14th Annual International Conference on Industrial Engineering and Operations Management
Dubai, United Arab Emirates (UAE), February 12-14, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/AN14.20240155
Published: February 12, 2024

Selection of an Artificial Intelligence Forecasting Approach for PV Panel Prices

Aisha Alhosani
Department of Industrial and Systems Engineering
Khalifa University of Science and Technology
Abu Dhabi, 127788, UAE
100052398@ku.ac.ae

Saed Amer
Faculty Member, Department of Management Science and Engineering
Khalifa University of Science and Technology
Abu Dhabi, 127788, UAE
saed.amer@ku.ac.ae

Andrei Sleptchenko
Associate Professor, Department of Management Science and Engineering
Khalifa University of Science and Technology
Abu Dhabi, 127788, UAE
andrei.sleptchenko@ku.ac.ae

Abstract

As the world moves towards renewable energy sources, solar energy is becoming more widely recognized. This impacts photovoltaic (PV) panel prices, which requires deeper analysis to shape strategic planning. This paper fills a gap in PV panel price forecasting by utilizing machine learning models for multivariate time series forecasting, offering a more robust, comprehensive, and adaptable approach. Based on literature review, feature analysis, and linear regression, PV panel prices factors were identified, and data was collected. The dataset consists of 26 monthly time series features covering January 2018 to July 2023. The study's methodology consists of two phases: Phase I involves applying machine learning models (CNN, KNN, Random Forest, SVR, XGBoost) to forecast prices, resulting in the selection of the Random Forest model as the most effective based on performance metrics and comprehensive evaluations. Phase II utilizes the aforementioned model, along with forecasted features from the SARIMA model, to predict future PV panel prices. Forecasts indicate a stabilized PV panel market for 36 months from July 2023, with slight price fluctuations. This stabilization hypothesis aligns with other research findings, suggesting industry agreement. As the solar industry is dynamic, regular market analyses and model calibrations are essential.Overall, the research enhances the understanding of solar energy price time series forecasting, merging data gaps, and outlining pivotal price-influencing factors, offering stakeholders direction towards a sustainable energy future. Recommendations include using a broader dataset, improving methodology, exploring more models, applying a hybrid model, and continually updating forecasting methods in alignment with industry dynamics.

Keywords

Renewable Industry, Solar Energy, Photovoltaic (PV) Panel Prices, Multivariate Timeseries Forecasting, Machine Learning.

1. Introduction

1.1 Motivation

Due to the development of the solar industry and its integration into global supply chains, PV panel prices began to fluctuate and increased by 15% in the second quarter of 2021 following a long period of price reductions (Anon 2021b). In which, solar photovoltaic (PV) module prices have reduced dramatically over the last few decades, falling by about 90% between 2007 and 2017 (Philipps 2023). These fluctuations and price increases made stakeholders, manufacturers, and anyone interested in solar energy question it. Therefore, it is crucial to understand the solar energy sector trend and its influences.

Solar-related products prices will be influenced by a wide range of factors, including technological advances (Haegel et al. 2017), governmental policies (Jiang et al. 2019), demand-supply dynamics in the market (Yue et al. 2016), global economic factors (Gross et al. 2016), and others. However, some of these factors are usually not considered when forecasting PV panel prices. Therefore, predicting the trajectory of solar product prices, especially PV panels, requires feature analysis. This can be done with statistical models such as time series analysis and regression models. The purpose of this step is to give more insight to researchers, stakeholders, and anyone interested in solar PV panels. To understand the dynamics and influences that might have a significant contribution to the PV panel prices. Also, it shapes manufacturers' strategies, distributors' plans, and governments' policies who aim to promote solar energy adoption.

In the energy sector, forecasting methods are widely used and papers can be found to understand energy related product prices. The adoption of those forecasting methods is usually found on oil and gas, electricity, natural gas, coal, carbon, and others. However, in the solar industry, few efforts are being made to understand and use forecasting techniques in predicting the price of renewable energy-related products (PV panels). This has made data collection and forecasting methods selecting a challenge, since only few data can be found and utilized. Therefore, collecting data and applying those forecasting methods on PV panel prices would give insights and help understand the trends in solar industry further.

In the past, energy product price forecasts were mainly based on statistical methods, such as time series analysis and regression models. While effective in simpler scenarios, the modern renewable energy market's dynamic nature, coupled with the massive amount of information we currently experience, pushes these methods to their limits. In this case, a Machine Learning (ML) model is more relevant. This technique is capable of analysing large datasets and detecting complex nonlinear patterns. This will allow it to provide more accurate forecasting, and be more adaptable. By implementing and applying machine learning models specifically tailored to forecasting prices of renewable energy-related products, stakeholders can potentially achieve more accurate and actionable insights in strategic planning and decision-making. As well, it is hoped that this collaboration will accelerate our collective shift toward a more sustainable energy vision, leading to a cleaner and more secure global energy future.

1.2 Problem Statement

A transition has occurred towards the use of renewable energy sources instead of conventional power generation sources. Throughout this transition, solar energy is becoming an increasingly important part of the global energy landscape, resulting in a profound impact on the prices of Photovoltaic (PV) panels. However, in the field of solar energy PV panels price forecasting, no significant contributions have been made. The complex price dynamics of the solar industry may not be captured by traditional forecasting techniques based on statistical methods such as time series analysis and regression models. Therefore, these deficiencies pose a potential threat to stakeholder's strategic planning and may prevent solar energy solutions from being adopted widely and optimally. In addition, the availability and utilization of data for forecasting PV panel prices is limited, making it challenging to predict market trends effectively.

1.3 Objectives

This paper will specifically examine historical price fluctuations for PV panels over the past decade, delving into an in-depth analysis of the factors influencing solar product prices. The research will initially begin with thorough data collection for these key factors, followed by a feature selection analysis using regression models to identify crucial variables impacting PV panel prices. Subsequently, the selected features will be employed in a time series multivariate forecasting method to predict prices. The study aims to develop, implement, and validate machine learning models using the Python programming language, tailoring them explicitly for multivariate time series forecasting of PV panel prices. Evaluation and analysis of these models will be conducted to identify the best-fit model for accurate forecasting of PV panel prices. Moreover, the chosen model will be applied to a real-life scenario for future price prediction. The research will result in providing recommendations based on forecasted future PV panel prices by the selected model and the study's findings, thereby addressing the knowledge gap in forecasting methodologies within the solar industry and offering a more precise, comprehensive, and adaptable approach to predicting PV panel prices.

2. Literature Review

2.1 Forecasting Methods

Forecasting is a common way of predicting the future. A forecast is anything we try to anticipate or estimate the outcome of future events-whether it's in business or personal life. Forecasting is among the most critical business functions since all other decisions are based on forecasts (Sanders n.d.). It has a wide range of applications such as weather forecasting, agriculture forecasting, business forecasting, price forecasting, energy use forecasting, technology forecasting, and many others (Sanders n.d.). This section presents a literature analysis on Artificial Intelligence forecasting methods, renewable energy related products while focusing on solar energy, and directly related works to the research topic.

2.1.1 Hybrid and Machine Learning Models

In the field of artificial intelligence, machine learning is a type of artificial intelligence that makes software applications better at predicting outputs without being explicitly programmed to do so (Cerjan et al. 2013) In machine learning algorithms, historical data is used to predict future values (Cerjan et al. 2013). Hybrid methods of time series forecasting combine the inherent benefits of both statistical and machine learning techniques, and the idea behind this is that it compensates for the weaknesses of one approach with the strengths of the other (Daniel Berberich 2020). The structure of hybrid models is extremely complex, including algorithms for decomposing and clustering data, feature selection, and combined forecasting models (Xiao et al. 2020). Recent advancements in hybrid and machine learning models have shown promising results in forecasting energy prices.

Niu et al. (2010) adopted the Self-Organizing Map neural network and Support Vector Machine models for electricity price forecasting. The use of SOM for data clustering was shown to enhance prediction ability by automatically filtering the training samples.

Wang and Wang (2016) proposed an ERNN model that merged Elman recurrent neural networks with stochastic time-effective functions, establishing a novel architecture. Their model presented impressive accuracy during significant price fluctuation periods.

Qin et al. (2019) utilized an ensemble empirical mode decomposition and local linear prediction model, EEMD-LLP, for energy price prediction. Their findings indicated the EEMD-LLP method's superior accuracy and computational efficiency when matched with historical data.

Jianwei et al. (2019) developed an Independent Component Analysis and Gated Recurrent Unit neural network model, referred to as the IGS model. The research demonstrated that the IGS model, when compared to other forecasting models, was more reliable and effective.

Wang et al. (2020) proposed a new hybrid data-driven model that incorporated IPSS-SVR-LSTM and an improved pattern sequence similarity search (IPSS). Their findings indicated that their hybrid model outperformed the traditional PS model in terms of forecast accuracy and prediction capability.

Wang and Wang (2020) presented a hybrid SW-GRU with EMD model to predict energy futures and spot prices. Their study showed that the SW-GRU combined with EMD produced accurate forecasts and maintained a high level of precision in complex energy price scenarios.

Cordoni (2020) explored multiple deep neural network architectures for energy spot price forecasting. The study emphasized the CNN-LSTM network's ability to produce highly accurate forecasts.

Ribeiro et al. (2020) presented a hybrid multi-stage heterogeneous ensemble model for electricity energy price forecasting. The proposed approach exhibited greater accuracy and efficiency than other compared models, especially as the forecast horizon extended.

Lin et al. (2022) introduced a hybrid model combining AR-IBLSTM-ELMAN with VMD for energy price forecasting. The model's adaptability and precision were highlighted, particularly when varying training segment lengths.

Overall, the fusion of hybrid and machine learning models offers a dynamic approach to energy price forecasting. The above-reviewed studies indicate the potential of these models to adapt to various scenarios, ensuring precise and reliable predictions.

2.2 Renewable Energy Related Products

The global shift towards sustainable and environmentally friendly energy sources has inspired significant interest and research in the domain of renewable energy. This literature review delves into renewable energy-related products particularly focusing on solar power, its types, applications, prices trends, and the factors that affect those prices.

2.2.1 Solar Energy

Over the past few decades, the focus on reducing carbon footprints has accelerated renewable energy product exploration and development. From wind turbines to biofuels, the renewable energy market has witnessed the development of technologies aimed at utilizing nature's power (Smith et al. 2015). However, solar energy stands out among these because of its widespread availability and potential scale (Anon 2021a) **.**

Solar energy, being one of the cleanest forms of renewable energy has experienced rapid technological advancements. Starting with passive solar building designs, the journey transitioned to advanced photovoltaic (PV) cells. According to Brown et al. (Brown et al. 2018), the efficiency of solar panels has almost doubled in the past two decades, making them more affordable and accessible.

Solar energy is an abundant and unlimited resource. Every hour, the sun radiates more energy onto Earth than the entire human population uses in a year (Bradford 2006). Utilizing even a fraction of this energy has the potential to reshape global energy landscapes. In addition to being abundant, solar energy is also environmentally friendly. Its adoption significantly reduces the carbon footprint, as it produces no greenhouse gases during operation. This quality makes it an essential tool in mitigating climate change (Jacobson and Delucchi 2011).

3. Methods

This research paper is structured into two phases. In Phase I, machine learning models are developed and applied to forecast PV panel prices, aiming to identify the best-suited model for this task. Phase II involves the application of the selected model to future forecasted features to predict subsequent PV panel prices.

3.1 Phase I: Machine Learning Models for Forecasting

To build and apply machine learning models, a literature review was conducted. From this review, the features affecting PV panel prices were identified, and time series data was gathered. The dataset was then analyzed, with correlation evaluations being conducted. Performance metrics were employed to check the dataset's accuracy and robustness. To enhance the dataset's robustness, additional features were incorporated. This dataset was then scaled and preprocessed before being divided into training and testing sets. The training data underwent validation using the k-fold cross-validation method. Models' accuracy was assessed through performance metrics.

The trained models were subsequently tested on the testing data. Their forecasting accuracy was verified using performance metrics. Models producing acceptable forecasts underwent further evaluations, such as Huber loss, histogram of residuals, and plots showing actual vs. predicted data points. In cases where forecasts were sub-optimal, hyperparameters were adjusted within the k-fold cross-validation, and the models were retrained. This cycle was repeated until forecasts reached the desired criteria. An analysis was then conducted to compare the performance of each model to identify the most effective one for predicting PV panel prices in the renewable energy sector. The methodology is illustrated in the flowchart shown in Figure 1.

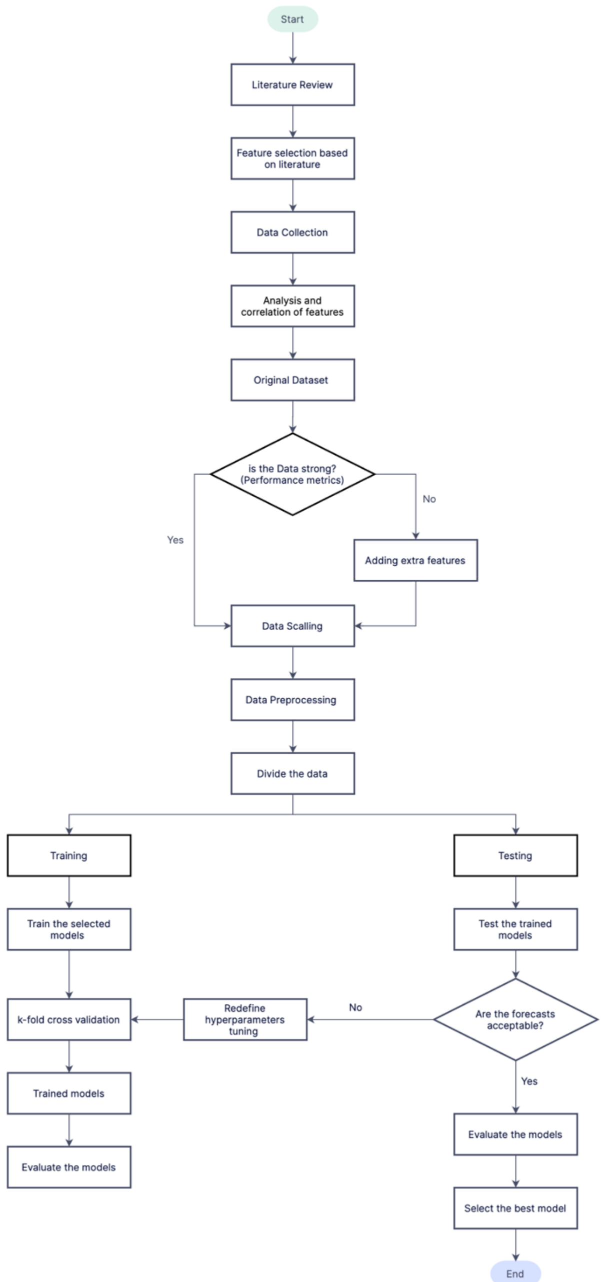

Figure 1. Methodology flowchart for the models built to forecast the PV panel price.

3.2 Phase II: Future Forecasting of Prices

The methodology outlined in Figure 2 focuses on using the selected model from Phase I to predict future prices. A new set of future data points for the features was generated. The initial step was to analyze the seasonality within the time series data. The Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model was employed for each feature, considering seasonality. For each feature, key parameters were adjusted to capture and represent inherent patterns and dependencies within the time series data effectively.

After the modeling phase, performance metrics were analyzed, and the output was visualized to evaluate the model's effectiveness in capturing patterns and dependencies for each feature. In cases where the results weren't as per the expectations, parameter adjustments were made, and the modeling process was repeated. Upon obtaining satisfactory results, the forecasted data points for the features were combined into a single dataset spanning 36 months. This dataset was then fed into the Phase I selected model, resulting in predictions for future PV panel prices.

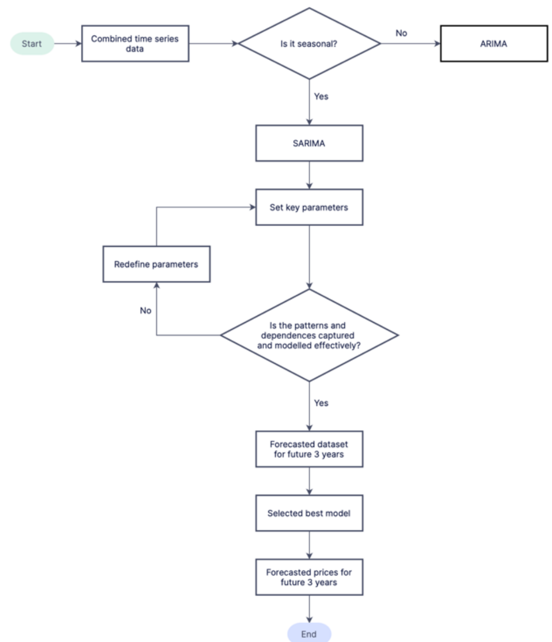

Figure 2. Methodology flowchart for applying forecasting future PV panel price.

4. Data Collection

The dataset in this study was collected from online sources. The data consists of a blend of features that were identified and prioritized from the literature based on their impact on PV panel prices. While the primary data revolved around the prices of materials influencing PV panel pricing, the secondary data was composed of less commonly considered factors that could, nonetheless, influence panel pricing.

4.1 Primary Data

A set of 12 critical features that influence PV panel prices were manually collected, representing monthly time series data from January 2018 through July 2023, specifically related to China. These data points were obtained from website graphical representations. Moreover, for the scope of this research, our response variable or "y" is denoted by the PV Solar Cell Mono price – a panel fabricated from monocrystalline solar cells (Mike n.d.).

4.2 Secondary Data

To enhance the dataset, several additional features, often overlooked in conventional studies, were systematically integrated. Each feature underwent comprehensive evaluation, ensuring enhanced dataset robustness. As a final step, 14 additional features (Anon n.d.) were selected and combined with the primary set to form a comprehensive 26 distinctive features list.

5. Results and Discussion

5.1 Feature Analysis

In the process of strengthening the data, further feature analyses were done. In which, Linear Regression was employed on the dataset to identify error loss and determine robustness. Performance metrics such as SMAPE,

MAPE, MAE, and RMSE were used. After this analysis, individual features were integrated into the dataset to boost its forecasting capabilities. Any feature that reduced the error loss was subsequently added to the original data. To summarize, Table 1 offers a comparison between the loss from the original data and the post-feature addition data. A noticeable reduction across all loss metrics as shown in Figure 3, including Symmetric Mean Absolute Percentage Error (SMAPE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), indicates the enhanced robustness of the dataset. Based on the findings from the article titled "Electricity price forecasting on the day-ahead market using machine learning" (Tschora et al. 2022), the addition of new features dramatically improves model performance for the majority of datasets and models. Therefore, the focused efforts in feature analysis and the subsequent optimization of the dataset have resulted in forecasts that are both accurate and reliable.

Table 1. Comparison of Loss Metrics: Original Data vs. with New Features

Error Loss	Original Data	With New Features
SMAPE	0.784657	0.263079
MAPE	0.519982	0.211954
MAE	0.423332	0.170015
RMSE	0.480076	0.211754

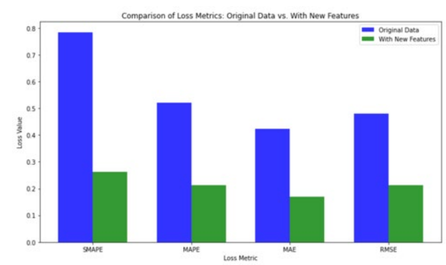

Figure 3. Comparison of Loss Metrics: Original Data vs. with New Features.

5.2 Phase I: Machine Learning Models

A comprehensive evaluation of the performance of the five models investigated in this study which are Convolutional Neural Network (CNN), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Regression (SVR), and eXtreme Gradient Boosting (XGBoost). As well as patterns and tendencies, are conducted to select the most appropriate model. Starting with the Analysis of Actual Vs. Predicted Graphs, Residual Distribution Analysis, and followed by Performance Metric Analysis with SMAPE, MAPE, MAE, RMSE, Mean Squared Error (MSE), and Huber loss as shown in Table 2.

Table 2. Comparison of Performance Metrics Analysis

Models	**SMAPE**	**MAPE**	**MAE**	**RMSE**	**MSE**	**Huber loss**
CNN	0.1529	0.1578	0.1329	0.1807	0.03264	0.01632
KNN	0.1785	0.1870	0.1569	0.2122	0.04506	0.02253
RF	**0.1303**	**0.1374**	**0.1160**	0.1693	0.02868	0.01434
SVR	0.1410	0.1414	0.1188	**0.1464**	**0.02144**	**0.0107**
XGBoost	0.1579	0.1597	0.1318	0.1838	0.03379	0.01690

Based on the extensive analysis conducted, the Random Forest and XGBoost models emerge in the evaluation of actual versus predicted performance. Both the Random Forest and the XGBoost models showed strong alignment with actual data values, suggesting they are reliable time series forecasters. Although both have demonstrated their ability to track data trends, XGBoost has a slight edge when it comes to capturing short-term fluctuations.In terms of performance metrics, the Random Forest model is the best performer, having the lowest SMAPE and an impressive RMSE, followed closely by the SVR model. Furthermore, residual distribution analysis highlights Random Forest's ability to effectively capture the underlying patterns of the dataset without significant biases.

Considering the comprehensive evaluations across various aspects such as trend tracking, volatility navigation, performance metrics, and residual analysis. The Random Forest model is chosen as the best-fit model for PV panel price time series forecasting. It offers stable, consistent, and accurate performance, making it ideally suited for scenarios where predictability and reliability are of high importance. To conclude, while XGBoost, SVR, and other models presented specific strengths and were competitive in certain aspects, the Random Forest model's overall balanced performance across multiple evaluation criteria makes it the most appropriate choice for the given dataset and forecasting requirements. Future applications may consider deploying the Random Forest model for similar time series forecasting tasks while ensuring periodic model retraining to adapt to new data patterns.

5.3 Phase II: Future Forecasting of the Price

To apply the best selected model for forecasting time series PV panel prices, to a real-life scenario of the future. The 26 features were forecast each alone for 36 months using the SARIMA model and combined in a new dataset. This dataset was utilized using the Random Forest Model from phase I. The observed and predicted prices from the model are illustrated in Figure 4. The observed prices represented by the blue line, are the combined data from Phase I. They exhibit some volatility and sharp fluctuations at the beginning. This is followed by a more stabilized declining trend around the midpoint around time point 40. After this phase, observed prices stabilize further, with some minor fluctuations. The forecasted prices represented by the red line, appear to remain relatively flat in Figure 4. This suggests that the model does not predict any significant price changes going forward. However, Figure 5 shows a zoomed-in view of the forecasted PV panel prices, showing more detailed fluctuations. While the variations are minor, there's a pattern of slight peaks and troughs, indicating some variability in future prices. Although there is some variation in the forecasted prices, they remain within a relatively consistent range. This indicates a stable prediction of the future prices of PV panels.

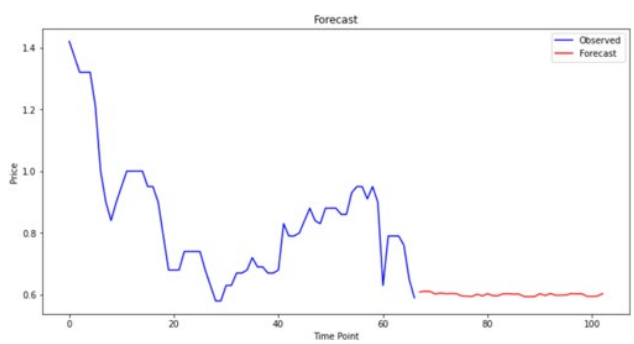

Figure 4. Random Forest Model Future Predictions.

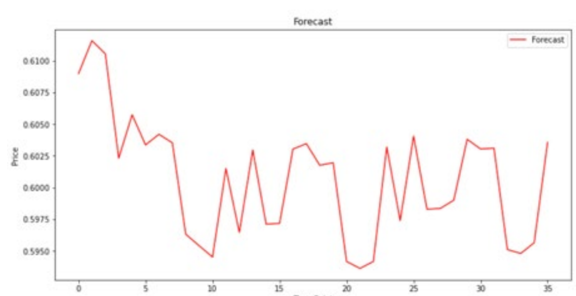

Figure 5. Random Forest Model Future Predictions Zoomed in.

These results can be corroborated by Figure 6, which displays the forecasted prices of PV panels sourced from an online platform (Mike n.d.). from July 2023 to December 2024. From this data, it's evident that the prices predicted by both models closely align. This similarity suggests that either future prices will stabilize or that the models used to forecast PV panel prices may require further refinement.

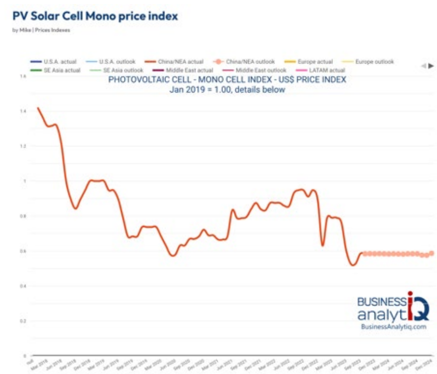

Figure 6. PV Solar Cell Mono price.

Further corroborating this observation, the Silicon Industry Branch of China's Nonferrous Metals Industry Association reported a relatively minor decline in polysilicon prices, which is ranged between 0.27% and 1.12%. This reduced rate of decline compared to previous rates, is further indication of the emerging stabilization in the market. Similarly, InfoLink Consulting's findings on the average price of polysilicon monocrystalline corroborate this trend. (Wang 2023)

Moreover, there is a noticeable alignment in the price metrics provided by different research institutions, suggesting a broader industry agreement on price stabilization. Also, the narrowing price gap between products from top-tier and lower-tier manufacturers indicates that the market is reaching a critical point in price reductions. This trend can be related to the fact that prices have reached a low point, with some even falling below the break-even threshold. This phenomenon leaves limited room for further reductions, confirming the ongoing stabilization of the market (Wang 2023). Furthermore, the inventory pressures that formerly constrained polysilicon manufacturers seem to be easing. This is evident from the increased sales outputs of leading companies and consistent inventory metrics. On the demand side, a stabilized pricing environment across the industry chain is influencing large-scale project developers to initiate construction. (Wang 2023)

Combining these observations with the earlier discussed predictions from the Random Forest and online source models, it is evident that the solar PV industry is set for a period of pricing stability. However, constant monitoring is required to identify any significant deviations or irregularities that might arise. This indicates the need for model adjustment or market reevaluation. Given the nature of the observed data, it would be crucial to evaluate the accuracy and reliability of the forecasted features by the SARIMA model. Considering that some features might need a more enhanced model or even a machine learning model, since they can be affected by other factors. Also, models like Random Forest tend to generalize based on the patterns they recognize in the training data. If the most recent data exhibits lower volatility, the model might forecast a smoother future trend. Therefore, periodically re-evaluating the model with updated data would be essential to ensure its predictive accuracy.

6. Conclusion
6.1 Contributions
This study makes several critical contributions to the field of solar energy price forecasting. A variety of factors are explored, including technological advances, governmental policies, and market dynamics and global economic indicators, to provide a comprehensive review of the factors that affect solar product prices. Addressing the existing gaps in data availability, this research collects the most relevant data, and through a robust feature analysis, identifies the most significant variables for PV panel prices.

Moreover, this study introduces machine learning models tailored to PV panel price forecasting. This innovative approach ensures the capture of the complexities of price dynamics existing in the solar industry. A significant aspect of this study's contribution lies in the evaluation of the machine learning models. By running extensive analyses and comparisons, the research identifies the most efficient model for forecasting andapplies it to future scenarios. This bridges the gap between theoretical advancements and practical applications. With the forecasted insights from the best-fit model, practical recommendations are provided. These are aimed at aiding stakeholders in their strategic planning, ensuring the optimal adoption of solar energy solutions, and driving the industry towards a sustainable energy future.

Lastly, the combination of these contributions hopes to facilitate the global transition to cleaner energy solutions, encouraging a future where solar energy plays a dominant role in global energy consumption. Therefore, this research serves as an innovative approach to solar energy price forecasting. Through its multifaceted approach and in-depth analyses, it offers a comprehensive framework for future studies and industry practices, with significant implications for the broader energy sector.

6.2 Final Outcomes

In conclusion, this paper analyzed and determined factors that influence PV panel prices, through literature review and correlation analysis. To refine the data, a feature analysis using linear regression optimized error loss. As a result, 26 features over 67 months were selected.

Phase I: The study applied five machine learning models: CNN, KNN, Random Forest, SVR, XGBoost for multivariate time series forecasting. Each model underwent a detailed evaluation including actual vs. predicted graph analysis, various performance metrics (SMAPE, MAPE, MAE, RMSE, MSE, and Huber loss), and residual distribution analysis. All models showed impressive performance. Based on the comprehensive analysis conducted, Random Forest and XGBoost models emerge from the comparison of actual and predicted performance. Among the performance metrics, the Random Forest model performs best, with the lowest SMAPE and impressive RMSE, followed closely by the SVR model. Furthermore, residual distribution analysis demonstrates that Random Forest can effectively capture underlying patterns without introducing significant bias.While XGBoost, SVR, and other models were competitive in certain aspects and had specific strengths, the Random Forest model offers the best overall performance across multiple evaluation criteria and is most suitable for the given dataset and forecasting requirements due to its balanced performance.

Phase II: Implementing the Random Forest model and integrating forecasted features from the SARIMA model, predictions indicate relative price stability in PV panels for the next 36 months starting July 2023. Although there are minor fluctuations in the predicted prices, these changes remain within a consistent range, indicating a stabilized market. The observation is further supported by comparable findings from an online platform and other research institutions, which indicate the industry is in agreement that prices will stabilize. Furthermore, the decline in PV panel prices, the narrowing price gap between products of different manufacturing tiers, and the ease of inventory pressures all point to a stabilizing market. However, the dynamic nature of the solar industry requires regular market analysis and model reassessments. Given that models like Random Forest rely heavily on recent data trends, periodic updates with new data are essential.

This study provides insight into the complexities of solar energy time series price forecasting, bridging gaps in data and highlighting factors that influence PV panel prices. By integrating tailored machine learning models, the study also provides valuable insights for stakeholders, steering them towards a sustainable energy future.

6.3 Recommendations and Future Work

In this study, 26 features have been analyzed for 67 months, which is a small sample size for time series forecasting. Therefore, using a larger dataset will enhance and improve forecasts. In which, instead of monthly data it would be more beneficial to use weekly, daily, or hourly datapoints. Data collection was limited due to the few available sites that give data for free. Therefore, it is advised to find a company or online source that gives valid data for a reasonable price. This will ensure a longer dataset. Also, expanding the study's geographical scope beyond China could offer a more comprehensive view of PV panel prices.

Although the five models used showed impressive performance, it would be beneficial to use or explore more machine learning models for forecasting multivariate timeseries data. Furthermore, improving the method of evaluation might also improve their performance. Consider alternative approaches such as LOOCV (Leave One Out Cross Validation) instead of k-fold cross validation and use more evaluation metrics. For future scenarios, it would be valuable to investigate the forecasted features further as some features may not be captured accurately by the SARIMA model because other factors influence them. In this regard, capturing their performance with another model would enhance the data.

Furthermore, when tailoring a statistical or machine learning model for forecasting multivariate timeseries data, a lot of factors must be taken into consideration. It would therefore be more accurate and efficient to develop and apply a hybrid model that combines both statistical and machine learning characteristics. To conclude, since changes in the solar industry are sudden, constant monitoring is necessary to detect any deviations or irregularities. Also, models like Random Forest tend to generalize based on the patterns they recognize in the training data. In order to ensure the model's predictive accuracy, it should be re-evaluated periodically with updated data.

Acknowledgements
We would like to express our appreciation and gratitude to Khalifa University of Science and Technology. We are equally grateful to our thesis committee members, Dr. Ahmad Mayyas and Dr. Mecit Simsekler, we express our deep appreciation for their time, detailed reviews, and insightful questions. Their feedback has been extremely valuable, and their shared expertise has significantly contributed to the depth and scope of this research.

References
Anon, Annual Capacity Additions of Solar PV, Wind and Other Renewables, Main and Accelerated Cases, 2020-2026. *Iea*. Available: https://www.iea.org/data-and-statistics/charts/annual-capacity-additions-of-solar-pv-wind-and-other-renewables-main-and-accelerated-cases-2020-2026, Accessed on October 23, 2023.

Anon, US Solar PV System Costs Increase in 2021. *Wood Mackenzie*. https://www.woodmac.com/news/opinion/us-solar-pv-system-costs-increase-in-2021/, Accessed on October 23, 2023.

Anon, China Statistics. *National Bureau of Statistics of China*. http://www.stats.gov.cn/english/, Accessed on October 23, 2023.

Bradford, T., Solar Revolution: The Economic Transformation of the Global Energy Industry, *MIT Press*, 2006.

Brown, L., Y. Zhao, and M. Green., The Efficiency Evolution of Solar Panels: A Review, 2018.

Cerjan, Marin, Ivana Krzelj, Marko Vidak, and Marko Delimar., A Literature Review with Statistical Analysis of Electricity Price Forecasting Methods, pp. 756–63 in *Eurocon 2013*. IEEE. 2013.

Cordoni, F., A Comparison of Modern Deep Neural Network Architectures for Energy Spot Price Forecasting, *Digital Finance,* vol. 2, no. 3–4, pp. 189–210, 2020. doi: 10.1007/s42521-020-00022-2.

Daniel Berberich, Hybrid Methods for Time Series Forecasting. *Inovex GmbH*. Available: https://www.inovex.de/de/blog/hybrid-time-series-forecasting, 2020. Accessed on October 23, 2023.

E, Jianwei, Jimin Ye, Lulu He, and Haihong Jin., Energy Price Prediction Based on Independent Component Analysis and Gated Recurrent Unit Neural Network. *Energy,* vol. 189, pp. 116-278, 2019. doi: 10.1016/j.energy.2019.116278.

Haegel, Nancy M., Robert Margolis, Tonio Buonassisi, David Feldman, Armin Froitzheim, Raffi Garabedian, Martin Green, Stefan Glunz, Hans-Martin Henning, Burkhard Holder, Izumi Kaizuka, Benjamin Kroposki, Koji Matsubara, Shigeru Niki, Keiichiro Sakurai, Roland A. Schindler, William Tumas, Eicke R. Weber, Gregory Wilson, Michael Woodhouse, and Sarah Kurtz., Terawatt-Scale Photovoltaics: Trajectories and Challenges. *Science,* vol. 356, no. 6334, pp. 141–43, 2017. doi: 10.1126/science.aal1288.

Jacobson, Mark Z., and Mark A. Delucchi. Providing All Global Energy with Wind, Water, and Solar Power, Part I: Technologies, Energy Resources, Quantities and Areas of Infrastructure, and Materials. *Energy Policy,* vol. 39, no. 3, pp. 1154–1169, 2011. doi: 10.1016/j.enpol.2010.11.040.

K. C. Gross, R. P. Broadwater, and P. P. Wung., A Study on the Impact of Time of Use Rates on the Cost and Valuation of Distributed Solar PV. *IEEE Transactions on Power Systems,* vol. 31, no.5, pp.3954–3962, 2016.

L. Yue, Q. Wang, J. Zhang, and J. Golbeck., The Review of the Impact of Silicon Supply Chain on PV Power Generation Cost. *IEEE Journal of Photovoltaics,* vol. 6, no. 6, pp. 1587–1593, 2016.

Lin, Yu, Qin Lu, Bin Tan, and Yuanyuan Yu., Forecasting Energy Prices Using a Novel Hybrid Model with Variational Mode Decomposition. *Energy,* vol. 246, pp. 123-366, 2022. doi: 10.1016/j.energy.2022.123366.

Mike, PV Solar Cell Mono Price Index. *Business Analytiq*, Available: https://businessanalytiq.com/procurementanalytics/index/pv-solar-cell-mono-price-index/, Accessed on October 23, 2023.
Niu, Dongxiao, Da Liu, and Desheng Dash Wu.,A Soft Computing System for Day-Ahead Electricity Price Forecasting, *Applied Soft Computing,* vol. 10, no. 3, pp. 868–875, 2010. doi: 10.1016/j.asoc.2009.10.004.
Philipps, Simon. 2023, Photovoltaics Report - Fraunhofer ISE, *Fraunhofer Institute for Solar Energy Systems*, Available: https://www.ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html, Accessed on October 23, 2023.
Qin, Quande, Kangqiang Xie, Huangda He, Li Li, Xianghua Chu, Yi-Ming Wei, and Teresa Wu., An Effective and Robust Decomposition-Ensemble Energy Price Forecasting Paradigm with Local Linear Prediction, *Energy Economics,* vol. 83, pp. 402–414, 2019. doi: 10.1016/j.eneco.2019.07.026.
Ribeiro, Matheus Henrique Dal Molin, Ramon Gomes da Silva, Cristiane Canton, Naylene Fraccanabbia, Viviana Cocco Mariani, and Leandro dos Santos Coelho., Electricity Energy Price Forecasting Based on Hybrid Multi-Stage Heterogeneous Ensemble: Brazilian Commercial and Residential Cases, in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, pp. 1–8, 2020.
S. Jiang, Y. Gong, and J. Wang., Supply Chain Collaboration for Mitigating the Price Volatility and the Bullwhip Effect in the Photovoltaic Industry, *Applied Energy,* vol. 238, pp. 123–133, 2019.
Sanders, N. n.d. *Forecasting Fundamentals.*
Smith, A., B. Jones, and C. Watson.,Renewable Energy Products: An Overview, 2015.
Tschora, Léonard, Erwan Pierre, Marc Plantevit, and Céline Robardet., Electricity Price Forecasting on the Day-Ahead Market Using Machine Learning, *Applied Energy,* vol. 313, 2022. doi: 10.1016/j.apenergy.2022.118752.
Wang, Bin, and Jun Wang., Energy Futures and Spots Prices Forecasting by Hybrid SW-GRU with EMD and Error Evaluation, *Energy Economics,* vol. 90, pp. 104-827, 2020. doi: 10.1016/j.eneco.2020.104827.
Wang, Jianliang, Changran Lei, and Meiyu Guo., Daily Natural Gas Price Forecasting by a Weighted Hybrid Data-Driven Model, *Journal of Petroleum Science and Engineering,* vol. 192, pp. 107-240, 2020. doi: 10.1016/j.petrol.2020.107240.
Wang, Jie, and Jun Wang., Forecasting Energy Market Indices with Recurrent Neural Networks: Case Study of Crude Oil Price Fluctuations, *Energy,* vol. 102, pp. 365–374, 2016. doi: 10.1016/j.energy.2016.02.098.
Wang, Vera., Polysilicon Prices Stablize after Months of Decline, *SOLARBE GLOBAL*, 2023. Available: https://www.solarbeglobal.com/polysilicon-prices-stablizes-after-months-of-decline/, Accessed on October 23, 2023.
Xiao, Jin, Yuhang Tian, Ling Xie, Xiaoyi Jiang, and Jing Huang., A Hybrid Classification Framework Based on Clustering, *IEEE Transactions on Industrial Informatics,* vol. 16, no. 4, pp. 2177–2188, 2020. doi: 10.1109/TII.2019.2933675.

Biographies

Aisha Alhosani is a Graduate from Khalifa University of Science and Technology. She earned B.S. in Mechanical Engineering from Khalifa University of Science and Technology, UAE, and Masters in Systems Engineering and Management from Khalifa University of Science and Technology, UAE.

Saed Amer is a highly qualified academic with a Doctorate of Philosophy in Computer and Information Systems Engineering. He earned this degree in August 2012 from Tennessee State University in the United States. Currently, he is a faculty member in the Department of Industrial and Systems Engineering at Khalifa University. Dr. Amer's research is focused on several areas, including Health Safety and Environment (HSE) engineering, training, and education. He is also involved in Human Factors simulation and validation, seat comfort analyses, and the integration of humans into Industry 4.0. Dr. Amer possesses expertise in the field of research related to fire protection, industrial hygiene, system safety, quality health, safety, and environment (HSE), as well as human factors pertaining to HSE.

Andrei Sleptchenko is an Associate Professor at the Department of Management Science and Engineering at Khalifa University. His primary research interests are in the modeling and optimization of various stochastic service systems with limited resources. Before joining Khalifa University, Dr. Sleptchenko served as a faculty member at the Department of Mechanical and Industrial Engineering at Qatar University. Prior to that, he worked for several years for different consulting companies in the area of Supply Chain and Transportation and as an Assistant Professor and a researcher at various Universities in the Netherlands and the USA. The current email address of Dr. Sleptchenko is andrei.sleptchenko@ku.ac.ae, and his homepage is https://www.ku.ac.ae/college-people/andrei-sleptchenko