

The Impact of Big Data on Climate Change

Adish Chandra Shrivastava,
Department of ITCA
Madan Mohan Malaviya University of
Technology
Gorakhpur, India
aadishshrivastav@gmail.com

Mohd. Umar
Dept. of Electronic and Communication
Madan Mohan Malaviya University of
Technology
Gorakhpur, India
umammmut@gmail.com

Abstract— The precariousness and impacts of climate change are spreading worldwide, resulting in economic, environmental, and social damages. Due to this damage, some irregular variations in temperature, rainfalls, and sea level are being experienced. Such variations in temperature, rainfall patterns, or sea level rise are irrelevant. Variations in rainfall intensity will significantly impact water level and water quality. Rising sea levels will impact land usage and development in the coastal area. Big climate data analytics implementations have focused on climate change because it is a new issue, and extensive study has been done on various subjects. Rainfall prediction is the most crucial aspect of Big data techniques in climate change. Since most individuals worldwide depend on agriculture, this may benefit farmers and the general populace. The study illustrates numerous methods by which big data has aided in creating predictions regarding the problem of climate change. It can assist farmers in making wise judgments on crop yield. Studying the timing of floods or droughts simultaneously may be possible.

Keywords— Big Data, Climate Change, Data Analytics, Global Warming, Regression, Statistics.

I. INTRODUCTION

Over the past several decades, the consequences of change in climate on "Earth" and the impact on living conditions and life have been documented. The apparent rise in possibly irreversible climate change consequences, such as mass extinctions of flora and fauna, global warming, falling water tables, and melting ice caps, has recently sparked considerable anxiety among the world's governments. To counteract these consequences, researchers have looked for fresh perspectives and strategies.

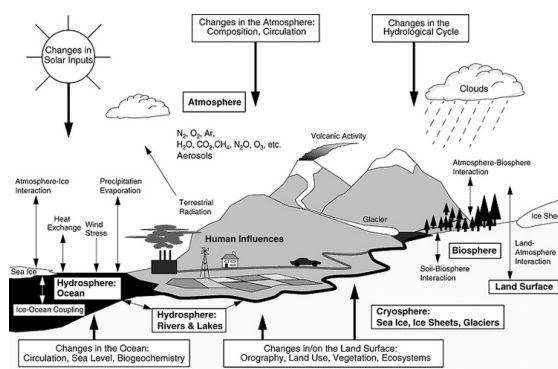


Fig 1- Diagram showing the elements of the global climate system (in bold), their interactions and processes (in thin arrows), and some potential changes brought on by global warming (bold arrows).

It is possible to gain valuable insights into any process by collecting relevant data, collating it, and analyzing it. Internet-connected sensors give real-time data, yet, the sharp volume of data points created by sensors, intelligent gadgets, satellite information, and in-situ devices is overwhelming. Big Data refers to data collections that are so vast that typical Database Management Systems (DBMS) find it difficult to process them. These data sets are too enormous to be acquired, saved, managed, or analyzed using specific software tools and storage methods in a reasonable amount of time.

The traditional definition of climate change is shifting in regional or global weather patterns that are mostly attributable to the creation of CO_2 gas from fossil fuel combustion. The abnormal climate fluctuations, which have progressively increased throughout the 20th century, may impose harm to habitable ecosystems. Human activity is responsible for a wide range of climate changes, including the rise in temperature, uneven rainfall distribution, ice caps defrosting, retreating glaciers, dropping water levels, altered river courses, flash floods, rising sea levels, contrary agriculture outputs, and the extinction of native flora and fauna.

Without analysis to extract information, data is merely a sequence of quantities and is therefore worthless. Big Data is increasing independently and significantly advancing our understanding of data-dependent methodologies because of the exponential development of big data caused by technological advancements, research approaches, and the number of connected devices. Early climate data was collected through observation; measurements must be brutally recorded and confirmed. Nowadays, data is usually accessible in real-time and is growing at a faster pace. The gathering and analysis of climate data have been revolutionized by sensors and satellites.

Current developments in big data tools for climate research, the Internet of Things, intelligent analytics platforms, and other cutting-edge technologies have brought even more attention to the potential implications of big data for advancing big data science and climate science. Big data has been used in various ways to analyze climate data.

In order to combat climate change, current research combines big data analytics to enhance the energy efficiency, intelligent agriculture, careful urban design, weather forecasting, and natural disaster management. Even though this is not a new topic, much literature on big data and climate change exists. Additionally, innovative

breakthroughs and technical advancements proliferate each day.

Experts in big data have spent the previous two decades focusing on climate change, a developing concern that employs diverse data. Understanding climate change requires a substantial amount of data. The Internet of Things, cloud computing, and noteworthy climate data have all been explored and researched in order to stop climate change. However, because the world is an extensive, dynamic system, we are still learning about it. We do not fully comprehend its workings like other professions do when employing big data. The goal of this paper is to add to the body of previously published work by analysing the role that cutting-edge technologies, like big data, have played in studies of the environment and climate change. It also provides a detailed evaluation of the big data resources and their analytics techniques that have been used in the study of climate data.

In this article, we look at some of the major big data problems that researchers face when examining climate data, as well as the possible impacts of big data on climate change. The most important concepts will be discussed and illustrated with examples from recent climate and data science studies. This will aid in preparing the next generation of data scientists to tackle the enormous challenges of climate change.

II. IMPACT OF BIG DATA ON CLIMATE CHANGE

We now understand massive climate data, how they are collected, and what they may be used for, thanks to abundant resources. But these hardly scratch the surface of big data analytics. When combined with climate science, big data has also played a crucial role in prediction, such as weather forecasting, monitoring and early warning of natural disasters, forecasting energy use, traffic patterns, etc. Applying the appropriate data mining techniques enables knowledge discovery of possible linkages and causal inferences, further modeling, and forecasting [14].

Extremely rapid changes can occur in the atmosphere, such as the hydrological cycle, cryosphere, land surfaces, ocean, and weather patterns. These changes are unpredictable. Without warning, these changes may affect a small geographic region or perhaps a whole continent. To better comprehend and address climate problems, data analysis of weather pattern shifts can lead to optimal yield, risk avoidance, and improvised emergency action [5]. In collaboration with climate research, big data enables forecasting energy consumption, traffic, early warning systems for natural disasters, and the weather, among other phenomena [6].

Big data is used to formulate trends, whereas AI can identify anomalies or patterns that are invisible to the human eye. Accurate forecasts can be beneficial in developing adaptive policies to lessen the effects of climate change.

Previous methods relied solely on tracking and observation. As the volume and complexity of data from many sources increase, new learning algorithms and methodologies are being developed. Using AI and ML, intelligent systems can recognize patterns independently with less human effort.

Environmental change research has undergone a complete transformation due to this new technique.

III. RELATED WORK

Nikhil Sethi and Kanwal Garg built a multiple linear regression using climate factors by employing the climatic data set for the Rajasthani city, Udaipur. He can compute the coefficient values using the linear regression coefficient formula.

H. Hasani presented a polynomial regression-based model for predicting the growth of height. This approach can help in researching child development.

H. Nguyen describes a methodology that uses historical sales data, economic indicators, artificial neural networks, and regression to forecast industry sales.

Imran Ahmad et al. reported a model for predicting rainfall using multiple regressions in the Coonor area of the Nilgiris district, Tamil Nadu. He could suggest a brand-new formula for figuring out regression coefficient values.

Ismail used linear regression to develop the forecasting model to estimate the price of gold. Anything he can implement using the variables money and inflation, it is possible to conclude that MLR outperforms the Naive prediction approach. Data mining scientists and others can assess the forecasting methodologies.

IV. BIG DATA FRAMEWORK

The two pillars of big data applications in climate change are extensive resources for climate data and thorough data analytics techniques. We categorise these studies according to the particular application issue they deal with and the value they provide. The four primary ways big data aids in developing value in the context of climate change are observation and comprehension; prediction and optimization; other innovative uses; and other methods. Five categories have been used to categorize the applications:

- Smart farming, agriculture, and forestry;
- sustainable urban planning and infrastructure
- evaluation of natural disasters and diseases
- other cutting-edge applications

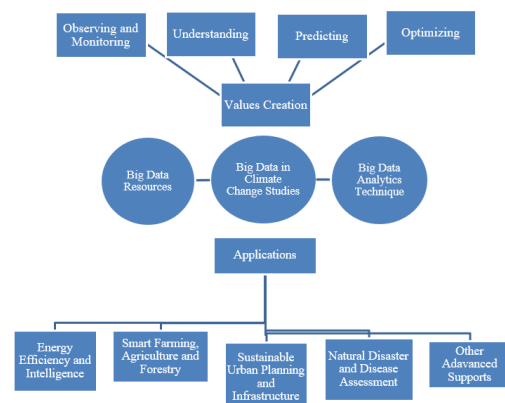


Figure 2- Big data Framework in Climate Change

To Observe and Monitor

The thorough revelation of the truth generated from the enormous amount of data acquired is one of the insights that big data may offer. The primary resource for extensive data analysis is extraordinarily vast data sources, including much helpful information. The climate system must be monitored to understand the system's interactions and driving forces in a better way. Additionally, it is advantageous for us to be aware of potential changes brought on by global warming [3]. Following that, it can be said that big data's essential contribution to climate change research is observation and monitoring.

Over the past few decades, earth observation technology has been vital in obtaining comprehensive observation and climate change parameters [8]. With the aid of a multidimensional, vast data system that has been created and is currently undergoing rapid evolution, we have been able to detect and track changes in a variety of earth and climatic parameters globally. [3]

To Understand

For renowned data researchers to investigate and study even before going on to data mining, climate science's 3Hs feature (high dimension, high complexity, and high uncertainty), which analyses the large world-scale changes of numerous observations/parameters, has made it a fascinating playground. Big data has sped up the development of data-intensive analytics in research pertaining to climate change and produced a multidimensional system for accumulating and tracking climate change data. Here, we clearly group them into the parts of understanding, forecasting, and optimizing.

The wealth of resources mentioned above has made it possible to learn more about big climate data, including what they are, how they are collected, and what they may be used for. These, though, only begin to scrape the surface of big data analytics. When combined with climate science, big data may be used to predict the weather, energy usage, traffic, early warning of natural disasters, and more.

To predict

Although big data's value generation emphasizes optimizing data, also accurate forecasts can help develop climate-adaptive policy. Understanding and making predictions based on enough information from big data or generalizations across many cases or applications is a relatively simple process.

To optimize

The optimal model or infrastructure must be set up to optimize performance, efficiency, and usefulness or, in certain cases, to ensure sustainable development. To do this, optimization calls for a solid theoretical foundation and strong data analytics abilities. Several useful applications for managing use of resource, energy efficiency, the smart grid, smart agriculture, etc., have been implemented recently.

V. RESOURCES OF BIG DATA

Essential Climate Variable (ECV) values are recorded by using the Earth's observation data, which is gathered through analysis of sensors or satellite data and atmospheric discernment (over sea, land, and ice), oceanic and terrestrial elements [8]. A physical, chemical, or biological factor or group of anomalies that determine the Earth's atmosphere and climate is known as an ECV. ECV databases are needed [22] to comprehend and forecast climate change, direct mitigation, adapt actions, and estimate risks.

The 4V characteristics of big data from Earth observation, namely volume, variety, veracity, and velocity, are vital for comprehending climate change [4]. Before the invention of earth observation equipment, scientists did most of their work manually using observational techniques.

Technology advancements have brought about sensors that use a tiny amount of electricity, transmit data instantly, and operate on batteries that can last for extended periods. Similarly, synchronous satellites deliver ongoing information about aerial and ground conditions. The Framework of the UN UNFCCC's Climate Change Protocol provides thirty-four essential climate variables combined (ECVs) that in 2010 call for contributions from Earth observation. Twenty-five of the Global Climate Observing System Important climate factors are constantly observed and simultaneously captured to generate climatic information motifs [9].

VI. TECHNIQUES OF BIG DATA ANALYTICS

To fully utilize the advantages of data, it is crucial to share and disseminate information effectively and to analyze data for insights. By making big data an underlying technology instead of just a for-profit solution, "Where models are found through the algorithmic search process of data exploration," as per [11], big data will become more than just a commercial solution. Including Apache Hadoop, R, Storm, and Massive Online Analysis, citation [12] reviewed and summarized the most popular big data solutions.

Four fundamental subcategories were selected to organize the fundamental approaches [13]: clustering, classification, association rules, and regression. Each tool or platform has advantages and disadvantages.

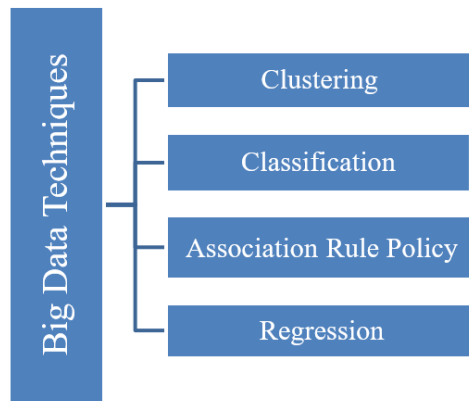


Figure 3- Big Data Key Techniques

With the advent of IoT technology, the digital imprint of our more interconnected world is growing exponentially. Online, people share all sorts of personal details. Along with more people's online presence, social networks and social media are now the newest sources of information hotspots. Because of this, social network analysis is crucial. Analysis of social networks is a method in the study of big data applicable in climate data, given that historical learnings have shown a rise in internet searches for information on flash floods or forest fire floods, and other comparable catastrophes in the afflicted area. Reference [6] has discussed the leading extensive data analysis topics. Approaches in their article include social network analysis, Climate change, and big data.

A. Clustering

Machine learning tasks that are unsupervised include clustering. Using a clustering algorithm entails providing a considerable amount of unlabeled input data and allowing it to find any groupings in the data. The items in those sets form clusters. Classifying data points called clusters is possible based on their similarities and proximity. Cluster analysis has several practical uses, including feature engineering and pattern discovery.

The clustering method classifies the data points as a specific group to distinguish between various items [14]. The distance measure is the fundamental idea behind clustering, encompassing several methods. Other evolutionary techniques follow the list of clustering techniques like Centroid Based Clustering, Hierarchical Clustering, k-means Clustering, Density-based clustering, etc.

B. Classification

Classification is a technique for data mining that organizes unstructured data into classes and groups. It is one of the most straightforward techniques for analyzing massive quantities of unstructured data [15]. Some of the best methods for categorizing data are detailed here. Decision trees, which ask dataset a series of questions to zero in the

right category traits, are the most used method. Support vector machines divide things into two classes to reduce classification errors. The Bayes Rule is used to determine the likelihood of each class. Using the kNN technique, the closest neighbour for the observed object is found and categorized.

- Naïve Bayes Rule Decision Trees
- Support Vector Machine (SVM)
- Decision Trees
- k-Nearest Neighbors (kNN)
- Neural Networks

C. Association Rule

An “association rule mining” technique analyzes data sets from relational databases, transactional databases, and other data repositories to identify recurring patterns, correlations, links, or causal structures.

Association rules employ the if-then clause to show how likely it is that particular pieces of data are contained throughout big data sets in multiple databases. Various association, and rule-mining applications, are frequently employed in value-based or clinical data files to help identify sales correlations.

Association rule mining's fundamental techniques are enumerated below-

- Apriori-based algorithm
- Event-based algorithm
- Parallel methods
- Fuzzy association rule mining
- Genetic algorithm

D. Regression

Regression analysis is a set of statistical methods for examining the association between a dependent (or target) variable and a set of independent (or predictor) variables. It helps determine the degree of association between the variables and projects and how they will interact.

Machine learning and regression analysis both have applications in prediction and forecasting. On the other hand, it is also employed in time series modeling and discovering causal links between various variables.

Some important methods of regression are listed below-

- Regression tree
- Linear - Non-linear regression
- Logistic and Lasso regression.

It can be used in predicting the weather. Hence it is beneficial in predicting climate change; for example- Rainfall prediction using SLR

The Simple Linear Regression Algorithm

The simple linear regression representation will be like this-

$$Y = \alpha + \beta X \quad (1)$$

Where Y= dependent variables, X= independent variable,
 α, β = regression coefficients

Because it will be comparable to a slope-intercept mathematical equation, we will implement Slope and Intercept in basic linear regression. One may assess the strength and direction of the association between the two variables using the regression coefficient formula. Similar to this, statistical and mathematical evolution processing uses additional correlation coefficient formulae.

The mathematical expression for r is-

$$r = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (2)$$

Error Calculation-

After computing the predicted values, the error may be determined, and the difference between the actual and expected values can be determined. The method for figuring out the error is

$$RMSE = \frac{\sqrt{\sum(y_i - \hat{y}_i)^2}}{n}$$

Where Y_i = actual value, \hat{Y}_i = predicted value, n= the total number of samples.

RMSE is known as Root Mean Square Error. It can be used for measuring the difference between sample and population values.

The structure in Figure 4 will assist you in completing the challenge of predicting rainfall in a logical order.

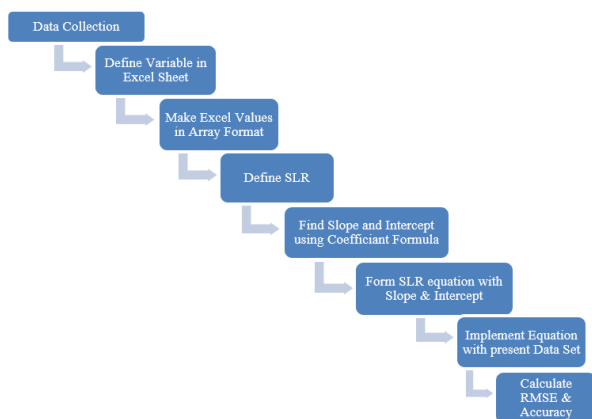


Figure 4- Regression process flow

Step 1: Perform (CC, VP, A, B, X, Y, F1)

Step 2: Gather all the data and enter it into Excel.

Step 3: Open the new file (E:\sample.xls) and read the values from the excel sheet.

Step 4: CC = [j] = (a. obtain contents () in step four.

Step 5: VP = [j] = (b. obtain contents () in step 5.

Step 6: Calculate the correlation coefficient value in step six.

Step 7: B is the correlation coefficient, where $Y = A+Bx$ value.

Step 8: We must determine where $A = Y-Bx$ is. / We can obtain a worth.

Step 9: SLR Equation can be formed using "A" and "B" values.

Step 10: Using the given equation and the selected data set, We can get the anticipated number that is a most closely dependent element

Step 11: We determined the RMSE.

Step 12: Using confusion, we can compute the accuracy. Matrices.

*RSME= Root Square Mean Error

*SLR= Simple Linear Regression

VII. WEATHER PREDICTION- A COMPARATIVE ANALYSIS OF VARIOUS TECHNIQUES

Prediction of Weather

Climate change is a complicated phenomenon with far-reaching ramifications for our planet's future. It is hard to foresee the extent and impact of climate change, and various methodologies have been attempted to study and anticipate its impacts. This comparative study will look at some of the most prevalent approaches for predicting climate change and their advantages and disadvantages.

Climate Models:

Climate models are the most often used methods for predicting climate change. Temperature, precipitation, ocean currents, and even atmospheric gasses are all considered in such models as they replicate the Earth's climate system. The model represents these aspects and forecasts future climate change using mathematical equations.

Strengths: Climate models are comprehensive and can account for a wide range of climate-related elements. They may also be used to run various scenarios and assess the effects of certain policy measures.

Weaknesses: Climate models have limits, and the accuracy of their projections is frequently affected by errors in data and model assumptions. Moreover, processing power and data availability constraints might restrict the complexity of the models.

Paleoclimate Analysis

Paleoclimate analysis studies the Earth's climate history via geological and biological data. This strategy can help academics comprehend how the Earth's climate has evolved over millions of years and what causes have contributed to those changes.

Strengths: Paleoclimate study can give important insights into long-term climate trends and the Earth's reaction to

natural and anthropogenic climatic change. It can also assist researchers in better understanding the mechanisms underlying climate change.

Weaknesses: Paleoclimate analysis has data availability and accuracy restrictions. Geological and biological processes that hide or remove historical climatic signals can impact data quality. Moreover, the data sets may not accurately represent Earth's climatic history.

Big Data:

Big Data allows scientists to construct more complete and accurate climate change models by providing more accurate and extensive information about the climate system.

Strengths: Big Data technologies enable real-time monitoring of climate-related phenomena such as temperature variations, weather patterns, and sea level rise, resulting in early warnings and better resource management. Big Data analysis may assist decision-makers in identifying and evaluating various climate-change mitigation alternatives, such as generating sustainable energy sources, lowering emissions, and adjusting to changing weather patterns.

Weaknesses: The quality of Big Data in climate change may vary based on the information's source and accuracy. Inaccurate analysis and decision-making might result from poor data quality.

Access to Big Data may be restricted, particularly in poorer nations or locations with inadequate infrastructure, limiting the capacity to create comprehensive and reliable climate models

Comparative Analysis of Proposed Model and Previous Models			
Features	Big Data	Climate Models	Paleoclimate Analysis
Prediction	Short-Term	Long-Term	Long-Term
Datasets	Large	Small	Small
Accuracy	High	Low	Low
Complexity	Low	High	High

Figure-5 Comparative Analysis of Proposed Models and Previous Models

VIII. IMPACT OF BIG DATA IN CLIMATE CHANGE: OPPORTUNITY

One of today's most urgent problems is climate change. Its repercussions can be seen everywhere and have already begun by interrupting people's regular lives. Based on data, new insights into the results can be gained using research methods on climate change and encouraging prevention measures. As Earlier, we noted that significant climatic data typically originates from Earth's observational data, four recorded sources, and model results with paleoclimatic data. As of now started, in 2016, around 130 earth observation missions tracked many climate variables operationally. A study of Monitoring and observation of 50 ECVs is required for climatic data, as outlined by the current Global Climate Observing System (54 [9]); The knowledge is utilized to comprehend, foresee, and optimize climate processes to reduce confusion.

A. There is a lot of inevitable big data.

Every digital device and network leaves behind an unavoidable digital trail, which opens up several possibilities for discovering new, more affordable data sources. By using the related characteristics from a study of digital footprints established, traditional data sources like surveys and other types of research can be replaced.

B. Substitutes for Random Sampling

Big Data is an automated record of everything that occurs in the real world; in contrast to conventional data collection methods, it captures values in real-time or at predetermined intervals. Computers can examine all of the data that has been captured in this scenario, sampling $n = \text{universe } N$, and create more precise information without the need for a random sample.

C. Real-time accessibility

Currently, almost everyone has access to a smartphone, laptop, or another connected device, and everyone utilizes at least one social networking site. Due to the prevalence of social media in developing nations and the wealth of context it offers, it is particularly crucial [5]. It is monitored so that the internet trends depending on location, may be seen.

D. Fusion Data

The term "data fusion" refers to merging information from several different sources; data fusion leaves behind a fragmented digital footprint that contains missing values, often compensated by redundant data from other sources [5]. The objective is to acquire more exact, trustworthy, and applicable data than from a data source.

E. Big Data Analytics

Big data includes analyses of gathered data for wise decision-making and increasing the quantity and quality of data. The major paradigm shift is the systematic positioning of data analysis at the center of intelligent decision-making, regardless of the specific petabyte, exabyte, or zettabyte scale [17].

IX. IMPACT OF BIG DATA IN CLIMATE CHANGE: CHALLENGES

The two major categories of big climate data can be broadly classified into two parts, data challenges, method challenges, with technique challenges [18]. The 4Vs of big data, or the size, speed, diversity, and quality issues, are referred to as "data challenges," whereas "method challenges" refers to the difficulties encountered when studying data, or more specifically, the constraints of the data analysis methodologies. Technology problems are another aspect that requires investigation; even perfect technology is susceptible to the effects of time. Research on climate change is still in its infancy, and additional technical

advancements are anticipated and awaited to limit its consequences.

A. Data Challenges

Naturally, any data science expedition is propelled by the data, which goes without saying. In fact, before deploying their arsenal of data mining tools, most big data practitioners perform exploratory data analysis to comprehend the diversity and complexity of the data. This is particularly important in climate science, where data scientists are mostly excluded from data collection. There are four different types of climate data sources: in situ, remote sensing, model output, and paleoclimatic. Before beginning any data-driven research, it is essential to understand the advantages, limits, and suitable usage of each data source. See Faghmous and Kumar (2013) [7] for a more thorough overview of climate data sources.

- *Data That Is Constantly Changing*

The proliferation of sensors and data storage obtained from those sensors on the cloud has resulted in a plethora of data. However, climate data scientists use the observation system in a state of perpetual change, which presents a considerable challenge [19]. Due to the high volatility of climate variables, it is necessary to regularly upgrade and replace older, less effective programs with newer, more effective ones to acquire more accurate readings.

- *Empathy in Data*

Data empathy is the study of a dataset's origins to inform the selection of the most appropriate analysis for that dataset's research. Exploring data collection, methodology, and data analysis aid in detecting the inherent biases of the approach and allows you to change the outcomes.

- *Representation of Data*

Most data science learning algorithms require attribute value inputs, but many climatic events cannot be quantified [20]. It is crucial to remember that most climate phenomena cannot be represented in binary values because they often occur gradually and are recorded in spatiotemporal data sets.

- *Short Observational Record*

Big Climate Data availability is one of the main issues. Before the 1970s, most climate measurements were made using observational methods. Since then, satellite-based climate monitoring has become more and more common. However, because of recent developments in big data and cloud storage, many datasets of climatic variables are already roughly 20 years old. This challenge can present problems since data with high spatial resolution data spanning a wide area. Still, low temporal resolution or data only covering a brief period cannot be used for spatiotemporal analysis. Paleoclimatic information is gathered via creative techniques such as examining tree rings, corals, and ice cores.

B. Method Challenges

Although the ultimate goal of analyzing climate data is the construction of reliable and accurate prediction models, the current emphasis is on understanding rather than forecasting. The spatial and temporal dimensions of climate data present unique challenges. The two most common types of data analysis are regression and categorization.

- *Sample Bias*

When a dataset that is not randomly selected from the population of interest is used, it introduces bias into the sample, also known as sampling bias. If the model is constructed with such a data set, it will provide an incorrect hypothesis. To build a reliable data analysis model, it is necessary to use samples from each population.

- *Challenges in Problem Definition*

When studying climate data, it might be challenging to identify the goal function because the analysis is exploratory.

Given the complexity of earth systems, various factors may be to blame for a given climate phenomenon. It is also feasible to research using the same climatic dataset and reach different conclusions [18].

- *Challenges in Evaluation*

The goal of the data-driven approaches to climate research is to maximize objective data metrics that can be used to evaluate the efficacy of a given algorithm. Such metrics include revenue, churn rate, and incorrect classification. The uncertainty of how much a change in a single parameter will affect the climatic phenomenon can cause concern when translating model performance.

C. Technological Challenges

Even while our knowledge of climatic systems constantly expands, specific gaps remain. Big data has challenges, including many real-time, high-speed data transfers, diverse data, and data security [21], as well as cloud storage and quick access to that data.

Data scientists are learning more about the intricate processes that take place on earth, thanks to artificial intelligence (AI)-based algorithms based on environmental changes. AI is embracing and supporting this study.

X. CONCLUSION

The main contributors to climate change are rapid urbanization, excessive exploitation of natural resources, and other human development endeavors. One of the most significant problems of our day is climate change, which imperils human, plant, and animal life. The ecological balance has been impacted, and some species have been driven to extinction due to rising water and air pollution, declining forest cover, and habitat degradation. Although it

is still in its early stages, climate data analysis is helping to contain the adverse effects of climate change.

Climate science's threat to our world makes it one of our era's most significant problems. Additionally, it offers a massive opportunity for considerable data research because climate science poses new problems for data, methodologies, and evaluation. Big data analytics, by themselves, however, are insufficient to reliably and insightfully examine climate data. Data science techniques that combine the strength of big data analytics with the prudence of scientific theory and fundamental principles are required.

Future climate difficulties may be predicted using predictive analytics models, and knowing when a climate disaster will occur can assist plan for mitigation. Nevertheless, the benefits of big data research also have their own unique set of drawbacks. Advances in data collecting and processing techniques are assisting data scientists in creating more precise models. Although many ongoing studies have already materialized in mitigation measures, we still have a long way to go until we fully understand how the climate works.

REFERENCES

- [1] Faghmous JH, Kumar V. A big data guide to understanding climate change: The Case for Theory-Guided Data Science. 155–163. [10.1089/big.2014.0026](https://doi.org/10.1089/big.2014.0026).
- [2] Hassani, H.; Huang, X.; Silva, E. Digitalisation and Big Data Mining in Banking. *Big Data Cogn. Comput.* 2018,2, 18.
- [3] Hassani, H.; Huang, X.; Silva, E. Big-Crypto: Big Data, Blockchain, and Cryptocurrency. *Big Data Cogn. Comput.* 2018,2, 34.
- [4] Götz, M.; Richerzhagen, M.; Bodenstein, C.; Cavallaro, G.; Glock, P.; Riedel, M.; Benediktsson, J.A. On scalable data mining techniques for earth science. *Procedia Comput. Sci.* 2015,51, 2188–2197.
- [5] Pang-Ning, T.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Pearson: London, UK, 2014.
- [6] Hilbert, M. (2016), *Big Data for Development: A Review of Promises and Challenges*. *Dev Policy Rev*, 34: Pages 135-174
- [7] H. Hassani, X. Huang, and E. Silva, "Big Data and Climate Change," *Big Data and Cognitive Computing*, vol. 3, no. 1, p. 12, Feb. 2019.
- [8] Faghmous JH, Kumar V. Spatio-temporal data mining for climate data: Advances, challenges, and opportunities. In Chu W, editor., Ed., *Data mining and knowledge discovery for big data*. Springer, Heidelberg, Germany, (2013), pp. 83–116.
- [9] Mangal, P.; Rajesh, A (2020), *Addressing Climate Change – Making a case for Big Data based Decision Making Framework*, IJSPM.
- [10] OECD (2016), *Space and Innovation*, OECD Publishing, Paris. DOI: <https://dx.doi.org/10.1787/9789264264014-en>, pages 95-99.
- [11] Bojinski, Stephan & Verstraete, Michel & Peterson, Thomas & Richter, Carolin & Simmons, Adrian & Zemp, M. (2014). The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bulletin of the American Meteorological Society*. 95. 1431. [10.1175/BAMS-D-13-00047.1](https://doi.org/10.1175/BAMS-D-13-00047.1)
- [12] Hua-Dong Guo, Li Zhang, Lan-Wei Zhu. Earth observation big data for climate change research. *Advances in Climate Change Research*, Volume 6, Issue 2, June 2015, Pages 108-117.
- [13] Global Climate Observing System (2020), <https://public.wmo.int/en/programmes/global-climateobservingssystem?name=EssentialClimateVariables>.
- [14] Hassani, H.; Saporta, G.; Silva, E.S. Data mining and official statistics: The past, the present, and the future. *Big Data* (2014), 2, 34–43
- [15] Fan, W; Bifet, A. Mining big data: Current status, and forecast to the future. *ACM SIGKDD Explore. News.* (2013), 14, 1–5.
- [16] Chen, C.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques, and technologies: A survey on Big Data. *Inf. Sci.* (2014), 275, 314–347.
- [17] Kotsiantis, S.B.; Zaharakis, I.; Pinellas, P. Supervised machine learning: A review of classification techniques, *Emerg. Artif. Intell. Appl. Comput. Eng.* (2017), 160, 3–24.
- [18] Hilbert, Martin. (2016). *Big Data for Development: A Review of Promises and Challenges*. *Development Policy Review*. 34. 135-174. [10.1111/dpr.12142](https://doi.org/10.1111/dpr.12142).
- [19] James H. Faghmous, Vipin Kumar, A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science (2014), *Big Data*. Sep2014.155-163. <http://doi.org/10.1089/big.2014.0026>.
- [20] Trenberth K.E. et al. (2013) Challenges of a Sustained Climate Observing System. In: Asrar G., Hurrell J. (eds) *Climate Science for Serving Society*. Springer, Dordrecht.
- [21] Langley P. The changing science of machine learning. *Mach Learn* 2011; 82:275–279.
- [22] Khan, Nawsher & Yaqoob, Ibrar & Hashem, Ibrahim & Inayat, Zakira & Kamaleldin, Waleed & Alam, Muhammad & Shiraz, Muhammad & Gani, Abdullah. (2014). *Big Data: Survey, Technologies, Opportunities, and Challenges*. *The Scientific World Journal*. 2014. 18. [10.1155/2014/712826](https://doi.org/10.1155/2014/712826)