

## PREDICTIVE ANALYTICS OF SOLAR POWER GENERATION USING DATA SCIENCE

VENKATESH P. & K. MURUGAN

*Asso. Professor, Sri Sairam Institute of Management Studies, Sri Sairam Engineering College, Chennai, India*

### ABSTRACT

*By 2040, India's share of the world's energy consumption is predicted to quadruple to 11%, making it imperative to boost energy security and independence in terms of electricity generation without raising environmental costs. As a result of this increase in demand for power, India's reliance on coal, oil, and natural gas as energy sources is probably going to increase. Understanding the elements that contribute to time delays and coming up with strategies to reduce them while increasing productivity are the main objectives of this study as they relate to the steel detailing industry. There have been several successful research on how to increase solar power generation utilising a variety of models and strategies, taking into account how employee happiness with their jobs, workplace culture, and motivational variables, as well as their training and development, changes. The goal of this project is to determine how data science techniques may be applied to forecast solar plant output more accurately. Using historical solar power generation and weather data, machine learning techniques like linear regression can be used to forecast solar power generation based on the analysis of the identified issue.*

**KEYWORDS:** *Climate Data, Productivity Rate, Science Techniques, Job Satisfaction, Motivation Factors, Work Culture, Training And Development, Solar Plant Production*

**Received:** Dec 02, 2022; **Accepted:** Dec 22, 2022; **Published:** Dec 29, 2022; **Paper Id:** IJBMRJUN20232

### INTRODUCTION

Sun-based energy is the conversion of light energy into electrical energy, either directly using photovoltaics (PV) or indirectly using concentrated sunlight-based electricity. Photovoltaic cells convert light into an electric current by using the photovoltaic effect. Concentrated solar power systems use lenses, mirrors, and solar tracking devices to focus a lot of sunlight on a hot area, usually a steam engine.

Solar energy is the heat and light from the sun that is captured using a range of methods, including solar engineering, solar thermal energy (such as solar water heating), and solar electricity production. It is a significant source of renewable energy, and depending on how solar energy is captured, disseminated, or converted into electricity, its methods are typically characterised as either passive solar or active solar. Active solar energy strategies include using photovoltaic systems, concentrated solar power, and solar water heating. Designing areas with natural airflow, choosing Orienting a building toward the Sun and using materials with advantageous thermal mass or light-dispersing qualities are examples of passive solar approaches. To shield the solar energy sector from foreign competition, lower international prices that hamper R&D efforts, exchange rate fluctuations, price risk, quality imports, imported products without quality certification, and unforeseen price swings, India should place more emphasis on domestic manufacturing.

Background research typically focuses on predicting solar power generation in particular, taking into account elements such as past meteorological data, regional irradiance, and previous solar power generation. Various studies have been conducted.

This research focuses primarily on understanding the factors causing the time delay and finding solutions to minimize the time delays and improve the productivity rate with respect to the steel detailing sector. Employees vary with respect to job satisfaction, motivation factors, work culture, training & development, and other factors. Machine learning, predictive modelling, and other statistical techniques that use data mining, predictive modelling, and predictive analytics are collectively referred to as "predictive analytics." analyze current and historical data and make predictions about events that may occur in the Solar energy is dependent on sunlight, a natural resource that is only available occasionally. Because of this, the ability to predict power output is essential for incorporating solar photovoltaics into the conventional grid system. Solar plant owners and operators can optimize inspections, schedule maintenance, enhance the operational performance of their equipment, and maximize the environmental benefit of their renewable energy investments with accurate estimates of solar energy production and insights into solar equipment performance. Using previous and present data, create predictions about potential future events. Sunlight, a scarcely accessible natural resource, is a prerequisite for solar energy. Because of this, integrating solar photovoltaics into the current grid system requires the ability to estimate power output. With precise estimates of solar energy production and insights into the performance of solar equipment, solar plant owners and operators may schedule maintenance, improve the operational performance of their equipment, and maximize the environmental impact of their investments in renewable energy.

## **PROBLEM STATEMENT**

The purpose of this study is to investigate how data science methods can be used to improve predictions of solar plant production. The findings demonstrate that, for certain applications, that predictions from modern physical models perform worse than those from hybrid machine learning models based on historical data, temperature, and physical model information; device dependability can be affected by lightning, temperature, and grid operating conditions, among other environmental factors.

## **OBJECTIVES OF THE STUDY**

- The main goal of this study is to forecast or predict solar power generation using data science.
- Examining historical weather information and solar output

## **REVIEW OF LITERATURE**

**Mohana Alanazi, Mohsen Mahoor, and Amin Khodae.** At the North American Power Symposium in 2017, hybrid day-ahead solar forecasting was presented in two stages (NAPS). In the not-too-distant future, renewable generation is predicted to overtake other sources of generation. Power generation from renewable resources is expanding globally. Numerical simulations performed over the course of three test days under varied weather circumstances show the effectiveness of the suggested two-stage approach.. **S. D. Kumar et al., 2022**), The main goal of the study was to determine how electronic entertainment affected young people's development and access to employment prospects in the Indian state of Tamil Nadu. The continuous survey used a non-erratic checking approach and had 560 respondents. Young adults who are considering or have potential make up the responders. According to the report, there is a link between

young people's use of virtual games and future employment chances in areas like enrollment and employment data. **Kumar, S. D, et.al.** ( 2022) The study's primary objective was to ascertain how electronic entertainment impacted young people's growth and access to employment opportunities in Tamil Nadu, an Indian state. 560 people participated in the continuous survey, which employed a non-erratic checking approach. The respondents are young individuals who are thinking about or have potential. In areas like enrollment and employment data Nadu, there is a correlation between young people's use of virtual games and future employment prospects. The financial sector also contributes to capital compliance, creation, and adaptability in addition to monetary approach. To ensure a strong financial foundation and efficient moderation, it is imperative to carefully assess and examine the presentation of banks. There's no denying that it's difficult to survey the Indian financial industry. There are so many distinct factors that need to be considered in order to distinguish between excellent banks and poor banks. In order to assess the financial sector's presentation, we choose the CAMEL model since it calculates the banks' exposure from each of the key parameters, including Capital Adequacy, Mean Quality, Management Efficiency, Earning Quality, and Liquidity. The current review's main objective is to assess how well two public area banks in Tamil Nadu promote themselves. **(Krishnaswami, Soundarapandiyar, & Dinesh Selvaraj and K. Maran and coworkers,** The development of technology has altered every area of our lives, including the way we work, shop, and communicate with others. As a result of technological innovation and the explosive growth of the internet, more individuals are remaining in and only leaving the house to buy basics. According to **Kausar Suraiya Quraishi, Dr. Salma Ahmed, and Dr. K. Maran (2019),** The development of technology has altered every area of our lives, including the way we work, shop, and communicate with others. As a result of technological innovation and the explosive growth of the internet, more individuals are remaining in and only leaving the house to buy basics. According to **V Suresh, K Prabhakar, K Santhanalakshmi, and K Maran (2016)** established in their study a theoretical framework that includes the core ideas of TAM, specifically perceived utility and ease of use. Trustworthiness and customised data were two more external characteristics employed. According to the overall research model, each of the aforementioned factors has an effect on the amount of information exchanged via social media and the acceptance of technology by out-patients in private hospitals, either directly or indirectly. Using information from 200 outpatients in Chennai City, exploratory factor analysis was utilised to determine the variables that affect the acceptance of technology in the private hospital outpatient sector.

## RESEARCH METHODOLOGY

Discovering, interpreting, and assessing approaches for projecting solar generations under specific conditions is the main goal. Research does not necessarily necessary to follow the scientific approach. In this study, the research design, data collection, sampling, analysis, and interpretations are all part of the research methodology. In market research, there are two types of sampling: probability sampling and non-probability sampling. Data are collected from publicly accessible websites like kaggle. Let's examine these two sample methods in more detail. The study's sample strategy was decided upon using the Simple Random sampling approach, which is a part of probability sampling. The entire data set, which includes values for a one-year period, is taken into account and is significant to solar forecasting.

Python is used to analyze the various secondary source data using a variety of analysis models, including the following:

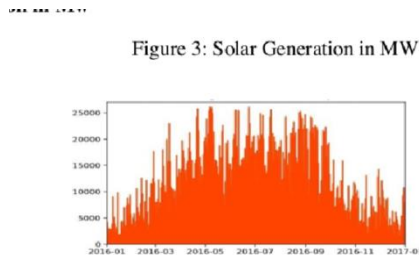
The Time Series Regular Regression Enterprise Intelligence

## DATA ANALYSIS

**Time Series of Generation Data**

The present platform, which contains data from 37 nations, is filtered out based on the data collected for Open Power System Data with regard to various site locations within the same region. Time series are plotted against solar generation over a 12-month period because only solar generation data is relevant to this study. Only actual solar generation data is taken into account.

Solar Generation in Mw



**Figure 3: Solar Generation in MW**

As was to be predicted, the middle months of the year showed much higher solar generation.

**LINEAR REGRESSION**

The input is a linear function of the output of a linear regression algorithm:

$$: \mathbb{R}^n \quad \mathbb{R}, \hat{y} \quad f(x) \quad \beta^T x + \beta_0$$

Where

$$\beta = (\beta_1, \dots, \beta_n) \in \mathbb{R}$$

is a parameter vector. Finding the parameters that reduce mean squared error is the goal.

$$\operatorname{argmin}_{\beta, \beta_0} \frac{1}{N} \sum_{i=1}^N (y_i - \beta^T x_i - \beta_0)^2$$

Linear Regression from the Python scikit-learn module can be used to do this.

**SOLAR POWER GENERATION PREDICTION**

We build the features matrix x solar with the values of the temperature, total horizontal radiation from the ground, and horizontal radiation from the top of the atmosphere in order to estimate the solar generation.

$$R^2 = 1 - \frac{\sum_i (y_i - f(x_i))^2}{\sum_i (y_i - \bar{y})^2}$$

The formula: is used to calculate it formally.

The result is:

average = 0.9245405855731855

We achieve an even higher R2 value. We are able to anticipate the solar generation in the chosen region for the specific year using only the temperature, top-of-the-atmosphere radiation, and ground radiation.

The amount of horizontal radiation or climate has an impact on the generation of solar power.

Solar energy production can be integrated with the current electric grid by forecasting or predicting it.

The dataset must be free of any disturbances and contain at least one year's worth of historical data for a forecast or prediction to be relevant.

The operator is able to make important business decisions based on forecasts and historical data.

Due to the fact that the output of a solar power plant is also influenced by other factors like the upkeep and operation of equipment, forecasting solar power generation based on historical solar power generation data and weather data will only provide a limited level of accuracy. With precise estimates of solar energy production and insights into the performance of solar equipment, solar plant owners and operators may schedule maintenance, improve the operational performance of their equipment, and maximise the environmental impact of their investments in renewable energy. Therefore, by taking these aspects into consideration, plant-level solar power generation can still be boosted.

## CONCLUSIONS

Using machine learning techniques like linear regression and historical data on solar power generation, it is possible to forecast solar power generation based on the study of the detected issue.

## REFERENCES

1. Mohana Alanazi, Mohsen Mahoor, and Amin Khodae. Two-stage hybrid day-ahead solar forecasting. In *North American Power Symposium (NAPS)*, 2017.
2. Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short- term solar power forecasting. *Solar Energy*, 83:1772–1783, 2009.
3. Hector Beltran, Endika Bilbao, Enrique Belenguer, Ion Etxeberria-Otadui, and Pedro Rodriguez. Evaluation of storage energy requirements for constant production in PV power plants. *IEEE Transactions on Industrial Electronics*, 60:1225 – 1234, 2012.
4. Songjian Chai, Zhao Xu, and Wai Kin Wong. Optimal granule-based PIs construction for solar irradiance forecast. *IEEE Transactions on Power Systems*, 31:3332–3333, 2016.
5. Chi Wai Chow, Bryan Urquhart, Matthew Lave, Anthony Dominguez, Jan Kleissl, Janet Shields, and Byron Washom. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85:2881– 2893, 2011.
6. Chi Wai Chow, Bryan Urquhart, Matthew Lave, Anthony Dominguez, Jan Kleissl, Janet Shields, and Byron Washom. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy*, 85:2856– 2870, 2011.
7. Yang Dazhi, Panida Jirutitijaroen, and Wilfred M. Walsh. Hourly solar irradiance time series forecasting using cloud cover

- index. Solar Energy*, 86:3531–3543, 2012.
8. Alex Dobbs, Tarek Elgindy, Bri-Mathias Hodge, and Anthony Florita. *Short-term solar forecasting performance of popular machine learning algorithms. In International Workshop on the Integration of Solar Power into Power Systems, 2017.*
  9. Eric Fidler. *Asset performance management helps oil and gas companies increase asset availability, improve uptime and empower more intelligent decision making. In Offshore Technology Conference, 2009.*
  10. Jorge M. Filipe, Ricardo Bessa, Jean Sumaili, R. Tome, and J. N. Sousa. *A hybrid short-term solar power forecasting tool. pages 1–6, 2015.*
  11. *Pollution Reduction Mechanism Model in Urban India using Solar Assisted Bicycle, SA Puviyarasu - Indian Journal of Science and 2016*
  12. Karen Smiley, Xiao Qu, Travis Galoppo, Eric K. Harper, Alok Kucheria, Mithun P. Acharya, and Frank Tarzanin. *Managing solar asset performance with connected analytics. ABB Review, pages 34–41, 2019.*
  13. Mike K. Williams, Ananth Seshan, and Karen Smiley. *Asset Performance Management (APM) 2.0 - guidelines for goal setting and implementation planning. Technical report, 2017.*
  14. Kumar, S. D., Soundarapandiyam, K., & Meera, S. (2022). *Sentience of Career Opportunities and Career Development using Social Media—A Study with Reference to Tamil Nadu. Journal of Big Data Technology and Business Analytics, 1(1), 7-14.*
  15. Kumar, S. D., & Soundarapandiyam, K. (2022). *The Effectiveness of Product Placement in Tamil movies: A Study with Reference to the State of Tamil Nadu, India. South Asian Journal of Social Sciences and Humanities, 3(5), 28-41.*
  16. Soundarapandiyam, Krishnaswami & Selvaraj, Dinesh. (2022), *Performance Analysis Using the Camel Approach Reference to Select Banks in Tamil Nadu during COVID-19 – An Empirical Study. South Asian Journal of Social Studies and Economics. 54-61.*
  17. Kausar Suraiya Quraishi, Dr. Salma Ahmed, Dr. K. Maran(2019), *China's Growth in Residential Solar PV-A Booster for India, Global Journal of Management And Business.*
  18. K. Maran, P. Priyadarshini, L. Jenifa, C. R. Senthilnathan and P. Venkatesh, "Data Analysis on Mobile Payment Technology with Reference to Users' Behaviour of Retail Goods in India," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), 2021, pp. 267-272, doi: 10.1109/ICCCT53315.2021.9711823.
  19. V Suresh, K Prabhakar, K Santhanalakshmi, K Maran (2016) "Applying technology acceptance (TAM) model to determine the factors of acceptance in out-patient information system in private hospital sectors in Chennai city", *Journal of Pharmaceutical Sciences and Research, Vol.8, Issue.12. P.1373*
  20. Immaculate, P.S., Maran K. (2017), *An study of women perception towards worklife balance in various manufacturing sectors in Chennai, Journal of Advanced Research in Dynamical and Control Systems, Spl.Issue.13, PP.700-704.*