Temperature Anomaly Prediction and Analysis Dashboard

1. Introduction

1.1 Problem Statement

Climate change is a critical global issue, and accurately predicting temperature anomalies is essential for understanding future climate trends. This project aims to build a **machine learning**-based model to predict **temperature anomalies** and **visualize** historical and predicted trends using Tableau..

1.2 Objective

The main objectives of this project are:

- ✓ To clean and preprocess historical temperature anomaly data.
- ✓ To perform feature engineering for improved model accuracy.
- ✓ To train machine learning models (Random Forest & XGBoost) for anomaly prediction.
- ✓ To deploy predictions in a Streamlit application for interactive forecasting.
- ✓ To visualize historical and predicted trends in Tableau for insights.

2. Data Preprocessing

2.1 Data Collection

The dataset used in this project consists of historical global temperature anomalies collected from climate research sources. The data includes temperature deviations from the historical mean across different seasons and years.

2.2 Data Cleaning and Preprocessing

- ✓ **Text Cleaning:** Handled missing values and inconsistencies and analysed the data.
- ✓ **5- Year MA:** Calculated 5-Year Moving Average to capture long term trends.
- ✓ **Reshaped data:** Reshape wide-format data into a long format.
- ✓ **Data Encoding:** Applied data encoding on seasons.
- ✓ **Normalization:** Applied normalization on data.

3. Model Development

3.1 Machine Learning Models

Implemented in train.py-

- ✓ **Linear Regression** – Basic Statistical model for trend analysis
- ✓ **Random Forest** – Captures non-linear relationships in data.
- ✓ **XG Boost Regressor** –Advanced boosting technique for enhanced accuracy.

3.2 Model Training & Evaluation

Each model was trained on historical temperature anomaly data and evaluated using:

- ✓ **Mean Absolute Error (MAE)** - Measures average error magnitude.
- ✓ **Mean Squared Error (MSE)** - Penalizes larger errors more heavily.
- ✓ **R^2 Score** - Measures how well the model explains variance in data.

3.3 Model Performance Comparison

Model	**MAE**	**MSE**	**R2 Score**
Linear Regression	1.06e-15	1.69e-30	1.0000
Random Forest	1.96e-03	1.62e-05	0.9999
XGBoost	7.43e-04	2.84e-06	0.9997

3.4 Insights from Model Performance

- ✓ Linear Regression achieved an R^2 score of 1.0, indicating overfitting due to potential collinearity in features.
- ✓ Random Forest provided stable predictions, capturing non-linear temperature variations well.
- ✓ XGBoost had the lowest MAE and MSE, making it the most accurate model for forecasting temperature anomalies.

4. Deployment

Implemented in deploy.py:

- ✓ Developed an interactive Streamlit application to predict temperature anomalies.
- ✓ Allowed users to select a future year and month to get anomaly predictions.

- ✓ Displayed confidence indicators for anomaly predictions.

5. Data Visualization Using Tableau

The processed data and machine learning predictions were exported using tableau_data.py for Tableau dashboard visualization.

5.1 Dashboard Components

- ✓ Historical temperature trends (Monthly, Seasonal, and Yearly).
- ✓ Actual vs. Predicted Temperature Anomalies.
- ✓ Heatmap of anomalies to visualize warming and cooling trends.

5.2 Key Insights

- ✓ Significant warming trends observed in recent decades.
- ✓ Future predictions indicate continued warming, with anomalies exceeding 2.5°C in the coming decades.
- ✓ Seasonal variations impact temperature anomalies, especially in winter and summer months.

6. Future Enhancement

- ✓ Include real-time temperature updates from live datasets.
- ✓ Improve model accuracy using LSTMs or Transformer-based models.
- ✓ Deploy as a cloud-based API for public access.

7. Conclusion

This project successfully built and deployed machine learning-based climate prediction models. The combination of Linear Regression, Random Forest, and XGBoost provided robust insights into historical and future temperature anomalies. The integration of Tableau allowed for interactive visualization, enabling better understanding and analysis of global climate trends.