RENEWABLE ENERGY

Mina Carbajal
Carmen Morales-Mier
Gaayathri Rayapaneni
Ignatius Jonathan Sugijono

IFI 8420

Table of Contents

Problem Statement

Can we confidently forecast solar energy growth to considering environmental risks and policy changes, to support strategic financial and operational decisions?

- Assess whether the U.S. solar energy sector will see continued growth in the near future.
- Evaluate how external risk factors like severe weather and public investment affect energy output.
- Support data-driven financial decisions for investment planning, capacity expansion, and policy response.

- Key Business Questions
- How will solar energy generation evolve in the next 5-10 years?
- What are the primary drivers and risks to solar output?
- How can predictive models assist in financial and operational decision-making?

DATA

EIA Monthly Energy Generation

→ Solar energy output by U.S. state and producer type (commercial, residential, etc.).

→ 51,000 rows covering every state and energy source monthly

PV Plant & Extreme Weather Dataset

→ Tracks weather-related disruptions at solar plants: wind speed, rainfall, duration of maintenance delays.

→ 38 columns, 51,505 rows of detailed plant performance data.

Investing in America (IIA) Public Investments Data

→ Federal/state funding by project type, location, and sector.

→ Used to assess investment impact on solar adoption by region.

U.S. Renewable Energy Consumption Data

→ Longitudinal data by sector (residential, industrial, transportation).

→ Key to identifying evolving demand trends for solar integration.

EIA Open Data API

→ Used pagination and filtering to pull >220,000 rows of energy data across multiple years and metrics.

→ Dynamic data extraction (220,000+ rows).

Data Engineering

- Filtered, cleaned, and merged datasets by state and date.
- Removed incomplete and zero-output entries.
- Created derived metrics:
- Storm duration, wind speed, investment intensity.
- Aggregated monthly and annual trends for analysis.

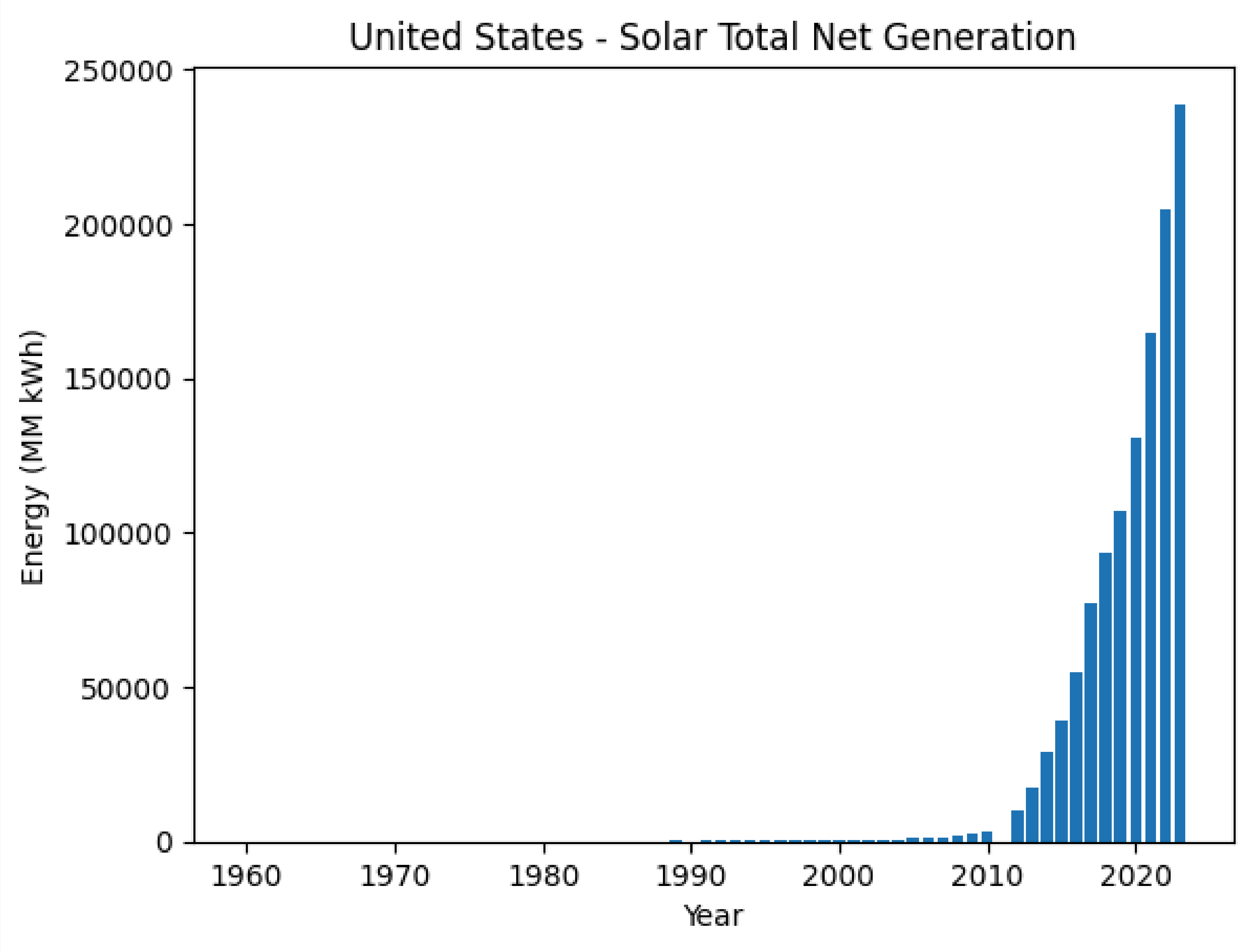
Key Risk Factors

- Weather Disruption: Storms and wind reduce output.
- Policy Volatility: Changes in investment levels shift growth patterns.
- Data Gaps: Grid capacity and infrastructure not fully captured.
- Forecast Horizon: Accuracy declines beyond 18 months.

Analytical Approach

- Linear Regression – Baseline trend analysis
- Random Forest Regression – Non-linear relationships
- Support Vector Regression (SVR) – Stability under volatility

SOLAR ENERGY PRODUCTION



→ Solar energy remained minimal until the late 2000s, with almost no significant contribution before 2005.

→ A sharp rise in production began around 2010, driven by falling solar panel costs and federal incentives.

→ Year-over-year growth accelerated rapidly from 2015 onwards, reflecting major policy shifts and technological adoption.

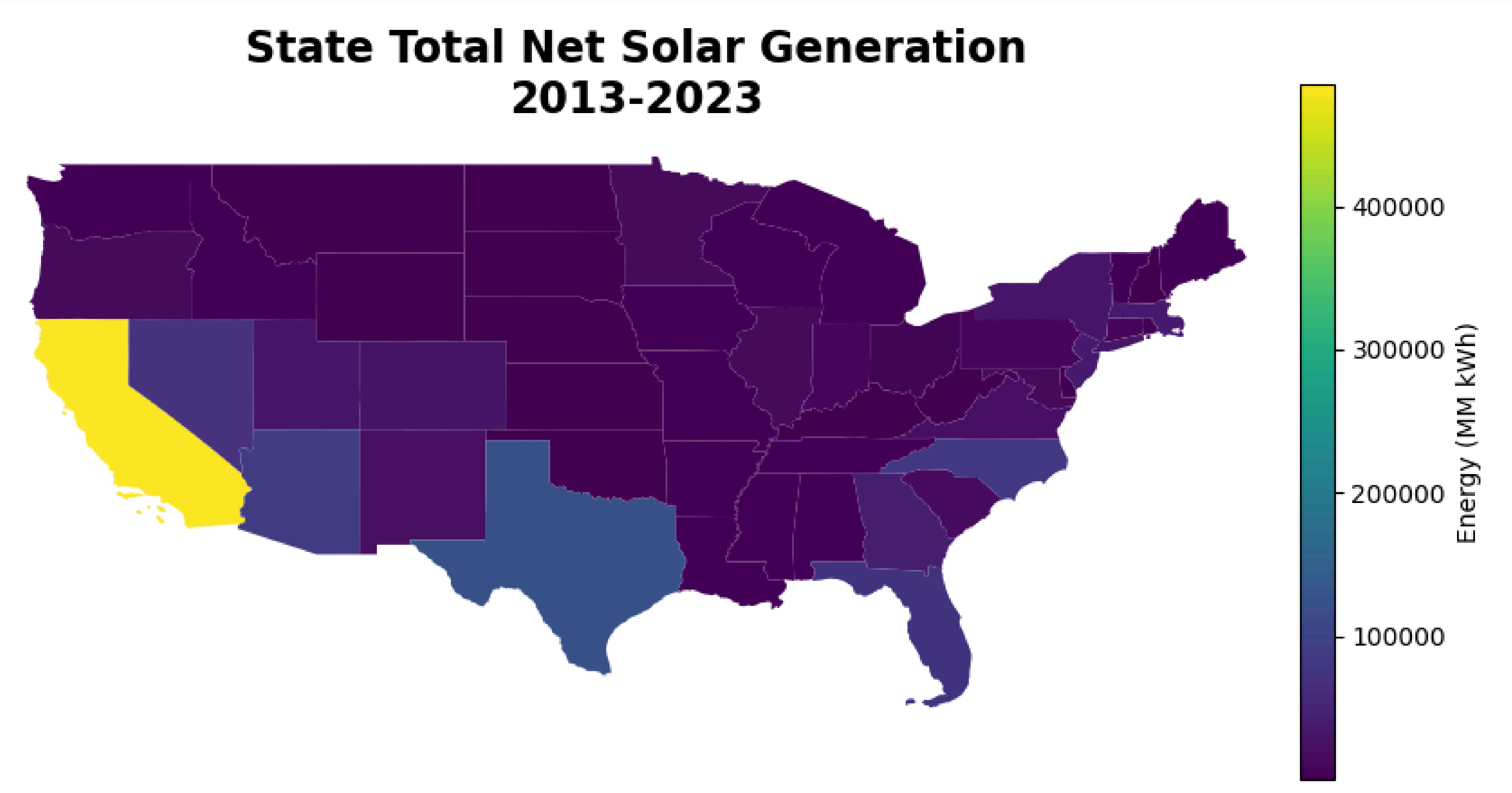
→ By 2023, total net generation exceeded 240,000 MM kWh, marking a major milestone in renewable energy contribution.

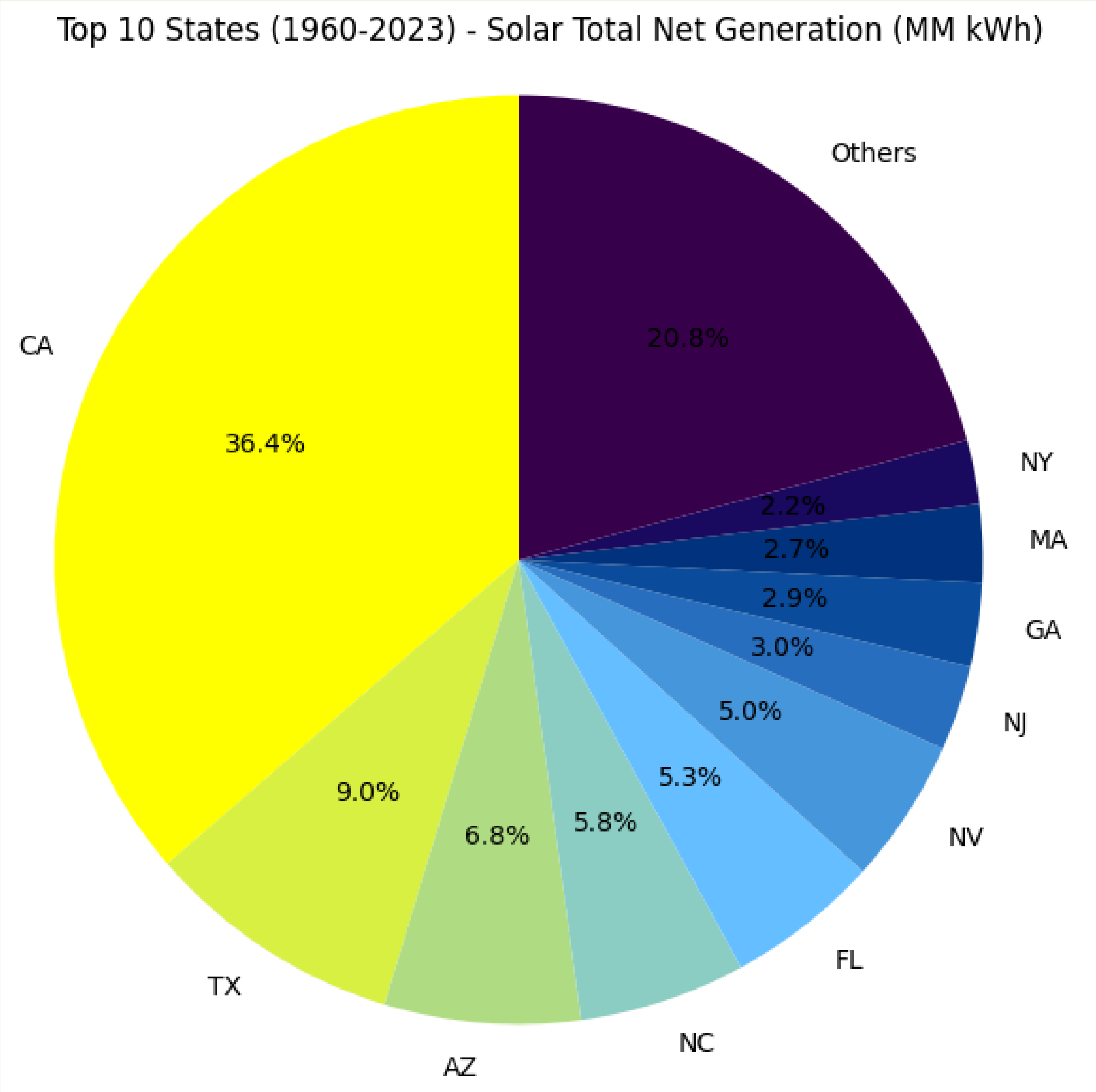
→ The growth trend suggests compounding investments and increasing efficiency in solar infrastructure.

→ Future forecasts must account for this exponential growth pattern, as linear models may underestimate ongoing expansion.

SOLAR PRODUCTION BY STATE

California leads U.S. solar generation with 36.4%, followed by Texas (9%), Arizona (6.8%), and North Carolina (5.8%).





The Top 10 states produce nearly 80% of total solar energy, while many others contribute minimally, showing uneven adoption.

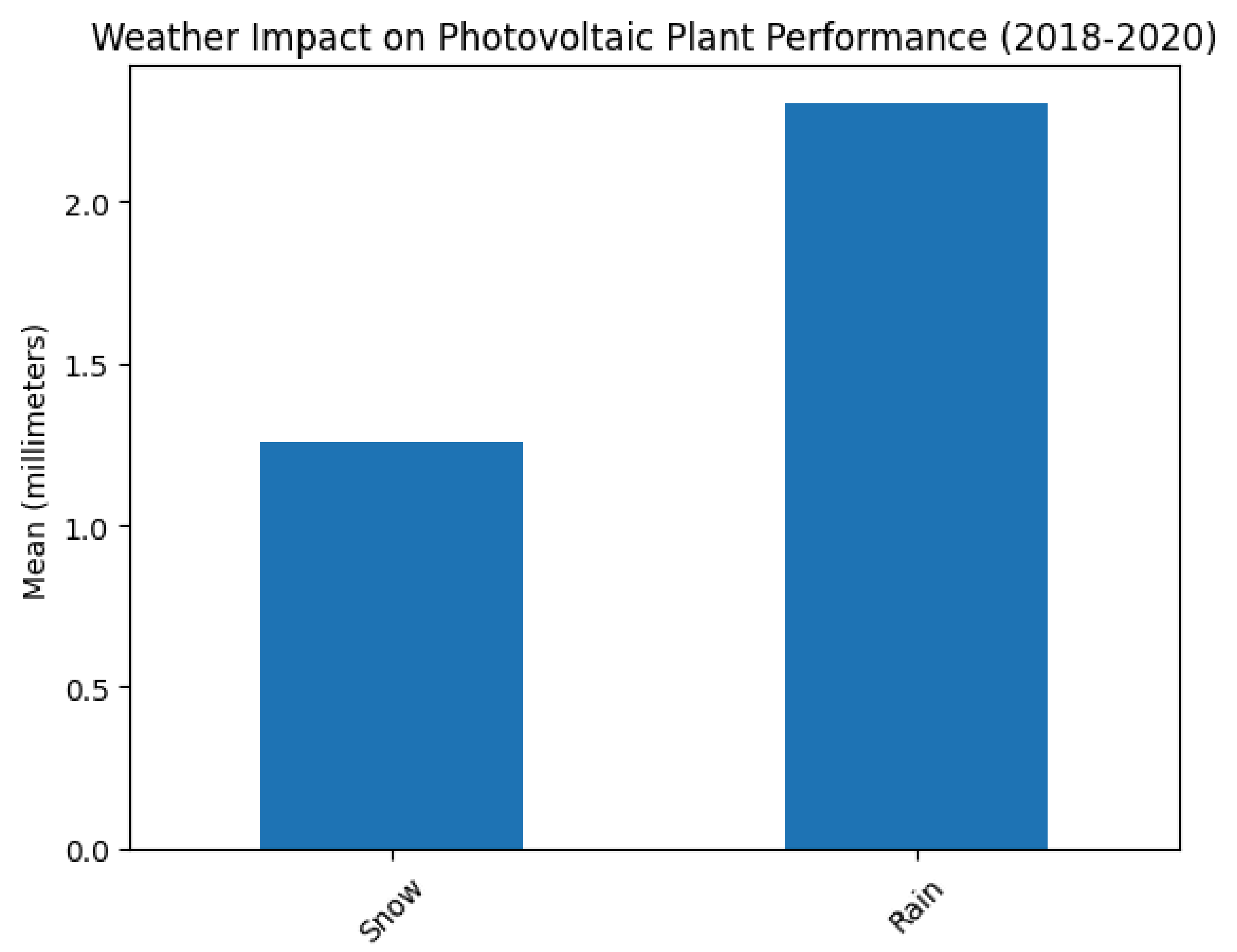
State-level policies and incentives play a key role in driving solar energy growth across regions.

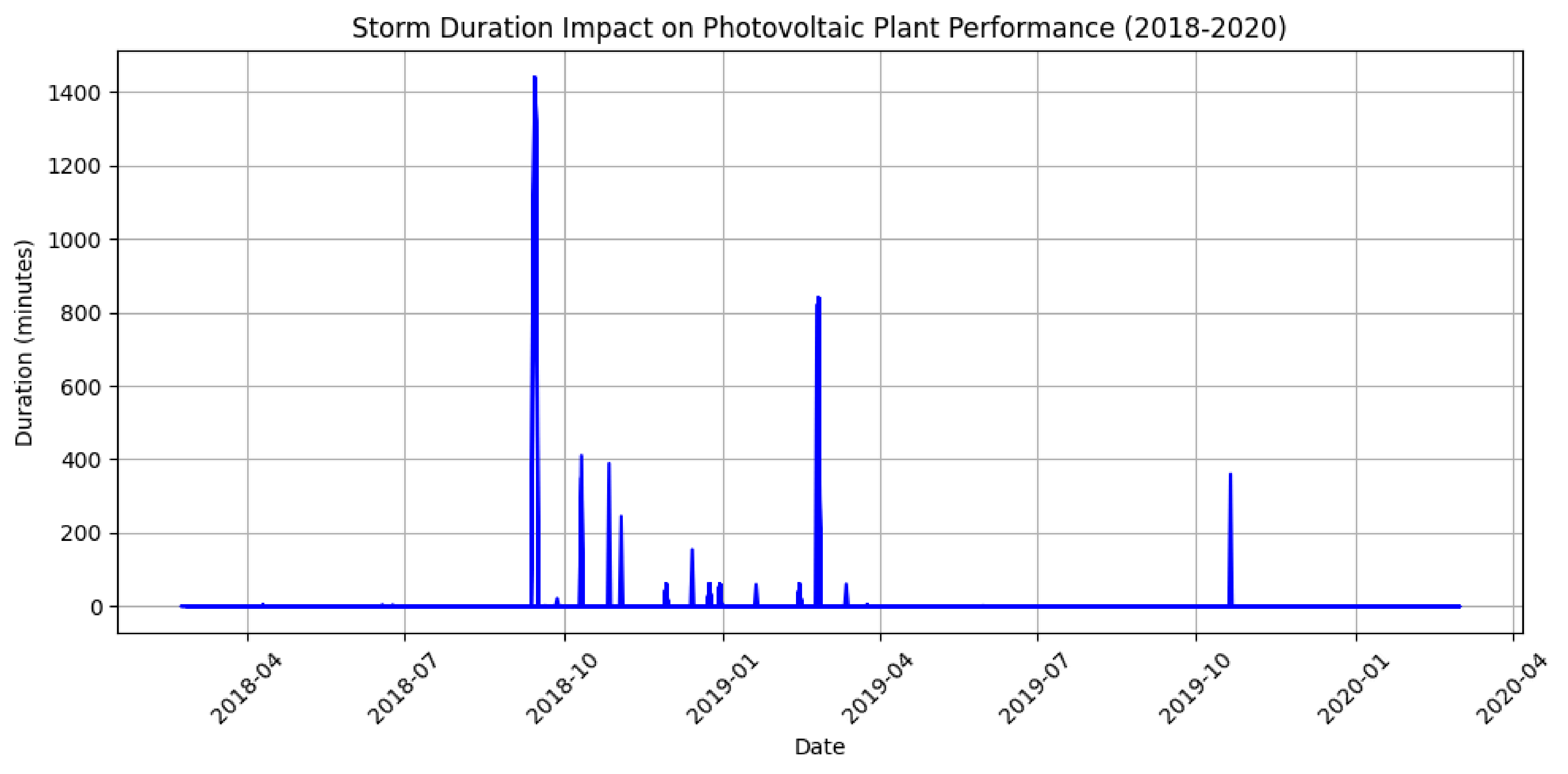
WEATHER IMPACT

Storm durations varied widely, with severe events lasting over 1,400 minutes, especially in late 2018, disrupting plant output.

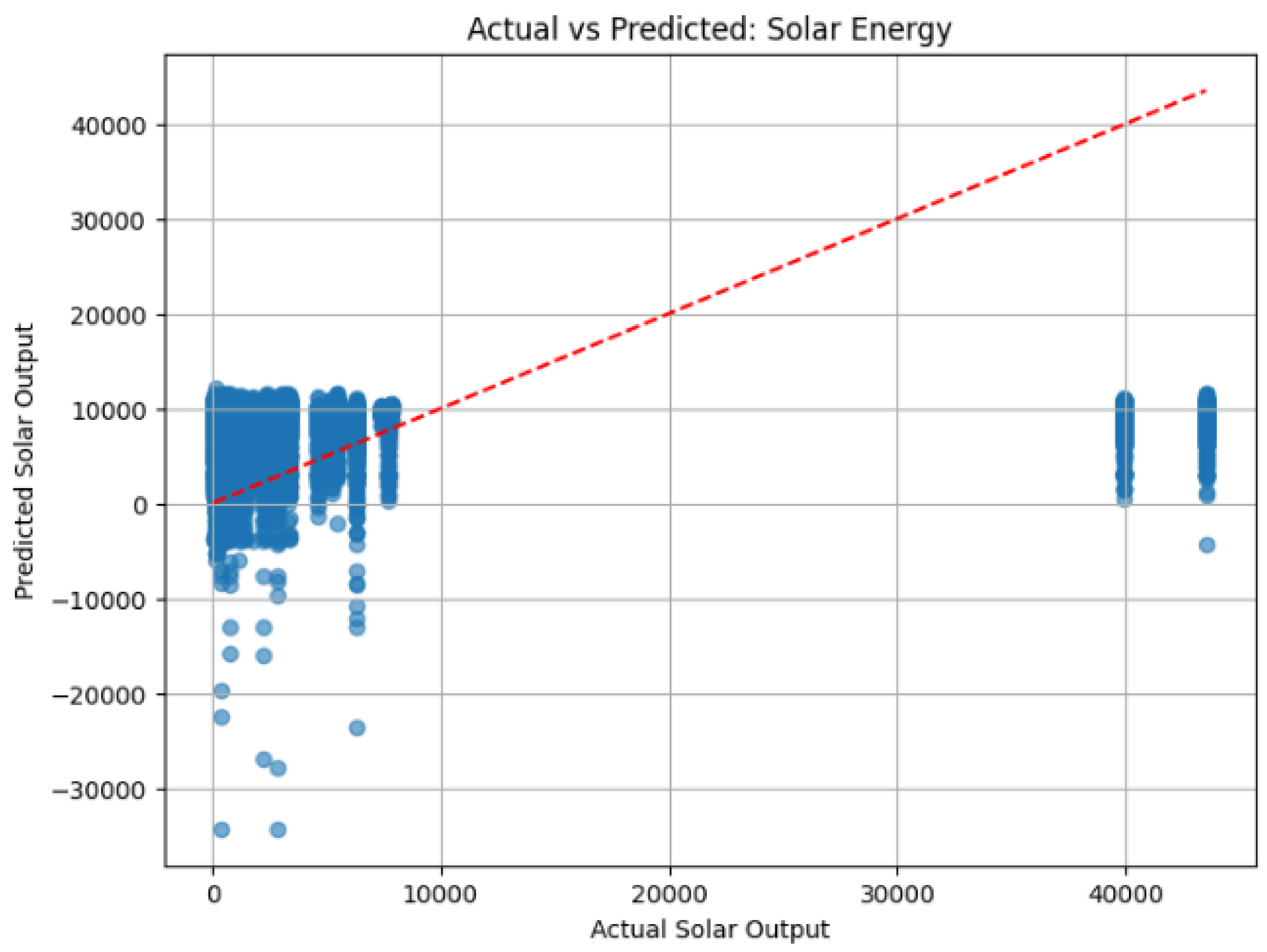
Frequent weather events, even short ones, consistently reduced solar efficiency, emphasizing the need to include weather risks in energy forecasting.

Rain had a stronger average impact than snow, showing higher mean precipitation affecting photovoltaic performance.





LINEAR REGRESSION SOLAR



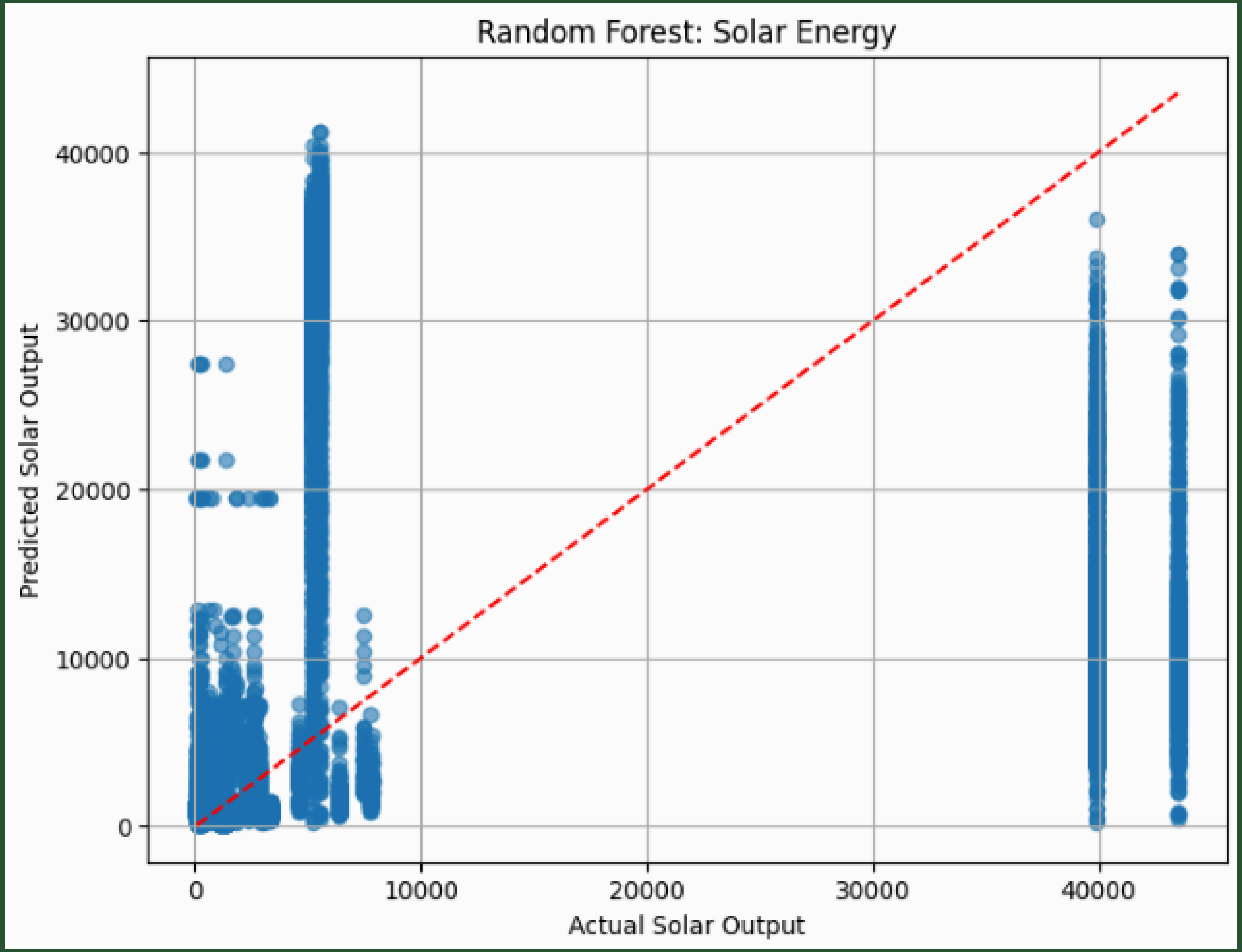
R^2 Score: 0.03661651691218493
Mean Absolute Error (MAE): 10081.298277645034
Mean Squared Error (MSE): 195908899.7776904
Root Mean Squared Error (RMSE): 13996.746042480388

- Detected a weak and inconsistent trend in solar output.
- Failed to adjust to variations caused by extreme weather and investment shifts.
- R^2 = 0.0366 | RMSE = 13,996.75

Insights:

- Solar energy's growth could not be captured reliably with a linear approach.
- Oversimplified model fails to incorporate key external drivers like storms or government incentives.
- Not reliable for forecasting solar trends, especially in dynamic or risk-sensitive environments.

RANDOM FOREST REGRESSION SOLAR

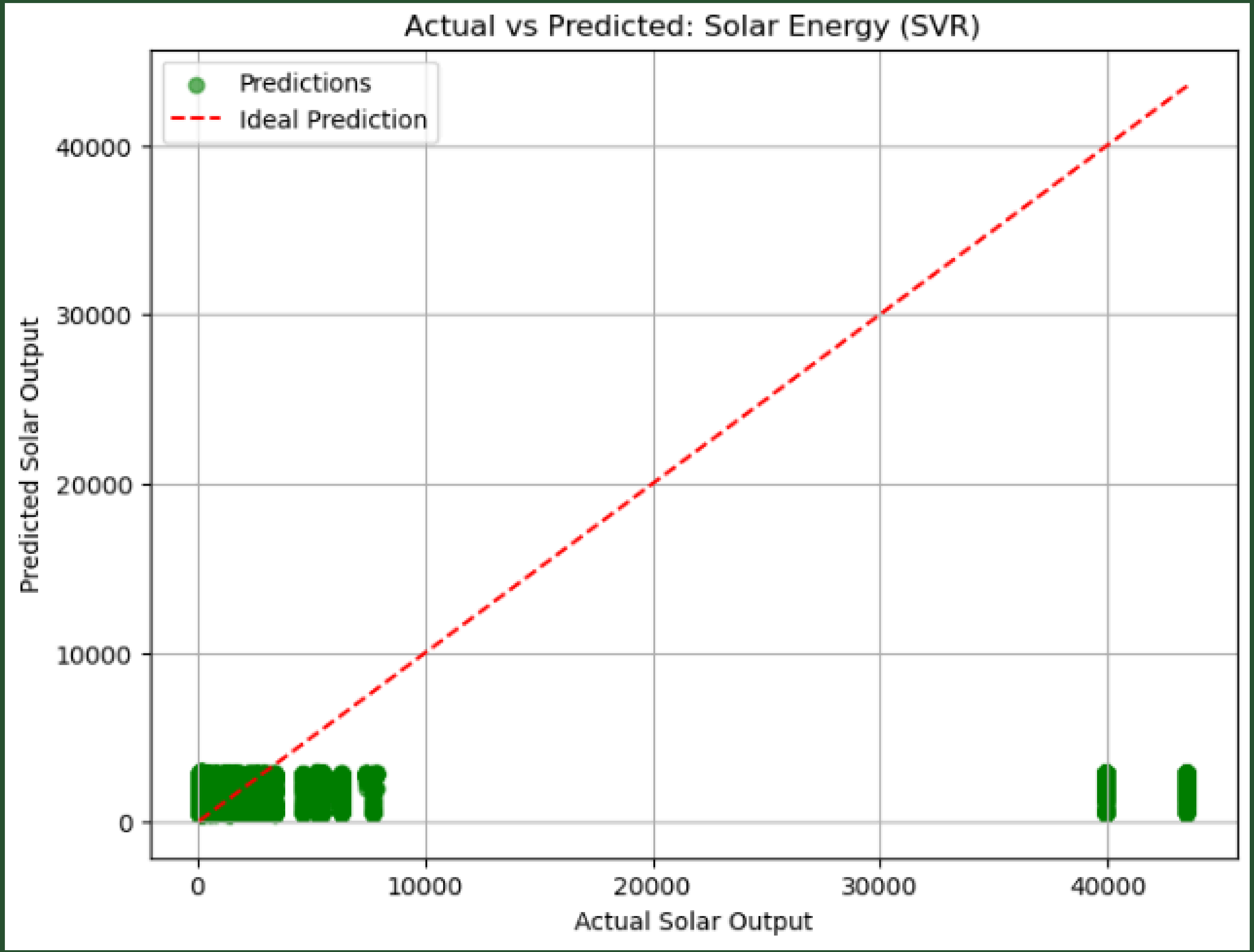


R^2 Score: -0.19017009725193423
Mean Absolute Error (MAE): 9533.320277655836
Mean Squared Error (MSE): 242027103.84196982
Root Mean Squared Error (RMSE): 15557.22031218848

- Failed to capture meaningful trends in solar energy output.
- Model overfit to noise and underperformed even in relatively stable conditions.
- R^2 = -0.1901 | RMSE = 15,557.22

Insights:
- Complexity of the model did not translate to better predictions.
- Likely caused by insufficient or misaligned features with solar dynamics.
- Current form is not effective; requires refined features or additional data for solar forecasting.

SUPPORT VECTOR REGRESSION SOLAR



R^2 Score: -0.15756512301787007
Mean Absolute Error (MAE): 7582.696959659339
Mean Squared Error (MSE): 235396717.56110683
Root Mean Squared Error (RMSE): 15342.643760483616

- Performed poorly in modeling wind output, overfitting to noisy data.
- Unable to generalize wind variability, leading to large forecast errors.
- R^2 = -0.5194 | RMSE = 8,873.86

Insights:

- Model failed to adapt to rapid and irregular wind fluctuations.
- Current configuration cannot be used for wind energy forecasts.
- Requires significant rework, possibly including temporal or lag-based features.

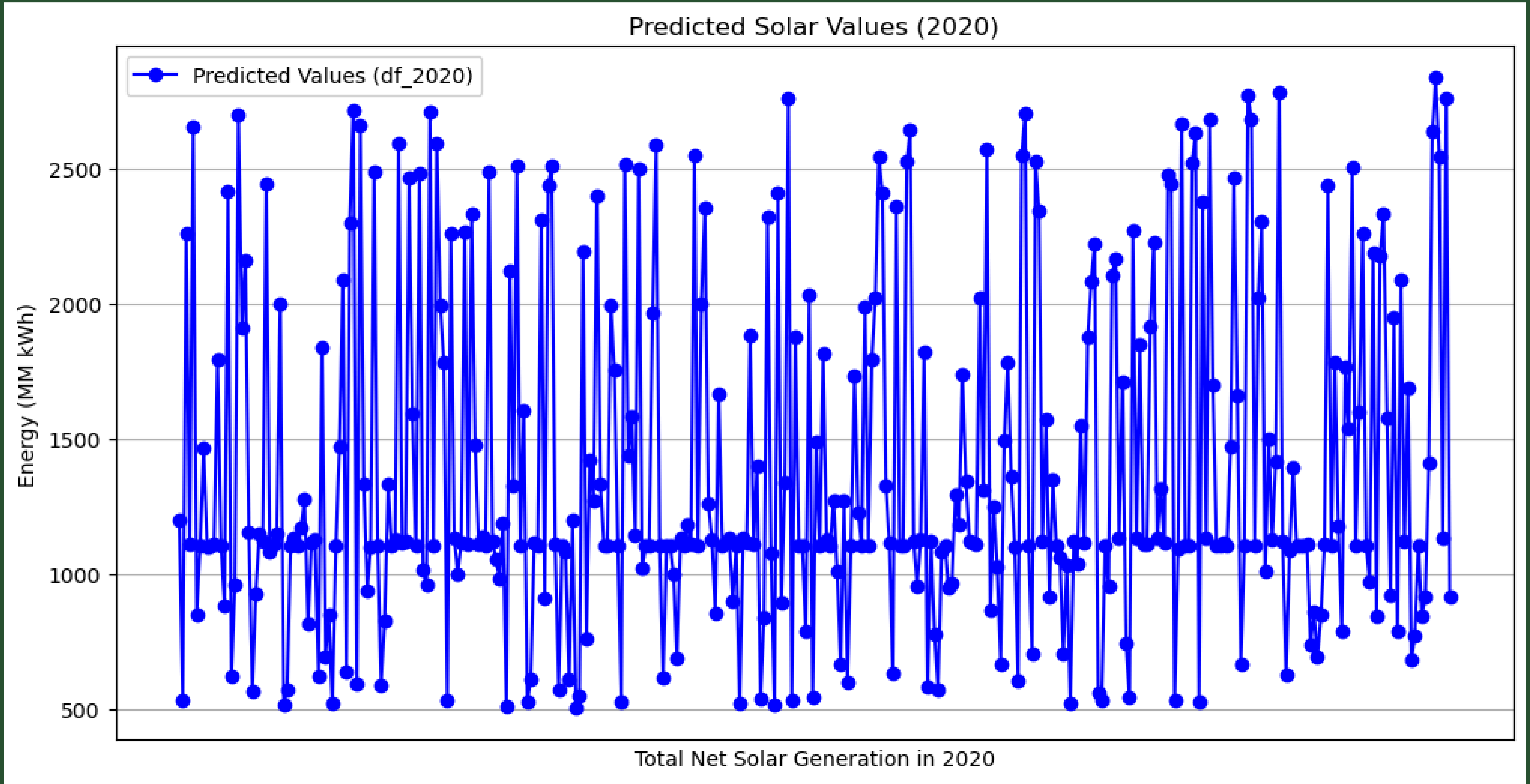
SOLAR VS WIND PERFORMANCE

Model	Solar R^2	Wind R^2	Best Performing Metric
Linear Regression	0.0366	0.0091	Neither – failed across both
Random Forest	-0.1901	-0.5194	Underperformed, overfitting issues
SVR	-0.1575	-0.3229	Least error for wind, still limited

Key Takeaways

- SVR for wind was the most consistent but still inadequate for accurate forecasting.
- No current model sufficiently predicts solar or wind growth for financial decision-making.
- Next Steps: Refine features, explore advanced methods (time series, deep learning), and improve data quality.

FORECAST MODEL PERFORMANCE



FORECAST MODEL PERFORMANCE

→ The model's predictions fluctuate widely, with no consistent trend, indicating poor learning of solar generation patterns.

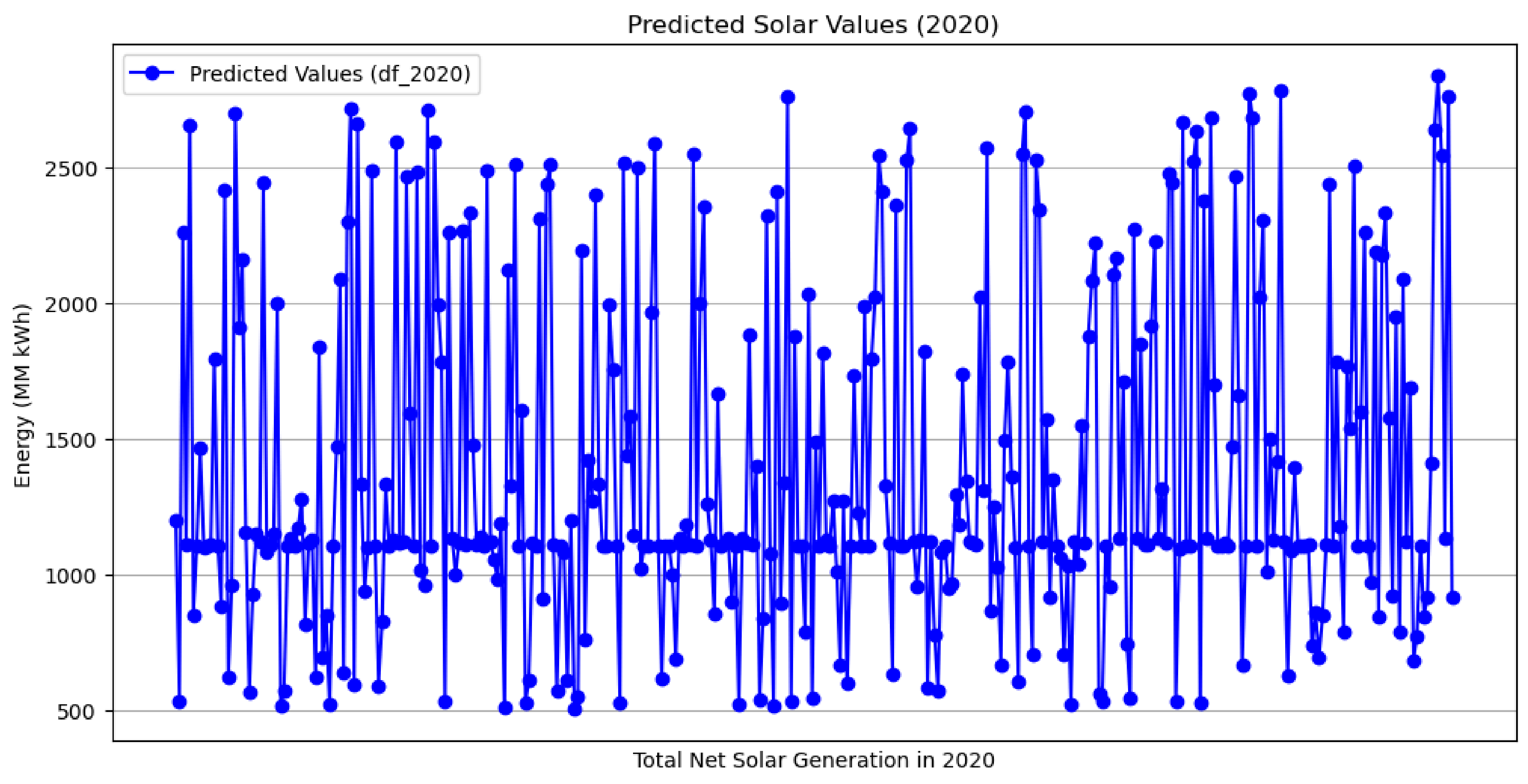
→ The model shows high variance, with prediction differences reaching up to ~2,200 MM kWh between consecutive points.

→ Over 40% of predicted values deviate by more than ±500 MM kWh from the model's central trend (~1,000 MM kWh).

→ Sharp prediction spikes, reaching up to ~2,700 MM kWh, suggest overfitting to noise rather than capturing real patterns.

→ Lack of smoothness or clear seasonality shows the model fails to generalize well on unseen data.

→ This forecast demonstrates low precision and control, requiring improved feature selection and better model tuning.



ISSUES & IMPROVEMENTS

Model Performance Summary

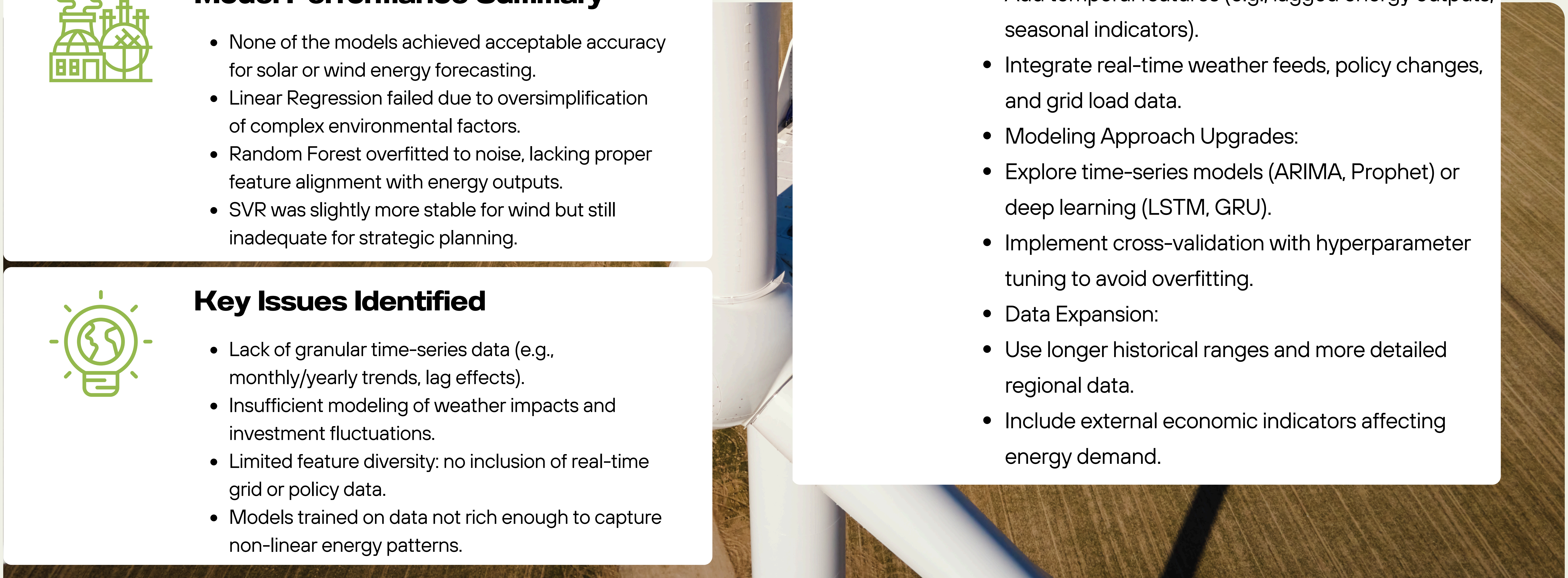
- None of the models achieved acceptable accuracy for solar or wind energy forecasting.
- Linear Regression failed due to oversimplification of complex environmental factors.
- Random Forest overfitted to noise, lacking proper feature alignment with energy outputs.
- SVR was slightly more stable for wind but still inadequate for strategic planning.

Key Issues Identified

- Lack of granular time-series data (e.g., monthly/yearly trends, lag effects).
- Insufficient modeling of weather impacts and investment fluctuations.
- Limited feature diversity: no inclusion of real-time grid or policy data.
- Models trained on data not rich enough to capture non-linear energy patterns.

Future Suggestions

- Feature Engineering Enhancements:
- Add temporal features (e.g., lagged energy outputs, seasonal indicators).
- Integrate real-time weather feeds, policy changes, and grid load data.
- Modeling Approach Upgrades:
- Explore time-series models (ARIMA, Prophet) or deep learning (LSTM, GRU).
- Implement cross-validation with hyperparameter tuning to avoid overfitting.
- Data Expansion:
- Use longer historical ranges and more detailed regional data.
- Include external economic indicators affecting energy demand.

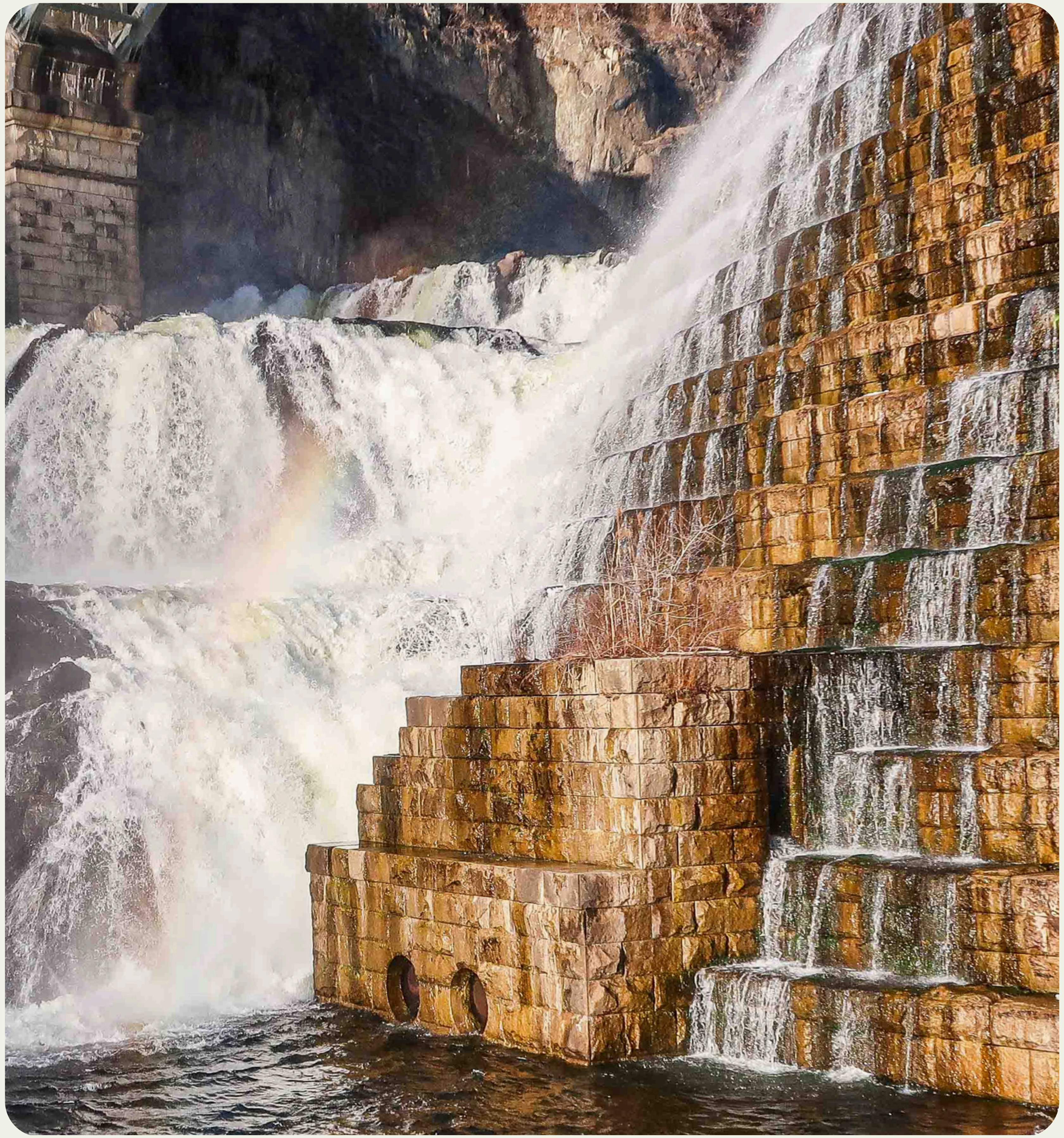


CONCLUSION

- Current models are not reliable for accurate forecasting of solar or wind trends.

- Substantial model refinement and data enhancement are needed to produce actionable insights.

- Predictive energy modeling remains a critical tool for guiding future infrastructure and investment decisions, but requires more advanced techniques and richer data to be effective.



THANK YOU