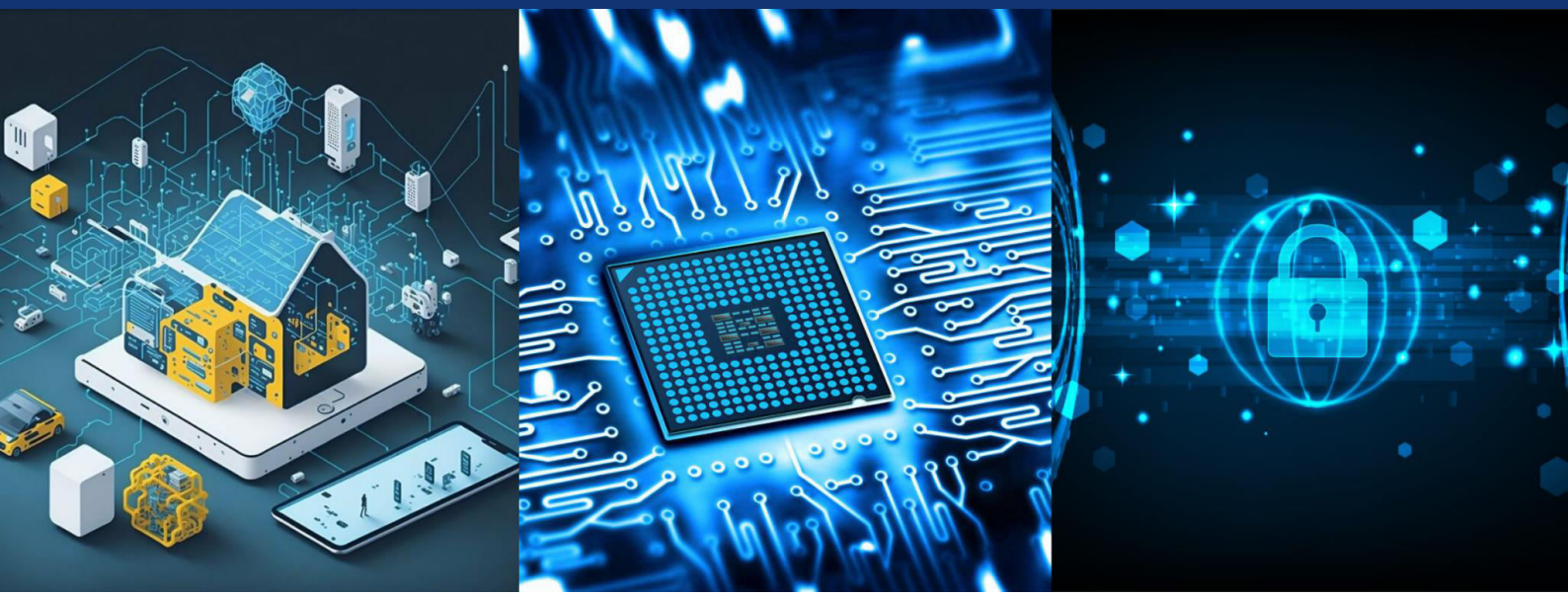
IJIRCCE

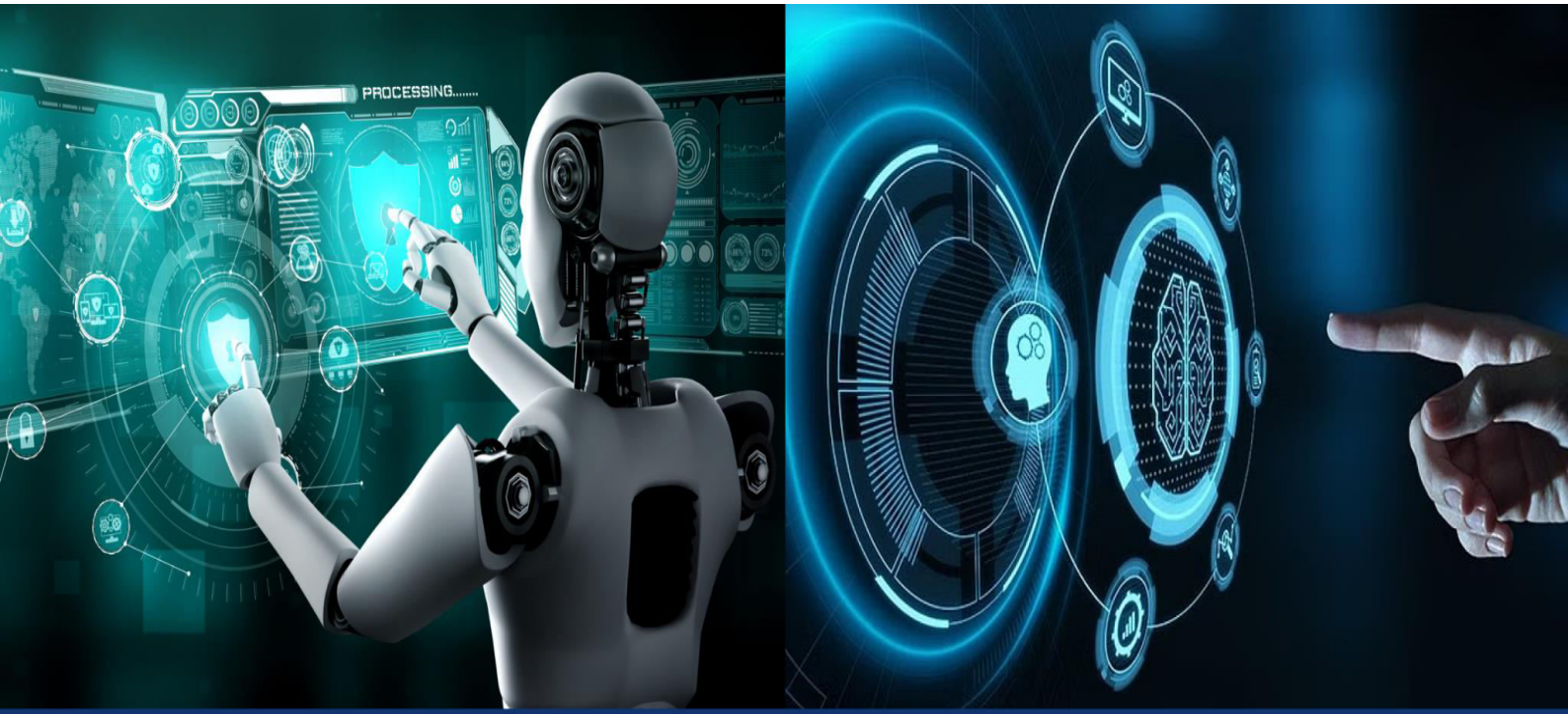
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Weather Forecasting using Big Data Map Reduce Prediction Approach

M. Susmitha[1], G.Vagdevi[2], M.Nagapranathi[3], K.Lakshmankumar[4], Dr. V. Sreenivas[5]

UG Student, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[1-4]

Professor, Dept. of CSE-DS, SRK Institute of Technology, Enikepadu, Vijayawada, Andhra Pradesh, India[5]

ABSTRACT: Weather forecasting using plays a vital role in daily routine, businesses and their decisions. The process of weather forecasting is developing as the effect of advancement in technology right from the realization of increasing size of data, Weather forecasting was found to be based on big data. The researchers have taken review with the objective to study the current forecasting process and methods, and the need of a data structure is recognized for handling the weather data, which is bigger in size, used for the process of weather forecasting. This paper presents a big data analysis framework for weather data set based on Map Reduce Algorithm, and offers not only weather data set analysis, but also various analytic capabilities on huge amounts of data. However, this work establishes a guideline for researchers and industrial practitioners on how to analysis big data. Weather forecasting, as an important and indispensable procedure in people's daily lives, evaluates the alteration happening in the current condition of the atmosphere. Big data analytic is the process of analyzing big data to extract the concealed patterns and applicable information that can yield better results. Nowadays, several parts of society are interested in big data, and the meteorological institute is not excluded. Therefore, big data analytic will give better results in weather forecasting and will help forecasters to forecast weather more accurately. In order to achieve this goal and to recommend favorable solutions, several big data techniques and technologies have been suggested to manage and analyze the huge volume of weather data from different resources.

KEYWORDS: Big Data, Data Analysis, Weather forecasting, Map Reduce and prediction framework.

I. INTRODUCTION

Studying historical data, for the purpose of future prediction and planning, has been a core concept in the big data analysis and decisions making. In this paper, we focus on studying historical weather data from the National Oceanic and Atmospheric Administration (NOAA). The data is collected over the span of about 11 years; from 1997 to 2007. Big Data is a term refer to describe the exponential growth for data, both structured and unstructured, because data coming from everywhere such as social media, videos, digital pictures, sensors etc., and that make it difficult to use software tools to capture, Analysis, manage and process data within a tolerable elapsed time . Big data have three characteristics high Volume, high Velocity and high Variety. According to Bryon "Weather is the original big data problem". It has been discussed earlier though any approach is followed; weather forecasting is initial value problem. Size of initial data increases, accuracy of forecasting increases . According to Nancy Grady the velocity of weather data plays role in the development of economy as a case. This weather data can be used by combining it with other disciplines which can generate new opportunities to weather prediction .Prediction based on temperature is important to agriculture and commodity markets. Therefore, temperature prediction is used by utility companies to estimate demand over coming days. On an everyday basis, people use weather forecasts to determine what to wear on a given day . Since outdoor activities are severely curtailed by wind chill, heavy rain and snow. Moreover; weather forecasting can be used to plan activities around these events and to plan ahead and survive them. In order to predict weather in a very effective way and to help overcome all such problems we have proposed A Prediction Framework of Weather with Big Data Using Map Reduce Algorithm and the advantage with big data has over other weather prediction method is the big data minimizes the error us in various algorithms and gives us a predicate value .

II. RELATED WORKS

Weather forecasting is a critical application area for data analytic, machine learning, and big data technologies. Traditional weather prediction methods rely on complex mathematical models and numerical simulations, which require extensive computational resources. However, the emergence of big data technologies has provided new

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

opportunities for improving the accuracy and scalability of weather forecasting systems. weather data is generated in high volume, velocity, and variety from multiple sources such as satellites, weather stations, and sensors. The use of big data techniques enables the storage, p reprocessing, and analysis of this massive data for improved forecasting accuracy. Map Reduce is a widely used programming model for processing large-scale data in distributed computing environments. Hadoop Map Reduce was utilized to handle historical weather datasets for distributed processing, making weather prediction faster and more efficient. weather prediction model using Hadoop Map Reduce for parallel processing of meteorological data. Their study demonstrated that Map Reduce reduces the computational time significantly compared to traditional methods.

Some researches about big data analytics have been done. This section presents some survey papers in the big data analytic field and the descriptions of their important factors and limitations. Sahasrabuddhe and Jam sandekar first introduced weather forecasting, including basic processes and different approaches, and then big data. They then explained a number of data structures used for big data and weather forecasting with an overview of a number of papers. This study did not have classification, a research methodology, and a discussion, and this paper was not a systematic review.

Priya described weather forecasting, big data, and rain forecast. The author reviewed several papers with the issue of predicting rainfall using data mining techniques. The author summarized them in a table and then began to expound the papers. The limitation was explained. This paper did not have a research methodology, the reviewed works, taxonomy, and discussion sections, so it was not a systematic review.

S.NO	PAPER_INFORMATION	DESCRIPTION	LIMITATION
1.	Cheng et al and Cramer et al	Metrological data attribute based on genetic algorithms And proposing a Hybrid genetic Programming.	Programming with coding the fitness function and not testing other regression and classification techniques.
2.	K vinge et al	Dimension-driven data analysis using a statistics dimensionlity-reduction technique called the kappa profile	Low reliability Unable to estimate the minimum Embedding Dimension.
3.	Buszta and Mazurkiewic z	Weather forecasting using visual data presentation techniques and neural networks.	Low ability to detect temperature externs manual location choice human interaction needs.
4.	Mahmood et a	Forecasting climate change based on data mining techniques.	Low scalability
5.	Azimi et al	Accurate wind power forecasting. Based on data mining	Low prediction Accuracy

III. BACK GROUND

1. Big Data

Relative to computational technology, the volume of data grows very quickly, and big data definitions are numerous. One of these concepts says that big data requires using new tools, analytic, and technical architectures to make high value sources for businesses and to gain vast hidden information in analytic. Five main features characterize big data: variety, variability, volume, velocity, and value.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

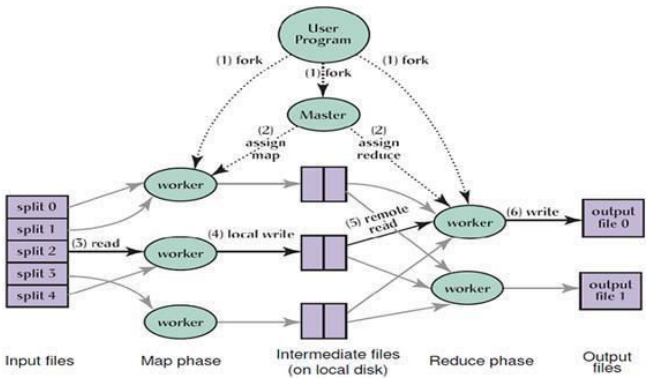
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1.1 Map Reduce

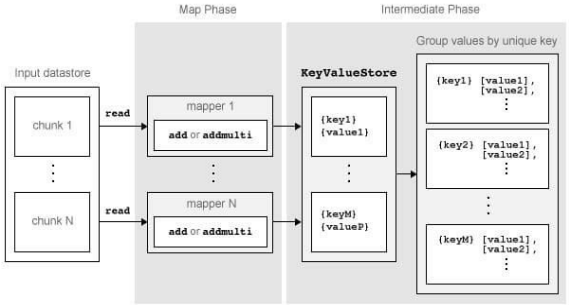
Google has generated Map Reduce to process a big volume of data, such as online records and site application request reports on large service node clusters .The Map Reduce paradigm is an effective programming model for large-scale data-intensive computing applications . Master, Map function and Reduce function are three parts of Map Reduce.



Map Reduce Architecture

1.1.1 Mapper Function

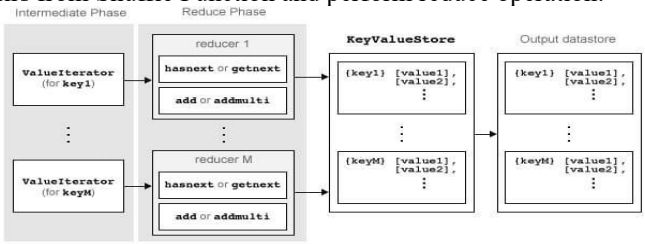
Map Reduce Algorithm uses the following three main steps Map Function, Shuffle Function and reduce Function. Map Function is the first step in Map Reduce Algorithm. It takes input tasks (Datasets) and divides them into smaller sub-tasks. Then perform required computation on each sub-task in parallel. This step performs the following two sub-steps: Splitting step takes input data set from source and divides into smaller Sub Datasets.



Mapper Function

1.1.2. Reduce Function

Reduce Function it is the final step in MapReduce Algorithm. It performs only one step: Reduce step. It takes list of <Key, List<Value>> sorted pairs from Shuffle Function and perform reduce operation.



Reduce Function

1.1.3 Shuffle Function

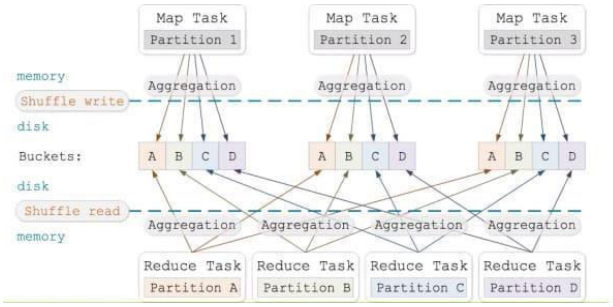
Shuffle Function It is the second step in MapReduce Algorithm it takes a list of outputs coming from “Map Function” and performs these two sub-steps on each and every key-value pair. Merging step combines all key value pairs which have same keys (that is grouping key value pairs by comparing “Key”).

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

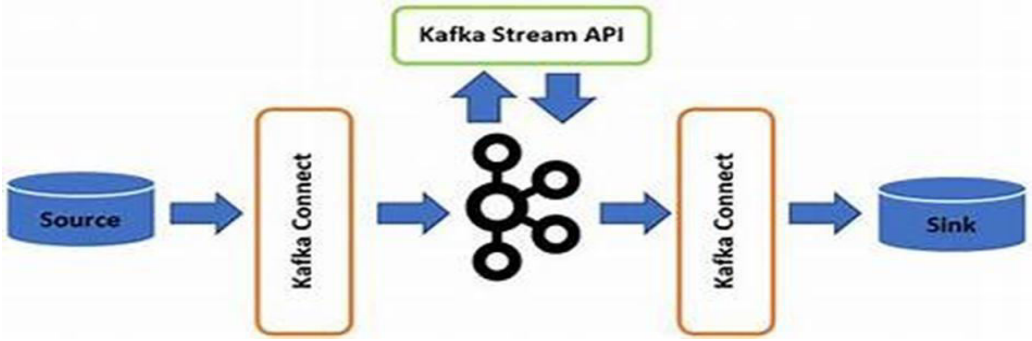


Shuffle Function

2. Big Data Tool

2.1 Apache Kafka

Kafka is a distributed streaming platform. It handles streaming and operating data via in-memory computational strategies for real-time decision-making. Supporting distributed computing, concurrent data loading in Hadoop, and high-throughput are the major attributes of Kafka. The activity (clients' activities) and operational (servers' execution) data have been widely used in recent years to procure functional websites. Knowing how to interpret operating data is essential for controlling real-time activities. Kafka combines offline and online computations to contribute to real-time computing and provide an ad hoc approach for these two forms of data .



Kafka work flow

3. Data Set

A weather forecast dataset with the following attributes are taken for analysis to perform the prediction as shown in Table.

A	B	C	D	E	F	G	H
_id	city	emperatur	humidity	pressure	weather	timestamp	
661bca12f	New York	16.4	72	1012	light rain	2025-04-04	12:00:00
661bca22f	New York	15.7	74	1010	overcast c	2025-04-03	12:00:00
661bca34f	New York	14.9	76	1008	clear sky	2025-04-02	12:00:00
661bca46f	New York	14.2	78	1006	few clouds	2025-04-01	12:00:00
661bca57f	New York	13.5	80	1004	scattered (	2025-03-31	12:00:00
661bca67f	Chicago	12.8	70	1007	broken clo	2025-04-04	12:00:00
661bca77f	Chicago	13.2	68	1005	clear sky	2025-04-03	12:00:00
661bca87f	Miami	28.4	82	1011	thunderstc	2025-04-04	12:00:00
661bca97f	Seattle	10.1	85	1013	light rain	2025-04-04	12:00:00
661bcaa7f	Los Angele	21.6	60	1015	sunny	2025-04-04	12:00:00

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

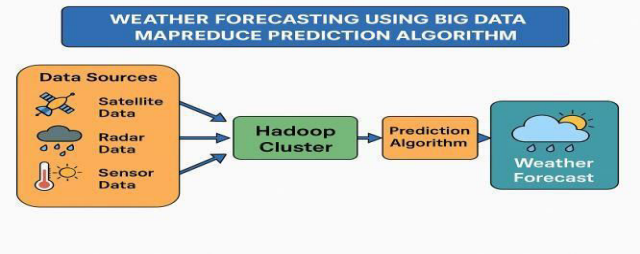
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. PROPOSED SYSTEM

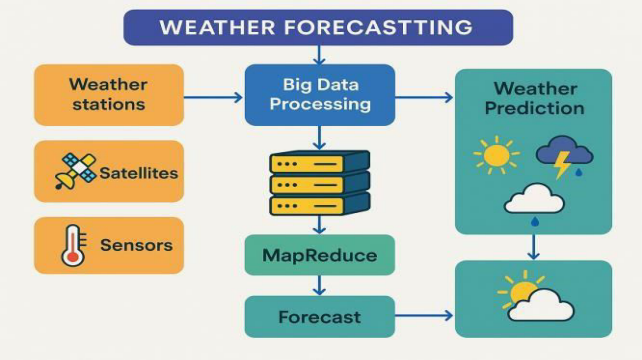
1. Architecture

The architecture of Weather Forecasting using Big Data MapReduce Prediction Algorithm. It explains how large-scale weather data is collected, processed, and used for forecasting. Weather data is collected from multiple sources such satellite provides images, temperature, humidity, and cloud patterns from space. Radar data Provides precipitation data, rainfall intensity, and storm tracking. Captures environmental parameters like temperature, pressure, humidity from IoT-based weather sensors. Hadoop is a Big Data framework used to handle large datasets. Collected weather data is stored in Hadoop Distributed File System (HDFS). MapReduce processes this large dataset in parallel, making the system scalable and faster. After processing, machine learning or statistical prediction algorithms are applied. Algorithms analyze historical patterns and current data to predict future weather conditions like temperature, rainfall, storms, or humidity.



KMR Architecture

1. Work Flow



In this system, weather data is collected from various sources such as weather stations, satellites, and sensors. These sources provide large volumes of raw weather data including temperature, humidity, wind speed, rainfall, and atmospheric pressure. This collected data is then sent for Big Data Processing, where advanced tools and technologies handle and manage the huge amount of data effectively. After processing, the MapReduce technique is applied to analyze, filter, and summarize the data. This technique divides the large data into smaller chunks (Map phase) and then processes them to produce accurate results (Reduce phase). The processed data is then used for Weather Prediction, which helps to forecast different weather conditions like rainfall, sunshine, storms, and cloudy environments. Finally, the forecast results are generated and shared for public and organizational use, helping in disaster management, agriculture, and daily weather reports.

3. Data Collection and Preprocessing

Weather data is collected from multiple sources to ensure diversity and accuracy:
Ground-Based Stations: Provide data on temperature, humidity, pressure, and wind speed.
Satellites and Radar Systems: Offer large-scale atmospheric data, including cloud coverage, storm development, and precipitation patterns.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304145

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

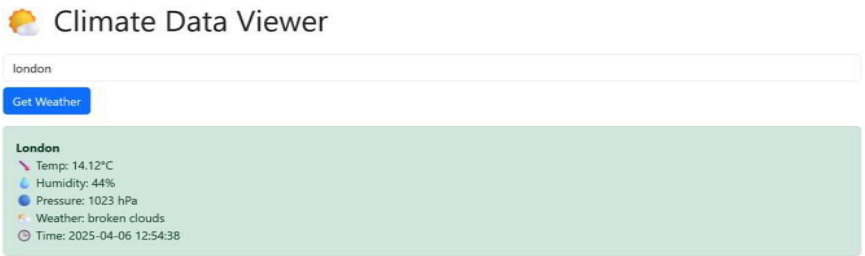
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

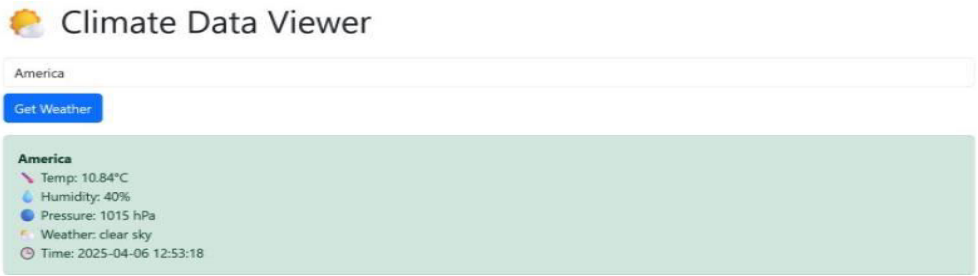
IoT and Sensor Networks: Real-time environmental monitoring is enabled by IoT devices, which provide granular data from urban and rural areas.
Historical Weather Data Repositories: Past weather data is stored and used to train machine learning models for improved predictions.

V. RESULT

Accuracy, precision, and recall: In order to evaluate a big data method, to compare it with earlier methods to determine a more efficient algorithm, and also to discover its pros and cons, we need QoS factors. Precision, accuracy, and recall are some of these factors used in various articles. In the following, we represent the most important QoS factors for big data analytics in weather forecasting. Accuracy, precision, and recall are defined. Accuracy is the percentage of correct prediction. The correct positive prediction percentage is called precision. The percentage of occurrences predicted as positive is called recall.

From the data set of observations presented, we study the temperature factors which are important to agriculture, and therefore to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days The Algorithm are consists from two stages a map function and a reduce function, and when a function called the below steps of actions take place. The first procedure is to preprocess the input weather data and split into a number of pieces of a specified size.





VI. CONCLUSION

In study, propagation big data analysis is used for predicting the temperature based on the MapReduce algorithm to the big data. Through the implementation of this framework, it is illustrated, how an intelligent system can be efficiently integrated with big data prediction framework to predict the temperature. This method proves to be a simplified

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

conjugate gradient method. When implementation into the framework in big cluster the performance of MapReduce was satisfactory as there were not substantial number of errors in data processing. Weather forecasting with big data approach for temperature forecasting is capable of yielding good results and can be considered as an alternative to traditional meteorological approaches. This approach is able to determine the nonlinear relationship that exists between the historical data (temperature, wind speed, humidity, etc.,) supplied to the system during the training phase and on that basis, make a prediction of what the temperature would be in future. However; the limitation of this study proposed only uses the structured data instead of using both (structured data and unstructured data) for efficiency temperature prediction accuracy.

REFERENCES

1. Zhao, z., & liu, w. (2020). Weather forecasting based on machine learning and deep learning algorithms. International journal of meteorology, 12(2), 45-58.
2. Kong, x., & li, q. (2020). Big data analytics for weather prediction. Proceedings of the international conference on big data, 234-243.
3. Lin, d., & liu, y. (2020). Map reduce-based weather prediction using big data. Journal of computational meteorology, 29(4), 103-115.
4. Guerra, j., & tufekci, s. (2021). Weather prediction models using machine learning and big data. Weather forecasting journal, 25(3), 98-110.
5. Wang, h., & chen, z. (2021). Cloud computing for weather data storage and analysis. International journal of cloud computing, 14(1), 22-39.
6. Zhang, j., & xu, j. (2021). Real-time weather forecasting with iot and cloud computing. Ieee internet of things journal, 5(3), 789-801.
7. Jain, s., & gupta, n. (2021). Enhancing weathe Predictions with deep learning and artificia Intelligence. Journal of artificial
8. Izhao, w., & wang, y. (2022). Ai-based weather forecasting models and their Applications. International journal of ai research, 36(1), 12-25.
9. Gonzalez, p., & williams, d. (2022). Challenges in data collection and quality in weather forecasting.Weather and climate analysis, 9(2), 101-110.
10. Chen, w., & tang, h. (2022). Optimization of map reduce algorithms for weather prediction. Computational methods in meteorology, 13(1), 99-107.
11. O'reilly, t., & zhang, l. (2023). Machine learning for weather forecasting. Springer international publishing.
12. Kashani MH, Jahanshahi M (2023) A new method based on memetic algorithm for task scheduling in distributed systems. Int J Simul Syst Sci Technol 10.
13. Niu b, wang h (2023) bacterial colony optimization. Discrete dyn nat soc 2012:28 .
14. Kashani mh, jahanshahi m (2023) using simulated annealing for task scheduling in distributed systems. In: 2009 international conference on computational intelligence, modelling and simulation. Pp 265–269 .
15. Dasgupta d, ji z, gonzalez f (2023) artifcial immune system (ais) research in the last fve years. In: The 2023 congress on evolutionary computation, 2023. Cec '03., vol 1. Pp 123–130 .
16. jameii sm, kashani mh, karimi r (2024) laspea: Learning automata-based strength pareto evolutionary algorithm for multi-objective optimization. Int j comput sci telecommun 6(9):14–19 .
17. Yang x-s (2024) a new metaheuristic bat-inspired algorithm. In: Nature inspired cooperative strategies for optimization (nicso 2023).
18. Yang x-s. Bat algorithm for multiobjective optimisation. Arxiv e-prints, accessed 01 mar 2024. Arxiv:1203.6571y .
19. Krishnanand kn, ghose d (2024) glowworm swarm optimization for simultaneous capture of multiple local optima of multi modal functions. Swarm intell 3(2):87–124 .
20. Sarvizadeh r, kashani mh, zakeri fs, jameii sm (2024) a novel bee colony approach to distributed systems scheduling. Int j comput appl 42(10):1–6 .

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details