Centre for Development of Advanced Computing (CDAC)

Post Graduate Diploma in Big Data Analytics (PG-DBDA)

Mumbai Centre – USM VITA

Academic Year: Feb – Aug 2025

Project Synopsis

Big Data-Driven Weather Analysis and Visualization Framework for Indian Cities

Guided By:

Ms. Tejal Mate

Submitted By:

Group 7

Title:

Big Data-Driven Weather Analysis and Visualization Framework for Indian Cities

Team Members:

The project team consists of seven dedicated and skilled members:

- Gourav Sharma
- Abhishek Chandel
- Laxmi Thakre
- Pranali Raul
- Sakshat Bankar
- Sarvesh Kamble
- Yash Jadhav

Problem Statement:

India lacks an efficient system for analyzing city-level weather patterns due to the massive volume and variety of weather data. This project aims to develop a scalable big data solution that can efficiently analyze historical weather data from 4000+ Indian cities and generate accurate, city-wise weather analysis to support informed decision-making.

Introduction:

Weather analysis and climate monitoring play a crucial role in various sectors such as agriculture, disaster management, transportation, and daily life planning. The growing availability of large-scale weather data from satellites, sensors, and weather stations presents an opportunity to improve the understanding of weather and climate patterns. However, traditional methods struggle with processing and analysing this vast amount of data efficiently.

This project focuses on leveraging big data technologies to analyse historical weather data from **over 4,000 Indian cities**. The primary goal is to build a robust system for insightful analysis of climate patterns. By combining scalable data processing with advanced analytics, the project aims to support timely and informed decision-making for multiple applications.

Additionally, the project aims to explore temporal and spatial variations in weather parameters to detect trends and anomalies with potential environmental and economic

impacts. Ultimately, this initiative seeks to support smarter resource management and disaster preparedness by providing stakeholders with reliable, data-driven insights into India's diverse climate.

Objective:

- To collect and process large volumes of historical weather data from over 4,000 Indian cities using big data tools and frameworks.
- To perform comprehensive data cleaning, transformation, and time-series analysis to ensure high-quality input for modelling.
- To identify and analyse seasonal patterns, extreme weather events, and temperature trends to understand climate behaviour.
- To build interactive and user-friendly visual dashboards for real-time monitoring of weather metrics including temperature, humidity, wind speed, and precipitation.
- To demonstrate the practical application of big data technologies in environmental monitoring.
- To enable decision-makers in tourism, disaster management, and government sectors to utilize weather insights effectively.
- To improve the scalability and efficiency of weather data processing for large-scale city-level climate analysis.

Data Dictionary & Description:

The dataset used for this project is titled **"**Indian Cities Weather 2010–2024**"**, sourced from Kaggle.

https://www.kaggle.com/datasets/mukeshdevrath007/indian-5000-cities-weather-data/data

Weather Data Folder:
This folder includes individual CSV files for more than 4,000 cities. Each file contains daily weather records from 2010 to 2024 for a specific city. For example, the file for Mumbai contains the following weather parameters:

- date: The specific date of weather observation
- temperature_2m: Air temperature at 2 meters above ground (°C)
- relative_humidity_2m: Relative humidity at 2 meters (%)
- dew_point_2m: Dew point temperature (°C)
- apparent_temperature: Feels-like temperature (°C)
- precipitation: Total precipitation (mm)
- rain: Rainfall amount (mm)

- snowfall: Snowfall amount (mm)
- snow_depth: Snow depth on ground (cm)
- pressure_msl: Mean sea-level pressure (hPa)
- surface_pressure: Atmospheric pressure at the surface (hPa)
- cloud_cover: Overall cloud cover (%)
- cloud_cover_low, mid, high: Layered cloud cover (%)
- wind_speed_10m / 100m: Wind speed at 10m and 100m heights (m/s)
- wind_direction_10m / 100m: Wind direction (degrees)
- wind_gusts_10m: Wind gusts at 10m (m/s)

These columns provide a rich, multidimensional dataset suitable for descriptive analysis (such as trends in temperature, rainfall, humidity, wind patterns, etc.).

Using this data, we aim to:

- Analyse long-term weather trends and seasonal variations across cities
- Detect extreme weather events such as heatwaves, heavy rainfall, or storms
- Visualize patterns and anomalies through interactive dashboards

This dataset offers a strong foundation for building a city-wise weather analysis system using big data technologies.

Expected Outcomes:

- A clean and processed weather dataset ready for analysis.
- Insightful visualizations showing regional and temporal weather trends.
- A scalable big data solution for detecting climate patterns.
- A deployable dashboard for public or organizational use.
- Improved understanding of how weather impacts key sectors such as tourism and disaster response.

Use Cases:

- **Tourism:** Helps identify ideal travel periods through seasonal weather analysis.

- **Disaster Management:** Enables early warning systems for heatwaves, storms, and floods.
- **Government Bodies:** Assists in climate-resilient urban planning through weather trend analysis.

Conclusion:

This weather analysis project applies big data techniques to solve real-world meteorological challenges. By combining large-scale data processing and data visualization, it delivers powerful insights that can support climate research, policy decisions, and commercial planning.