iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54601

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Addressing Big Data with Various Application - A Review

Remya Stephen
Computer Science, BCM College for Women, Kerala, India

Abstract: ***The term big data refers to data that is diverse, huge, and incredibly fast. Aside from the volume, big data is also complex, making conventional data management tools unable to handle it. There is an explanation of big data in this abstract, as well as various applications of this technique.***
Big data is precious and an important dogfight for all sectors moment. It's used by nearly all associations and has numerous use cases.
Keywords: Big-data, Characteristics, challenges, Applications, AI and Big data

I. INTRODUCTION

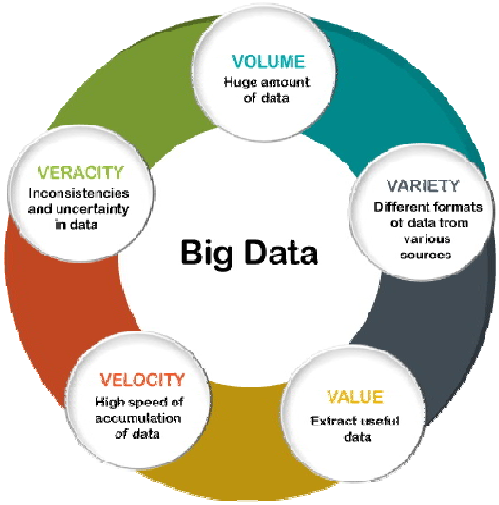
Large Quantities of data are stored in an systematized manner on a big data platform. Added up data sets are generally stored on big data platforms, which combine tackle and software tools for data operation.
Multitudinous sophisticated and largely scalable pall data platforms have surfaced to store and reuse the continuously growing volume of data from colorful sources because the data affluence is patient and only anticipated to increase in intensity. Big data platforms are the name given to these kinds of platforms.
A big data platform organizes and stores this volume of data in a way that makes it easy to draw perceptive conclusions. Big data platforms use a combination of tackle and software tools for data operation to aggregate data on an enormous scale, generally onto the pall.
The constant influx of information from colorful sources is becoming more violent, especially with advances in technology. And this is where big data platforms come into the store and dissect the ever- adding a mass of information.
A big data platform is an intertwined computing result that combines multitudinous software systems, tools, and tackle for big data operation.
It's a one- stop armature that solves all the data requirements of a business, anyhow because of the volume and size of the data at hand. Due to their effectiveness in data operation, enterprises are decreasingly espousing big data platforms to gather tons of data and convert it into structured, practicable business receptivity.
Presently, the business is swamped with multitudinous Open source and commercially available big data platforms. They boast different features and capabilities for use in a big data terrain.

A. Characteristics of big data

Big data generally has a multidimensional structure and can be characterised by five.

1) *Volume:* Big Data is a vast' volumes' of data generated from numerous sources daily, similar as business processes, machines, social media platforms, networks, mortal relations, and numerous further.
2) *Velocity:* Velocity plays an important part compared to others. Velocity creates the speed by which the data is created in real-time. It contains the linking of incoming data sets pets, rate of change, and exertion bursts. The primary aspect of Big Data is to give demanding data fleetly.
3) *Veracity:* Veracity means how important the data is dependable. It has numerous ways to filter or restate the data. Veracity is the process of being suitable to handle and manage data efficiently. Big Data is also essential in business development.
4) *Variety:* Big Data can be structured, unshaped, andsemi-structured that are being collected from different sources. Data will only be collected from databases and wastes in the history, But these days the data will comes in array forms, that are PDFs, Emails, audios, SM posts, prints, vids,etc.

©IJRASET: All Rights are Reserved



II. METHODS AND TOOLS USED IN BIG DATA

With the elaboration of technology and the increased multitudes of data flowing in and out of associations daily, there has come a need for faster and more effective ways of assaying similar data. Having piles of data on hand is no longer enough to make effective opinions at the right time. Similar data sets can no longer be fluently anatomized with traditional data operation and analysis ways and architectures. thus, there arises a need for new tools and styles specialized for big data analytics.

In this context, the Big-Data Analytics and Decisions framework is proposed, which incorporates big-data analytics tools and styles into decision-making. In this frame, different big data storage, operation, and processing tools, analytics tools, and visualization and evaluation tools are mapped to the different phases of the decision-making process. Data and analytics recycling, as well as big data analyses that can be applied for knowledge discovery and informed decision-making, are three main areas associated with big data analytics.

A. Big Data Warehouse and governance

The conventional styles of structured data storehouse and reclamation include relational databases, data marts, and data storages.Although, the big data domain calls for magnetic, nimble, Deep analysis chops, which differ from the aspects of a traditional Enterprise Data Warehouse(EDW) terrain. Consequently, several results, ranging from distributed systems and Massive Pa-rallel Processing(MPP) databases for furnishing high query performance and platform scalability, to non-relational or in-memory databases, have been used for big data.

On the other hadoop is a frame for performing big data analytics which provides trustability, scalability, and manageability by furnishing an perpetration for the MapReduce paradigm. The Hadoop platform consists of two main components: HDFS for storage and MapReduce for analytics. As a redundant and reliable distributed file system, HDFS is optimized for large files, dividing each file into blocks and distributing them across cluster nodes. In addition to this, a replication mechanism ensures the data's availability and reliability even if one of the nodes fails. HDFS nodes can be categorized into two types: Data Nodes and Name Nodes. Across multiple Data Nodes, data is replicated in file blocks, and the Name Node acts as a regulator between the client and the Data Node, directing the client to the Data Node that contains the requested information.

B. Analyzing large amounts of data

There are four essential aspects for efficient big data processing. A key factor is a fast loading speed as disk and network traffic can disrupt query executions during data loading. Another critical requirement is the quickness of query processing, to be able to meet the demands of heavy workloads and on-time requests. Response time is a must for any queries. Consequently, the data placement structure must be able to maintain rapid query processing speeds as the number of queries significantly grows. In addition, efficient utilization of storage space is a critical requirement for effective big data processing. With the fast-growing user activities demanding scalable storage capacity and computing power, it is essential that storage space be managed effectively during processing due to limited disk space availability.

Ultimately, the fourth requirement is robust adaptivity to enormously dynamic workload styles. The underlying machine should be relatively adaptive to unexpected dynamics in data processing, and not precise to sure workload styles.

III. VARIOUS APPLICATION OF BIGDATA

Industries,, academics, and other prominent stakeholders certainly agree that big data is revolutionizing most, if not all, modern industries in recent years. As big data continues to permeate our daily lives, the focus has shifted significantly away from the hype surrounding big data to finding real value in using big data.

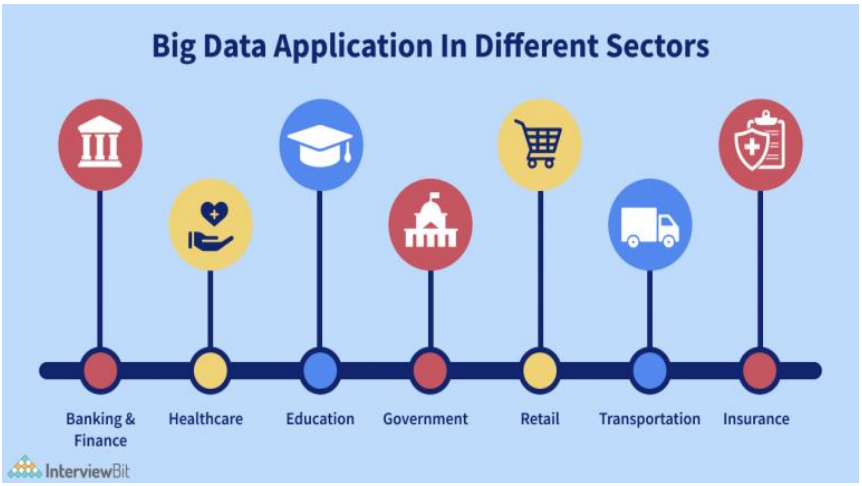


Figure 2: Application of bigdata

A. Big Data in Banking and Securities Industry

Big Data has made banks more efficient in every industry. The use of technology has overcome user difficulties, increased bank revenues, and made insights more transparent and understandable than ever before. From detecting fraud to analyzing and streamlining trade processing to better understanding users, perfecting trade processing, and improving the user experience, Big Data has a wide range of uses[2].

The purpose of Big Data in banking is that people expect banks to provide realistic and practical solutions. They expect a faster and more streamlined process, personalized advice, accurate, complete and up-to-date information on topics that are informative and relevant, such as applying for a traditional loan to buy a house[2].

Advanced analytics and the use of cutting-edge technologies such as artificial intelligence and machine learning can leverage data in areas such as personalization and convenience, protection, and the creation of relevant offers. Artificial intelligence and machine learning will play an important role in the future of finance[2]. They will help banks analyze data, predict customer behavior and personalize financial services.

1) *Personalization and convenience*: When it comes to loan applications, big data will enable banks to assess potential credit risk more quickly (through faster response). Big data provides a framework for understanding what kinds of interactions customers prefer with their banks.
2) *Protection:* Big data helps banks detect potential fraud before irreparable damage is done, supports monitoring suspicious transactions, and enhances investment safety and data protection of customer accounts to do.
3) *Create relevant offers:* With the help of big data, banks can predict their customers' financial goals and create offers that meet them. It also helps advisors and sales representatives to cater to individual customers and offer specific discounts based on their individual needs and interests.

B. Big Data in Media and Entertainment Industry

As the current strategy of using gadgets to consume online content becomes the latest trend, enthusiasm for traditional media consumption methods is slowly fading. Big data is winning in this industry due to the enormous amount of data generated. From predicting what audiences want in terms of genres, music and content by age group, to providing insights into customer churn, big data has made media houses' jobs much easier[6].

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Organizations in this industry simultaneously analyze customer and behavioral data to create detailed customer profiles that can be used to:

- Create Content for Different Audiences
- Recommend Content on Demand
- Measure Content Performance

Spotify, an on-demand music service, uses Hadoop big data analytics to collect data from millions of users around the world and uses the analyzed data to make informed music recommendations to individual users[9].

Amazon Prime, which aims to provide a superior customer experience by offering a one-stop shop for videos, music, and Kindle books, is also making heavy use of big data[9].

Another apt example of big data playing a key role in transforming media platforms would be Netflix. Technology not only affects the series Netflix invests in, but also how the series is presented to viewers. Your search and viewing history (including where you paused videos) influences everything from personalized thumbnails to shows you watch on Netflix's popular section.

C. Big Data in the Healthcare Sector

The health care sector has access to vast amounts of data, but is plagued by deficiencies in using data to contain rising health care costs and inefficient systems that hinder prompt and better health care overall. Due to the high medical costs, and health ignorance, the health care industry is in crisis[9]. We need a system that collects real-time data from patients and predicts disease risk. Such a system can uses efficient machine learning algorithms and big data analysis technology to enable patients to remotely monitor their health status anytime, anywhere[4].

In developing countries like India, healthcare is a major concern, including elderly health, maternity, diabetes, asthma, cardiac arrest, and cancer. Due to the high cost of treatment, there is a growing need for big data analytics for healthcare. This cost is much higher and has been increasing for 20

years. only 20% of individuals are wealthy enough to afford such medical care. But the remaining 80% of people who can't even think of going to multiple reputable specialty hospitals for treatment [4]. With this technology, the system could be made intelligent enough to predict the frequency of disease outbreaks and inform patients early about their condition.

Even though, there are some challenges in big data One of the major obstacles to big data analytics in healthcare is that the data is spread across many sources and assimilating these data sources requires the development of new strategies and policies. However, the healthcare industry faces tremendous challenges around data storage, security, visualization, management, reporting, querying, sharing, and many data integrity concerns, to name a few.

Advantage of big data in health care

Big data analytics have the potential to significantly improve the patient experience, including treatment quality, satisfaction, cost savings and timely predictions. With Medical Data Analytics. Here, we could get,

- Expanding patient care
- Expediting treatment
- Reducing electricity bills:
- Fraud detection
- Reducing hospitalizations and readmissions
- Finding cures for illness

D. Big Data in Education

Big data is increasingly being used in education to improve teaching and learning, improve student achievement, and inform decision-making. Here are some ways to use big data in education[9].

1) *Personalized Learning:* Big data analytics can create personalized learning experiences for students. By analyzing individual student data such as progress, preferences, and learning pace, the education platform can provide tailored content, adaptive assessments, and personalized recommendations to meet each student's specific needs.
2) *Predictive Analytics:* Educational institutions can use predictive analytics to identify students who may be at risk of dropping out or under performing. By analyzing various factors, such as attendance, grades, and engagement, predictive models can help identify early warning signs and enable targeted interventions to support struggling students.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) *Curriculum Development:* Big data can inform curriculum development by analyzing student performance data, educational standards, and real-time feedback. It helps educators identify areas where students may be struggling, revise and improve instructional materials, and align curriculum with the evolving needs of students.
4) *Research and Evaluation:* Big data can support educational research and evaluation efforts. Researchers can analyze large datasets to gain insights into factors affecting student achievement, educational policies, and teaching methodologies. This can lead to evidence-based practices and the development of more effective educational strategies.
5) *Adaptive Assessments:* Big data analytics can be applied to adaptive assessments, which dynamically adjust the difficulty and content of questions based on the student's responses. By analyzing vast amounts of data on student performance, these assessments can provide real-time feedback and identify areas where students need further support or challenge.

E. Big Data in Transportation

Big data plays a key role in making transportation more perfect and efficient. Whether managing revenue generated, maintaining earned reputation, or pursuing strategic marketing, big data is making an impact in this area.[3] It also helps in planning routes according to user needs, efficiently managing waiting times, and identifying accident-prone areas to increase traffic safety levels.
A perfect example of the use of big data in the transportation industry is Uber. The platform creates and uses a variety of data about vehicles, drivers, locations, and travel by each vehicle, and tests that data for use in predicting demand, supply, and accurate driver locations and fares[9].

Here are some of the tools listed that big data using in transportation

1) *Traffic Management:* Big data analytics is used to monitor and manage traffic flow by collecting and analyzing vast amounts of data from various sources such as traffic sensors, GPS systems, and social media.[3] This data helps predict traffic patterns, identify congestion points, and optimize signal timing to reduce congestion and improve overall efficiency. Tools used for traffic management include:
2) *Traffic Management Systems:* These systems collect real-time data from sensors and cameras to monitor traffic conditions and make informed decisions.
3) *Intelligent Transportation Systems (ITS):* ITS leverages big data to manage traffic
4) *Demand Prediction and Planning:* Big data is utilized to predict and plan transportation demand, helping transportation agencies and companies optimize their services. By analyzing historical data, weather conditions, events, and other factors, predictive models can forecast demand and enable proactive planning. Tools used in demand prediction and planning include
5) *Demand Forecasting Models:* These models analyze historical data, seasonality, and external factors to predict future demand patterns.
6) *Data Visualization Tools:* These tools help present and analyze large datasets visually, allowing stakeholders to understand trends and patterns easily.

Big data analytics is employed to improve the efficiency and quality of public transportation systems. By analyzing passenger data, fare collection records, and scheduling information, transportation authorities can optimize routes, adjust service frequency, and enhance overall passenger experience. Tools used in public transportation optimization include:

a) *Automated fare Collection Systems:* These systems collect data on passenger journeys, boarding and alighting patterns, and payment information, enabling transportation authorities to analyze passenger flows and make service adjustments.
b) *Real-time Passenger Information Systems:* These systems provide travelers with real-time updates on arrival times, service disruptions, and alternative routes, improving the overall passenger experience.

IV. HOW BIG DATA RELATED WITH AI

Big data and artificial intelligence have a synergistic relationship. AI requires vast amounts of data to learn and improve decision-making processes, and big data analytics leverages AI for better data analysis.
AI can assist users at every stage of the big data cycle, or the process of aggregating, storing, and retrieving different kinds of data from different sources. These include data management, pattern management, context management, decision management, behavior management, goal management and risk management.
AI can identify data types, find possible connections between data sets, and recognize knowledge using natural language processing. It can be used to automate and speed up data preparation tasks such as data model generation to aid in data exploration.

You can learn patterns of common human errors and identify and correct potential information errors. It also learns by observing users interact with analytics programs, quickly discovering unexpected insights from massive data sets. AI can also learn semantic nuances and context-specific nuances to help users better understand numerical data sources. It can also alert users to anomalies and unexpected patterns in data, pro actively monitor events, and identify potential threats using system logs, social networking data, and more.

Here's how big data and AI are interconnected:

1) *Real-time Analytics:* Big data combined with AI enables real-time analytics, where large volumes of data are processed and analyzed in real-time. This allows for quick decision-making, fraud detection, dynamic pricing, and other time-sensitive applications.
2) *Recommendation Systems:* Big data is essential for building recommendation systems powered by AI. These systems analyze user behavior, preferences, and historical data to generate personalized recommendations. The more data available, the more accurate and relevant the recommendations can be.
3) *Data Preprocessing:* Big data often requires preprocessing before it can be used effectively by AI systems. Preprocessing involves cleaning, transforming, and organizing the data to remove noise, handle missing values, and standardize formats. This step ensures that the data is suitable for AI algorithms to process.
4) *Machine Learning:* Big data plays a crucial role in machine learning, a subset of AI. Machine learning algorithms process and analyze vast amounts of data to identify patterns, extract features, and create models. These models are then used to make predictions, classify data, or generate insights.
5) *Training Data:* AI algorithms require large amounts of data to learn patterns, identify correlations, and make accurate predictions. Big data serves as the training data for AI models. The more diverse and extensive the dataset, the better the AI system can learn and generalize from it.

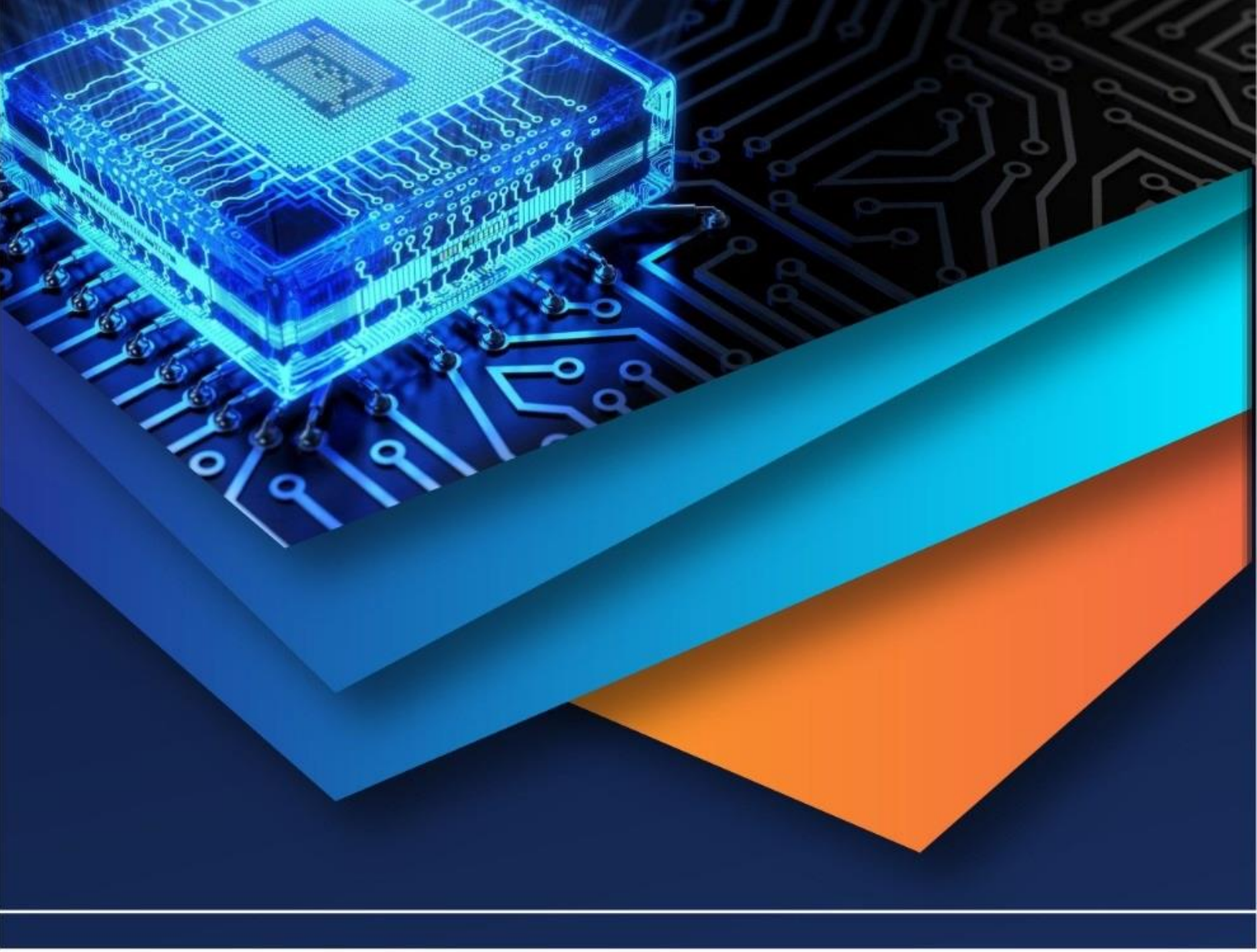
Overall, big data provides the necessary raw material for AI systems to learn, adapt, and make intelligent decisions. The abundance of data, along with advancements in AI algorithms, has led to significant breakthroughs in various fields, driving innovation and improving outcomes.

V. CONCLUSION

In conclusion, big data has become a powerful tool with widespread applications across various industries. Its ability to process and analyze vast amounts of data has transformed the way businesses operate, governments make decisions, and researchers gain insights. As big data continues to grow in volume and complexity, its applications are expected to expand further, transforming industries and driving innovation. However, it is important to address privacy concerns, ensure data security, and adhere to ethical practices when utilizing big data for various purposes.

REFERENCES

[1] Yasmeen Shaikh, S. R. Biradar and V. K. Parvati, “Role of IoT and Bigdata Analytics in Healthcare for Disease Prediction” 2020 IEEE International Conference on Convergence to Digital World – Quo Vadis (ICCDW 2020)
[2] Anku Jaiswal and Purrushottam Bagale, 3rd “A Survey on Big Data in Financial Sector” 2017 International Conference on Networking and Network Applications
[3] Nii Attoh-Okine “Big Data Challenges in Railway Engineering” 2014 IEEE International Conference on Big Data
[4] Margaret. Rouse, ‘Remote Patient Monitoring (RPM)’, April 2019.[Online].Available:https://searchhealthit.techtarget.com/definition/remote-patient-monitoring-RPM. [Accessed: 13- Jun- 2019]
[5] Md Ileas Pramanik, Raymond Y.K. Lau, Haluk Demirkan, Md.AbulKalam Azad, “Smart health: Bigdata enabled Health paradigmwithin Smart Cities”, An international journal of Expert Systems with Applications 87 (2017) 370–383, ScienceDirect, Elsevier.
[6] Jagadish, H.V. et al. Big Data and Its Technical Challenges.Communications of the ACM, July 2014.
[7] Chen, C.L. P. and Zhang, C. Y . Data-intensive applications, challenges,techniques and technologies; A Survey on Big Data. Information Science, Vol 275, 2014, 314-327.
[8] https://www.qlik.com/us/augmented-analytics/big-data-ai#:~:text=are%20transforming%20BI-,Relationship%20between%20AI%20and%20Big%20Data,AI%20for%20better%20data%20analysis.
[9] https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications
[10] https://www.interviewbit.com/blog/applications-of-big-data/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)