© May 2022| IJIRT | Volume 8 Issue 12 | ISSN: 2349-6002

Climate Extremes: Observations, Visualization, Prediction and Impacts

Giriraj Patil[1], Basvapriya Kodge[2], Kiran Kanasnale[3], Konapur Anand[4], Krishnakant Sonji[5]
[1,2,3,4,5] *Guru Nanak Dev Engineering College, Bidar-585128*

***Abstract*—Climate science as a data-intensive subject has been overwhelmingly affected by the era of data science and relevant technological revolutions. The big successes of data science analytics in diverse areas over the past decade have also prompted the expectation of data science and its efficacy on the big problem—climate change. This discrepancy stems from the complex nature of climate data as well as the scientific questions climate science brings forth. This article introduces a data science audience to the challenges and opportunities to mine large climate datasets. climate change has been at the forefront of the big climate data analytics implementations and exhaustive research has been carried out covering a variety of topics. This paper aims to present an outlook of data science in climate change studies over the recent years by investigating and summarizing the current status of data science applications in climate change-related studies. It is also expected to serve as a one-stop reference directory for researchers and stakeholders with an overview of this trending subject at a glance, which can be useful in guiding future research and improvements in the exploitation of climate data.**

I. INTRODUCTION

Data science has a key role to play in climate change. The science, policy, and communication practices around data science, machine learning, and artificial intelligence have important implications for the climate crisis and the solutions society will utilize in the future. From machine learning to data visualization, data science techniques are used to study the effects of climate change on marine biology, land use and restoration, food systems, patterns of change in vector-borne diseases, and other climate-related issues. Data science is a powerful tool to help researchers understand the uncertainties and ambiguities inherent in data, identify interventions, strategies, and solutions that realize co-benefits for humanity and theenvironment, and evaluate the multiple–and sometimes conflicting–goals of decision-makers.

This article discusses some of the major big data challenges researchers face when mining climate data and how being mindful of such intricacies can have a significant impact on a socially relevant and commercially viable domain. We will use examples from existing research in climate and data science to demonstrate and discuss key concepts, with the goal of preparing a new generation of data scientists with the tools and processes for data science to have the highest impact on momentous challenges facing our society due to climate change.

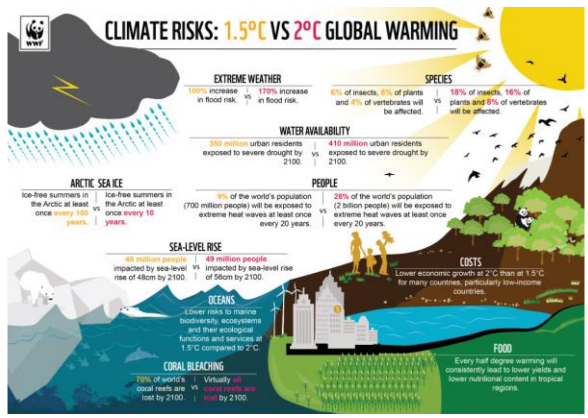

Figure. (1.1) Extreme Climate Changes between 1.5℃ VS 2℃

II. CHALLENGES

- Traditionally, many data-driven explorations tend to boil down to regression or classification. However, the focus of climate science is more about understanding than predicting. Yes, there is interest in predicting individual events, but a greater focus remains on a system-level view rather than individual events.

- This difference comes with two consequences: First, there is a greater emphasis on model interpretability rather than model flexibility. Thus, ‘‘black box’’ models that tend to do very well in regression or classification tasks are not adopted because their accuracy cannot be related back to physical processes. Second, existing data science conventions have focused on a narrow form of data representation that does not abstract spatiotemporal phenomena. This means that the way the data are represented relies heavily on attribute-value representation without any notion of contextual information (in space or time).

- This representation was reinforced by the popularity of the UC Irvine Machine Learning repository—a set of small public datasets that researchers are developing novel learning algorithms could use to test their methods. However, not all problems lend themselves to a set of attributes that are generally assumed to be independent. Finally, the spatiotemporal nature of the data presents unique challenges.

III. PROBLEM STATEMENT

- Climate change is a heated topic. It’s on everyone’s mind. Unfortunately, the global warming controversy still rages on--especially in politics--and this is why data visualization is one of our best allies in this debate. We’ve scoured the Web to find the best visualizations on climate change facts from the last few years. So, our main objective here, considering all the factors is to find how exactly will earth look like, its resources, and other domains in the next coming years. A great advantage traditional data science has is the clear definition of learning tasks (regression, classification, etc.).
- In climate science, however, the objective function can be harder to define. For example, one of the most feared impacts of global climate change is drought. Yet, the very notion of drought is ambiguous from a data-driven perspective. First, there are numerous types of droughts: agricultural, metrological, and hydrological. However, in general, droughts may be defined as ‘‘the prolonged absence, or marked deficiency, of precipitation.’’ Second, even if we can agree on a definition, how to represent such ‘‘deficiency’’ is unclear. For instance, droughts may be quantified in absolute or relative terms, and depending on which data representation is chosen, one might arrive at different conclusions. Two recent studies published in Nature Climate Change and Nature came to opposing conclusions about whether there are noticeable changes in drought trends as a result of climate change.
- One study further highlighted the disparate results by pointing out (among other reasons) that each study relied on a different precipitation dataset. There are other instances where two studies used the same data but arrived at opposing conclusions. When examining the changes in sea surface temperatures and hurricane occurrences, one study found that the observed hurricane trends were just part of larger oscillations, whereas another study concluded that hurricane counts were increasing along with sea surface temperatures. Both groups used the same data but instead relied on different data analysis techniques to arrive at opposing conclusions.

IV. DATASET DESCRIPTION

We have repackaged the data from a newer compiled by the Berkeley Earth, which is affiliated with Lawrence Berkeley National Laboratory.

In this dataset, we have included several files:

- Global Land and Ocean-and-Land Temperatures (GlobalTemperatures.csv)
- Date: starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures
- LandAverageTemperature: global average land temperature in Celsius
- LandAverageTemperatureUncertainty: the 95% confidence interval around the average
- LandMaxTemperature: global average maximum land temperature in Celsius
- LandMaxTemperatureUncertainty: the 95% confidence interval around the maximum land temperature
- LandMinTemperature: global average minimum land temperature in Celsius
- LandMinTemperatureUncertainty: the 95% confidence interval around the minimum land temperature
- LandAndOceanAverageTemperature: global average land and ocean temperature in Celsius

Other files include:

- Global Average Land Temperature by Country (GlobalLandTemperaturesByCountry.csv)
- Global Average Land Temperature by State (GlobalLandTemperaturesByState.csv)
- Global Land Temperatures by Major City (GlobalLandTemperaturesByMajorCity.csv)
- Global Land Temperatures by City (GlobalLandTemperaturesByCity.csv)

V. DESIGN

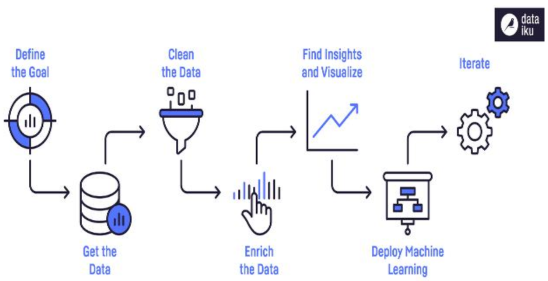

Figure. (1.2) Data Science Lifecycle

VI. IMPLEMENTATION

Machine Learning algorithms can be used to monitor potentially catastrophic weather events, and predict damages.

To compute precise predictions regarding the future climate we will use regression algorithms: -

1. **Linear Regression** - Simple linear regression is useful for finding relationships between two continuous variables. One is a predictor or independent variable and the other is a response or dependent variable.
2. **Polynomial Regression** - In simple words, we can say that if data is not distributed linearly, instead it is an nth degree of polynomial then we use polynomial regression to get desired output.

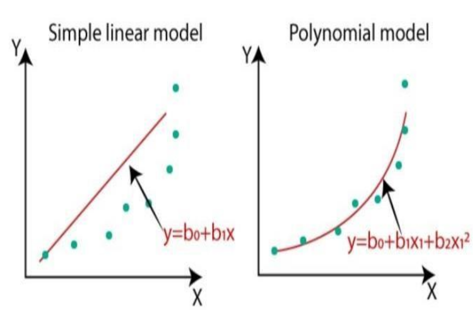

Figure. (1.3) Linear Regression & Polynomial Regression

VIII. DISCUSSION

This paper described the essential need for research and development objectives to realize and manage the complex issues of climate change through Data science tools. Data-driven applications were reviewed through the co-occurrence analysis of keywords, which showed the widespread application of Data Science technologies and tools, however, comprehensively utilized and integrative analyses are less prevalent.

It can be claimed that the exclusive analysis of climatic factors cannot bring about sufficient strategic adaptation by itself, rather the socio-environmental factors must be integrated into the climate change models. Mitigating the impacts of climate change and successful adaptation requires effective climate change strategic planning by countries worldwide whose decision-making requires complex models and sources of information. The Data Science toolkit enables the systematization, processing, and evaluation of heterogeneous data and information sources, which is unfeasible with traditional disciplinary analysis tools. The harmonization of the ever-expanding scientific knowledge and diversified data sources related to climate change may be one of the most urgent tasks for researchers in the future. This research presented Data Science analytics tools and their contribution toward exploring the characteristics of climate change as well as climate action-related counterparts such as sustainability and social sciences that are essential for the successful development and implementation of strategies.

IX. CONCLUSION

Researchers have generally realized the big impact of data-intensive research. have summarized the milestones of Big Earth Data development and the big challenges that follow. In this paper, we present the most up-to-date overview of big data analytics in climate change covering over 100 research papers. Although there is a lot of existing literature on big data and smart cities, the sustainable aspect that links to climate change still stays in the early stage of its development.

However, big data analytics alone are not enough to insightfully and accurately exploreclimate data.There is a need for theory-guided data science methods that blend the power of big data analytics with the caution of scientific theory and first principles.

It is also worth noting the popularity of integrating the Internet of Things, the architecture of a platform that enables efficient, real-time, in memory cloud computing and storage of complex big data with the most advanced analytics or data mining techniques, as well as integrating the Internet of Things is the key to future research.

REFERENCES

[1] https://www.nasa.gov/audience/forstudents/k-4/stories/nasa-knows/what-is-climate-change-k4.html

[2] https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data/code?datasetId=29

[3] Chang EKM, Guo Y. Is the number of North Atlantictropical cyclones significantly underestimated prior to the availability of satellite observations? Geophys Res Lett 2007; 34:5.

[4] Holland GJ, Webster PJ. Heightened tropical cyclone activity in the North Atlantic: Natural variability or climate trend? Phil Trans R Soc A 2007; 365:2695–2716.

[5] Schneider DP, Deser C, Fasullo J, Trenberth KE. Climate data guide spurs discovery and understanding. Eos Trans Am Geophy Union 2013; 94:121–122.

[6] Ginsberg J, et al. Detecting influenza epidemics using search engine query data. Nature 2009; 457:1012– 1014.

[7] Lazer D, Kennedy R, King G, Vespignani A. The parableof Google flu: Traps in big data analysis. Science 2014; 343:1203–1205.

[8] https://medium.com/@stallonejacob/data-science-scaling-of-data-in-python-ec7ad220b339#:~:text=Scaling%20data%20is%20the%20process%20of%20increasing%20or,but%20it%20definitely%20is%20a%20good%20practice%20.

[9] https://datagy.io/sklearn-train-test-split/

[10] https://medium.com/mlearning-ai/train-and-test-data-split-152bad0afbb2

[11] https://www.sciencedirect.com/topics/computer-science/validation-dataset

[12] https://www.qlik.com/us/data-analytics/data-insights

[13] https://dotnettutorials.net/lesson/data-pre processing-in-data-science/

[14] https://www.geeksforgeeks.org/python-plotly-tutorial/

[15] https://www.datascience-pm.com/data-science-process/

[16] https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372 b4 92#:~:text=5%20Steps%20of%20a%20Data%20Science%20Project%20Lifecycle.,4.%20Model%20Data.%205%205.%20Interpreting%20Data.%20

[17] https://plotly.com/

[18] https://climate.nasa.gov/news/2743/the-scientific-method-and-climate-change-how-scientists-know/

[19] https://css.umich.edu/factsheets/climate-change-science-and-impacts-factsheet

[20] https://datascience.columbia.edu/research/focus/climate/#:~:text=%20Data%20science%20has%20a%20key%20role%20to,the%20solutions%20society%20will%20utilize%20in%20the%20future.

[21] https://www.wired.com/sponsored/story/data-science-on-the-front-lines-of-battling-climate-change/

[22] https://iq.opengenus.org/polynomial-regression-using-scikit-learn/

[23] https://joblib.readthedocs.io/en/latest/

[24] https://seaborn.pydata.org/

[25] https://science2017.globalchange.gov/

[26] https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=2046&context=gs_rp