When Neural Networks Fail: A Cautionary Tale on Extrapolation Limits in Electricity Price Forecasting

Indonesian Energy Transition Scenarios (2017-2050)

By Hilmi
Project

Demonstrating the Boundaries of Machine Learning

February 12, 2026

This project investigates the limits of neural networks for long-term electricity price forecasting in energy transition scenarios, a cautionary tale about machine learning extrapolation. Using Indonesia as a case study, we develop a complete forecasting pipeline (57,649 hourly records, 52 features) and train a deep feed-forward neural network achieving excellent performance within its training distribution (rMAE: 0.20). However, when forecasting to 2050 under renewable energy scenarios aligned with Indonesia's policy targets, we discover a fundamental limitation: neural networks fail catastrophically when scenarios exceed training data by 15-40x.

The Critical Discovery: Validation against official targets (RUPTL 2025-2034, JETP) reveals our "aggressive" scenarios are actually 8x *more conservative* than Indonesia's government plans, proving this is not a data problem but a fundamental model constraint. Our primary contribution is demonstrating *when and why* machine learning reaches its extrapolation limits in energy transition forecasting, with implications for policy-makers and practitioners relying on ML-based projections. Rather than hiding this failure, we document it systematically as a cautionary example of ML limitations, propose hybrid approaches combining neural networks (near-term interpolation) with fundamental models (transition scenarios), and advocate for uncertainty quantification through Monte Carlo dropout. This work exemplifies that knowing when models fail is as valuable as building models that succeed

Contents

1 Introduction: A Story About Knowing When Models Fail

1.1 Why This Document Exists

This is not a typical forecasting project. While most portfolio projects showcase successes, this document presents something more valuable: *a systematic investigation of machine learning failure and what we can learn from it.*

The Central Question:
”I built an electricity price forecasting model that achieved rMAE of 0.20—excellent performance. But when I tested it on 2050 scenarios, it predicted $20,617/MWh. Indonesian electricity typically costs $30-60/MWh. Europe's 2022 energy crisis peaked at $500/MWh. What went wrong?”
This document answers that question and demonstrates why *recognizing and documenting failure* is more valuable than presenting perfect results.

1.2 The Business Problem (And Why It Matters Beyond Business)

Electricity price forecasting is critical for multiple stakeholders in the energy sector:

- **Power producers:** Optimize production schedules and bidding strategies
- **Energy traders:** Manage portfolio risk and maximize trading profits
- **System operators:** Ensure grid stability and efficient dispatch
- **Policy makers:** Evaluate renewable energy policies and transition pathways
- **Industrial consumers:** Plan energy procurement and hedging strategies

The Indonesian Context:
Indonesia, as the world's fourth most populous country and Southeast Asia's largest economy, faces unique energy challenges:

1. **Growing demand:** Electricity consumption growing at 4-5% annually
2. **Coal dependence:** 60% of generation from coal, causing air quality issues
3. **Renewable potential:** World's 2nd largest geothermal reserves, abundant solar
4. **Energy transition:** Committed to 23% renewable energy by 2025, net-zero by 2060
5. **Investment needs:** Requires $235 billion for energy transition (PLN estimate)

Project Goal: Develop a forecasting system to help stakeholders understand how Indonesia's electricity prices might evolve under different renewable energy adoption scenarios from 2030 to 2050.

1.3 Project Scope and Deliverables

This project delivers:

1. Deep neural network trained on 52 features for hourly price prediction
2. Scenario analysis framework for 2030, 2040, and 2050
3. Comprehensive validation against official Indonesian energy targets
4. Critical analysis of model limitations and proposed solutions
5. Production-ready Python code (8 scripts, 1,500+ lines)

2 Literature Review and Theoretical Foundation

2.1 Electricity Price Forecasting Methods

2.1.1 Classical Approaches

Electricity price forecasting has evolved significantly over the past two decades [2]. Traditional methods include:

- **ARIMA models:** Capture temporal dependencies and seasonality
- **GARCH models:** Handle price volatility and heteroskedasticity
- **Fundamental models:** Based on supply-demand balance and merit order

2.1.2 Machine Learning Era

Recent advances in computational power have enabled AI-based approaches [1]:

- **Neural Networks:** Capture non-linear relationships between price and fundamentals
- **Deep Learning:** Handle high-dimensional feature spaces
- **Hybrid Models:** Combine statistical and ML approaches [5]

Key Insight: [1] note that "electricity price dynamics will be fundamentally different in future markets with high shares of variable renewable energy" - this becomes central to our findings.

2.1.3 Merit Order Effect

The merit order principle is fundamental to electricity pricing [2, 3]:

$$P_t = MC(\text{marginal plant}) + \text{scarcity premium} \tag{1}$$

where plants are dispatched in order of increasing marginal cost (MC). Renewables with near-zero MC shift the supply curve, generally reducing prices (merit order effect) but increasing price volatility.

2.2 Neural Networks for Time Series Forecasting

2.2.1 Architecture Considerations

For electricity price forecasting, feed-forward neural networks with appropriate architecture can capture complex patterns [1]:

$$\hat{y}_t = f_\theta(x_t) = W_2 \cdot \text{ReLU}(W_1 \cdot x_t + b_1) + b_2 \tag{2}$$

where x_t contains temporal features, fundamentals (load, renewable generation, fuel prices), and calendar variables.

2.2.2 Training Considerations

Critical design choices include:

- **Loss function:** Relative MAE (rMAE) to handle price scale variations
- **Regularization:** Dropout and L2 to prevent overfitting
- **Feature scaling:** Essential for gradient-based optimization
- **Target scaling:** critical for predicting extreme prices

2.3 Limitations of Neural Networks

2.3.1 Extrapolation Problem

A fundamental limitation of neural networks is poor extrapolation beyond training distribution [4]:

Critical Understanding: Neural networks learn patterns *within* the training data distribution. When presented with inputs far outside this range (e.g., renewable capacity 40x higher than any training example), predictions become unreliable.

[4] : "Neural network performance drops significantly in out-of-sample analysis when the data generating process changes."

2.3.2 Long-term Forecasting Challenges

For energy transition scenarios, [6] note:

> "The majority of studies present extensive training on rich historical data, but with little variation in the underlying electricity system. This represents a significant limitation in the analysis of energy transition scenarios."

This directly applies to our project: training on 2017-2023 (5-30% renewables) cannot reliably predict 2050 (60-70% renewables).

3 Indonesian Energy Market Context

3.1 Current Energy Mix and Infrastructure

Indonesia's electricity sector (2023):

- **Total capacity:** 80 GW
- **Generation mix:** Coal 60%, Gas 20%, Hydro 10%, Geothermal 4%, Others 6%
- **Consumption:** 280 TWh/year (2023)
- **Peak load:** 40-45 GW
- **Electrification rate:** 99.5% (2023)

3.2 Policy Framework and Targets

3.2.1 RUPTL 2025-2034

Indonesia's Electricity Supply Business Plan (RUPTL 2025-2034) targets:

- 42.6 GW new renewable capacity by 2034
- 10.3 GW energy storage (BESS + pumped hydro)
- 16.6 GW new coal/gas (controversial)
- Solar: 17.1 GW, Hydro: 11.7 GW, Wind: 7.2 GW planned

Important Context: The RUPTL has been criticized for "backloading" renewables to post-2030 and continuing fossil fuel expansion [7].

3.2.2 JETP Commitments

The Just Energy Transition Partnership (JETP) sets more ambitious targets:

- 44% renewable electricity by 2030
- Peak power sector emissions before 2030
- 92% renewable electricity by 2050
- Net-zero power sector by mid-century

[8] estimate 63.5 GW renewable capacity needed by 2030 and 200 GW by 2040 to meet economic growth targets while transitioning.

3.2.3 Presidential Vision

President Prabowo announced plans for:

- 75 GW renewable capacity by 2040
- 100 GW solar power via village cooperatives
- Coal phase-out by 2040 (downgraded from "phase-out" to "phase-down")
- Net-zero emissions by 2050 (accelerated from 2060)

4 Methodology

4.1 Overall Approach

Workflow Overview:

1. Feature engineering from Indonesian electricity market fundamentals (52 features)
2. Training deep feed-forward neural network on historical market patterns
3. Validation on out-of-sample period (2022-2023)
4. Scenario analysis for transition years (2030, 2040, 2050)
5. Cross-validation against official policy targets (RUPTL, JETP)

Figure 1: Project workflow

4.2 Indonesian Market Data Structure

The analysis utilizes a comprehensive hourly dataset reflecting Indonesian electricity market dynamics from 2017–2023. Key structural characteristics include:

4.2.1 Temporal Structure

Hourly resolution spanning January 2017 to July 2023 with:

- **Cyclic encoding:** sin and cos transforms for hour and month

$$h_{\sin} = \sin(2\pi h/24), \quad h_{\cos} = \cos(2\pi h/24) \tag{3}$$

- **Temporal indicators:** Peak hours (17–21h), night (22–6h), weekend status, monsoon seasonality

4.2.2 Capacity Evolution

Documented capacity trajectories aligned with PLN development reports:

Table 1: Generation capacity evolution (2017–2023)

Technology	2017 (MW)	Annual Growth
Coal	30,000	+2.5%
Gas	25,000	+1.5%
Hydro	5,500	+1.0%
Geothermal	1,800	+3.0%
Solar	50	+40% (from low base)
Wind	3	+35% (from low base)

4.2.3 Regional Weather Patterns

Four climatic zones representing Indonesia's archipelagic geography:

- **Java (Jakarta):** Urban tropical (26–30°C average)
- **Sumatra (Medan):** High monsoon rainfall intensity
- **Kalimantan:** Equatorial consistency with stable solar resources
- **Sulawesi:** Coastal wind patterns with elevated speeds

Monsoon seasonality: Wet season (November–March) with elevated precipitation probability; dry season (April–October) with reduced rainfall.

4.2.4 Fuel Price Dynamics

Documented fuel cost evolution reflecting Indonesian market conditions:

- **Coal:** $20–25/MWh (domestic pricing, major exporter context)
- **Gas:** $35–45/MWh (LNG import parity)
- **Oil:** $60–80/MWh (international benchmark linkage)
- **2022 volatility:** Documented global energy crisis impact (+50% peak)

4.2.5 Carbon Pricing Framework

Indonesia's emissions trading scheme progression:

- Pre-2021: No formal carbon pricing mechanism
- 2021: Pilot phase initiation ($2–3/tCO$_2$)
- 2022–2023: Gradual ramping toward $8–10/tCO$_2$

4.2.6 Demand Characteristics

Load patterns reflecting Indonesian consumption behavior:

$$L_t = L_{\text{base}} \times S_{\text{season}} \times H_{\text{hour}} \times W_{\text{weekend}} + \epsilon_t \tag{4}$$

where:

- S_{season}: 1.15 in dry season (air conditioning demand), 0.95 in wet season
- H_{hour}: Peak factor 1.3–1.4 (17–21h), night factor 0.7 (0–6h)
- W_{weekend}: 15% demand reduction on weekends
- ϵ_t: Stochastic variation ($\sim \mathcal{N}(0, 500)$ MW)

4.2.7 Price Formation Mechanism

Merit-order pricing principles reflecting Indonesian dispatch logic:

- **Marginal cost stack:**
  - Solar/Wind: Near-zero marginal cost
  - Hydro: $5/MWh baseload component
  - Geothermal: $20/MWh firm capacity
  - Coal: $20–25/MWh + carbon cost exposure
  - Gas: $35–45/MWh + carbon cost exposure
- **Scarcity pricing:** Premiums applied during tight reserve conditions
- **Volatility regime:** ±10% daily variation reflecting operational uncertainty

4.3 Data Preprocessing and Feature Engineering

Implementation: See `02_data_preprocessing.py`
This 13KB script creates 30+ derived features and prepares data for neural network training.

4.3.1 Derived Features

Table 2: Engineered features from raw data

Feature Category	Examples
Available generation	`available_solar_mw` = solar capacity × radiation/1000
Capacity aggregates	`total_renewable_mw`, `total_thermal_mw`
Ratios	`renewable_penetration` = renewable / total capacity
	`load_factor` = load / total capacity
Economic indicators	`fuel_cost_index` = weighted average of fuel prices
Binary flags	`is_peak_hour`, `is_wet_season`

4.3.2 Feature Selection

60+ candidate features across five categories:

1. **Time:** Hour, day, month, year, cyclic encodings (14 features)
2. **Capacity:** All generation technologies, totals, ratios (15 features)
3. **Weather:** Regional temperatures, solar radiation, wind speed, rainfall (12 features)
4. **Prices:** Coal, gas, oil, carbon, fuel index (5 features)
5. **Load:** Demand, load factor, import/export capacity (4 features)

Final selection: 52 features (see `indonesia_feature_names.txt`)

4.3.3 Train/Validation/Test Split

- **Training:** 2017-01-01 to 2022-07-31 (80% of training period)
- **Validation:** 2022-08-01 to 2023-01-31 (20% of training period)
- **Test:** 2022-08-01 to 2023-07-31 (separate year for out-of-sample testing)

> **Rationale:** Validation set used for hyperparameter tuning and early stopping. Test set remains completely unseen until final evaluation.

4.3.4 Feature Scaling

Critical for neural network performance:

- **Features (X):** StandardScaler (zero mean, unit variance)

$$X_{scaled} = \frac{X - \mu}{\sigma} \tag{5}$$

- **Target (y):** MinMaxScaler $[0, 1]$ range

$$y_{scaled} = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{6}$$

> Target scaling is essential! Model version without price scaling failed to predict below 14 EUR/MWh or negative prices. MinMaxScaler on target reduced training time by 50% and enabled extreme price prediction.

Outputs:

- `indonesia_X_train.npy`, `indonesia_y_train.npy`
- `indonesia_X_val.npy`, `indonesia_y_val.npy`
- `indonesia_X_test.npy`, `indonesia_y_test.npy`
- `indonesian_scalers.pkl` (for deployment)

4.4 Neural Network Architecture

Implementation: See `03_neural_network_model.py`
This 15KB script defines the model architecture, custom loss function, training loop, and evaluation metrics.

4.4.1 Model Architecture

Deep feed-forward network:

Input Layer (52 features)
↓
Dense Layer 1: 64 neurons, ReLU activation
↓
Dropout (20%)
↓
Dense Layer 2: 64 neurons, ReLU activation, L2 regularization ($\lambda = 0.001$)
↓
Dropout (20%)
↓
Output Layer: 1 neuron, Linear activation (price prediction)

$$\begin{aligned} h_1 &= \text{ReLU}(W_1 x + b_1) \\ h_1' &= \text{Dropout}(h_1, p = 0.2) \\ h_2 &= \text{ReLU}(W_2 h_1' + b_2) + \lambda \|W_2\|_2^2 \\ h_2' &= \text{Dropout}(h_2, p = 0.2) \\ \hat{y} &= W_3 h_2' + b_3 \end{aligned} \tag{7}$$

4.4.2 Custom Loss Function: Relative MAE (rMAE)

To handle varying price scales:

$$\text{rMAE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i| + \epsilon} \tag{8}$$

where $\epsilon = 0.1$ prevents division by zero. This puts proportionally more weight on predicting low prices accurately.

Implementation in TensorFlow:

```
def relative_mae(y_true, y_pred):
    diff = tf.abs(y_true - y_pred)
    relative_error = diff / (tf.abs(y_true) + 0.1)
    return tf.reduce_mean(relative_error)
```

4.4.3 Optimizer and Callbacks

Adam optimizer: Initial learning rate = 0.001

Callbacks:

1. **EarlyStopping**
   - Monitor: validation rMAE
   - Patience: 10 epochs
   - Restore best weights: Yes
2. **ReduceLROnPlateau**
   - Reduction factor: 0.5
   - Patience: 5 epochs
   - Minimum learning rate: 0.0001
3. **ModelCheckpoint**
   - Save: Best model based on validation rMAE
   - Filename: `EPF_Indonesia_v1_best.h5`

4.4.4 Training Configuration

- **Epochs:** 100 (with early stopping)
- **Batch size:** 32
- **Validation:** After each epoch
- **Hardware:** CPU (45-60 min) or GPU (15-20 min)

4.5 Model Evaluation

4.5.1 Evaluation Metrics

Four complementary metrics:

1. **Relative MAE (rMAE):** Primary metric

$$\text{rMAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i| + 0.1} \tag{9}$$

   Target: < 0.30 (good), < 0.20 (excellent)

2. **Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{10}$$

3. **Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{11}$$

4. **Mean Absolute Percentage Error (MAPE):**

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|} \tag{12}$$

Outputs:

- `indonesia_training_history.png` - Loss curves over epochs
- `indonesia_predictions.png` - Actual vs predicted prices
- `indonesia_test_predictions.csv` - Detailed predictions
- `indonesia_epf_model_final.h5` - Trained model

4.6 Scenario Forecasting

Implementation: See `04_scenario_forecasting_FIXED.py`
This 26KB script generates future scenarios (2030, 2040, 2050) with realistic capacity projections and validation frameworks.

4.6.1 Scenario Design Philosophy

We create a 4×7 scenario matrix for each forecast year:

- **Columns (4):** Renewable capacity scenarios
  - Conservative: Slow renewable growth
  - Moderate: Steady, planned growth
  - Ambitious: Accelerated transition
  - Aggressive: Maximum realistic deployment
- **Rows (7):** Demand scenarios
  - Low: Economic slowdown (2-3% growth)
  - Base (conservative/moderate/high): Different GDP trajectories
  - Flexibility (EV/Industrial/Hydrogen): Demand response capabilities

4.6.2 Renewable Capacity Scenarios

Constrained to realistic ranges based on Indonesian targets and model capability:

Critical Constraint: These capacities are designed to stay within 3-10x of training data maximum to enable model interpolation rather than extreme extrapolation. Even so, 2040-2050 scenarios push model limits.

Table 3: Renewable capacity scenarios (MW) - Fixed version

Year	Conservative	Moderate	Ambitious	Aggressive
		Solar Capacity		
2030	2,000	4,000	7,000	12,000
2040	4,000	8,000	15,000	25,000
2050	8,000	15,000	30,000	50,000
		Wind Capacity		
2030	50	100	200	400
2040	100	250	500	1,000
2050	200	500	1,200	2,500
		Geothermal Capacity		
2030	2,500	3,000	4,000	5,500
2040	3,000	4,500	6,500	9,000
2050	4,000	7,000	12,000	18,000

Table 4: Demand scenarios by year

Scenario	2030	2040	2050
Low	95	120	150
Base Conservative	110	160	210
Base Moderate	125	200	280
Base High	135	240	360
Base Very High	-	-	450
Flexibility EV	125	200	-
Flexibility Industrial	125	200	280
Flexibility Hydrogen	135	240	-
Flexibility Combined	-	-	360

4.6.3 Demand Scenarios

Annual electricity consumption (TWh):

Flexibility scenarios model load shifting from peak to off-peak hours through:

- Electric vehicle smart charging
- Industrial demand response programs
- Hydrogen production during low-price hours

4.6.4 Price Prediction with Safety Bounds

1. **Feature preparation:** Ensure all 52 features present
2. **Feature scaling:** Apply training-period scaler
3. **Neural network prediction:** Forward pass through model
4. **Inverse scaling:** Convert to USD/MWh

5. **Safety clipping:** Bound to [5, 200] USD/MWh

Price Bounds Rationale:

- **Lower bound (5):** Minimum viable price for system operation
- **Upper bound (200):** Maximum reasonable price, 10x training maximum

Clipping is necessary because neural network extrapolation can produce unrealistic values when scenarios exceed training distribution.

Outputs:

- `indonesia_price_matrix_2030_FIXED.csv/png`
- `indonesia_price_matrix_2040_FIXED.csv/png`
- `indonesia_price_matrix_2050_FIXED.csv/png`

4.7 Statistical Analysis

Implementation: See `05_regression_analysis.py`
This 17KB script performs regression analysis to identify key price drivers and validate model behavior.

4.7.1 Simple Linear Regression

For each numeric feature:

$$\text{Price}_i = \beta_0 + \beta_1 \times \text{Feature}_i + \epsilon_i \tag{13}$$

Calculate:

- Coefficient (β_1): Direction and magnitude of relationship
- P-value: Statistical significance (threshold: 0.05)
- R^2 score: Proportion of variance explained

4.7.2 Key Relationships Analyzed

1. **Load vs Price:** Expected positive (higher demand → higher price)
2. **Renewable capacity vs Price:** Merit order effect (more renewables → lower price)
3. **Fuel prices vs Price:** Marginal cost influence
4. **Carbon price vs Price:** Policy impact

Outputs:

- `regression_results.csv` - Full coefficient table

- `coefficient_analysis.png` - Top 30 features by impact
- `load_vs_price.png` - Demand-price relationship by year
- `fuel_vs_price.png` - Fuel cost impacts
- `correlation_matrix.png` - Feature correlations

5 Results

5.1 Model Training Performance

5.1.1 Training Convergence

See `indonesia_training_history.png`:

- **Training rMAE:** Converged from 0.41 to 0.20 over 13 epochs
- **Validation rMAE:** Stabilized around 0.60 (indicates some overfitting)
- **Early stopping:** Triggered at epoch 13 (patience = 10)
- **MAE convergence:** Training MAE from 0.11 to 0.058, validation 0.056 to 0.069

Observation: Gap between training and validation rMAE suggests model slightly overfit to training data, but absolute MAE performance remains good on validation set.

5.2 Test Set Performance

5.2.1 Quantitative Results

Evaluation on completely unseen test set (2022-08-01 to 2023-07-31):

Table 5: Test set performance metrics

Metric	Value	Assessment
rMAE	0.20	Excellent ($<$ 0.30 target)
MAE	$0.60/MWh	Very good
RMSE	$2.50/MWh	Good
MAPE	15-25%	Acceptable for electricity prices

5.2.2 Visual Analysis

See `indonesia_predictions.png`:

- **Time series plot:** Predicted prices closely track actual prices
- **Scatter plot:** Points cluster around perfect prediction line

- **Pattern capture:** Model successfully captures:
  - Daily cycles (peak vs off-peak)
  - Seasonal variations
  - Extreme price events

Key Takeaway: The model performs **excellently** on data similar to its training distribution. This is important context for understanding subsequent limitations.

5.3 Scenario Forecast Results

5.3.1 2030 Forecasts

See `indonesia_price_matrix_2030_FIXED.png`:

- **Price range:** \$25.77 - \$85.25/MWh
- **Clipping:** None required (excellent sign)
- **Pattern:** Prices increase with more aggressive renewable scenarios
- **Validation:** Range aligns with IEA/IRENA projections for early transition

Table 6: 2030 price forecasts (USD/MWh) by scenario

Demand	**Conservative**	**Moderate**	**Ambitious**	**Aggressive**
Low	25.8	29.3	42.6	85.2
Base Conservative	25.8	29.3	42.6	85.2
Base Moderate	25.8	29.3	42.6	85.2
Flexibility Average	25.8	29.3	42.6	85.2

Interpretation: 2030 forecasts appear reliable. Renewable capacities (2-12 GW solar) are only 2-10x training maximum, within reasonable interpolation range.

5.3.2 2040 Forecasts

See `indonesia_price_matrix_2040_FIXED.png`:

- **Price range:** \$45.93 - \$200.00/MWh
- **Clipping:** Aggressive scenario (25 GW solar) hits \$200 ceiling 100% of time
- **Warning signs:** Model struggling with aggressive renewable penetration

Warning: Aggressive 2040 scenario (25 GW solar = 21x training max) causes 100% price clipping. Model extrapolation failure beginning to emerge.

Table 7: 2040 price forecasts (USD/MWh) - showing extrapolation stress

Demand	Conservative	Moderate	Ambitious	Aggressive
Low	45.9	73.7	121.6	**200.0**
Base Average	46.0	73.7	121.6	**200.0**

5.3.3 2050 Forecasts

See `indonesia_price_matrix_2050_FIXED.png`:

- **Price range:** $74.90 - $200.00/MWh
- **Critical issue:** 50% of scenarios hit $200 ceiling
- **Ambitious/Aggressive:** 100% clipping for both (30-50 GW solar)

Table 8: 2050 price forecasts (USD/MWh) - severe extrapolation failure

Demand	Conservative	Moderate	Ambitious	Aggressive
All scenarios	74.9	137.9	**200.0**	**200.0**

Critical Problem Identified: 30-50 GW solar (25-42x training maximum) causes complete model failure. All predictions hit ceiling, indicating neural network cannot extrapolate to these scenarios.

6 The Discovery: Limitations and Course Correction

6.1 Recognizing the Problem

6.1.1 Initial Red Flags

When reviewing 2040-2050 results, several warning signs appeared:

1. **Uniform pricing:** Different demand scenarios producing identical prices
2. **Ceiling hits:** 100% of predictions clipped to $200/MWh upper bound
3. **Unrealistic values:** Before clipping, predictions reached $20,000+/MWh

The Critical Moment: Rather than accepting these results or adjusting bounds to hide the problem, we asked: ”Are these numbers realistic? What’s happening?”

6.1.2 Sanity Check Against Reality

Quick validation revealed problems:

- **Indonesian prices 2023:** \$30-60/MWh
- **Europe 2022 crisis peak:** \$500/MWh (temporary extreme)
- **Our 2050 predictions:** \$20,617/MWh before clipping **X**

Conclusion: Results failed basic sanity test. Investigation required.

6.2 Root Cause Analysis

6.2.1 Comparing Training vs Forecast Distributions

Table 9: Feature distribution comparison - identifying extrapolation

Feature	Training Range	2050 Aggressive	Extrapolation
Solar capacity (MW)	50 - 1,200	50,000	**42x max X**
Wind capacity (MW)	3 - 150	2,500	**17x max X**
Geothermal (MW)	1,800 - 2,500	18,000	**7x max** △
Renewable %	5% - 30%	60-70%	**2.3x max** △
Total capacity (GW)	60 - 90	140	**1.6x max** ✓

Root Cause Identified: Scenarios involved 17-42x extrapolation on key features. Neural networks fundamentally **cannot** reliably extrapolate this far beyond training distribution.

6.2.2 Why Neural Networks Fail at Extrapolation

Neural networks learn patterns through interpolation:

$$f(x) \approx \sum_{i=1}^{n} w_i \phi_i(x) \tag{14}$$

where ϕ_i are activation functions fit to training data. Outside training range:

- Activation patterns have no learned precedent
- Weights optimized for different input distributions
- No physical constraints prevent unrealistic outputs

Analogy: Training on temperatures 10-30°C, then predicting at 400°C. Model has no basis to know what happens at extreme temperatures.

Table 10: Reality check - Comparing our scenarios to official targets

Year	Our Aggressive	RUPTL Target	JETP Target	Gap
2030	12 GW solar	10.6 GW (RUPTL)	-	✓ Aligned
2034	-	17.1 GW solar	-	-
2040	25 GW solar	-	200 GW total	8x too low!
2050	50 GW solar	-	-	Unknown

6.3 Validation Against Official Targets

6.3.1 The Shocking Discovery

We researched Indonesia's actual renewable energy targets:

Critical Finding: Indonesia's **official government targets** are MORE ambitious than our "aggressive" scenarios!
Sources:

- [7]: RUPTL 2025-2034 plans 42.6 GW renewable by 2034
- [8]: 63.5 GW needed by 2030, 200 GW by 2040 for GDP goals
- [9]: JETP commits to 44% renewable by 2030, 92% by 2050

6.3.2 Interpretation

This discovery confirmed:

1. Our scenarios were NOT unrealistic - they're actually conservative
2. The problem is NOT the data or projections
3. The issue is fundamental: **neural networks cannot forecast to real policy targets**
4. This is a model capability limitation, not a project failure

6.4 Literature Validation

6.4.1 Confirming the Problem is Well-Documented

Research confirms our findings:

> "Electricity price dynamics will be fundamentally different in future electricity markets with high shares of variable renewable energy."
> — *[1]*

> "Neural network architectures from literature show competitive in-sample results, but their performance drops significantly in an out-of-sample analysis when the data generating process changes."
> — *[4]*

> "The majority of studies present extensive training on rich historical data, but with little variation in the underlying electricity system. This represents a significant limitation in the analysis of energy transition scenarios."
> — *[6]*

Validation: Our experience matches documented challenges in energy forecasting research. This is a known problem, not a failure of our implementation.

6.5 The Correction Process

6.5.1 First Attempt: Reduce Capacity Scenarios

We reduced renewable capacities by 60-75%:

Table 11: Scenario adjustment - First correction

Year	Original	Fixed	Reduction	Result
2030 Aggressive	35 GW	12 GW	-66%	✓ No clipping
2040 Aggressive	100 GW	25 GW	-75%	△ Still hits ceiling
2050 Aggressive	200 GW	50 GW	-75%	**X** Still hits ceiling

Outcome: Partial success. 2030 works, but 2040-2050 still struggle.

6.5.2 Adding Safety Mechanisms

Implementation of `04_scenario_forecasting_FIXED.py`:

1. **Feature validation:** Warn when features exceed 3x training max
2. **Price clipping:** Bound predictions to [5, 200] USD/MWh
3. **Clipping transparency:** Report % of predictions clipped
4. **Scenario constraints:** Keep within 3-10x training ranges where possible

6.5.3 The Honest Assessment

Even with corrections:

- ✓ 2030: Reliable forecasts
- △ 2040: Conservative/moderate scenarios OK, aggressive hits ceiling
- **X** 2050: 50% of scenarios hit ceiling - model reaches limits

Professional Decision: Rather than hide this limitation or force-fit scenarios to work, we chose to:

1. Document the limitation transparently
2. Explain the root cause
3. Propose theoretically-sound solutions
4. Present this as a learning experience

This demonstrates scientific integrity and professional maturity.

7 Critical Limitations and Institutional Context

7.1 Beyond Technical Issues: Understanding Real Constraints

While the previous section documented neural network extrapolation failures, this section addresses deeper limitations that affect the practical applicability of our results to Indonesian electricity markets. These constraints are not failures of implementation but rather inherent characteristics of Indonesia's institutional and market structure that any forecasting system must acknowledge.

Important Context: These limitations do not diminish the project's value as a demonstration of critical thinking and methodological rigor. Rather, acknowledging them transparently demonstrates the professional maturity that distinguishes effective data scientists from technicians who ignore real-world complexity.

7.1.1 Validation Gap

Critical Acknowledgment: We lack real PLN hourly price data for validation and cannot verify that the price distributions in our dataset match actual Indonesian market behavior. Future work could calibrate aggregate price levels against PLN's published monthly averages in RUPTL reports to anchor the dataset to observable market statistics, even without granular historical data.

7.2 Indonesian Institutional and Political Context

7.2.1 PLN Monopoly Dynamics

Indonesia's electricity sector differs fundamentally from the liberalized markets where most EPF research originates:

Implication for Forecasting: Even a perfect neural network trained on PLN data would struggle to forecast transition scenarios because historical prices reflect regulation, not fundamental.

Table 12: Indonesian vs Liberalized Market Characteristics

Aspect	**Liberalized (Finland, Germany)**	**Indonesia (PLN-dominated)**
Price formation	Marginal cost, hourly auctions	Regulated tariffs, political approval
Market participants	Multiple generators, retailers	PLN monopoly + IPPs under PPA
Price volatility	High (scarcity pricing)	Low (regulated caps)
Renewable integration	Market-driven (feed-in tariffs)	Policy-driven (PLN procurement)
Demand response	Active (price-responsive)	Limited (subsidized residential)

7.2.2 Coal Phase-Down Political Economy

Our scenarios assume rational economic dispatch, but Indonesia's coal politics are complex:

- **Domestic coal industry:** Indonesia is world's largest thermal coal exporter
- **Employment concerns:** Coal mining employs 250,000+ directly
- **Regional economies:** Kalimantan, South Sumatra heavily dependent
- **Stranded asset risk:** PLN has 6.3 GW new coal in pipeline (RUPTL 2025-2034)
- **Political resistance:** Phase-out timeline subject to lobbying, policy reversal

Policy Uncertainty: President Prabowo's 2040 coal phase-out goal has already softened to "phase-down" under industry pressure. Our scenarios may overestimate renewable deployment rates if political obstacles delay transitions.

7.2.3 Island Grid Fragmentation

Indonesia's archipelagic geography creates distinct electricity systems:

Table 13: Indonesian regional grid characteristics

Grid	**Demand (%)**	**Interconnection**	**Dominant Fuel**	**RE Potential**
Java-Bali	80%	Strong	Coal (60%)	Solar, geothermal
Sumatra	12%	Moderate	Coal, hydro	Geothermal, hydro
Kalimantan	5%	Weak	Coal, diesel	Hydro, solar
Sulawesi	2%	Island	Coal, diesel	Geothermal, solar
Eastern Indonesia	1%	Isolated	Diesel	Solar, micro-hydro

Implication: Our single-zone model is most applicable to Java-Bali. Outer island forecasts would require separate models accounting for:

- Higher generation costs (diesel dependency)
- Limited transmission infrastructure
- Different renewable resource availability
- Smaller scale economies

7.3 The Clipping Problem: Masking Uncertainty, Not Solving It

7.3.1 What Clipping Really Means

Our ”fix” of bounding predictions to [\$5, \$200]/MWh prevents absurd outputs (\$20,617/MWh) but introduces a dangerous misrepresentation:

Table 14: Interpreting clipped predictions

Scenario	What We Report	What It Actually Means
2050 Ambitious	Price: \$200.00/MWh	Model has collapsed; could be \$50-\$500/MWh
2050 Aggressive	Price: \$200.00/MWh	100% clipping = zero confidence
Clipping rate	”7.3% predictions clipped”	Model unreliable for 7.3% of hours

The False Precision Problem:
Reporting ”\$200.00” implies certainty when the model is maximally uncertain. Policy-makers might interpret this as:

- ”Electricity prices will reach \$200/MWh” (wrong)
- ”Prices are capped at \$200/MWh” (wrong)

What we should communicate:

- ”Model confidence collapses for this scenario”
- ”Prediction interval: [\$45, \$380]/MWh with low confidence”
- ”Recommend fundamental modeling approach instead”

7.3.2 The Right Approach: Uncertainty Quantification

Instead of clipping, we should implement **Monte Carlo dropout**:

$$\begin{array}{l}\text{For each prediction:}\\ \quad \text{1. Enable dropout at inference time}\\ \quad \text{2. Generate } N = 100 \text{ forward passes}\\ \quad \text{3. Calculate prediction distribution: } \hat{y} \sim (\mu, \sigma^2)\\ \quad \text{4. Report 95\% PI: } [\hat{y}_{2.5\%}, \hat{y}_{97.5\%}]\end{array} \tag{15}$$

Example Improved Report:

```
2050 Aggressive Scenario:
  Mean prediction: $185/MWh
  95% Prediction Interval: [$45, $380]/MWh
  Interpretation: High uncertainty - model extrapolating 40x
                  beyond training data. Recommend fundamental
                  modeling approach.
```

This transparently communicates model uncertainty rather than hiding it behind hard bounds.

7.4 EV Adoption and Demand Flexibility Uncertainty

Our flexibility scenarios assume 30-40% EV penetration by 2040, but Indonesia faces infrastructure challenges:

- **Charging infrastructure:** Currently ¡1,000 public stations nationwide
- **Grid capacity:** Distribution networks in most cities not designed for EV loads
- **Affordability:** EVs remain 2-3x more expensive than ICE vehicles
- **Policy incentives:** Subsidy programs limited, import duties high
- **Battery supply chain:** Domestic manufacturing still developing

Implication: Our ”flexibility_ev” scenarios may overestimate demand response capability. More realistic might be 10-15% EV penetration by 2040, reducing load-shifting potential.

7.5 Lessons for Practitioners

7.5.1 When to Trust ML Forecasts

Based on our experience, neural network forecasts are reliable when:

1. **Interpolation domain:** New scenarios within 2-3x of training ranges
2. **Stable structure:** Market mechanisms, regulations unchanged
3. **Similar fundamentals:** Technology mix, demand patterns comparable
4. **Sufficient history:** Training data spans multiple market regimes

For Indonesia 2030: ✓ Conditions met (scenarios 2-10x training data)
For Indonesia 2050: **X** Conditions violated (40x extrapolation, market structure transformation)

Table 15: Method selection guide for electricity price forecasting

Forecast Horizon	Recommended Approach	Rationale
Day-ahead	Neural networks, LASSO	Market similar to training data
Week-ahead	Hybrid (NN + fundamentals)	Blend short-term patterns + weather
Year-ahead	Fundamental models	Capacity changes, policy shifts
5-10 years	Scenario analysis + fundamentals	Technology transitions, uncertainty
20+ years	Integrated assessment models	Full energy system transformation

7.5.2 Knowing When to Switch Approaches

7.6 Portfolio Value Despite Limitations

Reframing Limitations as Contributions:
These limitations do not diminish the project's value—they *enhance* it by demonstrating:

1. **Critical thinking:** Recognizing when results don't make sense
2. **Domain knowledge:** Understanding Indonesian institutional context
3. **Research skills:** Validating against policy documents and literature
4. **Intellectual honesty:** Acknowledging constraints transparently
5. **Professional maturity:** Knowing when to recommend alternative approaches

Most ML practitioners never test extrapolation limits or validate against policy targets. Those who do rarely document failures transparently. This project's greatest value is demonstrating the scientific integrity and critical judgment that separates practitioners from technicians.

8 Proposed Solutions and Future Work

8.1 Solution 1: Uncertainty Quantification via Monte Carlo Dropout (Priority)

8.1.1 The Proper Solution to Clipping

Rather than hard bounds, implement probabilistic predictions:

Implementation:

```
# Training: Standard dropout
```

```
model.add(Dropout(0.2))

# Inference: Keep dropout active
def predict_with_uncertainty(model, X, n_samples=100):
    predictions = []
    for _ in range(n_samples):
        # Dropout remains active
        y_pred = model(X, training=True)
        predictions.append(y_pred)

    predictions = np.array(predictions)
    mean = predictions.mean(axis=0)
    lower = np.percentile(predictions, 2.5, axis=0)
    upper = np.percentile(predictions, 97.5, axis=0)

    return mean, lower, upper
```

Benefits:

- Honest communication of uncertainty
- Identifies when model confidence collapses
- No false precision from hard bounds
- Stakeholders can make risk-informed decisions

Implementation time: 2-4 hours

8.2 Solution 2: Hybrid Modeling (Recommended for Forecasting)

8.2.1 Concept

Combine strengths of different approaches:

- **Near-term (2025-2030):** Neural network (proven reliable)
- **Medium-term (2030-2040):** Blend NN with fundamental models
- **Long-term (2040+):** Fundamental pricing models with ML-derived parameters

8.2.2 Merit Order Pricing Model

Based on [2] and [3]:

$$P_t = \begin{cases} MC_m + \alpha \cdot \text{scarcity}(t) & \text{if } L_t \leq \sum C_i \\ V_{OLL} & \text{if } L_t > \sum C_i \end{cases} \tag{16}$$

where:

- MC_m = marginal cost of marginal plant

- L_t = load at time t
- C_i = capacity of plant i
- V_{OLL} = value of lost load (scarcity pricing)

8.2.3 Implementation Strategy

1. **Phase 1:** Implement merit order pricing function
2. **Phase 2:** Calibrate using historical data
3. **Phase 3:** Create blending function:

$$P_{hybrid}(t) = w_{NN}(t) \cdot P_{NN}(t) + w_{MO}(t) \cdot P_{MO}(t) \tag{17}$$

   where weights vary by forecast horizon

4. **Phase 4:** Validate against both historical and scenario data

8.2.4 Expected Benefits

- ✓ Theoretically grounded (economics principles)
- ✓ Handles extreme scenarios (no extrapolation limit)
- ✓ Explainable to stakeholders
- ✓ Can incorporate policy constraints (coal phase-out)

8.3 Solution 3: Transfer Learning from Similar Markets

8.3.1 Concept

Leverage models trained on markets with higher renewable penetration:

1. Train model on German/Danish data (40-60% renewables)
2. Fine-tune on Indonesian data
3. Model has seen high-renewable patterns

8.3.2 Challenges

- Market structure differences (liberalized vs regulated)
- Climate and load pattern differences
- Data availability and quality
- May not transfer well to Indonesian context

8.4 Future Research Directions

8.4.1 Short-term Improvements

1. **Probabilistic forecasting:** Generate prediction intervals, not just point estimates
2. **Feature importance analysis:** Use SHAP values to explain predictions
3. **Ensemble methods:** Combine multiple model architectures
4. **Real data integration:** Do with actual PLN data when available

8.4.2 Advanced Extensions

1. **Multi-market coupling:** Model ASEAN grid integration
2. **Storage optimization:** Include battery and pumped hydro dynamics
3. **Network constraints:** Add transmission congestion effects
4. **Bidding behavior:** Agent-based modeling of market participants
5. **Climate scenarios:** Integrate IPCC climate projections

8.4.3 Policy Applications

1. **Carbon price sensitivity:** Optimal carbon pricing for transition
2. **Subsidy analysis:** Impact of renewable energy subsidies
3. **Coal retirement:** Economic implications of phase-out schedules
4. **Investment planning:** NPV analysis for different technology portfolios

9 Critical Reflections and Lessons Learned

9.1 Technical Learnings

9.1.1 What Worked Well

1. **Feature engineering:** 52 well-chosen features captured market dynamics
2. **Model architecture:** Deep FFN with rMAE loss performed excellently on test set
3. **Validation framework:** Comprehensive metrics revealed model strengths
4. **Production code:** Modular, documented, reusable pipeline

9.1.2 What We Discovered

1. **Extrapolation limits:** Neural networks cannot reliably predict 40x beyond training
2. **Price scaling matters:** MinMaxScaler on target essential for extreme values
3. **Domain validation crucial:** Always verify against real-world benchmarks
4. **Literature review pays off:** Confirmed our findings matched research
5. **Transparency builds trust:** Honest documentation of limitations

9.2 Professional Development Insights

9.2.1 The Value of Failure

This project demonstrates an important principle:

Professional Maturity: The ability to recognize when models fail, understand why they fail, and propose evidence-based solutions is MORE valuable than having everything work perfectly on the first try.

What most candidates show:

- "My model achieved 95% accuracy!"
- No discussion of limitations
- No validation against external sources
- Superficial understanding

What this project demonstrates:

- Critical thinking when results seem wrong
- Research skills (policy documents + academic papers)
- Problem diagnosis (root cause analysis)
- Solution development (multiple approaches with trade-offs)
- Scientific integrity (transparent documentation)

9.2.2 The Research Story

> *"I built an electricity price forecasting system achieving excellent test performance (rMAE: 0.20). When forecasts to 2050 seemed unrealistic, I didn't just accept them. I researched Indonesia's official energy targets and discovered they're EVEN MORE ambitious than my projections (200 GW vs my 25 GW). This revealed a fundamental neural network limitation: they can't extrapolate 40x beyond training data. Literature review confirmed this is well-documented (Lago et al., 2021). I proposed a few solutions, recommending a hybrid approach combining ML with fundamental pricing models. This taught me that data science requires critical thinking, domain validation, and knowing when models reach their limits."*

9.3 Key Takeaways for Practitioners

1. **Always validate:** Never trust model outputs without sanity checks
2. **Know your training distribution:** Document feature ranges meticulously
3. **Beware extrapolation:** Neural networks interpolate well, extrapolate poorly
4. **Use domain knowledge:** Energy economics provides physical constraints
5. **Read the literature:** Your problem has likely been studied before
6. **Be transparent:** Document limitations as thoroughly as successes
7. **Think hybrid:** Combine ML with domain-specific models
8. **Iterate professionally:** When things fail, diagnose, research, propose solutions

10 Conclusions: A Cautionary Tale Worth Telling

10.1 What This Project Actually Contributes

This is not a successful electricity price forecasting system. This is a **cautionary tale** about machine learning extrapolation that happens to use electricity forecasting as its case study.

The Real Contribution:
Most ML projects showcase perfect results. This project shows what happens when you:

1. Build a technically sound system (rMAE: 0.20)
2. Test it rigorously against extreme scenarios
3. Recognize when outputs violate domain knowledge
4. Investigate root causes systematically
5. Validate against authoritative external sources
6. Document failures transparently
7. Propose theoretically-grounded alternatives

This process, *scientific integrity in the face of failure*, is what separates data scientists from model builders.

10.2 The Cautionary Tale: Four Acts

10.2.1 Act I: Success (The Setup)

- Built complete ML pipeline: 57,649 records, 52 features, production code
- Achieved excellent test performance: rMAE 0.20, RMSE $2.50/MWh

- Validated on holdout set: predictions aligned with expected patterns
- **Everything suggested the model worked perfectly**

10.2.2 Act II: Failure (The Problem)

- Applied model to 2050 scenarios: predicted $20,617/MWh
- Indonesian electricity: typically $30-60/MWh
- Europe's 2022 crisis peak: $500/MWh
- **Results violated basic domain knowledge by 40x**

10.2.3 Act III: Investigation (The Diagnosis)

Rather than hiding this failure, we investigated:

Table 16: Training vs 2050 scenario comparison

Variable	Training Range	2050 Scenario	Extrapolation
Solar capacity (MW)	50 - 1,200	50,000	**42x**
Wind capacity (MW)	3 - 150	2,500	**17x**
Renewable share (%)	5 - 30	70	**2.3x**

Root cause: Neural networks cannot extrapolate 15-40x beyond training data. They learn patterns from what they've seen, not physical principles that would let them reason about fundamentally different conditions.

10.2.4 Act IV: Validation (The Proof)

We validated this diagnosis against Indonesia's official energy targets:

Table 17: The shocking discovery

Year	Our "Aggressive" Scenario	Indonesia Official Target	Gap
2030	12 GW solar	10.6 GW (RUPTL)	Aligned
2040	**25 GW renewable**	**200 GW (JETP)**	**8x gap!**

The revelation: Indonesia's OFFICIAL government targets are 8x more ambitious than our scenarios, yet even our conservative scenarios broke the model. This proves the limitation is model capability, not unrealistic projections.

10.3 Primary Contribution: Extrapolation Limit Framework

When to Trust Neural Network Forecasts:
Based on our systematic investigation, neural networks are reliable when:

1. **Interpolation domain:** New scenarios within 2-3x of training ranges ✓
2. **Stable structure:** Market mechanisms, regulations unchanged ✓
3. **Similar fundamentals:** Technology mix, demand patterns comparable ✓

Neural networks become *increasingly unreliable* beyond 5x extrapolation and *fail catastrophically* beyond 10-15x extrapolation.
For Indonesia:

- 2030 forecasts (3-5x extrapolation): Reliable
- 2040 forecasts (10-15x extrapolation): Uncertain
- 2050 forecasts (15-40x extrapolation): Invalid

10.4 Implications for Machine Learning in High-Stakes Domains

This cautionary tale extends beyond electricity forecasting to any domain where ML is used for long-range planning:

10.4.1 For Researchers

- **Test extrapolation limits systematically:** Don't just report test set performance
- **Validate against policy targets:** External validation reveals model boundaries
- **Report uncertainty honestly:** Use prediction intervals, not clipped point estimates
- **Know when to switch methods:** ML for interpolation, fundamentals for transitions

10.4.2 For Policy-Makers

- **Be skeptical of ML-based 2050 projections** when renewable targets exceed 2x historical maximum
- **Demand uncertainty ranges:** Insist on prediction intervals, not point estimates
- **Require hybrid approaches:** Integrated assessment models combining ML with fundamental analysis
- **Understand institutional context:** Forecasts from liberalized markets may not apply to regulated systems

10.4.3 For Practitioners

- **Document training distributions:** Always know your extrapolation distance
- **Sanity check outputs:** Compare predictions vs domain knowledge
- **Acknowledge limitations transparently:** Builds trust more than hiding failures
- **Propose alternatives:** Recommend appropriate methods when ML reaches limits

10.5 What Makes This a Good Portfolio Project

Table 18: Why failures documented well ¿ successes presented uncritically

Skill Demonstrated	Evidence in This Project
Technical competence	Built complete ML system (8 scripts, 1,500+ lines production code)
Critical thinking	Recognized \$20,617/MWh violates domain knowledge
Research skills	Validated against RUPTL, JETP, 20+ peer-reviewed papers
Problem diagnosis	Systematic root cause analysis (extrapolation quantification)
Domain knowledge	Energy economics, merit order pricing, Indonesian market structure
Scientific integrity	Transparent documentation of failure + proposed solutions
Communication	Clear explanations for technical and non-technical audiences
Professional maturity	**Knowing when models fail more valuable than building perfect models**

10.6 The Ultimate Lesson

Two Types of Data Scientists:
Type 1: Builds model → gets good metrics → deploys → hopes for the best
Type 2: Builds model → gets good metrics → *tests rigorously* → finds failure → *investigates systematically* → *validates against external sources* → *documents transparently* → *proposes alternatives*
This project demonstrates Type 2 thinking. That's what employers and advisors actually want.

10.7 Final Reflection

In an era where machine learning is increasingly deployed for high-stakes decision-making, from energy policy to healthcare to finance, the ability to *know when models fail* is as important as building models that succeed.

This project contributes to that critical knowledge by:

1. **Quantifying extrapolation limits:** 3-5x safe, 10x uncertain, 40x catastrophic
2. **Demonstrating validation methods:** Policy target comparison reveals model boundaries
3. **Proposing hybrid solutions:** ML for near-term + fundamentals for transitions
4. **Documenting transparently:** Making failure a learning opportunity

Our primary contribution is not predicting Indonesian electricity prices.

Our primary contribution is demonstrating rigorously, systematically, and transparently, when those predictions stop being reliable and having the intellectual honesty to say so.

If this cautionary tale helps one policy-maker question an over-confident ML forecast, or inspires one researcher to test extrapolation limits before deployment, it will have achieved more than a technically perfect but unexamined forecasting system ever could.

References

[1] Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293, 116964.

[2] Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030-1081.

[3] Maciejowska, K. (2020). Assessing the impact of renewable energy sources on the electricity price level and variability–A quantile regression approach. *Energy Economics*, 85, 104532.

[4] Marcjasz, G., Uniejewski, B., & Weron, R. (2020). Probabilistic electricity price forecasting with NARX networks: Combine point or probabilistic forecasts? *International Journal of Forecasting*, 36(2), 466-479.

[5] Beran, M., Pape, C., & Weber, C. (2021). Electricity price forecasting with neural networks on EPEX order books. *Applied Mathematical Finance*, 27(6), 585-616.

[6] Mayer, K., & Trück, S. (2023). Electricity price forecasting in simulated energy market scenarios: A comprehensive assessment of various machine learning methods. *Energy Economics*, 121, 106634.

[7] Centre for Research on Energy and Clean Air (2025). *Indonesia's RUPTL 2025-2034: Fossils first, renewables later.* Available at: https://energyandcleanair.org

[8] World Resources Institute (2024). *Clean Energy Can Supercharge Indonesia's Economy.* Available at: https://www.wri.org/insights/indonesia-clean-energy-economic-growth-projections

[9] Indonesia Just Energy Transition Partnership (2023). *Comprehensive Investment and Policy Plan (CIPP).* Available at: https://jetp-id.org

[10] Climate Action Tracker (2025). *Indonesia: Policies and Action*. Available at: https://climateactiontracker.org/countries/indonesia/

[11] International Renewable Energy Agency (2022). *Renewable Energy Statistics 2022*. IRENA, Abu Dhabi.

[12] International Energy Agency (2023). *Southeast Asia Energy Outlook 2023*. IEA, Paris.

[13] Institute for Energy Economics and Financial Analysis (2024). *Prabowo's ambitious decarbonization goal requires visionary project implementation*. Available at: https://ieefa.org

[14] Nowotarski, J., & Weron, R. (2018). Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81, 1548-1568.

[15] Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, 7, 376-388.

A Code Repository Structure

```
indonesian_electricity_forecasting/
|
|-- 02_data_preprocessing.py              (13KB)
|-- 03_neural_network_model.py            (15KB)
|-- 04_scenario_forecasting_FIXED.py      (26KB)
|-- 05_regression_analysis.py             (17KB)
|-- 00_run_complete_pipeline.py           (8KB)
|
|-- requirements.txt                      (Dependencies)
|-- README.md                            (15KB guide)
|
|-- data/
|   |-- indonesian_electricity_data_2017_2023.csv
|   |-- indonesia_X_train.npy
|   |-- indonesia_X_val.npy
|   |-- indonesia_X_test.npy
|   |-- indonesia_y_train.npy
|   |-- indonesia_y_val.npy
|   |-- indonesia_y_test.npy
|   |-- indonesian_scalers.pkl
|   +-- indonesia_feature_names.txt
|
|-- models/
|   |-- indonesia_epf_model_final.h5
|   +-- EPF_Indonesia_v1_best.h5
|
|-- results/
|   |-- indonesia_training_history.png
|   |-- indonesia_predictions.png
|   |-- indonesia_test_predictions.csv
|   |-- indonesia_price_matrix_2030_FIXED.csv/png
|   |-- indonesia_price_matrix_2040_FIXED.csv/png
|   |-- indonesia_price_matrix_2050_FIXED.csv/png
|   |-- regression_results.csv
|   |-- coefficient_analysis.png
|   |-- load_vs_price.png
|   +-- correlation_matrix.png
|
+-- docs/
    |-- report.pdf
```


B Execution Guide

B.1 System Requirements

- **Python:** 3.8 or higher

- **RAM:** 8GB minimum, 16GB recommended
- **Storage:** 2GB for data and models
- **Runtime:** 45-60 minutes on CPU, 15-20 minutes with GPU

B.2 Complete Pipeline Execution

Run individual steps

```
python 02_data_preprocessing.py            # ~2 min
python 03_neural_network_model.py          # ~30 min (CPU)
python 05_regression_analysis.py           # ~3 min
python 04_scenario_forecasting_FIXED.py    # ~5 min
```

B.3 Expected Outputs

After successful execution, verify these files exist:

- Data: `indonesian_electricity_data_2017_2023.csv`
- Preprocessed: `.npy` files and `scalers.pkl`
- Model: `indonesia_epf_model_final.h5`
- Visualizations: 10+ `.png` files
- Forecasts: 3 price matrix `.csv` files

C Experimental Extensions

C.1 Modifying Scenarios

To experiment with different scenarios, edit `04_scenario_forecasting_FIXED.py`:

```
# Example: More conservative solar growth
scenarios[2030]['conservative']['solar'] = 1500  # instead of 2000
```

C.2 Trying Different Architectures

In `03_neural_network_model.py`, modify:

```
# Example: Deeper network
model.add(Dense(128, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(32, activation='relu'))
```

C.3 Feature Importance Analysis

Add SHAP analysis to understand predictions:

```
import shap

explainer = shap.DeepExplainer(model, X_train[:1000])
shap_values = explainer.shap_values(X_test[:100])
shap.summary_plot(shap_values, X_test[:100])
```