Lasso Regression Modeling for Multi-Source Renewable Energy Forecasting using Climate and Atmospheric Data

Dr .(Mrs.) N.A.D.N.Napagoda
Department of Mathematical Sciences
Faculty of Applied Sciences
Wayamba University of Sri Lanka
Kuliyapitiya.

The lasso regression model is selected for its capability to perform variable selection and model regularization , improving the interpretability of results while minimizing overfitting high dimensional meteorological datasets. By integrating multi-source renewable datasets and optimizing predictive performance, the research aims to develop a data-driven forecasting model that contributes to efficient smart energy management systems

Introduction

- Renewable energy outputs fluctuate due to variable weather conditions

- Accurate forecasting is critical for grid stability

- Multi-source datasets improve prediction accuracy

Key Objectives

- Integrate multi-source renewable generation data (solar, wind, distributed PV, etc.) with atmospheric and climatic predictors.
- Design LASSO regression to select variables, avoid overfitting, and improve interpretability.
- Tune regularization (λ) using cross-validation and assess stability of selected variables.
- Compare Lasso against baseline models (OLS, Ridge, Elastic Net, and a simple tree/ensemble).
- Deliver interpretable predictor sets and decision rules for smart energy management.

Data Sources & Preprocessing

- Renewable Data: Solar PV, wind turbines
- Atmospheric Data: Irradiance, wind speed/direction, temperature, pressure, humidity
- Preprocessing: Cleaning, Standardization, feature engineering
- Multi-source integration for richer inputs

LASSO Regression Overview

- Linear regression with L1 penalty
- Benefits: Selects relevant features, reduces complexity, minimizes overfitting

Data Normalization

LASSO requires standardized predictors:

- Z-score normalization applied:

$$X_{std} = \frac{X - \mu}{\sigma}$$

Model Specification

The LASSO optimization objective is:

$$\hat{\beta} = \arg\min_{\beta} \left(\frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right)$$

Hyperparameter Tuning

Use **k-fold cross-validation** (k = 5 or time-series CV) to determine the optimal value of λ.

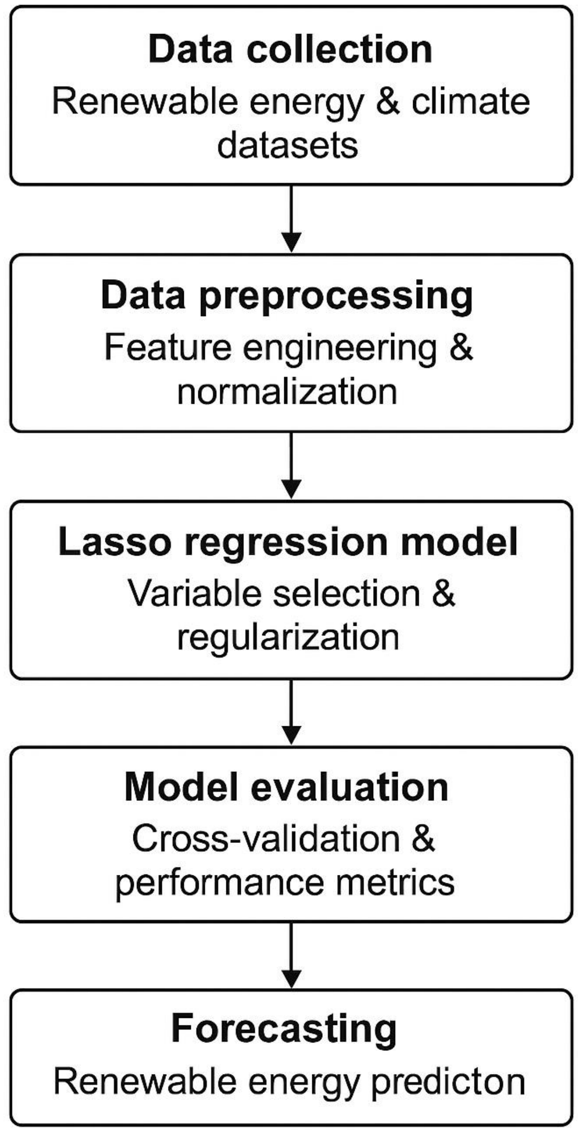
Two forms will be evaluated:

- **λ_min:** Value giving lowest CV error
- **λ_1SE:** Simpler model within one standard deviation of minimum error

Model Development

- 1. Merge and clean multi-source data
- 2. Feature engineering
- 3. Train/test split
- 4. Hyperparameter tuning
- 5. Model fitting
- 6. Validation

Project plan



Multi-Source Data Integration

- Incorporate heterogeneous inputs: meteorological forecasts, historical generation
- LASSO selects most informative predictors

Forecasting Performance

- Improves accuracy and robustness
- Sparse models focus on meaningful variables
- RMSE reduction, fewer samples, robust to anomalies

Expected Outcomes

- Identification of the most influential meteorological variables affecting renewable energy output.
- A robust and interpretable LASSO regression model capable of forecasting multi-source renewable energy with high accuracy.
- A unified data-driven framework suitable for integration into smart energy management systems.
- Enhanced decision-making for grid operators, energy storage planning, and load balancing.

Conclusion

- LASSO handles high-dimensional multi-source data
- Provides interpretable and robust forecasts
- Enables smart energy management
- Future work: ensemble models, real-time data

Ethical Considerations

- All datasets are open-source and publicly accessible.
- No personal or sensitive human data involved.
- Environmental research ethics maintained.

Thank You