DATA 606 CAPSTONE IN DATA SCIENCE

Project Title: BlizzardX: Cracking the Code of Snow & Temperature Swings

Team H:

- Lokeswar Kudumula
- Namruth Goud Thimmapuram
- Hema Pushpika Konduru
- Sree Sai Preetham Kadiyam

Research Question:

How does snowfall and snow depth influence temperature fluctuations? Can machine learning models predict temperature changes and extreme cold events with high accuracy?

Project Overview:

This project explores the relationship between **snowfall, snow depth, and temperature fluctuations** in a given area. The key objectives are:

- Understanding how snowfall influences maximum and minimum temperatures.
- Developing **predictive models for extreme cold events** using snowfall data.
- Enhancing insights into microclimate variations and providing useful forecasting tools.

Dataset Information:

Dataset Source:

- **Data**: Directory with ".dly" files for all of GHCN-Daily [https://www.ncei.noaa.gov/pub/data/ghcn/daily/all/]
- **ghcnd-inventory.txt**: File listing the periods of record for each station and element [https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-inventory.txt]
- **ghcnd-stations.txt**: List of stations and their metadata (e.g., coordinates) [https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt]

Dataset Description:

This dataset contains information about **snowfall, temperature fluctuations, and other relevant factors**.

- **Dataset Size**: Approximately **1,29,648 station records** across world.
- **Attributes**:
  - **Snowfall (SNOW)**: Daily snowfall amount.
  - **Snow Depth (SNWD)**: Depth of snow on the ground.
  - **Maximum Temperature (TMAX)**: Maximum daily temperature.
  - **Minimum Temperature (TMIN)**: Minimum daily temperature.
  - **Observed Temperature (TOBS)**: Actual temperature observed at the station.
  - **Station Identifier**: Unique identifier for each station.
  - **Year**: Year of the observation.
  - **Month**: Month of the observation.

  (Additional station metadata available in a separate CSV file):
  - **Station Name**: The name of the station.
  - **Coordinates (Latitude, Longitude)**: Geographical location of the station.
  - **Elevation**: Elevation of the station above sea level.

Planned Analysis & Methodology:

Exploratory Data Analysis (EDA):

- **Correlation Analysis**: Examining the relationship between snowfall and temperature fluctuations (TMAX, TMIN).
- **Seasonal Trends**: Investigating how snowfall impacts temperature across different months.
- **Geospatial Analysis**: Analyzing the impact of station coordinates (latitude, longitude) and elevation on temperature and snowfall patterns.
- **Visualizations**: Creating scatter plots and line graphs to compare temperature fluctuations on snowy vs. non-snowy days.

Machine Learning Models:

Regression Models (for Predicting Temperature Changes):

- **Gradient Boosting Regressor**
- **Long Short-Term Memory (LSTM)**

(Other regression models may be explored as the project progresses.)

Classification Models (for Predicting Extreme Cold Events, where temperature drops below a specific threshold):

- **Logistic Regression**
- **XGBoost**

(Additional classification models may be considered in later phases.)

These models will predict temperature variations based on snowfall and snow depth and classify days that are likely to experience extreme cold temperatures.

Performance Evaluation:

Regression Models (Evaluation of Temperature prediction Models)

- **R^2 Score**
- **Root Mean Squared Error (RMSE)**

Classification Models (Predicting Cold Events)

- **Accuracy**
- **F1-Score**
- **Confusion Matrix**

User Interface:

We are planning to develop an interactive web application to visualize the predicted temperature variations and extreme cold event forecasts, with real-time data fetching from NOAA.

Conclusion:

This project aims to improve our understanding of how **snowfall influences temperature fluctuations** and develop **machine learning models to predict extreme cold events**. The predictive insights gained from this study could be valuable for:

- **Weather forecasting systems**.
- **Emergency preparedness for severe winter conditions**.
- **Urban planning and infrastructure adaptation**.
- **Agricultural planning to mitigate frost damage**.

By leveraging **historical weather data, machine learning techniques, and geospatial analysis**, this project goes beyond traditional weather apps, providing **data-driven insights and predictive tools** that help forecast extreme cold events more accurately.