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 Prediction and Visualization Framework for Indian Cities

Guided By:

Ms. Tejal Mate

Submitted By:

Group 7

Title:

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

Team Members:

The project team consists of seven dedicated and skilled members:

- Gourav Sharma – Team Leader
- Pranali Raul – Team Member
- Laxmi Thakre – Team Member
- Sakshat Bankar – Team Member
- Abhishek Chandel – Team Member
- Sarvesh Kamble – Team Member
- Yash Jadhav – Team Member

Problem Statement:

India lacks an efficient system for analyzing and predicting 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 5000 Indian cities and generate accurate, city-wise weather predictions to support informed decision-making.

Introduction:

Weather forecasting 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 forecasting accuracy. 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 5,000 Indian cities. The primary goal is to build a robust predictive model that can accurately forecast weather conditions while providing 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 5,000 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 develop accurate predictive models for forecasting weather parameters such as temperature, humidity, rainfall, and wind speed.
- 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 and meteorological forecasting.
- To enable decision-makers in agriculture, disaster management, and transportation sectors to utilize weather forecasts 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. It comprises two main components:

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

1. **City Metadata File**:
   This file contains the names, latitudes, and longitudes of over 6,000 Indian cities. It serves as a geographical reference to map and visualize city-level weather data across the country.

2. **Weather Data Folder**:
   This folder includes individual CSV files for more than 5,000 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 both descriptive analysis (such as trends in temperature, rainfall, humidity, wind patterns, etc.) and predictive modelling.

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
- Forecast future weather conditions using machine learning models
- Visualize patterns and anomalies through interactive dashboards
- Correlate geographic data (latitude, longitude) with climatic conditions

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

Expected Outcomes:

- A cleaned and processed weather dataset ready for analysis.
- Insightful visualizations showing regional and temporal trends.
- A scalable solution using big data for climate pattern detection.
- Optionally, a model to forecast temperature or rainfall.
- A deployable dashboard for public or organizational use.
- Better understanding of weather impact on key sectors like agriculture and disaster response.

Use Cases:

- Disaster Management: Early detection of heatwaves, storms, and floods.
- Agriculture: Advising farmers on sowing based on weather patterns.
- Urban Planning: Preparing cities for changing rainfall or temperature zones.
- Tourism: Seasonal planning and prediction of ideal travel periods.
- Healthcare: Tracking humidity and temperature for disease spread.

Conclusion:

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