Climate Change Modeling Project Report

Name: Hardik Singh

Institution: Unified Mentor

Date: July 06, 2025

GitHub Repository

Dataset Link

Objective

The goal of this project was to predict and model various climate change indicators-such as temperature anomalies, precipitation patterns, and sea level changes-using historical climate data and machine learning techniques. The objective was to support awareness and proactive decision-making in environmental science through data-driven insights.

Dataset Information

- Source: Google Drive

- Shape: 522 rows × 5 columns

- Features Include:

  - date

  - likesCount

  - profileName

  - commentsCount

  - text

Workflow & Methodology

1. Data Cleaning:

   - Removed nulls and standardized date formats.

2. Exploratory Data Analysis (EDA):

   - Analyzed engagement metrics and textual sentiment.

3. Feature Processing:

- Extracted sentiment scores using NLP tools.

4. Encoding:

  - Applied necessary transformations to categorical data.

5. Modeling:

  - Used machine learning models to identify trends and correlations in climate-related posts.

6. Time Series Forecasting:

  - Applied ARIMA model for temperature anomaly predictions.

Techniques Used

- Sentiment Analysis using VADER

- Time Series Forecasting with ARIMA

- Named Entity Recognition using spaCy

- Topic Modeling using Gensim (LDA)

Tools & Technologies Used

- Python, Jupyter Notebook

- Libraries: Pandas, Sklearn, Matplotlib, re, vaderSentiment, spaCy, gensim

Challenges Faced

- Textual data required extensive preprocessing and cleaning.

- Selecting optimal parameters for ARIMA involved multiple iterations.

Conclusion

The project demonstrated that machine learning and NLP can be effectively applied to understand and forecast climate-related trends from both structured and unstructured data sources. Temperature anomaly forecasts using ARIMA showed reliable trends based on historical patterns.

Future Improvements

- Integrate deep learning models for enhanced sentiment analysis.

- Extend time series analysis to other environmental indicators.

- Deploy interactive dashboards for real-time updates.