Title: Time Series Forecasting & Optimization Using Temperature Trends

---

Objective: The primary aim of this project is to evaluate different time series forecasting models—Moving Average, ARIMA, and Holt-Winters—on historical temperature data and to implement prescriptive analytics using Decision Tree algorithms for temperature-related risk assessment. This exercise provides insights into predictive climate modeling and helps build a basic decision-support system for environmental and climate-based planning.

---

1. Dataset Description :- **Dataset Name:** Global Land Temperatures by Berkeley Earth

- **Source:** https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data
- **Attributes Used:**
  - Date: Date in YYYY-MM-DD format, resampled to monthly
  - Mean: Monthly mean temperature anomaly (difference from average temperature baseline)
- **Time Range:** Covers data from 1850s to recent years (after preprocessing and filtering)

Preprocessing Steps:

- Converted Date column to datetime format
- Set Date as the index for time series operations
- Handled missing and duplicate values
- Resampled to monthly frequency using asfreq('MS')
- Smoothed out irregularities with interpolation and removed outliers where necessary

---

2. Time Series Forecasting Models

A. Moving Average (MA)

- Smoothing technique to remove noise and identify trend
- We used a 12-month rolling window to observe seasonal trends
- Useful for identifying general direction of the temperature anomalies
- However, it lacks predictive power and adaptability to sudden changes

B. ARIMA (AutoRegressive Integrated Moving Average)

- ARIMA(p,d,q) is ideal for univariate time series forecasting
- ACF and PACF plots used to determine parameters (p=5, d=1, q=0)
- Differencing was applied to make data stationary
- The model showed decent short-term predictions but struggled with seasonality

C. Holt-Winters Exponential Smoothing

- Captures trend and seasonality using exponential smoothing
- Additive model used as seasonality in temperature is relatively stable
- Seasonal period: 12 (monthly data)
- This model was able to handle both trend and seasonality well and produced more accurate forecasts compared to MA and ARIMA

Model Visualization and Dashboard:

- A forecasting dashboard was developed using matplotlib and seaborn in Jupyter Notebook
- Plots include:
  - Raw Data with Moving Average Overlay
  - ARIMA Forecast vs Actual
  - Holt-Winters Forecast vs Actual
  - All Models Combined Comparison
- Dashboard allows visual inspection of the model performance over different timelines

Model Comparison (MAE & RMSE):

Model	MAE	RMSE
Moving Average	0.512	0.638
ARIMA (5,1,0)	0.423	0.557
Holt-Winters	0.338	0.446

Observation: Holt-Winters provided the lowest error rate and better seasonal adaptability.

---

3. Prescriptive Analytics: Decision Tree for Risk Assessment

Goal: Classify months into high-risk vs low-risk categories based on temperature anomaly.

Steps:

- Derived Year and Month features from Date
- Labeled rows with temperature anomaly above the 75th percentile as high-risk (1), others as low-risk (0)
- Features used: Year, Month
- Target variable: Risk (binary classification)

Model:

- DecisionTreeClassifier with max_depth=4
- Train-test split: 80:20
- Evaluation metrics: Accuracy, Confusion Matrix, Classification Report

Results:

- Accuracy: ~87%
- Precision and recall scores indicate strong performance in detecting high-risk months
- Visualized using plot_tree() for better interpretability

Risk Patterns Identified:

- Warmer months in recent years tend to be classified as high-risk
- Specific rules (e.g., Year > 2000 & Month between June-August = High Risk)

---

4. Visual Insights & Forecasting Dashboard

The final dashboard included multiple subplots:

- Time Series Line Plot of Actual Temperatures
- Moving Average Curve for trend smoothing
- ARIMA Forecast Overlay on Original Data
- Holt-Winters Forecast Overlay
- Comparative Line Chart of All Forecast Models
- Decision Tree Diagram for Prescriptive Analysis

All visualizations are interactive (if exported to Dash or Streamlit) or static using Matplotlib/Seaborn for offline reporting.

---

5. Conclusion

- **Best Forecasting Model:** Holt-Winters provided the best balance of accuracy and seasonal adaptability
- **Prescriptive Power:** Decision Trees delivered a clear set of rules for identifying risky months based on trends
- The combination of statistical forecasting and ML-driven classification enhanced decision-making capabilities

---