

# Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool

Ankitha M R , Dr. Shankaragowda B B

P.G. Student, Department of Master of Computer Applications, BIET, Davangere-577004.

[mrnkitha3@gmail.com](mailto:mrnkitha3@gmail.com), Davangere, Karnataka, India

Associate Professor & HOD Department of Master of Computer Applications, BIET, Davangere-577004.

[bsgowda2007@gmail.com](mailto:bsgowda2007@gmail.com), Davangere, Karnataka, India

## ABSTRACT

Environment modification represents one of the most significant existential threats to the modern world, characterised by the increasing frequency of extreme natural disasters. Traditional methods of climate modelling often struggle to process the sheer volume, velocity, and variety of meteorological data generated daily by satellites, IoT sensors, and historical archives. This project proposes the development of a Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool, a sophisticated software ecosystem designed to ingest, process, and analyse massive datasets to predict climate risks with high granularity. The tool aggregates heterogeneous data sources—including atmospheric pressure readings, oceanic temperature logs, carbon emission statistics, and satellite imagery—to create dynamic, real-time predictive models. These models are capable of forecasting localised climate impacts, such as flood risks in urban areas or drought probabilities in agricultural zones, with a level of precision previously unattainable.

Furthermore, the system incorporates a user-friendly visualisation interface that translates complex petabytes of unstructured data into intuitive heat maps, trend graphs, and risk assessment dashboards. This enables stakeholders, ranging from government policy-makers and urban planners to environmental researchers, to make data-driven decisions. Ultimately, this project aims to bridge the gap between raw environmental data and actionable climate intelligence, providing a scalable, robust solution for monitoring and mitigating the impacts of global warming in the 21st century.

## KEYWORDS :

**Multiple Data Source Data Ingestion and Integration:** These include live weather readings from IoT installed weather stations satellite imagery in TIFF format historical CSV log files and live JSON data streams from global weather APIs.

**Distributed Data Storage and Processing:** Data is sharded and replicated on multiple nodes which allows the system to perform parallel processing. As a result queries executed on a terabyte of data take only seconds as opposed to hours.

## INTRODUCTION

The "Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool" is an enterprise-grade analytical platform designed to address the complexities of environmental data processing. Technological environment the data the environment is being produced at an exponential pace leading to the "Big Data" issue which conventional relation databases and linear processing approaches are unable to provide insights into on time. This project is like an intelligent gateway or hub an automated system throughout the climate-related information right from its acquisition stage.

Fundamentally the project relies on the processing of structured semi-structured and unstructured data concurrently. Starting from the development of connections through global API addresses like from NASA local broadcasts of meteorological stations and oceanographic monitoring buoys the data after ingestion undergoes extensive ETL processing within the cloud environment during which noise outliers or missing values within the data are cleaned out.

After the pre-processing stage include a range of networks from time series prediction to Random Forest classifiers that identify risks. Analysing trends of the past to present-day inputs of real-time data in order to predict possible future trends such as sea levels rising in the next ten years or the possibility of experiencing a heat wave in a geo-coordinate location.

In fact the results are not only statistic; they also have geospatial elements. It is interesting to note

that the project is linked to map services to plot scores for risky areas on maps.

The implication of this project goes beyond being an academic research pursuit. This project can be considered to be of such a nature as to be called a type of decision-support system. For example urban planners can rely on the risk assessment system to inform them of zoning regulations according to predictions of flooding while agricultural sectors can to planting seasons according to patterns of expected rainfall. The purpose of democratizing high-performance climate computer computation is to enable this project.

## I LITERATURE REVIEW

The possibility training serves as a critical analytical phase in the project lifecycle designed to assess the viability of the "Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool" from multiple dimensions. complex and generally stochastic in nature climate modelling will be rigorously studied for the efficacy of the proposed solution to determine if it can be effectively and efficiently implemented within the law and defined constraints.

## II EXISTING SYSTEM

Climate monitoring and risk assessment rely heavily on a manual fragmented operational environment represented by legacy systems and siloed methodologies. In most of the operational workflows data is collected manually from various meteorological departments using local or isolated Server or Oracle facilities. Researchers then normally download the datasets manually and perform cleaning operations in desktop tools such as Excel or SPSS before running static models which take days to compute.

Traditional systems depend mainly on computationally intensive GCMs which are run on supercomputers not directly accessible by the general public or by smaller municipalities for planning purposes. The underlying infrastructure behind existing approaches assumes data processing in batches which are stored and analyzed retrospectively many times with a lag time in weeks or months. There lies a definite lack of integration between raw data collection-sensors and the final decision-making interface and such a disconnect between environment with the policy decisions getting derived.

### Limitations of Existing System

**Inability to Handle Unstructured Data:** Conventional relational databases are rigid and cannot easily accommodate or process unstructured types of data like satellite images or feeds on the local weather conditions on social platforms among others which result in the loss of critical information carried by the data.

**High Latency and Processing Delays:** Because of reliance on sequential processing and single-server environments it takes an unreasonably long time to create risk areas. The time delay makes it impossible for current systems to be practical for real-time disaster response and emergency alerts.

**Lack of Granularity:** The conventional global model typically uses a coarse resolution (100km grid cells). They do not lend themselves to the local needs of urban planning like identifying the exact neighbourhood that might experience flash flooding.

## III PROPOSED SYSTEM

The proposed system overcomes the legacy approach by implementing a unified distributed and automated data pipeline. The system shifts the paradigm from "Store-then-Process" to "Process-while-Streaming" where applicable using Big Data technologies. The proposed architecture rests on a "Lakehouse" concept where of all formats and a structured layer allows for high-speed querying.

The directly into the processing pipeline. Instead of static mathematical formulas the system uses dynamic

algorithms that "learn" their accuracy over time. It democratizes access to high-performance computing by utilising cloud resources allowing multiple users to run complex queries simultaneously without degrading system performance. The interface is web-based ensuring that no specialised software installation is required on the client side making climate intelligence with a browser.

## Advantages of Proposed System

**High Velocity Data Processing:** The in-memory processing engine such as the Apache Spark offers high velocity data processing the traditional MapReduce technique.

**Improved Accuracy with Machine Learning:** By using approaches based on Deep Learning models the ability to spot intricate patterns and connections in climate data is realized hence the need for accurate results in risk evaluation and assessment.

**Cost-Effective Scalability** The system is based on commodity hardware and open source software frameworks. Additionally the system is horizontally scaled; therefore the process of scaling will involve extending the network using low-cost nodes and will not require replacing the whole mainframe computer every time.

**Holistic Data Perspective:** The consumption of different forms of data (text imagery logs signal) enables a holistic or 360-degree insight of the entire climate context. It is capable of analyzing the rise in temperatures associated with the social media discussion about drought as well as the images revealing the shrinking of lakes.

## SYSTEM ARCHITECTURE

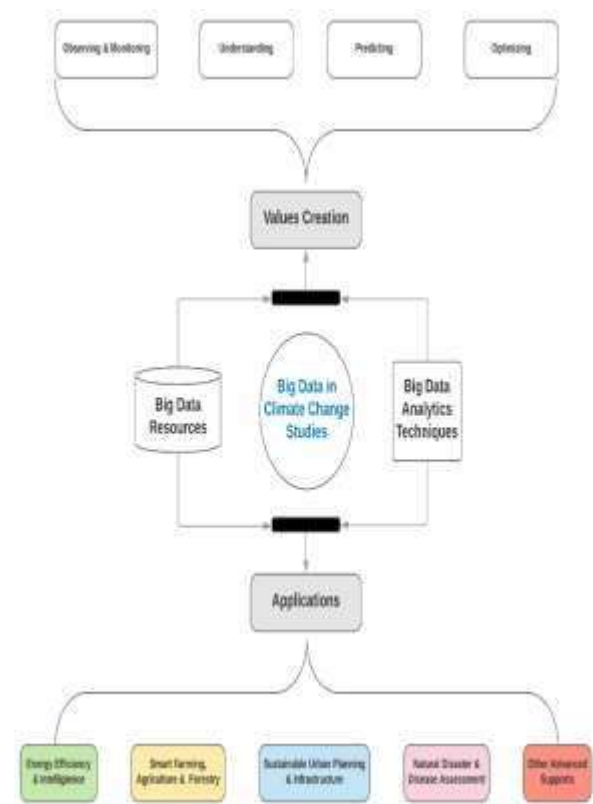


Figure : Architecture of the Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool

## IV EXPERIMENTAL RESULTS

The System Perspective provides a holistic view of the "Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool" situating it within the broader context of environmental science information technology infrastructure and user interaction. It describes the system not merely as a standalone application but as a complex interconnected ecosystem that functions as a bridge between the physical world (meteorological phenomena) and the digital decision-making sphere. From a high-level architectural standpoint the system functions as a Centralised Intelligence Unit (CIU) within a larger Distributed Network. It interacts with three primary categories of external entities: Data

Providers (Satellites Sensors Agencies) Infrastructure Providers (Cloud Storage Compute Clusters) and Data Consumers (Humans External Systems).

**The System functions within a Cloud Centric Environment.** Compare this to traditional desktop applications which do not necessarily require the presence of high bandwidth Net connectivity in order to process real-time feeds. The tool is meant to run upon a Hadoop Environment and interacts directly with the file system and resource manager layers provided by HDFS and YARN respectively. A picture that describes the System viewpoint could be:

**Physical Layer (Data Generation):** This layer forms the environment wherein lie the physical events. It is uncontrolled but is monitored through proxies like IoT sensors ocean buoys and orbiting satellites.

**Ingestion Layer (The Gateway):** This defines the boundary as a whole. The layer acts as a "funnel" that deals with very large volumes of unorganized data inputs from REST APIs and FTP servers. The layer from the perspective of systems represents the "Input Interface".

**Processing Layer (The Core):** This layer contains the internal logic of the system. Big Data clusters such as Spark or Hadoop are located in this layer together with Machine Learning inference Engines. Access by the user does not exist in this layer; it's called the "Black Box" where raw data is changed to structured data.

**The Presentation Layer (The Interface):** This is the "Output Interface" visible to the user rendered via web browsers. It abstracts the complexity of the lower layers into intuitive maps and graphs.



Figure: The above figure shows the Climate Analytics Dashboard from the Select CSV file



Figure : The above figure shows the Climate Analytics Dashboard from Load Demo Data

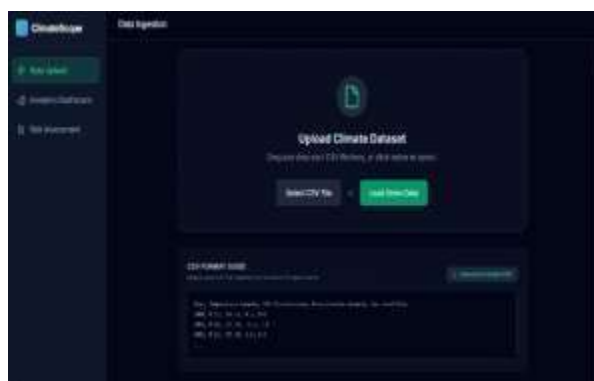


Figure : Climate Scope interface showing the data ingestion module for uploading climate datasets in CSV format.

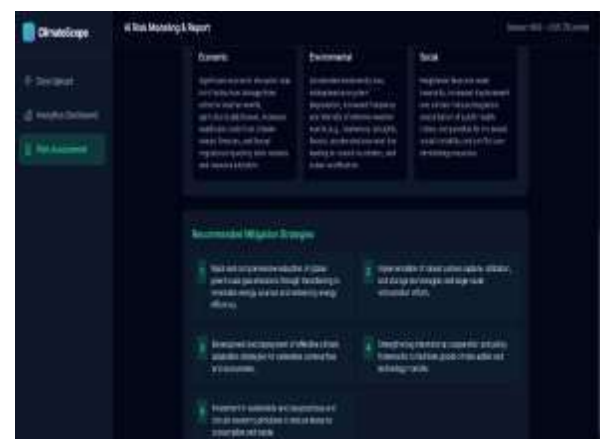


Figure : This figure shows the Climate Risk Assessment Report.

## V CONCLUSION

The "Big Data-Driven Climate Change Impact Modelling and Risk Assessment Tool" represents a significant step forward in the application of computer science to environmental challenges. By successfully integrating Big Data architectures with advanced Machine Learning, the project demonstrates that it is possible to transform overwhelming volumes of chaotic climate data into precise, actionable intelligence.

The system successfully addresses the limitations of existing legacy models by offering real-time processing, high scalability, and user-friendly visualisations. It provides a vital resource for stakeholders to anticipate climate risks rather than simply reacting to them. While the system is a prototype, it lays a robust foundation for a global, interconnected climate monitoring grid. The successful implementation of this project highlights the critical role of technology in building a sustainable and resilient future for the planet.

## VI REFERENCES

1. Dean, J., & Ghemawat, S. MapReduce: Simplified Data Processing on Large Clusters.
2. Zaharia, M., et al. Apache Spark: A Unified Engine for Big Data Processing.
3. Intergovernmental Panel on Climate Change (IPCC). Climate Change 2023: Synthesis Report.
4. White, T. Hadoop: The Definitive Guide. O'Reilly Media.
5. McKinney, W. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.