Utilizing big data in weather forecasting techniques

Theuri Patrick[a], David Wachapele, Patrick Gikunda and Samuel Muchina

[a]School of Computer Science and Information Technology, Dedan Kimathi University of Technology, Private Bag, Dedan Kimathi, 10143, Kenya

E-mail: theuri.patrick@dkut.ac.ke, david.wachepele@dkut.ac.ke, patrick.gikunda@dkut.ac.ke, samuel.muchina@dkut.ac.ke

Abstract

Today, several parts of the society leverage Big Data to transform unstructured data into actionable information. In the field of weather forecasting, results from Big Data analytics can provide weather forecasters with valuable insights regarding past, present, and future weather patterns. The purpose of this review is to explore the use of big data in building sustainable weather forecasting. A total of 12 peer-reviewed articles were examined. The findings suggest that the use of Big Data in weather forecasting is partly motivated by the fact that an abundance of remote sensing and satellite devices are continually monitoring the weather system and collecting huge amounts of atmosphere data all over the world. With more data available, meteorologists are now able to predict weather in much better ways than ever and with more accuracy. However, problems related to Big Data itself (volume, variety, velocity, veracity and value), the complexity of hardware and software involved, data detection challenges, architecture and algorithm, and issues with estimating statistical significance impact the use of Big Data in predicting the weather. Additional research is recommended to help meteorologists to overcome the challenges linked to using Big Data tools, applications, and technologies in forecasting.

Keywords: Big Data, remote sensing

Introduction

Severe weather conditions have been known to present devastating situations around the globe. Floods, typhoon, hurricanes, and big storms not only cause economic deaths, but also deaths and other calamities. According to Vaishnavi, Rizwan, and Dechakka (2020), natural disasters due to unpredicted, sudden changes in weather conditions impact more than a third of the world's population. For this reason, weather forecasting has become an indispensable and important procedure in the daily lives of people. Traditionally, human forecasters were at the center of the evaluation of the alteration occurring in the atmospheric conditions around the world. However, with the modern advancements in technology, it is possible to analyze weather and climate data to extract hidden patterns in the conditions of the atmosphere. As indicated by Fakherldin et al. (2019), this transition from slow, inaccurate weather predictions to hyper real-time forecasting methods is driven by the Big Data revolution with its use of supercomputers, massive data storage, and comprehensive analytic techniques.

Purpose & Objectives

The purpose of this review is to explore the use of big data in building sustainable weather forecasting. Thus, the objectives of the research are as follows:

1. To discuss the significances of using Big Data in modern weather forecasting
2. To explain some of the Big Data software technologies/applications used in weather forecasting.
3. To investigate the challenges associated with using Big Data in weather forecasting.

Research Questions

In line with the above objectives, this paper seeks to answer three research questions.

1. What are the significances of using Big Data in modern weather forecasting?
2. What are the key tools, applications, or software technologies used in Big Data-based weather forecasting?
3. What are the challenges associated with using Big Data in weather forecasting?

Justification

This work, therefore, is important as it evaluates several existing studies on the subject to uncover what research show about the significances of using Big Data in modern weather forecasting. Additionally, this review improves the understanding of existing studies and debates regarding Big Data technologies and how they aid weather forecasting efforts. Moreover, while research is not limited on the use of Big Data in weather forecasting, only a few studies have explored the challenges associated with Big Data when utilized in predicting the weather. In that regard, the present review contributes to the existing literature works on the top by summarizing key findings regarding the problems of using Big Data tools in forecasting.

Previous Works

Big Data plays significant roles in modern weather forecasting. Fathi et al. (2021) notes that results from Big Data analytics can provide weather forecasters with valuable insights regarding past, present, and future weather patterns. Consequently, Big Data is rapidly gaining popularity throughout the world in the context of weather forecasting and number of studies have investigated the use of big data in sustainable weather forecasting in the modern world. A study by Aljawarneh and Lara Torralbo (2021) illustrated that the central idea of existing literature works on meteorological big data is the roles Big Data plays in weather forecasting. Additionally, research on the utilization of big data in building sustainable weather forecasting have focused on meteorological technologies; specifically targeting the use of AI-powered applications and IT-based tools and software programs in environment change predictions as well as opportunities and challenges presented by modern Big Data analytics techniques (Ramya, Balaji & Girish, 2015; Fakherldin et al., 2019; Sahasrabuddhe & Jamsandekar, 2015).

Previous studies (Udayashankara, Murthy & Madhukar, 2016; Vaishnavi, Rizwan, & Dechakka, 2020) show that Big Data has found its way in weather forecasting partly because the complexity and size of weather patterns make it extremely challenging to forecast. Vaishnavi and his colleagues argue that massive volumes of data have to be generated or collected and analyzed to predict the weather. For Udayashankara, Murthy, and Madhukar (2016), this complexity means that utilizing weather data for forecasting involves a complex process especially considering that

numerous changeable parameters need to be carefully considered. Weather data are also disorganized and their parameters vary based on the weather conditions.

The systematic review by Fathi et al. (2021) and the study by Ramya, Balaji, and Girish (2015) both suggest that sustainable prediction of weather patterns involves the analysis of a large set of data (with numerous changing parameters) at an instantaneous rate. Thus, to make accurate and timely predictions, forecasting techniques and instrument must be refined and this is where the use of Big Data becomes vital. Sahasrabuddhe and Jamsandekar (2015) indicate that Big Data was introduced in weather forecasting as disorganized, heterogeneous, and enormous digital data that, if analyzed accurately, can enhance the prediction parameters such as accuracy and precision.

Sahasrabuddhe and Jamsandekar (2015) introduced weather forecasting, including basic different approaches and processes as well as how these aspects leverage Big Data. The authors documented that although prediction methods and processes have evolved over centuries, issues of lack of accuracy (or precision required to support historical weather data) remained significant for long-run forecasts. Big Data, platforms, however, promise to enhance the accuracy of forecasting. These findings are supported by Basvanth and Patil (2016) who explained the use Big Data to model how the weather is likely to unfold and to create a mathematical that describes the physics of the atmosphere. Subsequently, both Sahasrabuddhe and Jamsandekar (2015) as well as Basvanth and Patil (2016) agree that predicting weather patterns existing weather parameters and with preceding data for predictive interpretation commands a high level-computing algorithm in combination with agile query languages utilizing high cache memory and spatial data.

Many studies also suggest that building a sustainable weather forecasting through the use of Big Data involves highly advanced data tools and technologies. Fathi et al. (2021) and Fakherldin et al. (2019) show that conventional weather forecasting methods and infrastructure can no longer handle Big Data. One explanation for this is that sustainable forecasting extends beyond the standard meteorological parameters such as humidity, temperature, wind speed and direction (in terms of height and speed), precipitation, air density, and atmospheric pressure to more specific parameters; wave height, water level, and tide prediction (Selbesoglu, 2020). Thus, the utilization of Big Data in weather forecasting have required the adoption of new methods and techniques designed for handling, storing, analyzing, and interpreting Big Data.

Fathi et al. (2021) introduce five critical big data tools that are widely used by weather forecasters: Hadoop, Apache Spark, Apache Storm, Apache Mahout, and Apache Kafka. Of these technologies, Hadoop stands out as the most important tool for meteorologists' analysis of Big Data in weather forecasting. Basvanth and Patil (2016) as well as Fakherldin et al. (2019) investigated the use of Hadoop Map Reduce software in Big Data-based weather prediction. The latter describes Hadoop as an apache product for handling huge datasets in a distributed environment. According to the authors, Hadoop (which integrates Map Reduce) offers great benefits over fault-tolerant and scalable distributed processing technologies. Basvanth and Patil (2016), on the other hand, established that Map reduce, as a platform for highly distributed and parallel systems across huge datasets. Weather forecasters use this tool to analyze specific types of weather data and produce outputs that give insights into emerging weather conditions.

Fakherldin et al. (2019), similar to Ramya, Balaji, and Girish (2015), further explains that Hadoop and Map Reduce frameworks speed up the processing of huge data while Basvanth and Patil emphasize that Map reduce with Hadoop remove scalability bottleneck in weather forecasting.

Clearly, existing literature on this topic agree that Big Data has the potential to thoroughly reveal weather realities from huge chunks of recorded and processed data (Jain & Jain, 2017; Fathi et al., 2019; Aljawarneh & Lara Torralbo, 2021). An outstanding benefit of incorporating Big Data in prediction processes is that the exceptionally large volumes of data contain valuable information and is at the same time, critical assets in analyzing weather patterns (Hassani, Huang & Silva, 2019). Therefore, monitoring and observing the weather can be considered as the fundamental value that Big Data brings along when it is incorporated with climate change study. Overall, with these tools, Big Data can delivery forecasting accuracy in all most circumstances (Hassani, Huang & Silva, 2019).

However, the use of Big Data is weather forecasting has been linked to some challenges. Hassani and his colleagues, in their Big Data analytics studies in the context of weather forecasting, documented some of the problems that come along with the use of Big Data is predicting changes in weather patterns. According to the authors, forecasting with Big Data is associated with challenges linked to problems related to Big Data itself (volume, variety, velocity, veracity and value), the complexity of hardware and software, data detection, architecture and algorithm, and statistical significance. The problems linked to Big Data (also called 3Vs of Big Data) derive from the definition of Big Data. For instance, Taylor (2022) viewed Big Data as a collection of data that is huge in volume and complex in nature; yet growing exponentially with time. This means that Big Data cannot be managed (stored, analyzed, transferred, and processed) efficiently using traditional data management tools. Additionally, their paper discusses issues with the extant applications of data mining and statistical techniques for forecasting. Basvanth and Patil (2016), on their part, documented that Big Data often comprises datasets whose sizes (typically ranging from terabytes to many petabytes) exceed the ability of commonly used software tools to curate, capture, process, and manage weather data.

Further discussions of the problems associated with utilizing Big Data in weather prediction are provided by Fathi et al. (2021). In the literature review section of their paper, Fathi and his colleagues indicate that it is often difficult to use conventional data management techniques in processing Big Data. They suggest that convenient a data mining algorithm and high-performance platforms are required to analyze Big Data efficiently. Similar observations are presented by Jain and Jain (2017) who noted that high-tech hardware and software are required to process weather data (such as datasets of sunlight intensity, rainfall, humidity, atmospheric, air pressure, and radiation).

Although more data can be translated into improved predictions, the utilization of Big Data in weather forecasting is a new, yet rapidly evolving concept. In addition to the dynamic nature of Big Data, the complexity of weather patterns makes predicting changes in the weather a great challenge even if more data are available. As illustrated in this review, previous studies have shown how the presence of Big Data can help meteorologists to use data in describing the physics of the

atmosphere. However, additional research is recommended to help meteorologists to overcome the challenges linked to using Big Data tools, applications, and technologies in forecasting.

References

Aljawarneh, S., & Lara Torralbo, J. A. (2021, April). Meteorological forecasting based on big data analysis. In *International Conference on Data Science, E-learning and Information Systems 2021* (pp. 9-11).

Basvanth, R., & Patil, B. A. (2016). Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique. *International Journal of Advanced Research in Computer & Communication Engineering*, *5*(6), 1-6.

Fathi, M., Haghi Kashani, M., Jameii, S. M., & Mahdipour, E. (2021). Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering*, 1-29. https://e-tarjome.com/storage/panel/fileuploads/2022-01-15/1642249339_E15948.pdf

Fakherldin, M. A. I., Adam, K., Bakar, N. A. A., & Majid, M. A. (2019). Weather data analysis using Hadoop: applications and challenges. In *IOP Conference Series: Materials Science and Engineering* (Vol. 551, No. 1, p. 012044). IOP Publishing.

Hassani, H., & Silva, E. S. (2015). Forecasting with big data: A review. *Annals of Data Science*, *2*(1), 5-19.

Hassani, H., Huang, X., & Silva, E. (2019). Big data and climate change. *Big Data and Cognitive Computing*, *3*(1), 12. https://www.mdpi.com/2504-2289/3/1/12/pdf

Jain, H., & Jain, R. (2017, March). Big data in weather forecasting: Applications and challenges. In *2017 International conference on big data analytics and computational intelligence (ICBDAC)* (pp. 138-142). IEEE.

Ramya, M. G. P. G., Balaji, C., & Girish, L. (2015). Environment change prediction to adapt climate-smart agriculture using big data analytics. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *4*(5).

Selbesoglu, M. O. (2020). Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data. *Engineering science and technology, an international journal*, *23*(5), 967-972.

Sahasrabuddhe, D. V., & Jamsandekar, P. (2015). Data structure for representation of big data of weather forecasting: a review. *International Journal of Computer Science Trends and Technology (IJCST)*, *3*(6), 48-56.

Taylor, D. (2022). What is Big Data? Introduction, Types, Characteristics, Examples. https://www.guru99.com/what-is-big-data.html

Udayashankara, T. H., Murthy, B. S., & Madhukar, M. (2016). Impact of climate change on rainfall pattern and reservoir level. *Journal of Water Resource Engineering and Management*, *3*(1), 10-14.

Vaishnavi, P. N, Rizwan, M.A, Dechakka, M.P. (2020). Big Data in Weather Forecasting. *International Journal of Engineering Research & Technology* (IJERT). https://www.ijert.org/big-data-in-weather-forecasting