ADVANCED ENERGY & SUSTAINABILITY RESEARCH
Open Access
www.advenergysustres.com

A Machine Learning Frontier for Predicting LCOE of Photovoltaic System Economics

Satyam Bhatti, Ahsan Raza Khan, Ahmed Zoha, Sajjad Hussain, and Rami Ghannam*

In this research article, the objective is to determine the return on investment (ROI) of photovoltaic (PV) power plants by employing machine learning (ML) techniques. Special focus is done on the levelized cost of electricity (LCOE) as a pivotal economic parameter crucial for facilitating economic decision-making and enabling quantitative comparisons among different energy generation technologies. Traditional methods of calculating LCOE often rely on fixed singular input values, which may fall short in addressing uncertainties associated with assessing the financial feasibility of PV projects. In response, a dynamic model that integrates essential demographic, energy, and policy data, is introduced encompassing factors such as interest rates, inflation rates, and energy yield, which are anticipated to undergo changes over the lifetime of a PV system. This dynamic model provides a more accurate estimation of LCOE. The comparative analysis of ML algorithms indicates that the auto-regression integration moving average (ARIMA) model exhibits a high accuracy of 93.8% in predicting consumer electricity prices. The validation of the model is highlighted through two case studies in the United States and the Philippines underscores the potential impact on LCOE values. For instance, in California, LCOE values could vary by nearly 30% (5.03 cents kWh^{-1} for singular values vs 7.09 cents kWh^{-1} using our ML model), influencing the perceived risk or economic feasibility of a PV power plant. Additionally, the ML model estimates the ROI for a grid-connected PV plant in the Philippines at 5.37 years, in contrast to 4.23 years using traditional methods.

1. Introduction

The late Nobel laureate, Prof. Richard Smalley, once emphasized that "energy is the single most important issue we face today" and global energy demand is rapidly increasing fourfold.[1,2] Indeed, the majority of the world's energy is consumed by China and the United States, with India following closely behind.[3,4] To address these escalating energy needs, solar energy presents a viable solution, offering the potential to profoundly improve the lives of communities worldwide.[5–7] Photovoltaic (PV) systems, which convert sunlight directly into electricity, are a means to harness the sun's renewable, sustainable, and low-carbon energy source.[8] These systems often achieve high efficiency in their conversion process.[9] However, one of the key challenges in promoting the widespread adoption of this technology is the cost of solar electricity compared to conventional energy sources.[10]

Grid-connected PV systems are cost-effective renewable energy solutions that do not require batteries and mainly consist of a PV array generator and an inverter for converting direct current (DC) electricity to alternating current (AC).[11,12] Furthermore, they can be configured to supply energy to primary loads, with all excess electricity being sold to the grid or even bought back from the grid when PV supply is insufficient.[13] Alternatively, all the energy generated from the PV system can be sold directly to the grid.[14] In all cases, the difference between the price of buying and selling electricity from the grid is a substantial factor in determining the optimum size of grid-connected PV systems.[15,16]

Therefore, evaluating the economic feasibility of a PV system is extremely important.[17] For example, users need to know their expected return on investment (ROI), which is a measure of profitability, and funding agents need means to analyze proposed technology development.[18] Similarly, technology developers need to understand how they will compete relative to other technologies.[19] Moreover, regulators and policymakers (who help define the economics of energy production) require reliable information.[20] The capital cost of a PV system, its operation and maintenance costs, and its expected energy yield must be considered systematically so that a comparison with conventional fossil fuels can be made.[21] Consequently, one needs a method to compare energy costs fairly. Therefore, we developed a generalized framework for predicting the feasibility of a grid-connected PV system for utility-scale applications. Our prediction framework considers the amount of electricity consumption using several metrics, including gross domestic product (GDP), prices of electricity, population growth, and weather data.

Recent literature highlights key advancements in the adoption of solar PV systems, showcasing their diverse applications and economic considerations. The paper titled "Adoption of

S. Bhatti, A. R. Khan, A. Zoha, S. Hussain, R. Ghannam
James Watt School of Engineering
University of Glasgow
Glasgow G12 8QQ, UK
E-mail: s.bhatti.2@research.gla.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aesr.202300178.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aesr.202300178

Floating Solar Photovoltaics on wastewater management system: A Unique Nexus of water-energy Utilization, low-cost clean Energy Generation, and Water Conservation" (Clean Technologies and Environmental Policy) explores the innovative integration of floating solar PV with wastewater management, emphasizing the synergies between water and energy utilization, cost-effective clean energy generation, and sustainable water conservation practices.[22] Additionally, the study "Grid parity analysis of photovoltaic systems considering feed-in tariff and renewable energy certificate schemes in Hong Kong" (Renewable and Sustainable Energy Reviews) critically assesses the grid parity of PV systems in Hong Kong, considering factors like feed-in tariff mechanisms and renewable energy certificate schemes. These works contribute significantly to understanding the multifaceted benefits and economic viability of solar PV systems, offering valuable insights into unique applications and policy considerations that resonate with the objectives of our present study.[23]

The motivation of our manuscript stems from the imperative to address the dynamic and intricate nature of factors influencing the economic viability of PV systems. In addition, the traditional methods often rely on static inputs, neglecting the evolving landscape of variables such as population growth, inflation rates, and demographic shifts, however, the deployment of ML algorithms in predicting the levelized cost of electricity (LCOE) offers a pioneering approach, enabling us to capture the nuanced interplay of these dynamic elements. Moreover, the frontier of ML applications in the realm of PV system economics, our research seeks to bridge the gap between conventional singular input methodologies and the transformative potential of advanced analytics. By employing ML algorithms, we aspire to provide a more accurate and adaptive framework for LCOE predictions, thus contributing to a comprehensive understanding of the long-term financial viability of PV systems and advancing the discourse in renewable energy economics.

1.1. Organisation of the Article

Our article is divided into 7 Sections. Section 2 reviews the literature on calculating the LCOE of grid-connected PV systems. Section 3 describes our methodology for calculating the LCOE and ROI of a grid-connected PV system using various ML algorithms and briefly discusses our proposed model. Section 4 includes the data explanation and steps of data pre-processing. Section 5 presents the results of our energy prediction models using various ML techniques and provides comparisons with singular input demographic variables. Next, in Section 6 we discuss our results, and concluding remarks are presented in Section 7.

1.2. Key Contributions

The following are our key contributions to the existing body of knowledge on PV systems' economic evaluation: 1) We highlight the crucial impact of considering dynamic factors in estimating the LCOE. Our research demonstrates a significant difference between LCOE estimations using singular inputs versus those obtained using ML models that consider dynamic variables. This insight underscores the need for a more comprehensive approach to LCOE calculation, one that incorporates changing demographic and economic factors; 2) Development of a dynamic model for LCOE estimation, which incorporates important dynamic variables such as demographic, energy and policy data, including interest rates, inflation rates and energy yield, improving traditional methods relying on fixed singular input values; 3) Comparative analysis of ML algorithms for predicting consumer electricity prices; 4) Validation of the proposed model through case studies. Our work is supported by practical validation using two case studies from different geographic regions (the United States and the Philippines); and 5) Demonstrating a substantial difference in LCOE calculations using the traditional approach versus the proposed ML model. We show that the LCOE value increases substantially when employing the ML model compared to the traditional method using singular inputs. This underlines the importance of considering dynamic factors in such estimations for more realistic and accurate assessments of PV systems' long-term financial viability.

2. Literature Review

Numerous examples in the literature describe the statistical and probabilistic models for calculating the LCOE and ROI of PV systems. For example, K. Branker et al.[24] argued that there is a lack of understanding of the calculations involving assumptions and justifications for the estimation of LCOE, thus proving that poor assumptions lead to contradictory results for the calculations of energy ROI of a PV system. In their paper, they calculated the LCOE to reduce the assumptions-based model and represent a more accurate one. Nevertheless, their study was limited to singular inputs for calculating the LCOE.

A more detailed calculation of LCOE by Chul-Yong Lee et al.[25] presented a stochastic model for calculating LCOE for solar PV systems installed in the Philippines. Their results depicted that for a commercial solar panel, the LCOE ranged from a minimum of 10 to 18 ¢ kWh^{-1}. Moreover, they performed a sensitivity analysis to validate their results. However, their study lacked the optimized LCOE value and only discussed a range of possible LCOE values in their model.

Another study for a utility-based system installed in IESCO, Pakistan, conducted by Ahsan Khan et al.[26] showed an analysis for forecasting day ahead load demand using the auto-regressive (AR), moving average (MA), and auto-regressive integrated moving average (ARIMA) model for the statistical modelling for the load demand. In addition, they performed a comparative analysis using artificial neural networks (ANN) and bagged regression tree (BRT) models. Although their results for forecasting the load demand using various ML techniques are precise but lack the estimation of LCOE and hence, the energy ROI of the model.

Again, Geissmann et al.[27] showed a probabilistic approach for computing the LCOE of a nuclear plant and a gas power project. Furthermore, they implemented a Monte Carlo simulation to determine the dependency of singular input parameters on the model's final results. However, their study used singular inputs and lacked dependency on demographic variables. Further, Georgitsioti et al.[28] discussed the formula used to calculate the LCOE based on singular values for domestic PV systems in the UK, and the financial benefits that can be gained from

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a domestic PV system under the "Feed in Tariff (FiT)" PV supporting policy in the UK.

3. Methodology

The methodology section of our article outlines all the essential steps for calculating the LCOE and ROI of a grid-connected PV system. Furthermore, this section provides information on the ML algorithms employed for forecasting the LCOE using demographic variables. We also introduce our proposed model, which was used to estimate the LCOE and the ROI of utility-based grid-connected solar energy systems[29] installed in Sacramento (California) and Butuan City, Philippines.

3.1. Calculating the LCOE

A valuable parameter for comparing the cost of electricity production from any energy generation system over its lifetime is the levelized cost of electricity (LCOE).[30] According to the US's National Renewable Energy Laboratory (NREL), the LCOE is defined as the "net present value of the unit-cost of electricity over the lifetime of a generating asset".[31] It serves as a comparative measure of various power sources, aiming to provide a consistent basis for comparison. It offers an economic assessment of the average total cost to build and operate a power-generating asset over its lifetime, divided by the total energy output of the asset over that lifetime. This is typically defined as the average cost ($) per kWh of useful electrical energy generated by the power generating system throughout its years of operation. Mathematically, the LCOE can be calculated as follows

$$\mathrm{LCOE} = \frac{\text{System Lifetime Cost, } L_\mathrm{t}}{\text{Lifetime Energy Production Cost, } E_\mathrm{t}} \tag{1}$$

$$L_\mathrm{t} = I_\mathrm{t} + C_\mathrm{t} + S_\mathrm{t} \tag{2}$$

$$E_\mathrm{t} = E_0(1-d)^\mathrm{t} \tag{3}$$

where, I_t represents the initial costs, including expenses related to equipment, land and other setup necessities. The total costs paid at the beginning of the project, such as annual operation and maintenance costs, are denoted as C_t. Lastly, S_t signifies the salvage value, which is the use value of the project at the end of its lifetime. Similarly, the lifetime energy production cost (E_t) is defined as $E_\mathrm{t} = E_0(1-d)^t$, where E_0 refers to the initial rated energy output and the system degradation factor is represented by $(1-d)^\mathrm{t}$. This formula accounts for the gradual decrease in energy output over time due to factors such as aging and wear of the PV system components.[32]

Traditional methods of calculating LCOE relied on using singular input values for each of the variables above.[33] For instance, using the benchmark prices reported for 2017, a 50 MW utility-scale PV power plant installed in California would cost $56 million, corresponding to $1.12 W^{-1} (31% module, 69% balance of systems). This system would produce approximately 86 GWh of energy in the first year. Assuming that the discount rate is 5.5%, the federal tax rate is 30%, the state tax rate is 8%, the evaluation period is 25 years, and the system degradation is 0.5%, then the LCOE of this system is 5.83 c kWh^{-1}. A careful consideration of these numbers, as mentioned above, shows that many assumptions have already been made to determine the LCOE of this system.

Nevertheless, a case study of a PV system installed in Spain indicates that estimating the LCOE using traditional methods may lead to inaccurate estimations. This is particularly relevant as factors such as inflation rate, discount rate, degradation rate, and consumer price of electricity (CPE) are likely to vary during the lifetime of a PV project (typically 25 years). In Spain, LCOE analysis proved inadequate when an excessive number of projects were developed based on overly optimistic assumptions regarding panel failure rates and other performance factors.[34] A more comprehensive examination of the uncertainties associated with these assumptions might have averted significant losses.

Furthermore, the straight-line depreciation method is estimated for allocating the consumer price of electricity over its useful life.[35] Under this approach, the consumer price of electricity is divided by the estimated useful life of solar panels to determine the annual depreciation expenses.[36] The resulting annual depreciation amount remains constant throughout the asset's useful life. This method provides a systematic and simple way to allocate the cost of an asset over time, making it widely used for financial reporting and calculations like LCOE for renewable energy projects. Therefore, in our study, we employed straight-line depreciation to ensure a consistent and transparent approach to for accounting the depreciation of relevant assets in LCOE calculations.

3.2. Calculating the ROI

After determining the LCOE, the next step is to calculate the ROI for the PV system. To do this, it is necessary to compare the LCOE with conventional electricity prices.[37] However, this parameter is likely to change over the project's lifetime. As the input parameters continuously change and are strongly dependent on the system's location, we propose using ML techniques to accurately determine the ROI of a PV system. ML techniques can effectively capture the complex relationships between various factors and adapt to changing input conditions, making them well-suited for predicting the ROI of the PV system with higher accuracy.

Mathematically, the ROI of a PV system can be calculated using $\mathrm{ROI} = T_\mathrm{C}/B_\mathrm{I}$, where T_C represents the *total cost of the PV system*, and B_I denotes the annual benefit from the installation of the PV system. Here, the total cost of the PV system refers to the initial investment required for the PV system, including costs related to equipment, land, installation and other setup necessities. It is also sometimes called the capital expenditure (CAPEX) cost. Therefore, the ROI parameters estimate the number of years a client can expect to achieve an ROI for installing a PV system. A review of the literature reveals that the methods for calculating the ROI used by researchers worldwide often rely heavily on assumptions, leading to imprecise cost analysis estimations.[38]

3.3. ML Algorithms

Considering the factors previously discussed, we introduce ML techniques for calculating the LCOE and ROI of PV systems.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ADVANCED ENERGY & SUSTAINABILITY RESEARCH Open Access
www.advenergysustres.com

In this subsection, we explore the ML techniques that can be used to accurately forecast the LCOE.[39] ML, a branch of computer science, employs computational training algorithms to make predictions based on known input datasets.[40] ML can be broadly classified into three main categories: supervised, unsupervised, and reinforcement ML algorithms.[41] In supervised ML, the computer learns from labelled input data provided by the user, while in unsupervised ML, the computer identifies patterns within unlabelled data.[42] Additionally, reinforcement ML involves a trial-and-error learning approach where the agent (computer) iteratively makes decisions based on feedback it receives.[43] For this study, we focus on supervised ML, specifically employing regression techniques to predict the LCOE of PV systems. Subsequently, we calculate the ROI of the PV systems based on the predicted LCOE. Finally, we compare the LCOE values calculated using various approaches, incorporating fixed input values for different parameters to highlight the advantages of our proposed method.

3.4. Our Proposed Model

In a majority of previous studies, researchers have calculated the LCOE and ROI using singular input values, typically assuming that the CPE will increase by 5–10% over the lifetime of the solar plant. However, we argue that the estimation of LCOE should account for variations in CPE due to factors such as population growth, inflation rate, and interest rate over time.[44] Consequently, we propose an algorithm that accurately considers these dynamic variables. Using historical data, we apply ML algorithms to estimate the LCOE, taking into account the aforementioned factors. To validate our model, we extract historical data from two regions, California in the USA and Butuan City in the Philippines and compute the error function of various ML techniques to identify the most suitable ML model. **Figure 1** illustrates our proposed system, encompassing the required input parameters, relevant variables and ML algorithms used to determine the ROI of a PV plant.

The input parameters for our model were selected based on their significant influence on the LCOE. They include demographic, energy, and policy data, such as interest rates, inflation rates, and energy yield, which are expected to change over a PV system's lifetime. Each of these parameters carries a specific physical meaning in the context of LCOE calculations: 1) Interest rates: They influence the financial feasibility of PV projects by affecting the cost of capital; 2) Inflation rates: They impact the value of money over time, influencing the real costs of the PV system across its lifecycle; and 3) Energy yield: It represents the amount of energy produced by the PV system, a key determinant of its cost-effectiveness.

We appreciate that there could be other potential influencing factors. However, we focused on these parameters due to their direct and significant impact on LCOE. Moreover, including too many variables could increase the complexity of the model without necessarily improving its predictive accuracy. Also, it was essential for us to select parameters for which reliable historical

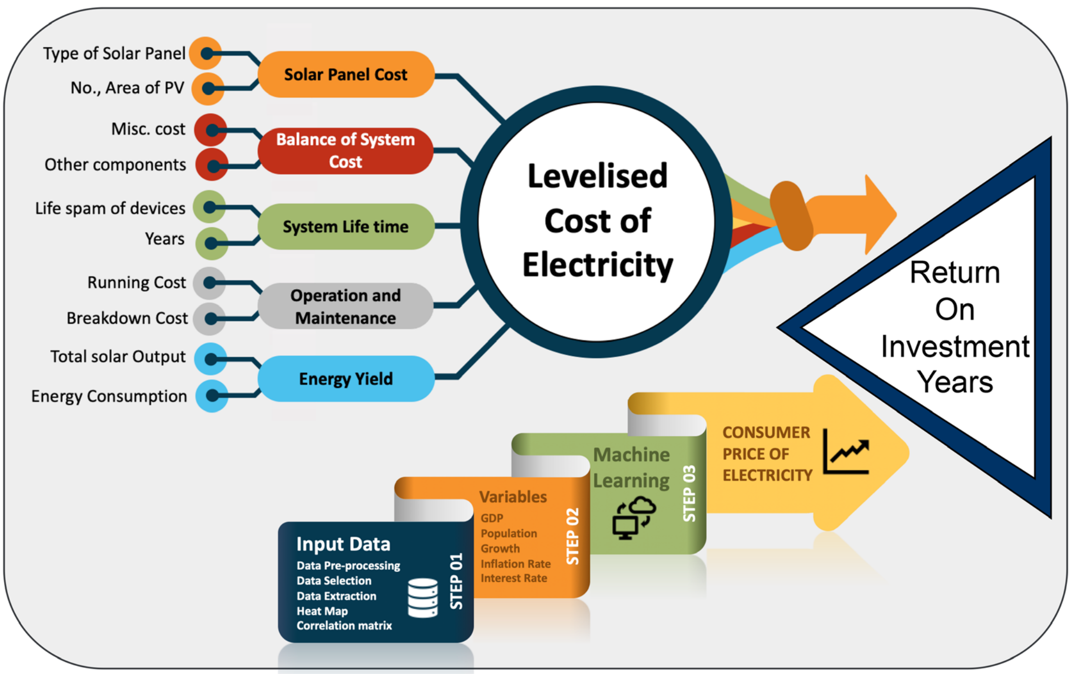

Figure 1. The figure demonstrates our proposed model for determining the LCOE and ROI for a utility-connected solar home system. The parameters required for estimating the LCOE, such as solar panel cost ($), the balance of system cost ($), system lifetime (years), operation and maintenance cost ($) and the energy yield ($ kWh^{-1}) are indicated. Moreover, the steps to evaluate the dependent variable (CPE) are mentioned using three steps. The first step involves determining the dataset and data preprocessing, followed by the second step, which discusses independent variables such as GDP, Population growth, inflation rate, interest rate, etc., that are used for estimating the defined dependent variable (CPE). The third step describes various ML techniques for accurately forecasting the energy ROI.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

data were available, as the ML model's performance depends on the quality of the training data.

4. Data Explanation

This section provides detailed information about the selected dataset values necessary for accurately estimating the LCOE and ROI of utility-based grid-connected PV systems. Initially, we discuss data extraction, plotting of datasets, and the heatmap for independent and dependent variables, which helps determine the correlation matrix.

4.1. Data Extraction

To predict the LCOE and ROI, it is crucial to obtain real-time data from reputable sources such as the environmental investigation agency (EIA),[45] international renewable energy agency (IRENA),[46] bureau of economic analysis (BEA),[47] and international energy agency (IEA).[48] Moreover, the CPE data was extracted from EIA, followed by the statistical data on population growth and the gross domestic product extracted from the websites BEA and IRENA. It is worth mentioning that some of the data were available on a quarterly or annual scale, however, to maintain the unity in the data comparison, the data was extrapolated using Python's generative adversarial networks (GAN) framework to obtain the complete data on an annual scale. For our study, we collected historical data for Sacramento, California, USA. The datasets comprise independent demographic variables such as the consumer price index (CPI) as a measure of the inflation rate (X_1), population growth (X_2), and gross domestic product (GDP) (X_3). In contrast, the dependent variable (Y) is represented by the average CPE (cents kWh^{-1}).[49] The timescale for the extracted dataset is monthly. **Table 1** presents the respective dependent and independent variables' dataset of demographic values, ranging from January 2005 to December 2021.

Additionally, it is worth mentioning the input parameter selection. For the various time series forecasting ML techniques, the selection highly depends on a particular domain and those parameters are only considered which are believed to influence the targeted variables, herein the CPE significantly. The independent variables have a meaningful relationship with the dependent variables. For instance, the input parameters, such as GDP, CPI, inflation rate, weather data, time of day, day of the week, and seasonal indicators, each of these parameters is expected to influence the energy consumption pattern over the system's lifetime.

Moreover, some of the other potential influencing factors are excluded due to the fact including too many variables or parameters in the ML models leads to the increased complexity of the model, overfitting, and results in reduced interpretability of our proposed model. Therefore, the selection of the parameters incorporates a critical balance between the meaningful factors and thus, avoids unnecessary complexity.

Exclusion of other potential influencing factors: The selection of input parameters is a crucial step in time series forecasting, and it requires careful consideration. While it's essential to include relevant factors, it may not be feasible or necessary to include all potential influencing factors. Including too many parameters can lead to overfitting, increased complexity, and reduced interpretability of the model. Therefore, the selection process involves striking a balance between including meaningful factors and avoiding unnecessary complexity.

Table 1. The table showcases an example of data extracted from various online websites such as EIA, IRENA, BEA, IEA, etc. The dataset consists of the dependent variable (CPE) and independent variables (Consumer price of the index, interest rate, GDP, population growth) over 15 years. The time resolution for the extracted dataset is in months.

Date	X1	X2	X3	Y
	Inflation	Population growth	Gross domestic product	CPE [¢ kWh^{-1}]
01/01/2005	200.35	294957.00	128234.50	12.19
01/02/2005	201.20	295167.33	128717.47	12.33
01/03/2005	201.85	295377.67	129200.43	12.12
01/04/2005	202.50	295588.00	129683.40	12.57
01/05/2005	201.85	295838.67	130635.33	13.4
01/06/2005	201.20	296089.33	131587.27	13.16
01/07/2005	202.10	296340.00	132539.20	13.43
01/08/2005	203.00	296588.67	133226.70	12.14
01/09/2005	204.45	296837.33	133914.20	11.3
01/10/2005	205.90	297086.00	134601.70	11.28
01/11/2005	204.65	297302.67	136682.93	12.8
01/12/2005	203.40	297519.33	138764.17	12.91
–	–	–	–	–
–	–	–	–	–
01/01/2021	303.67	331949.00	312120.20	21.43
01/02/2021	304.39	331973.00	310499.27	22.53
01/03/2021	306.90	331997.00	308878.33	23.37
01/04/2021	309.42	332021.00	307257.40	22.75
01/05/2021	309.46	332113.00	309934.50	23.11
01/06/2021	309.50	332205.00	312611.60	22.46
01/07/2021	310.33	332297.00	315288.70	23.34
01/08/2021	311.17	332392.67	319178.13	23.44
01/09/2021	312.22	332488.33	323067.57	21.97
01/10/2021	313.27	332584.00	326957.00	22.77
01/11/2021	314.54	332639.00	327936.67	23.83
01/12/2021	315.81	332694.00	328916.33	23.22

4.2. Statistical Representation of the Dataset

Initially, the collected raw dataset is non-uniform and contains noise, disturbances, irregularities, seasonality, trends, or patterns associated with it. Therefore, it is essential to understand these parameters before inputting them into our ML model. Subsequently, **Figure 2** depicts the plot of dependent and independent variables used to estimate the dependent variable, i.e., the CPE. The dataset plot shows that parameters such as population growth, the CPI, GDP, and CPE have a linear relationship.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH
26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ADVANCED ENERGY & SUSTAINABILITY RESEARCH
Open Access
www.advenergysustres.com

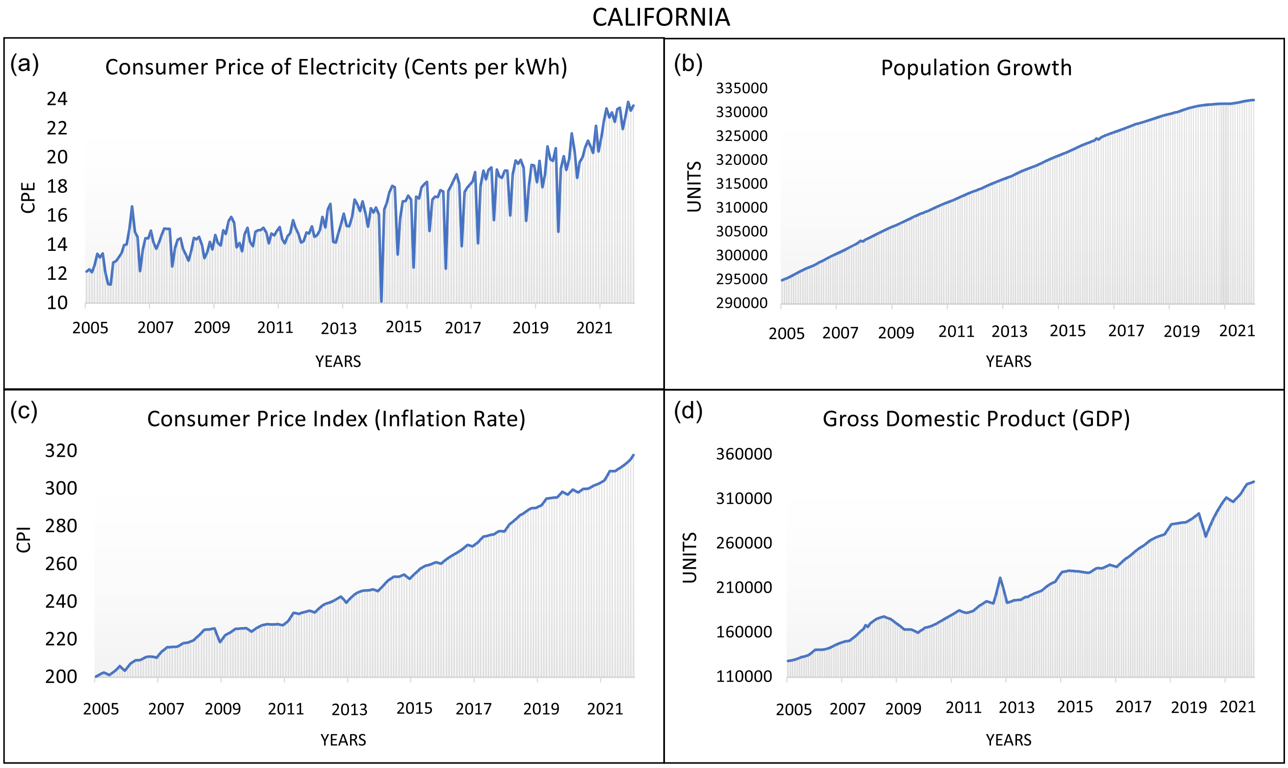

Figure 2. The statistical representation of the dataset for independent variables a) Consumer Price of Electricity (CPE), b) Population growth, c) Consumer Price Index d) Gross Domestic Product (GDP), seasonality and noise for the city of Sacramento in California, USA.

However, the CPE parameter has seasonality, noise, and irregularities associated with it. The chosen dataset ranges from January 2005 to December 2021.

4.3. Heat-Map for the Correlation Matrix

In ML, feature selection is a method in which we consider only those independent features in our model that contribute significantly to estimating the dependent variable. Accordingly, we use the heat map of the correlation matrix to distinguish between the independent variables and the dependent variables. **Figure 3** shows the heatmap for the parameters of the inflation rate, population, GDP, and CPE. The CPE and gross domestic product (GDP) have a correlation value of 0.91, indicating a strong relationship between these variables and their importance in estimating CPE. However, the population growth parameter exhibits a correlation value lower than 0.85, suggesting a weaker relationship with CPE. In fact, including the population growth data led to no change in the final outcome of our results, however, did lead to an increased computational time of the proposed model.

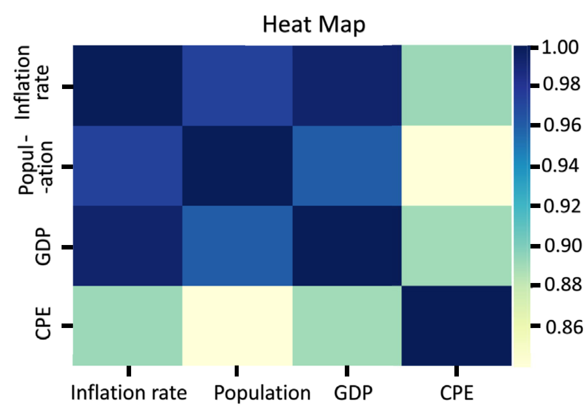

Figure 3. The correlation matrix showcases the heatmap for evaluating the inter-dependency of variables concerning each other. The chosen parameters for the correlation matrix are inflation rate, population growth, gross domestic product and CPE. The lighter colour (green) depicts the least significance, whereas the darker colour (blue) shows the most significant parameter.

5. Results

In this section, we discuss the implementation of various ML techniques using the aforementioned datasets and parameters to train and test our proposed model for accurately forecasting the LCOE and ROI of utility-based solar home systems. Additionally, we compare the results from our ML models with time series forecasting models such as autoregressive integrated moving average (ARIMA), long short-term memory (LSTM), and seasonal autoregressive integrated moving average (SARIMA). The results presented here focus on two locations: Sacramento, California, USA, and Butuan City, Philippines. We specifically chose these locations due to the availability of high-quality

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

datasets for demographic variables. Furthermore, parameters such as CPE are consistent in these regions and do not vary based on rates determined by the government or industry. This consistency allows for a more reliable evaluation of our proposed model and its performance in predicting the LCOE and ROI for solar home systems.

5.1. LR Model

For predicting the dependent variable CPE ($), we first applied the supervised learning ML model, specifically the linear regression (LR) model. The LR model determines the best-fit linear line between the independent and dependent variables. We have already defined the dependent and independent variables in the methodology section, and subsequently used the LR model with input (dependent) variables such as population growth, CPI, and GDP to calculate the dependent variable, i.e., CPE. **Figure 4** shows the scattered plot of actual values concerning the predicted CPE in a linear relation. Accordingly, the LR model predicts the output values for the CPE over the next ten years.

The actual versus predicted plot for the CPE showed an accuracy of less than 85%, and the error loss function showed a root mean square error (RMSE) value of more than 10%. According to the literature,[50] the accuracy should have a value of more than 90%, and RMSE should be less than 10% for the LR model to predict the values accurately. The limitation of such poor accuracy is that the input data for the independent variables was limited, and we only used three independent variables. To address this, we used high-quality data and added several independent variables.

5.2. LR Model with Multiple Variables

To improve the accuracy of our LR model, we incorporated multiple independent variables and increased the duration of each variable, i.e., from January 2005 to December 2021. It is worth mentioning that there were instances where the data was available in quarterly or annual resolutions. However, to enhance the accuracy of the LR model, the input data should be consistent and have the same time resolution. Accordingly, we employed a tool called generative adversarial networks (GAN), a subclass of ML in which two neural networks are considered.

As a result, the GAN model produces the best ML model among these two neural networks. One of the advantages of the GAN model is its ability to improve the quality of the model even with poor datasets. Additionally, to predict the dependent variable CPE, we divided the dataset into 80% and 20% to train and test the LR model with multiple variables. **Figure 5** showcases the results for the predicted values versus the actual values after executing the LR model with multiple variables. The overall accuracy for the LR model with multiple variables is 87%.

The accuracy achieved using the aforementioned model is within the limit of more than 85%. However, the model under consideration is not ideal for accurately forecasting the LCOE and ROI parameters of the utility-based solar home system, as it may still lead to ambiguity regarding the exact assumption of the ROI in terms of the year. Therefore, it is essential to identify an ML method with an accuracy of at least 90%.[51] In this regard, we implemented LSTM time series forecasting to improve the accuracy of the ML model and reduce the loss error function.

5.3. LSTM Model

Another ML model, the long short-term memory (LSTM) model, was applied to enhance accuracy. The LSTM method belongs to a subset of ANNs within the domains of AI and deep neural networks. Additionally, the LSTM model is a recurrent neural network (RNN) used for analyzing time series forecasting. In our model, we aim to predict the energy ROI, making time series forecasting crucial for accurately predicting the ROI of the

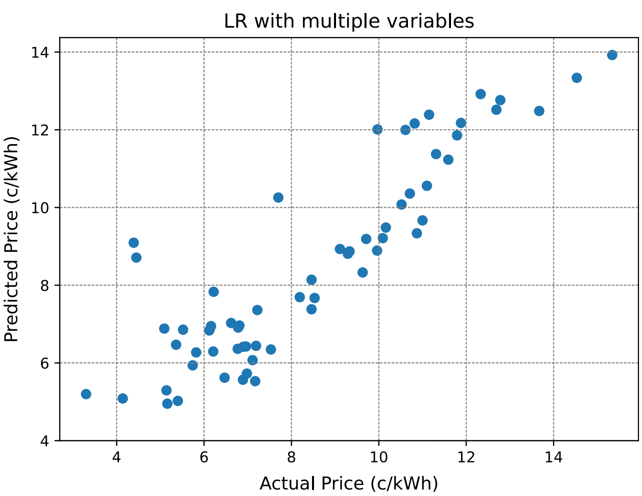

Figure 4. The scattered plot of the predicted values to the actual values for the consumer price electricity (¢ kWh^{-1}) using the supervised learning ML, subcategory, LR model. Herein, the dependent variable CPE is predicted linearly to the actual values of the parameter under consideration.

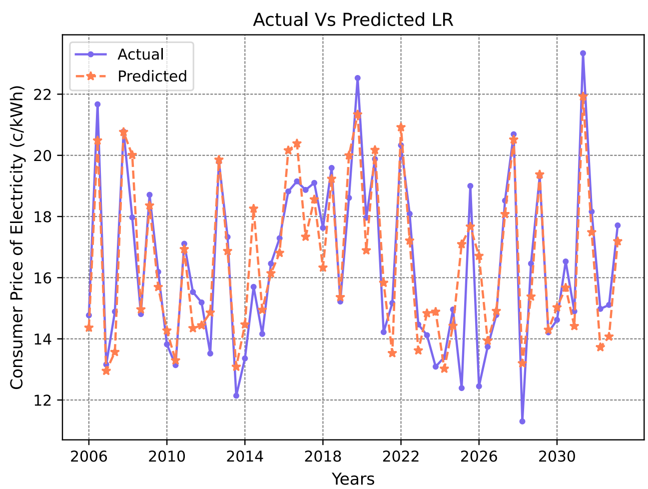

Figure 5. The figure shows the result of forecasting the dependent variable CPE (¢ kWh^{-1}) using the LR model with multiple input independent variables. The curve demonstrates the accuracy of the LR model with multiple variables. The curve in blue depicts the actual values of the CPE, whereas the curve in orange depicts the forecasted values. The curve is the relation of actual versus predicted values of the CPE over the period.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

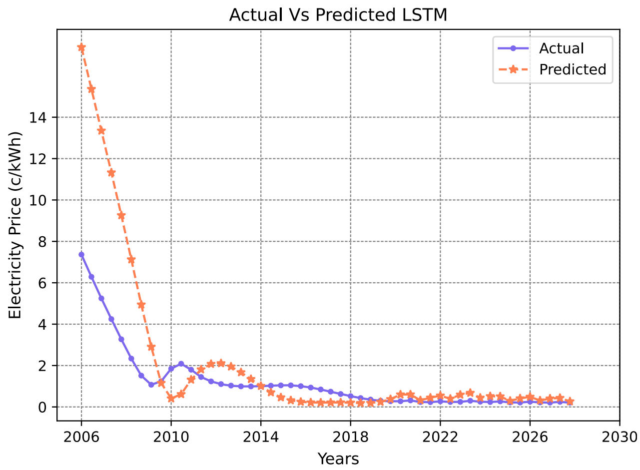

Figure 6. The curve describes the plot of actual versus predicted values of the CPE (¢ kWh^{-1}) using the LSTM model with multiple variables. Herein, the dataset is divided into train and test of 70% and 30%, respectively. The input data for the output result is for Sacramento, California, USA. Also, the blue curve here shows the values for the train, while the orange curve here shows the testing of the models.

installed system. Consequently, our results are extended by applying the LSTM model with multiple independent variables.

Our results, demonstrated in **Figure 6**, show the dataset divided into 70% for training and 30% for testing. The LSTM model with multiple variables achieves an RMSE of 3.237% and an accuracy of 91%. The larger error in the early years of the LSTM model can be attributed to several factors. As mentioned earlier, the LSTM model is a type of RNN designed to capture long-term dependencies in sequential data. Initially, the model may struggle to capture these dependencies, leading to higher errors in the early stages of the time series. As the LSTM model progresses through the time series and continues to train, it gradually learns the underlying patterns and relationships in the data. This learning process enables the model to better capture long-term dependencies and adapt to the time series' dynamics. Consequently, the model's predictions become more accurate over time, leading to a convergence of the error.

Figure 6 demonstrates the plot of the train and test of the LSTM model with multiple variables for predicting the dependent variable CPE (¢ kWh^{-1}) for Sacramento, California, USA. Moreover, the predicted values of the CPE from the model are incorporated to calculate the LCOE and ROI of the utility-based solar home system. Though the LSTM model with multiple variables achieved an accuracy of 91%, however, to obtain a more accurate model, we performed the ARIMA model test as discussed in the following subsection.

5.4. ARIMA Model

Next, we applied the ARIMA model to forecast the dependent variable, i.e., CPE. In general, an ARIMA model is a model that is fitted to the d^{th} order differenced time series, ensuring that the resulting differenced time series is stationary. A stationary time series is one in which the mean, variance, autocorrelation, and other statistical features remain constant across time. We will apply the ARIMA model for time series forecasting in the study under consideration. Additionally, we used a similar dataset as input to test our ARIMA model.

Apart from the high accuracy of the ARIMA model, there are several reasons for implementing it. First, ARIMA is a parametric model that offers interpretable coefficients that can be used to understand the underlying time series process. Second, the ARIMA model is highly flexible, as it can be applied to a wide range of time series data, including stationary, non-stationary, and seasonal data. Furthermore, ARIMA models can be extended to handle exogenous variables, making them valuable in forecasting scenarios where other factors may impact the time series. Lastly, ARIMA models are robust to missing data and outliers, and numerous libraries and software packages provide built-in ARIMA functions.

Before applying the ARIMA model, we tested the Dickey–Fuller algorithm, a parameter used to check the stationarity of the input dataset. The results of the Dickey–Fuller test indicate whether the dataset is stationary or not, based on the condition that the p-value (probability of the null hypothesis) should be very small. In our case, the model yielded a p-value of 0.23, which suggests that the dataset is stationary. In the ARIMA model, the autoregression (AR) part uses previous values to make future predictions, the moving average (MA) part uses past errors for making future predictions, and the integrated (I) component represents the difference between the AR and MA.

In order to weigh each factor under consideration, we also incorporated statistical tests such as the t-test and F-test for assessing the significance of the individual coefficients of the AR, MA, and the constant term for each factor. The t-test determines the t-value for each coefficient, which expresses how far from zero the coefficient is in terms of standard errors. Indicating that the coefficient is statistically significant at the chosen level of significance (often 5% or 1%), a high t-value (generally larger than 2 or 2.5) is required. On the contrary, the combined significance of a set of model coefficients is evaluated using the F-test. The F-test is specifically used to test whether a subset of the coefficients–typically all the coefficients in a particular order–are equal to zero. When the p-value is low (often less than 0.05), the null hypothesis can be rejected and the subset of coefficients is jointly significant.

Furthermore, the model runs a set of interactions based on the hit-and-trial method for calculating the most appropriate values for p (number of autoregressive terms), q (number of lagged forecast errors in the forecast equation), and d (number of nonseasonal differences required for stationarity). The results are analyzed using Akaike's Information Criterion (AIC), which helps determine the predictors for the regression model. Subsequently, the model searches for the minimum AIC score and the (p, q, and d) values. Using our input data in the ARIMA model resulted in a minimum AIC score of 3214.29 and (p, q, d) values of (1, 0, 1), respectively.

In addition, the dataset was split into training (70%) and testing (30%) portions, along with the order (1, 0, 1) to apply the ARIMA model. **Figure 7** demonstrates the actual (blue curve) vs predicted (orange curve) values for CPE, and the grey area highlights the confidence interval of 95% using these input values in our ARIMA model. It is worth mentioning that a

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

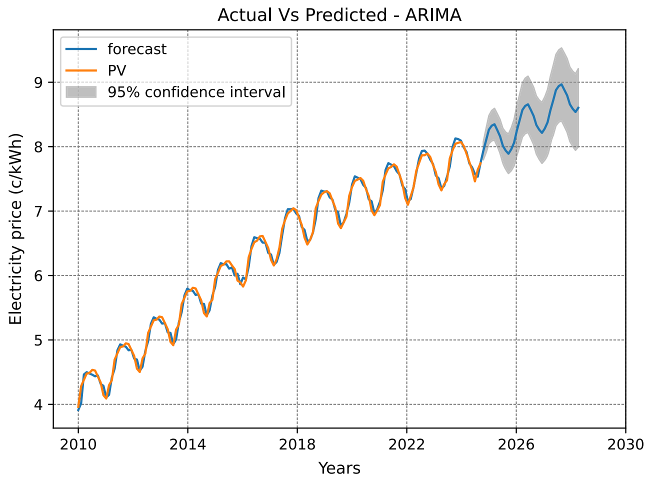

Figure 7. The curve depicts the actual vs forecasted values for CPE (¢ kWh^{-1}) for Sacramento, California, USA, using the Autoregressive integrated moving average (ARIMA) model. The area shaded with grey demonstrates the confidence interval of the forecasted values of CPE. The orange curve shows the predicted values, whereas the blue curve shows the actual values of the CPE and accordingly, the time resolution of the dataset is in months ranging between a period of 2010 till 2030.

confidence interval is a set of values surrounding a point estimate of a performance metric for a model (such as accuracy, precision, recall, etc.) that encapsulates the range of values in which the actual value of the performance metric is anticipated to reside with a given degree of confidence. The yellow-colored dotted lines indicate the range of the predicted values. We achieved an accuracy of 93.8% and forecasted CPE values up to 2030. Therefore, our proposed model achieved a maximum of 93.8% accuracy and was the most appropriate model for predicting the LCOE parameter among other ML techniques.

Accordingly, after determining the most appropriate model for predicting one dependent variable, i.e., CPE, we applied the same procedure for calculating the different other dependent variables (as mentioned in Figure 1), such as solar panel cost ($), the balance of system cost ($), system lifetime (years), operations and maintenance cost ($), energy yield ($/kWh), and incentives ($). Therefore, the dataset was collected for various independent variables such as the type of solar panel, number of solar panels, the area required, the life span of devices, energy consumption, solar energy generated, associated breakdown costs, etc. From the literature, we considered two case studies from Sacramento, USA, and Butuan City, Philippines to make a detailed comparison of our proposed model. The dataset of the demographic variables ranges from a duration between 2005 to 2021.

6. Discussions

The use of ML techniques for calculating the LCOE and ROI of a PV system is driven by the recognition of a significant difference in estimating these economic parameters when dynamic demographic variables are considered as opposed to relying on singular inputs. Several existing software tools estimate these economic parameters (e.g., PVWatts, EnergySage, SolarEstimate, Google Project Sunroof, RETScreen, HOMER Energy, Solar Design Tools, PVSOL, Clean Power Estimator). However, these tools often assume constant demographic variables over the system's lifetime, which we argue is not an accurate method for estimating the LCOE and ROI of a PV system. Our research aims to demonstrate the benefits of considering dynamic demographic variables in these estimates.

The specific advantages of ML over traditional methods are: 1) Adaptability: ML models can handle variables that change dynamically over time, allowing for a more realistic and accurate prediction of LCOE; 2) Complex interactions: ML can capture complex non-linear interactions between variables, facilitating a deeper understanding of how different parameters impact LCOE; and 3) Predictive power: ML models demonstrated better predictive performance in our case studies. The ARIMA-based ML model achieved an accuracy rate of 93.8% in predicting CPE, a significant improvement over traditional methods.

We implemented time-series forecasting ML techniques, such as ARIMA, to predict future values based on historical data. These techniques analyze patterns, trends, and seasonality within the data to make accurate predictions, which are not possible using traditional methods. While it is true that the theoretical derivation of LCOE/ROI with dynamic influencing factors could be done using multi-step calculations and approximations, this approach has limitations: 1) Accuracy: Approximations may lead to errors, which can accumulate over time; 2) Efficiency: ML techniques simplify the process and improve the efficiency of calculations, especially when dealing with large datasets or multiple variables; and 3) Model complexity: Multi-step calculations with dynamic parameters can result in complex models that are difficult to manage, whereas ML models can handle this complexity more efficiently.

The results of applying ML for estimating the CPE show that the ARIMA model yielded the highest accuracy. Accordingly, the independent variables for other parameters, such as solar panel cost, balance of system cost, system lifetime, operation and maintenance costs, and energy yield, were accurately forecasted for evaluating the LCOE of a utility-based grid-connected solar home system.[52] Furthermore, we extracted the dataset consisting of demographic variables for two locations: California (USA) and Butuan City (Philippines), and compared the results from our proposed model with the case studies available in the literature. It is worth mentioning that previously, case studies calculating the LCOE used single inputs rather than multiple variables, and the application of an ML approach is scarcely found in the literature. Therefore, to our knowledge, our proposed model is the first approach to accurately forecast the LCOE using an ML framework.

Moreover, here, we also emphasize the prerequisites for the adaptability of our proposed ML model. The LR model and LR with multiple variables model would perform better under the condition that it is not linked to the time series and have more independent variables. In addition, the data available in our proposed ML model plays a vital role in a way that leads to improved accuracy. While LR model performs without time series, the LSTM and ARIMA models perform better with time series forecasting as it allows the model to account for the trends, seasonality and noise in the dataset. Furthermore, it is essential to strike

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH
26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a balance between flexibility and stability to ensure that the model can handle changing data while avoiding needless retraining that may bring instability or overfitting.

The first case study under consideration involves a 25-year project lifetime for a 20 MW grid-connected PV utility system installed in the city of Sacramento, California, USA. The overall performance of the system is 197 peak watts per square meter. The estimated initial investment to install a PV system is $54 million, with a contribution of 2.7 $ W^{-1}, consisting of 65% modules and a 35% balance of systems. The direct purchase cost of the components involved in the PV utility system is $23.856 million, and subsequently, the calculated values for the operation and maintenance cost of the PV system are shown in **Table 2**, totaling $6.50 million over 25 years. These values are forecasted using the ARIMA model and ML techniques. The chosen PV module for the system is Monocrystalline-PERC (Passivated Emitter and Rear Cell), with an efficiency of 19.1%. Incorporating these input values into the equations mentioned in the methodology section, we calculated the net energy production as 738.537 GWh and the net present value of electricity as 383.169 GWh. The forecasted value of the Levelized total cost of electricity is $27.180 million. Therefore, considering all the input values from the literature but integrating the values of CPE from our proposed model, we obtain the LCOE to be 7.09 ¢ kWh^{-1}, whereas the LCOE using singular inputs gives a value of 5.83 ¢ kWh^{-1}. Similarly, using the equations mentioned in the methodology section, the forecasted ROI for the PV utility-based grid-connected system is 14 years.

Table 2. The table showcases the calculation of the LCOE and the ROI of the PV system for the duration of 25 years.

Year	Production [GWh]	NPV of electricity [GWh]	Direct purchase cost [million$]	Operation & maintenance cost [million$]	Levelized total cost [$]
0	–	–	23.856	–	23.856
1	31.343	29.569	–	0.26	0.245
2	31.187	27.757	–	0.26	0.231
3	31.032	26.055	–	0.26	0.218
4	30.878	24.458	–	0.26	0.206
5	30.724	22.959	–	0.26	0.194
6	30.571	21.552	–	0.26	0.183
7	30.419	20.231	–	0.26	0.173
8	30.268	18.990	–	0.26	0.163
9	30.117	17.826	–	0.26	0.154
10	29.967	16.734	–	0.26	0.145
11	29.818	15.708	–	0.26	0.137
12	29.670	14.745	–	0.26	0.129
13	29.522	13.841	–	0.26	0.122
14	29.376	12.993	–	0.26	0.115
15	29.229	12.196	–	0.26	0.108
16	29.084	11.449	–	0.26	0.102
17	28.939	10.747	–	0.26	0.097
18	28.795	10.088	–	0.26	0.091
19	28.652	9.470	–	0.26	0.086
20	28.509	8.889	–	0.26	0.081
21	28.368	8.344	–	0.26	0.076
22	28.226	7.833	–	0.26	0.072
23	28.086	7.353	–	0.26	0.068
24	27.946	6.902	–	0.26	0.064
25	27.807	6.479	–	0.26	0.061
Total	**738.537**	**383.169**	**23.856**	**6.50**	**27.180**

In addition, to validate our proposed model, we used 6the dataset from another case study for a solar PV farm in a specific location in Butuan City, Philippines.[53] Similar to the previous case study, we took the initial dataset, such as the power capacity of the solar farm being 5 MW with an investment of 300 million pesos (to make a comparison with the first study, all costs are converted to USD). Furthermore, we integrated the associated costs and energy yield to calculate the solar farm's LCOE parameter and ROI. According to the results of the case study for a duration of 20 years, the useful energy production is 4.18 kWh d^{-1} with an ROI of 4.23 years. However, their study used singular inputs to calculate these values. Consecutively, applying the ARIMA-based ML model in our proposed model, we predicted the LCOE value of 8.90 ¢ kWh^{-1} and the ROI was calculated as 5.37 years.

Accordingly, analyzing the two case studies reveals that the demographic variables of any country will undoubtedly change over time. Moreover, the discrepancy in values using singular inputs and our proposed model indicates that the LCOE and ROI calculated using singular inputs result in errors and miscalculated estimations of the ROI for solar home systems.

Nevertheless, there are limitations to our proposed model and framework. Firstly, the effectiveness of our ML model largely depends on the data's consistency with historical trends. When the model is applied to a new location, the essential prerequisite is to have a robust dataset for that specific region, covering demographic variables, energy yield, inflation rates and interest rates, among other things. The model performs optimally when data patterns in these variables remain similar to past trends.

However, situations where our model might not work as expected could include drastic policy changes or economic shifts in the region under study. For instance, sudden regulatory changes influencing energy costs or dramatic changes in inflation rates or demographic shifts can alter the data's underlying patterns significantly. In such situations, the model would need re-training to adapt to the new data trends.

Retraining the model is not difficult but does require an updated and comprehensive dataset that reflects the altered circumstances. The re-training process involves feeding this new data into the model and running the analysis again. While this process may take time and resources, it is an integral part of maintaining the model's accuracy and reliability over time.[54]

7. Conclusion

In conclusion, the CPE encapsulates the average electricity cost per unit (kWh) for consumers, influenced by various dynamic factors such as inflation rate, population growth, and gross

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

domestic product. In contrast, the LCOE calculates the average cost of electricity production per unit over the lifetime of a power generation system, such as a PV system. Our model establishes a relationship between CPE and LCOE, with CPE acting as a benchmark for estimating the LCOE of a PV system, providing context for cost-effectiveness comparisons with conventional electricity prices. By dynamically considering factors influencing CPE, our model aims to enhance LCOE estimation accuracy, thereby aiding in ROI determination for PV system installation and contributing to a better understanding of long-term financial viability.

However, it's crucial to acknowledge the limitations of our study. The model's reliance on historical data and assumptions about future trends introduces uncertainties. Additionally, the case study focus on California and the Philippines may limit the generalization of results to other regions. Furthermore, while our approach recognizes the dynamic nature of parameters such as population growth, past electricity costs, inflation, and gross domestic product, it may not encompass all potential variables influencing electricity costs.

The implications of our study are twofold. On a practical level, our model offers a valuable tool for decision-makers, assisting in informed choices regarding PV system investments. The theoretical contribution lies in the incorporation of ML techniques, showcasing the potential of advanced analytics in refining economic models for sustainable energy projects. Our results, demonstrating a significant difference in LCOE estimations between singular inputs and ML-based approaches, underscore the importance of considering dynamic factors in economic analyses. In essence, our study emphasizes the necessity of accounting for dynamic inputs for more accurate and reliable LCOE estimations, essential for evaluating the long-term financial viability of PV systems.

Therefore, we conclude that most existing studies rely on singular values. However, our argument emphasizes that many of these parameters are dynamic. Factors such as population growth, average past cost of electricity, inflation rate, gross domestic product, and other demographic variables significantly impact the cost of electricity. As a result, we developed a model that allows for the calculation of LCOE based on these dynamic input factors. Our results demonstrate a clear difference in estimating the LCOE of a PV system using singular inputs, yielding an LCOE of $5.83\,¢\,kWh^{-1}$. In contrast, when applying the ML model, the LCOE increases to a value of $7.09\,¢\,kWh^{-1}$. This comparison highlights the distinction between calculating the LCOE using singular inputs and employing ML and artificial intelligence-based algorithms. Ultimately, our study reveals a substantial difference in LCOE estimations, emphasizing the importance of considering dynamic factors.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

auto-regression integration moving average (ARIMA) model, levelized cost of electricity (LCOE), machine learning, photovoltaics

Received: August 26, 2023
Revised: January 14, 2024
Published online:

[1] R. E. Smalley, *MRS Bull.* **2005**, *30*, 412.
[2] P. Tillmann, K. Jäger, C. Becker, *Sustainable Energy Fuels* **2020**, *4*, 254.
[3] J. E. Payne, *Appl. Energy* **2010**, *87*, 723.
[4] A. S. Mundada, K. K. Shah, J. M. Pearce, *Renewable Sustainable Energy Rev.* **2016**, *57*, 692.
[5] W. Shen, X. Chen, J. Qiu, J. A. Hayward, S. Sayeef, P. Osman, K. Meng, Z. Y. Dong, *Renewable Sustainable Energy Rev.* **2020**, *133*, 110301.
[6] V. Kizilcec, C. Spataru, A. Lipani, P. Parikh, *Energies* **2022**, *15*, 857.
[7] N. AlQallaf, F. Ayaz, S. Bhatti, S. Hussain, A. Zoha, R. Ghannam, in *2022 29th IEEE Int. Conf. on Electronics, Circuits and Systems (ICECS)*, IEEE, Glasgow, UK **2022**, pp. 1–4.
[8] O. Schmidt, A. Hawkes, A. Gambhir, I. Staffell, *Nat. Energy* **2017**, *2*, 17110.
[9] V. Kondaiah, B. Saravanan, P. Sanjeevikumar, B. Khan, *J. Eng.* **2022**, *2022*, 665.
[10] S. B. Darling, F. You, T. Veselka, A. Velosa, *Energy Environ. Sci.* **2011**, *4*, 3133.
[11] A. Mosavi, A. Bahmani, Energy consumption prediction using machine learning; a review, **2019**.
[12] R. Ghannam, P. V. Klaine, M. Imran, in *Power Systems*, Springer, Singapore **2019**.
[13] D. W. Bian, S. M. Watson, N. C. Wright, S. R. Shah, T. Buonassisi, D. Ramanujan, I. M. Peters, A. G. Winter, *Desalination* **2019**, *452*, 265.
[14] W. Khan, S. Walker, W. Zeiler, *Energy* **2022**, *240*, 122812.
[15] G. M. Yagli, D. Yang, D. Srinivasan, *Renewable Sustainable Energy Rev.* **2019**, *105*, 487.
[16] S. Bhatti, H. U. Manzoor, A. Zoha, R. Ghannam, *arXiv* **2023**.
[17] V. S. B. Kurukuru, A. Haque, M. A. Khan, S. Sahoo, A. Malik, F. Blaabjerg, *Energies* **2021**, *14*, 4690.
[18] B. Del Fabbro, A. Valentincic, A. F. Gubina, *Sol. Energy* **2016**, *132*, 73.
[19] J. Y. Tao, A. Finenko, *Energy Policy* **2016**, *98*, 749.
[20] S. Shimura, R. Herrero, M. K. Zuffo, J. A. B. Grimoni, *Sol. Energy* **2016**, *133*, 294.
[21] N. Heck, C. Smith, E. Hittinger, *Electr. J.* **2016**, *29*, 21.
[22] A. Goswami, P. K. Sadhu, *Clean Technol. Environ. Policy* **2021**, *25*, 343.
[23] R. Zhang, M. Lee, L. Huang, *Renewable Sustainable Energy Rev.* **2023**, *181*, 113326.
[24] K. Branker, M. Pathak, J. M. Pearce, *Renewable Sustainable Energy Rev.* **2011**, *15*, 4470.
[25] C.-Y. Lee, J. Ahn, *Energies* **2020**, *13*, 3017.
[26] A. R. Khan, S. Razzaq, T. Alquthami, M. R. Moghal, A. Amin, A. Mahmood, in *2018 1st Int. Conf. on Power, Energy and Smart Grid (ICPESG)*, IEEE, Mirpur Azad Kashmir, Pakistan **2018**, pp. 1–6.
[27] T. Geissmann, O. Ponta, *Energy* **2017**, *124*, 372.
[28] T. Georgitsioti, N. Pearsall, I. Forbes, *Proc. Photovoltaic Science, Applications and Technology (PVSAT-9)*, IET Renewable Power Generation, Wiley, Swansea University, UK **2013**, pp. 9–12, https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-rpg.2013.0241.
[29] X. Yu, A. Manthiram, *Adv. Energy Sustainability Res.* **2021**, *2*, 2000102.
[30] B. Pillot, S. de Siqueira, J. B. Dias, *Renewable Energy* **2018**, *127*, 974.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[31] W. Short, D. J. Packey, T. Holt, in *A Manual for the Economic Evaluation of Energy Efficiency and Renewable Energy Technologies*, National Renewable Energy Laboratory, Golden, Colorado, USA **1995**.
[32] M. A. Ansari, P. Somdee, *Adv. Energy Sustainability Res.* **2022**, *3*, 2200063.
[33] J.-R. Rodriguez-Ossorio, A. Gonzalez-Martinez, M. de Simon-Martin, A.-M. Diez-Suarez, A. Colmenar-Santos, E. Rosales-Asensio, *Energy Rep.* **2021**, *7*, 199.
[34] D. Talavera, P. Pérez-Higueras, J. Ruz-Arias, E. Fernández, *Appl. Energy* **2015**, *151*, 49.
[35] T. R. Archibald, *J. Account. Res.* **1967**, *5*, 164.
[36] R. Vidal, J.-A. Alberola-Borràs, N. Sánchez-Pantoja, I. Mora-Seró, *Adv. Energy Sustainability Res.* **2021**, *2*, 2000088.
[37] M. Obi, S. M. Jensen, J. B. Ferris, R. B. Bass, *Renewable Sustainable Energy Rev.* **2017**, *67*, 908.
[38] N. Lukač, S. Seme, K. Dežan, B. Žalik, G. Štumberger, *Energy* **2016**, *107*, 854.
[39] S. Bhatti, H. U. Manzoor, B. Michel, R. S. Bonilla, R. Abrams, A. Zoha, S. Hussain, R. Ghannam, *Adv. Energy Sustainability Res.* **2023**, *4*, 2300004.
[40] T. M. Mitchell, T. M. Mitchell, in *Machine Learning*, Mcgraw-Hill, New York **1997**.
[41] M. I. Jordan, T. M. Mitchell, *Science* **2015**, *349*, 255.
[42] T. G. Dietterich, *Ann. Rev. Comput. Sci.* **1990**, *4*, 255.
[43] I. E. Naqa, M. J. Murphy, in *Machine Learning in Radiation Oncology*, Springer, Cham **2015**.
[44] M. Dong, Y. Li, D. Song, J. Yang, M. Su, X. Deng, L. Huang, M. Elkholy, Y. H. Joo, *Energy Convers. Manage.* **2021**, *229*, 113781.
[45] Environmental Investigation Agency (EIA), https://eia-international.org/ (accessed: December 2022).
[46] International Renewable Energy Agency (IRENA), https://irena.org/ (accessed: December 2022).
[47] U. D. of Economic Analysis, Bureau of Economic Analysis, https://www.bea.gov/ (accessed: December 2022).
[48] International Energy Agency (IEA), https://iea.org/ (accessed: December 2022).
[49] F. Egelioglu, A. Mohamad, H. Guven, *Energy* **2001**, *26*, 355–362.
[50] Y. S. Kim, *Expert Syst. Appl.* **2008**, *34*, 1227.
[51] M. Thompson, *Analyst* **1982**, *107*, 1169.
[52] S. Bhatti, A. Williams, *Renewable Energy Environ. Sustainability* **2021**, *6*, 25.
[53] J. T. Dellosa, M. J. C. Panes, R. U. Espina, in *2021 IEEE Int. Conf. on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, IEEE, Bari, Italy **2021**, pp. 1–6.
[54] S. Bhatti, A. R. Khan, S. Hussain, R. Ghannam, in *2022 29th IEEE Int. Conf. on Electronics, Circuits and Systems (ICECS)*, IEEE, Glasgow, UK **2022**, pp. 1–4.

© 2024 The Authors. Advanced Energy and Sustainability Research published by Wiley-VCH GmbH

26999412, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/aesr.202300178 by NHS Education for Scotland NES, Edinburgh Central Office, Wiley Online Library on [11/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License