# Leveraging Data Science to Improve Solar Plant Performance Management

## Business Value of Advanced Analytics

Effective SCADA systems incorporate advanced analytics and machine learning models, to evaluate performance and manage assets, yielding these benefits.

- Predict health and malfunction of site devices
- Optimize operational performance
- Proactively and cost-effectively schedule maintenance activities
- Maintain the optimal amount of spare inventory parts
- Reduce the likelihood of device malfunctions
- Diminish lost energy
- Reduce downtime
- Increase the lifetime of solar assets

## Overview

Solar power plants generate a vast amount of data which can be used to determine how a plant is performing and what is impacting its performance. Taking this a step further, a plant's performance can be optimized by leveraging advanced analytics. Advanced analytics includes formulating statistical matrices, implementing actionable performance alarms, and conducting artificial intelligent models to predict future performance.

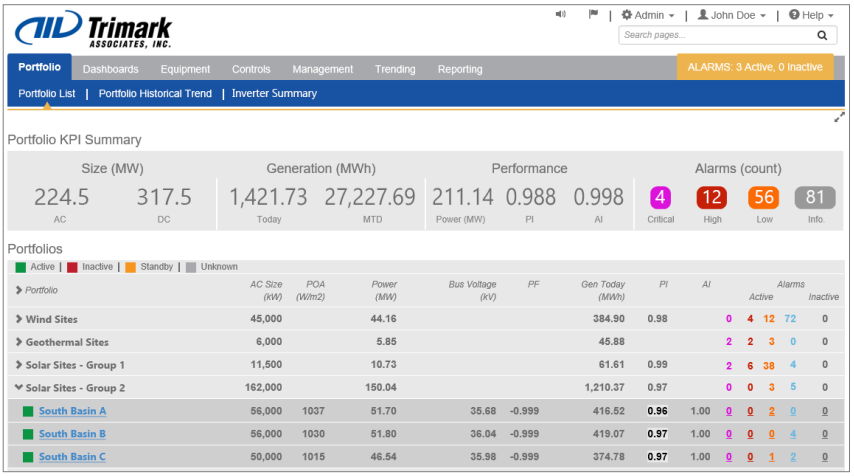
An effective SCADA system includes platforms and tools to evaluate plant performance and manage assets. Trimark's SCADA system does that and more, as it incorporates advanced analytics, along with machine learning and deep learning models to revolutionize the way owners and operators manage their renewable energy assets.

## Advanced Analytics

A Real-Time Weather Adjusted Performance Index measures the power and characteristics of a site (such as irradiance, back panel temperature, power generation, and DC and AC capability) to provide a real-time snapshot of the site's performance. Implemented at the device level and the site level, it is a valuable Key Performance Indicator (KPI).

By reviewing a site's KPIs, engineers can:

- Proactively troubleshoot an underperforming device and/or site
- Respond to a performance alarm based on a low KPI
- Forecast energy production by integrating one or more weather services, and
- Realistically predict the site's performance using machine learning (ML) techniques based on historical data and actual site conditions



## Inverter Energy Prediction

**Conventional Time Series Analysis**

By predicting how a plant or inverters will perform in the near term, operators and engineers can proactively take steps to ensure the expected energy production and to minimize the impact of failed devices. A relatively accurate energy prediction can be calculated by incorporating weather forecasts, a plant's historical performance, and plant and inverter design characteristics.

In *Figure 1*, an analysis of the total AC energy generated by an inverter at a solar facility shows daily oscillations. The figure also shows the daily total amount of AC energy produced by an inverter at a solar facility. The daily oscillations reflect when the sun rises and sets each day. Data transformation analysis, using the Fast Fourier Transformation technique, confirms this daily oscillation trend.

This illustrates that machine learning (ML) models can be used to predict energy generation by following the steps below.

1) Shift the energy data by a predefined amount of hours (24 hours) in the past for the learning targets. For example, to predict 10:00am today, use yesterday's 10:00am data.
2) Use the Random Forest ML model, a decision splitting tree model, to train the model until the desired energy generation predictability is met.
3) Generate predicted energy data using the trained ML model.

Trimark uses the Random Forest ML model to predict daily energy generation because it randomly splits the data features into subsets of trees, thus increasing the accuracy of the model.

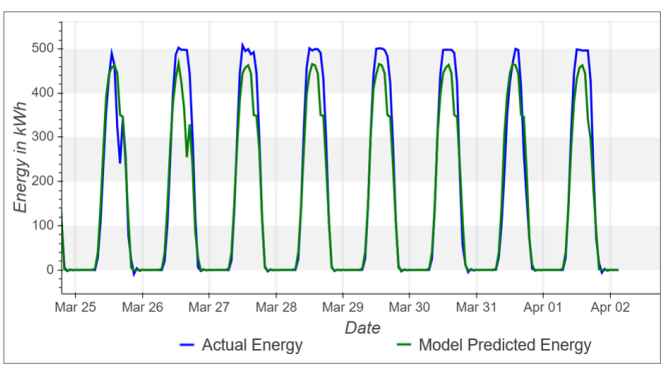



*Figure 1 – A Comparison of Actual to Projected Daily Energy Generation*

The predicted energy output (*green line in Figure 1*) follows the real energy output (*blue line in Figure 1*) most of the time, with a high model selection criteria of R square value (~0.85). This Point-to-Point regression model uses a single energy value in the present to predict a single energy value in 24 hours.

The model is reiterated many times to determine how far out it can reliably forecast energy production. *Figure 2* shows the energy projections as the hours increase. The model selection criteria R square value decreases as the number of forecasted hours increases, illustrating that the more days into the future the model predicts, the more it deteriorates.

However, the model generates a good prediction of the first three days, with R square value around 0.75. With this three-day window, owners and operators can plan in advance to schedule repairs and preventive maintenance.

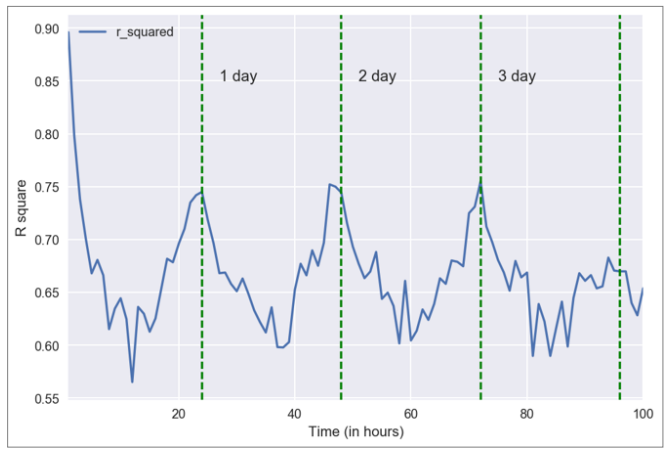



*Figure 2 – Energy Projections As Hours Increase*

**Artificial Neural Networks (ANN) Time Series Analysis**

Trimark is always looking to implement the latest and most effective machine learning models to predict solar energy generation. A model that is fueling artificial intelligence in solar asset management is the Artificial Neural Networks (ANN) model. Compared to a conventional machine learning model, this model's algorithm is superior in predicting energy performance.

**First ANN Approach: Fully Connected Perceptron**

Among the ANN deep learning algorithms, one of the basic models is the Multilayer Perceptron (MLP). MLP mimics the neurons of the human brain, learning the way our biological mind thinks and perceives subjects and behaviors. Then, just like our brain, MLP outputs results based on layers of input signal/information. Mathematical equations or models are used to simplify and represent the incredible architecture of our brains, resulting in more accurate energy predictions.

The mathematical model that is applied to the solar inverter data is called the Fully Connected Three Layers Perceptron model, containing one hidden layer composed of 1,024 neurons. As shown in *Figure 3*, this model increases predictability by 2.89%, compared to the conventional time series ML model. In addition, the predictability, especially during night time hours, is much smoother than the conventional time series ML model.

However, just like the conventional time series ML mode, this ANN model applies the same Point-to-Point energy prediction algorithm, which induces some unavoidable noise to the data results. For example, there is a period of data drop in the afternoon of March 25. Since the model uses a Point-to-Point modeling technique, it transfers this data drop from March 25 to March 26, even though the sky was clear on March 26. (*See Figure 3.*)

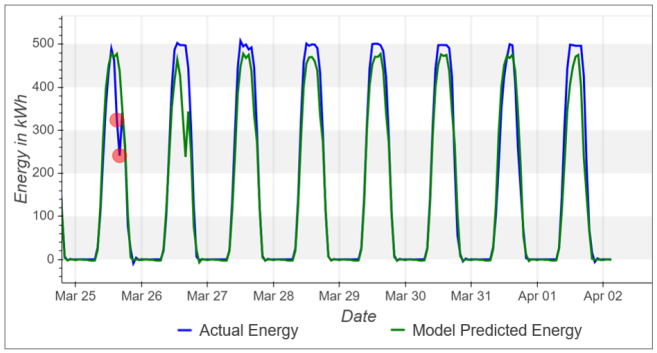



*Figure 3 – Fully Connected Three Layers Perceptron Model*

**Long Short-Term Memory (LSTM)**

Compared to Point-to-Point modeling, Sequence-to-Point (S2P) modeling is able to overcome the intrinsic noise. S2P uses a few consecutive points, instead of just one, to predict a single value into the future. Moreover, the input data is treated as sequences of data instead of data points by using the Long Short-Term Memory (LSTM) model from ANN.

Another significant advantage of ANN is its ability learn and perform feature engineering all by itself. No prior knowledge of the data is required, such as knowing the site's conditions (snow covered, soiled, shaded). Instead, ANN can learn through layers of information within the data.

To use the S2P model, the data needs to be transformed from Raw to Features as shown in *Figure 4.* The figure illustrates the data transformation process. A single time series data is transformed into a matrix so the data can be modeled using the Sequence-to-Point model to eliminate noise (as mentioned earlier).

In this data transformation, each row consisting of 24 consecutive points is a sequence. Then each sequence is trained by an LSTM ANN model to predict a single energy generation value 24 hours later. (*Target shown in Figure 4.*)

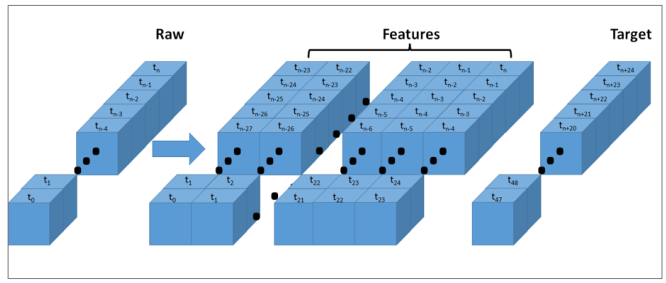



*Figure 4 – The Sequence-to-Point Model Needs Data Transformed from Raw to Features*

The prediction from this S2P model is less susceptible to noise as it does not follow previous day's target data behavior. (*See the green line within the purple dash in Figure 5.*)

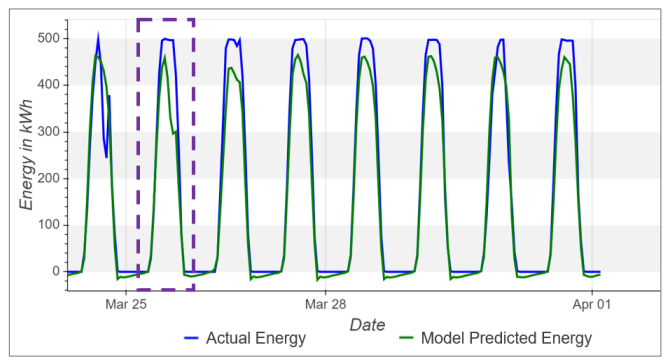



*Figure 5 – The Sequence-to-Point Model Is Less Susceptible to Noise than Other Models*

## Anomaly Detections and Monitoring

Traditional IEC standard performance index calculations, which use pre-defined formulas, do not account for the conditions at solar sites, such as shaded and soiled PV panels. A goal for more accurate energy prediction is the ability to pinpoint a device's performance anomalies in real time. To do this, an individual model can be built for each site, incorporating information about the given site's conditions into the model.

An outlier detection model, comprised of one class support vector machine technique, was built to monitor a device's performance by comparing the measured and predicted energy at a device/site level in real time. The abnormal energy generation is highlighted. (*See the red points in Figure 3.*) When an anomaly is detected, an alarm can be issued to alert the operator of the issue.

## Device Health and Malfunction Predictions

Trimark is leveraging machine learning models to predict the health and malfunctions of site devices. As a starting point, Trimark's SCADA databases contain a vast amount of detailed fault history information at many levels, from entire sites to the string level. This fault data provides essential "learning material" for machine learning modeling.

Trimark's initial focus is with inverters (since they represent the majority of device malfunctions) using the Sequence-to-Point Classification model. A sequence of consecutive device data right before the fault incidences serves as the features. The training targets consist of the types of fault incidences. A multiclass classifier uses either the conventional or the ANN ML models.

Detecting and predicting faults can positively impact operational performance. The predictive analysis can help to schedule maintenance activities, maintain the correct amount and types of spare parts in inventory, and reduce the potential for device malfunctions. Overall, this will reduce downtime, diminish lost energy, and increase the lifetime of a solar asset.

## Conclusion

An effective SCADA system should include platforms and tools to evaluate plant performance and manage assets. Incorporating advanced analytics and machine learning will revolutionize the way owners and operators manage their renewable energy assets.

A common aphorism in statistics says, "All models are wrong, but some are useful." Whether deep learning or shallow machine learning models are used, they can leverage SCADA data to produce meaningful results. As technology is evolving, Trimark's SCADA system continues to propel into more advanced business systems used by all areas of management.



**Trimark Associates, Inc.**
2365 Iron Point Road, Suite 100, Folsom, CA 95630
2500 Almeda Avenue, Suite 117, Norfolk, VA 23513
www.TrimarkAssoc.com | info@TrimarkAssoc.com | 916.357.5970

**SCADA | Metering | Energy Storage | Technical Support**