DATA CRUNCH

Climate Forecasting
for
Agricultural Sustainability
in
Harveston

By Scorpion X
University of Ruhuna

1. Problem Understanding & Dataset Analysis

Objective

Develop a time series model to predict five climate variables (Temperature, Radiation, Rain, Wind Speed/Direction) critical for Harveston's agriculture. The goal is to enable farmers to optimize planting cycles and resource management.

Key Findings

- **Missing Values**: Forward/backward filling addressed gaps in temperature data.
- **Year Format**: Corrected ambiguous years (e.g., "25" → 2025).
- **Temporal Patterns**: Lagged temperature features captured short-term trends.
- **Domain Relevance**: Variables like Radiation and Wind Direction were treated as independent, though physical correlations may exist.

Preprocessing Justification:

- **Forward/Backward Filling**: Ensured continuity in sequential data.
- **Date Parsing**: Unified date formats across kingdoms for consistency.
- **Lag Features**: Incorporated historical temperature trends to capture autocorrelation.

2. Feature Engineering & Data Preparation

Feature Creation

- **Lagged Variables**: Created 1-, 2-, and 3-day lags for temperature to model temporal dependencies.
- **Kingdom-Specific Processing**: Combined train-test tails for each kingdom to maintain context.

Transformations

- **Non-Stationarity**: Differencing and normalization were not applied, potentially limiting model performance on trends.
- **Limitation**: Only temperature lags were engineered; other variables (e.g., Radiation) lacked historical context.

3. Model Selection & Justification

Baseline Model

Random Forest (RF) Regressor was chosen for its ability to handle non-linear relationships and multi-output regression.

Advantages

- **Interpretability**: RF provides feature importance scores.
- **Speed**: Efficient for small-to-medium datasets.

Hyperparameter Tuning

- Limited to n_estimators=100; grid search or Bayesian optimization could improve accuracy.

Validation Strategy

- **Flaw**: Evaluated on training data, risking overfitting. A rolling-window cross-validation would better simulate real-world forecasting.

4. Performance Evaluation & Error Analysis

Metrics

- **sMAPE**: Chosen for scale invariance (critical for variables like Rain Amount with sparse non-zero values)

Results

Variable	sMAPE
Avg_Temperature	0.17
Radiation: 0.95	0.95
Rain_Amount	15.17
Wind_Speed	4.98
Wind_Direction	11.02

Error Insights

- **Wind Direction**: High sMAPE due to cyclical nature (0° ≈ 360°), unaddressed by the model.
- **Residuals**: Autocorrelation checks were omitted, leaving temporal dependencies unexplored.

5. Interpretability & Business Insights

Actionable Insights

- Farmers could align planting schedules with temperature/rain predictions.
- Wind forecasts aid in protecting crops during storms.

Improvements

- Integrate soil moisture or crop growth data for holistic planning.

6. Innovation & Technical Depth

Novelty

- **Kingdom-Specific Forecasting**: Tailored predictions using kingdom-level data splits.
- **Hybrid Approach**: Combined lag features with tree-based models for interpretability.

Limitations

- No advanced architectures (e.g., LSTM, Prophet) tested for long-term dependencies.

7. Conclusion

Summary

- Random Forest achieved moderate accuracy, with Rain and Wind Direction being the most challenging.
- Key hurdles included handling cyclical variables and temporal validation.

Future Work

- Test SARIMA or Transformer models for sequence modeling.
- Incorporate external data (e.g., satellite imagery) for richer context.