the Spanish influenza virus from remains preserved in Alaskan permafrost.

Malaria
Once prevalent in Europe and North America, malaria has been eradicated from these regions. It is believed that malaria may have contributed to the decline of the Roman Empire, and the disease was known as "Roman fever." Plasmodium falciparum, the deadliest form of malaria, became a major threat to settlers and indigenous populations in the Americas during the slave trade era. Malaria heavily impacted regions such as the Jamestown settlement and the South and Midwest of the United States. While malaria cases

have significantly reduced in these regions, the disease continues to be a public health concern in other parts of the world.

H5N1

Avian influenza, caused by the H5N1 virus, is not classified as a pandemic as it lacks sustained and efficient human-to-human transmission. While there have been some cases of transmission from birds to humans, confirmed instances of human-to-human transmission have been limited. Avian flu viruses attach to receptors deep in the lungs of humans, requiring close and prolonged contact with infected patients, thus limiting person-to-person transmission.

These current pandemics continue to pose significant challenges to global health, and efforts are ongoing to understand and control their spread.

LIMITATIONS AND FINDINGS

Limitations

- The paper provides a general overview of big data analytics algorithms and pandemic viruses, but it lacks in-depth technical details and specific examples.
- The information presented in the paper may become outdated quickly, as the field of big data analytics and pandemic viruses is rapidly evolving.
- The paper does not include a comprehensive analysis of the current state of big data analytics tools and techniques, which could provide more insights into their effectiveness and limitations.
- The discussion on pandemic viruses is limited to a few specific examples, and it does not cover other significant pandemics that may have occurred after the paper's cutoff date.

Findings

- Big data analytics algorithms play a crucial role in processing and analyzing large and complex datasets, enabling organizations to gain valuable insights and make data-driven decisions.
- Supervised algorithms like linear regression and logistic regression are effective for predictive modeling, while unsupervised algorithms like clustering and classification can help in discovering patterns and grouping similar data points.
- The paper highlights the importance of using advanced big data analytics techniques to handle the massive and diverse datasets that are generated in various industries.
- Pandemic viruses, such as HIV, COVID-19, and Malaria, have had significant global impacts, leading to millions of infections and deaths, underscoring the importance of public health preparedness and response.
- Understanding and analyzing pandemic viruses' characteristics and spread patterns can aid in the development of effective preventive measures and treatments.
- It is essential to stay updated with the latest advancements in big data analytics and virology to address new challenges and emerging threats effectively. Continuous research and collaboration among experts are crucial in tackling global health crises.

CONCLUSION

In conclusion, this paper highlighted some of the key big data analytics algorithms and advanced pandemic viruses. Big data analytics algorithms such as linear regression, logistic regression, K-nearest neighbors, and clustering play a crucial role in extracting meaningful insights from large and complex datasets. These algorithms have applications across various industries, including healthcare, finance, marketing, and more. Furthermore, the paper discussed some significant pandemic viruses like HIV/AIDS, COVID-19, and Malaria. These viruses have had a profound impact on global health and have caused widespread devastation. Efforts to control and manage these pandemics have been complex and challenging, requiring

global cooperation and coordination.

In the future, further research can be conducted to explore and develop more sophisticated big data analytics algorithms that can handle even larger and more diverse datasets efficiently. Additionally, advancements in machine learning and artificial intelligence can be leveraged to enhance the accuracy and predictive capabilities of these algorithms.

Regarding pandemic viruses, ongoing research and preparedness are essential to prevent and manage future outbreaks. The development of effective vaccines, improved surveillance systems, and better healthcare infrastructure are critical to mitigate the impact of pandemics on global health.

Moreover, the integration of big data analytics with the study of pandemic viruses can lead to better understanding and prediction of disease outbreaks. By analyzing vast amounts of real-time data, health authorities can make informed decisions and implement targeted measures to control the spread of infections effectively.

In conclusion, the intersection of big data analytics and pandemic virus research presents exciting opportunities for improving public health and safety. Continued research and collaboration among scientists, researchers, and data experts are essential to address the challenges posed by both big data and emerging infectious diseases. By harnessing the power of data-driven insights, we can work towards creating a safer and healthier world for all.

REFERENCES

1. Dwivedi, Y. K., Janssen, M., Slade, E. L., Rana, N. P., Weerakkody, V., Millard, J., Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modeling. Information Systems Frontiers, 19(2), 197-212. https://doi.oig/10.1007/sl0796-016-9675-5
2. SASSI, S. OUAFTOUH and S. ANTER, "Adaptation of Classical Machine Learning Algorithms to Big Data Context: Problems and Challenges: Case Study: Hidden Markov Models Under Spark," 2019 1st International Conference on Smart Systems and Data Science (ICSSD), Rabat, Morocco, 2019, pp. 1-7, doi: 10.1109/ICSSD47982.2019.9002857.
3. 4.Tang, Yan & Wang, Yu & Cooper, K.M.L. & Li, Ling. (2014). Towards Big Data Bayesian Network Learning - An Ensemble Learning Based Approach. Proceedings - 2014 IEEE International Congress on Big Data, BigData Congress 2014. 355-357. 10.1109/BigData.Congress.2014.58.
4. 3.Y. Zhao-hong, W. Hui-yu, Z. Bin, H. Zhi-he and L. Wan-lin, "A literature review on the key technologies of processing big data," 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, 2018, pp. 202- 208, doi: 10.1109/ICCCBDA.2018.8386512.
5. M. D. A. Praveena and B. Bharathi, "A survey paper on big data analytics," 2017 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, 2017, pp. 1-9, doi: 10.1109/ICICES.2017.8070723.
6. H. Y. Putra, H. Putra and N. B. Kurniawan, "Big Data Analytics Algorithm, Data Type and Tools in Smart City: A Systematic Literature Review," 2018 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung - Padang, Indonesia, 2018, pp. 474-478, doi: 10.1109/ICITSI.2018.8696051.
7. S. Bahri, N. Zoghlami, M. Abed and J. M. R. S. Tavares, "BIG DATA for Healthcare: A Survey," in IEEE Access, vol. 7, pp. 7397-7408, 2019, doi: 10.1109/ACCESS.2018.2889180.
8. Harshawardhan S. Bhosale, Devender P. Gadekar," A Review paper on Big Data and Hadoop", International Journal of Scientific and Research Publication, Vol 4, 2014
9. Cheikh Kacfah Emani, Nadine Cullot, Christopher, Nicolle, "Understandable Big Data: A Survey", Elsevier Computer Science Review (2015)
10. S. Vikram Phannendra,E. Madhusudhana Reddy,"Big Data – Solution for RDBMS Problems- A Survey", IJARCCE(International Journal Advance Research in computer and Communication Engineering), Vol 2 (2013).

11. Vibha Shukla, Pawan Kumar Dubey," Big Data: Moving Forward with Emerging Technology and Challenges", International Journal Advance Research in computer Science and Management Studies, Vol 2(2014)
12. Prity Vijay and Brigh Keshwani," Emergence of Big Data and Hadoop: A Review", IOSR Journal of Engineerinf, Vol 6(2016)
13. Raymond Gardiner Goss," Heading Towards Big Data ", IEEE (2013)
14. Poonam S. Patil and Rajesh N. Phursule," Survey Paper on Big Data and Hadoop Component", International Journal of Scientific and Research, Vol 3, 2014
15. Kuchipudi Sravanthi, T. Subba Reddy," Application of Big Data in Various Fields", International Journal of computer Science and Information Technologies, Vol 6(2015)
16. 1.S. Otoum, B. Kantarci and H. Mouftah, "Empowering Reinforcement Learning on Big Sensed Data for Intrusion Detection," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-7, doi: 10.1109/ICC.2019.8761575.
17. Kung-Hsiang, Huang (Steeve), Co-Founder & CTO, Rosetta.ai | Research Assistant, ISI |
18. J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques", 3rd Edition, MK Series, 2012.
19. Susmita Ray," A Quick Review of Machine Learning Algorithms", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT- Con), India, 14th -16th Feb 2019
20. S. Otoum, B. Kantarci and H. Mouftah, "Empowering Reinforcement Learning on Big Sensed Data for Intrusion Detection," ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-7, doi: 10.1109/ICC.2019.8761575.
21. Dr. O. Obulesu, M. Mahendra, M. ThrilokReddy "Machine Learning Techniques and Tools: A Survey", 2018 Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018)
22. IEEE Xplore Compliant Part Number: CFP18N67-ART;ISBN:978-1-5386-2456-2
23. Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004), "Detecting Student Misuse of Intelligent Tutoring Systems". Proceedings of the 7 th International Conference on Intelligent Tutoring Systems, 531-540.
24. Romero, C., Ventura, S., de Bra, P., & Castro, C. (2003)," Discovering prediction rules in aha! Courses". Proceedings of the International Conference on User Modelling, 25- 34.
25. Beck, J., & Woolf, B. (2000)." High-level student modelling with machine learning". Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 584- 593.
26. M. Ramaswami and R. Bhaskaran(2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, pp 10-18.
27. Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme(2009), "Improving Academic Performance Prediction by Dealing with Class Imbalance", Ninth International Conference on Intelligent Systems Design and Applications,.
28. Jiawei Han Michelin Kamber(2011), "Data Mining-Concepts and Techniques", Morgan Kaufmann Publishers.