Reading the Climate: An NLP-Powered Journey from Articles to Future Forecasts

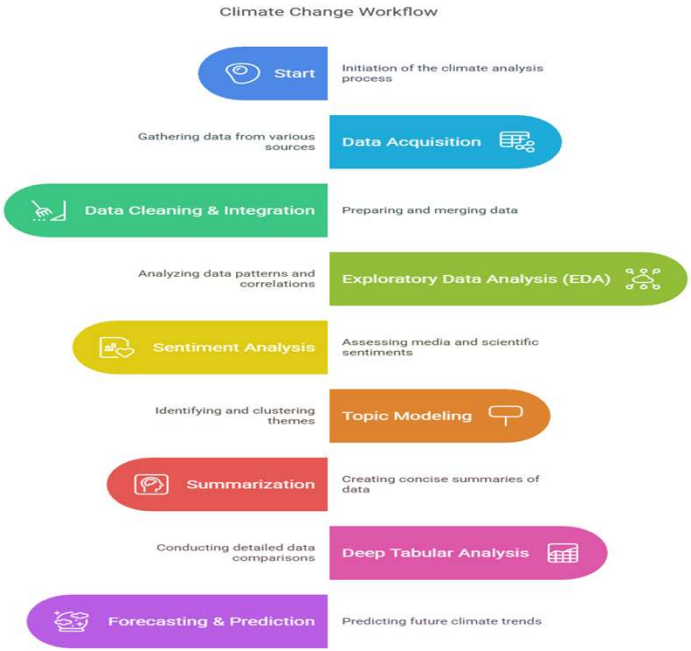
Krishna Khandelwal
Faculty Mentor: Arup Das & Dr. Jiacun Wang

MONMOUTH UNIVERSITY
SCHOOL of SCIENCE

Introduction: This project presents a comprehensive analysis of global climate change discourse by scraping data from key sources—*Nature Climate Change*, *The New York Times*, and other prime climate-focused platforms—using Firecrawl. Extracted articles are processed using NLP techniques including Exploratory Data Analysis (EDA), word clouds, sentiment analysis (VADER), and topic modeling with BERTopic (UMAP embeddings). A transformer-based summarizer (Flan-T5) is used to generate concise abstracts.

To enrich the analysis, external datasets such as global Air Quality Index (AQI), temperature anomalies, and disaster records are integrated for correlation insights. Finally, time series forecasting using RNN and LSTM models predicts future trends in AQI, natural disasters, and warming patterns over the next three years. This multi-source, multi-method approach bridges climate science, public discourse, and predictive analytics.

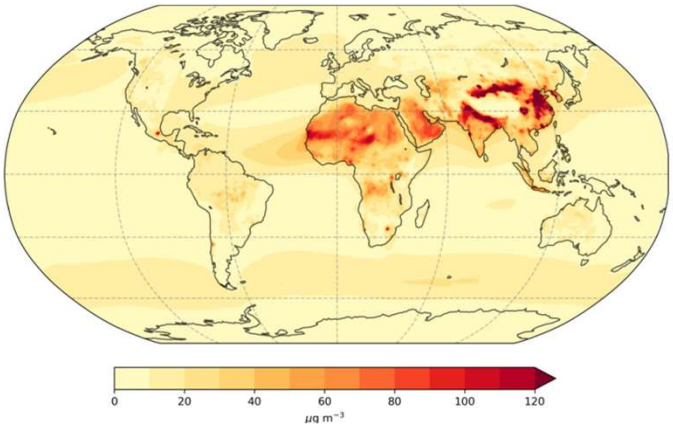
Key Terms in Climate Reports



Key Findings:

- Climate discourse shifted toward *resilience* and *carbon mitigation* post-2020
- Media sentiment showed greater fluctuation than scientific literature
- Forecasts predict increased coverage of extreme weather and adaptation themes
- Topic modeling revealed rising focus on *biodiversity*, *policy*, and *climate equity*

Temperature change over the past years



Thematic Links in Climate Science