The Impact of Data Analysis on Solar Power: A Study on how data analytics improves an Individual's solar solution in South Africa

Thesis submitted in fulfillment of the requirements for the degree **Post Graduate Diploma in Data Analytics** at The Independent Institute of Education, Varsity College.

Ashleigh Kingwill

[redacted]

Supervisor: Mr Renaldin Ponnan

07/ 11/ 2023

DECLARATION

I hereby declare that the research report submitted for the Post Graduate Diploma in Data Analytics to The Independent Institute of Education (The IIE) is my own work and has not previously been submitted to another University or Higher Education Institution for a postgraduate qualification.

Signature__  _____ Date_____07/11/2023_____

Abstract

This paper explored the feasibility of solar power systems for individual use. With the increase in load shedding and the lack of reliable electricity in South Africa, many households were looking into alternative sources of electricity. One of the main alternative sources was solar power. Many different factors could affect the amount of usable electricity that solar power systems can produce, including weather and the position of the panels. With the idea of solar power came many questions, with the biggest being whether solar power was feasible for individual households. With the help of data analytics, this question was answered by running a few simulations to determine if it was worth installing solar and, if it was, what the best way to set it up was. This was achieved by using a Python library that ran many different calculations and simulated the outcome using the given data. The library took in many different factors, including the coordinates, to determine the main weather patterns, the tilt of the panels, the number of panels, as well as start and end dates. This meant that data could be simulated for years to get the most accurate outcome. With all of this data, it could be used to determine if it would be feasible for an individual to use solar. This could be compared to a person's normal electricity usage to see if it would 1) generate enough electricity for the household and 2) what the optimal set up would be. This was done by creating a machine learning algorithm that took in all the information and generated the results for an individual household. It was found that while it might be able to sustain some households, enough power is generated to make a substantial difference.

Key Terms

Solar Power Systems, Alternative Sources of Electricity, Photovoltaic (PV) System

Acknowledgements

I would like to express my heartfelt gratitude to my research supervisor, Renaldin Ponnan, for their support and guidance throughout this research journey. Their expertise and encouragement have been instrumental in shaping the direction and quality of this study.

I am deeply appreciative of my loving family for their constant support and understanding. My mom, Michelle Kingwill, and my husband, Karel Schwab, have been my pillars of strength, offering patience and encouragement during the demanding phases of this research. Their belief in my abilities has been a driving force.

To all of these individuals, and to those whose support may not be explicitly mentioned, I am sincerely grateful for your contributions to this research endeavor.

Abbreviations

PV: Photovoltaic

IPP: Independent Power Producers

POA: Plane of Array

W: Watts

DC: Direct Current

AC: Alternating Current

kWh: Kilowatt hour

List Of Figures

Chapter One: Introduction

Background

In many countries around the world, fossil fuels such as coal and oil are still being used as the main source of electricity. Approximately 64% of the world's electricity comes from fossil fuels (Ritchie, Roser & Rosado, 2022). South Africa is a large contributor to this as about 94% of its electricity comes from fossil fuels (Ritchie, Roser & Rosado, 2022).

In South Africa, the main electricity provider, Eskom, has been unable to supply the county as a whole with constant and reliable electricity. Eskom has implemented a rotation system where different sections around the country have no electricity, known as load-shedding. In 2022, it was determined that the country had 2 881 hours of load-shedding by the 21st of November. This is equivalent to 120 days or 4 months of continuous power outages (Labuschagne, 2022).

Considering the circumstances, many South Africans are looking for different sources of electricity. One of these is solar power. Solar power is one of the best solutions for individual use as it is easy to set up a home compared to other sustainable solutions. However, with the diverse ecosystem of South Africa, solar is not a one-size-fits-all solution and there are many different factors that can play a role in how one sets up the system. One of the main factors to consider is the weather conditions. Along with the weather, there are other factors that come into play such as the maintenance of the system and the space that one has.

Eco Green Energy, a company founded in 2008 that distributes solar panels to over 72 countries, has determined the main factors that need to be taken into account when setting up solar panels. These factors included temperature, soiling, shading, mismatched panels, inverter efficiency and age (Staub, 2021).

Many of these factors can not be controlled such as the weather and temperature on any given day. However, other factors such as soiling, which is how clean the panels are, can be controlled by cleaning the panels regularly and ensuring that there is nothing that may be blocking the cells on the panels (Staub, 2021). It was determined that clean panels could output a minimum of 5% more power (Staub, 2021).

Other factors such as how much shade the panels get, will determine where the panels should be located to get the most sun possible. Along with this the age and mismatching panels can also lower the efficiency of the system (Staub, 2021).

Solar panels are not the only factor that plays a role in solar power. Once the power has been produced there needs to be a system in place that can store and convert the energy into usable electricity. This is done by installing an inverter. The inverter can also affect the efficiency of the system. For an inverter to be effective, the power coming into the system needs to be higher than the inverter's rated power (Staub, 2021).

Problem Statement and Objectives

Amidst the ongoing energy crisis in South Africa, where load-shedding has become commonplace due to Eskom's incapacity to supply steady and dependable electricity, more people and households are looking for alternative energy sources. Solar energy is one future alternative that looks very promising. But when South Africans investigate this option, they find a significant knowledge gap. Comprehensive information about the viability of installing solar power systems in individual households and the complexities of setting them up for maximum efficiency is hard to come by. There is a clear need for a simple and easy-to-use guide to evaluate the feasibility and usability of this kind of shift.

The goal of the study is to evaluate if solar power systems are feasible for particular homes, taking into account a number of variables such location, climate, and economic viability. Its main objective is to maximise solar system configurations by analysing inverter efficiency, panel location, and capacity. It examines how South Africa's varied topography and climate affect the operation of solar power systems, highlighting local differences in sun irradiance, temperature swings, and shading effects.

Research Questions

- Are solar power-based systems feasible for an individual household?

- What is the optimal setup for the system to ensure that it generates sufficient power?

Structure Of Report

This report will consist of 5 basic chapters including the introduction, literature review, research method, analysis and results, and finally conclusion and recommendation.

Each chapter will contain subheadings and all relative information that would be needed to understand the chapter and topic.

Chapter Two: Literature Review

There have been many studies done on how South Africa can move from using coal-powered electricity to using solar farms to supply the country with constant electricity, such as Repurposing South Africa's Retiring Coal-Fired Power Stations for Renewable Energy Generation: A Techno-Economic Analysis by Gugulethu Nogaya, Nnamdi Nwulu and Saheed Lekan Gbadamosi or Probabilistic solar power forecasting using partially linear additive quantile regression models: an application to South African data by Phathutshedzo Mpfumali. However, these studies are done on South Africa as a whole and not on how a small system can sustain an individual's needs. Very few studies have been done on the impact and feasibility of an individual's solar system in South Africa. One such study was done at the University of Fort Hare in Alice, South Africa where a system was set up on the roof of the university building and the researchers checked on the system and took readings regularly.

Solar Energy Resources Study in Alice, South Africa

O. Overen and E. Meyer from the University of Fort Hare in Alice, South Africa, set out to determine the feasibility of a residential stand-alone photovoltaic (PV) power system. Many different factors such as global horizontal irradiance, direct normal irradiance and diffuse horizontal irradiance were used to compare Alice to other parts of the world that have mega solar PV projects to determine if Alice would be a good fit for such a project (Overen & Meyer, 2022). The study discusses that the Department of Energy plans to implement Independent Power Producers (IPP) Solar PV project. These projects were discussed in 2017 and are still being planned (Overen & Meyer, 2022).

Overen and Meyer (2022) stated that according to the plans that have been released, the placement of these projects could be better. With the research done it was seen that provinces such as North West, which has the second highest Global Horizontal Irradiance only have one project set up while the Western Cape has the second hostess number of projects, however, they have the third-lowest daily solar radiation (Overen & Meyer, 2022). These placements are not ideal as the efficacy would be low and this could be solved by redetermining where the

projects are located according to the average daily solar radiation (Overen & Meyer, 2022).

The data that was collected was categorised based on if the skies were clear or cloudy (Overen & Meyer, 2022). It was then calculated how many of each day they had throughout the year. Overen and Meyer (2022) then ran many different calculations to determine if Alice was a good fit for a solar project. It was found that it would be a good fit and is comparable to other regions around the world where similar projects have been set up (Overen & Meyer, 2022).

While this study has many benefits, it, like many others, does not investigate how a solar system can be set up and improved for an individual in the long run. There are still many gaps in the knowledge base of solar power in South Africa in general. The study that follows will investigate how much power a solar system can generate over a long period of time and how the system can be improved to generate the most power. It will also be seen if it is feasible and if an individual will be able to move to solar-generated power on a permanent base. The study will also use data from many years and different locations around South Africa to get a more accurate and overall representation of the feasibility.

Theoretical Perspectives

Predictive Analytics

Predictive analytics is a type of analytics that uses historical data to determine events and outcomes of the future. It combines statistics, data mining and different machine-learning algorithms to complete this task (IBM, 2021). There are many different algorithms used for predictive analytics, the three main being classification, clustering and time series.

Classification models are used to categorise data. This can also be used to answer basic yes or no questions and is often used in fraud detection and credit evaluations. The different models used in classification include logistic regression, decision trees and random forest algorithms (IBM, 2021).

Clustering models are a form of unsupervised learning where the algorithm groups the data together based on different factors or attributes (IBM, 2021). This is most often seen when shopping online and the site suggests other products to buy based on what is already in the shopping cart. Some of these algorithms are k-mean clustering and historical clustering (IBM, 2021).

Finally, time series models use data inputs from different time-frequency. This is often used to determine different trends or weather behaviour (IBM, 2021). Some examples of algorithms used in these types of models include autoregressive and moving average models.

These can be used and implemented in the study to determine different weather patterns and how they affect the daily radiation in different locations. Historical data and simulated data will be used to determine these factors and how the locations are affected.

Optimization Theory

Optimization theory is a mathematical branch that is used to solve optimization problems. These problems are math functions where the value of the function must be maximised or minimised (Sinha, 2020). These problems are often found in data science and finding good solutions to these is what gives machine learning algorithms their high accuracy as they are built on numerical methods and find solutions that minimise error functions (Sinha, 2020).

There are 3 basic questions that summarise optimization theory. These are, what is the objective function? What is the set of variables of interest? And finally, what is the set of constraints? (Morgan, 2021).

Knowing the objective function is very useful and ensures that the correct algorithm is chosen, or at least it narrows down what algorithms may help solve the problem. It is also very important to know what the data looks like and how it can be interacted with (Morgan, 2021). This helps make the process of writing and training the model or algorithm much easier as there is already knowledge of how the data looks and can be used, causing fewer errors. Finally, knowing the

constraints is important to ensure that the algorithm is limited to the area it is needed ensuring that only feasible solutions are found (Morgan, 2021).

This will ensure that minimum values of daily radiation are taken into account for the study. This will help as it will be better to make these decisions in a worst-case scenario as the weather will not always be optimal for solar power harvesting.

Decision Theory

In simple terms, decision theory is the analysis of the decision-making process. This process evaluates how choices and decisions are made (Gordon, 2022). It is a study of how decisions are made in a system or this case algorithm where the decisions of the environment or variables are unknown (Gordon, 2022).

There are three main areas of decision theory, these are, prescriptive, descriptive and normative decision theory. Prescriptive decision theory is used to recommend how to make the best decision, including when there are unknown or uncertain environments (Gordon, 2022). Descriptive decision theory seeks to explain or describe how decisions are made. Finally, normative decision theory is used to give direction on how to make decisions (Gordon, 2022).

Chapter Three: Methodology and Data Collection Details

The main goal of the research that was conducted was an exploration of how data analytics can be used to enhance the performance and efficiency of an individual's solar power system. In the study, a critical realism paradigm was used when examining the quantitative data. An experimental research strategy was used as small changes were made in the data to determine the effect of it on the outcome.

Critical realism is a research tradition that acknowledges that an objective reality exists, however, a person's understanding of it is shaped by the person's perception and the social constructs that have been adapted and created by societies and the environment around them (Bal, 2020). In the study, a realistic outlook was used when conducting the research. While the real world can not be observed directly, the world that is observed by everyone can be studied. With this being the case, it is important to understand the events that occur leading up to the events that will be observed (Bal, 2020). Predictive data analytics served as a bridge between the reality of the system and how it is perceived. Grounded theory was also implemented, which means that while the data is being collected, the analysis is being performed (Noble & Mitchell, 2016), this was achieved by using simulated data.

Data Collection

The data for the study was simulated using pvlib library. The main reason for simulating data is that months or even years' worth of data can be generated in a small amount of time. This also means that multiple systems do not need to be set up around South Africa to gather data, instead it was all gathered and analysed in a much smaller time frame. Pvlib is a Python library that takes in many different variables and uses them to simulate how they may affect the outcome of the power that is generated (Holmgren, Hansen, and Mikofski, 2018). The following are some examples of information that was used to determine the outcome, the longitude and latitude of the location, the tilt of the panel, the cardinal direction that the panel will be facing, the start and end date of the simulation and finally how many panels will be set (Holmgren, Hansen, and Mikofski, 2018). Pvlib then ran all the calculations needed to determine how much power the set-

up would generate. This was done multiple times while making small changes to the data generated to see the best outcome.

Pvlib is a well-known and established library that is well-maintained and uses real data to run the calculator. This means that the outcomes are more realistic as the weather patterns and other information come from past data that has been collected. This makes the outcomes much more reliable. If, for example, a simulator was used that uses predicted weather patterns based on user assumption of the weather, the data become unreliable and therefore the outcomes are no longer as accurate. With global warming, the weather has become more unpredictable, however, pvlib uses historical data to predict the outcomes of the weather, whereas a person can not make the same prediction that would be as accurate.

Ethical Considerations

Careful adherence to ethical principles is essential to the validity and integrity of the study, extreme care was taken while conducting this research. Despite the fact that this study does not directly involve human subjects, it does involve a number of ethical issues that are critical to guaranteeing the reliability and validity of the findings.

The approach to data confidentiality and privacy has remained constant. The materials and simulation tools that were made publicly available provided the data used in this study. Strict standards for data privacy and confidentiality were followed even though no sensitive or private information was involved. To protect privacy and legal compliance, all data sources have been properly cited, and all data handling has been done in accordance with applicable data usage and sharing policies.

One of the main ethical principles supporting this research is transparency. To ensure that the procedures are understandable and reproducible by other researchers, data simulation and analysis has been carried out in a transparent manner. This dedication to openness is essential as it permits the transparent analysis and validation of the research methodology and conclusions.

Ensuring scientific integrity has been paramount throughout the entire investigation. There has been no fabrication, manipulation, or misrepresentation of any kind in the research process. The study's findings are accurately represented in the results, which have not been changed or twisted in any way.

The approach to recognizing and granting credit for the contributions of other researchers, pertinent literature, and the data simulation library has been strictly followed. In addition, crediting other people's contributions is a requirement of ethical research, so proper citations and acknowledgments were essential to the process.

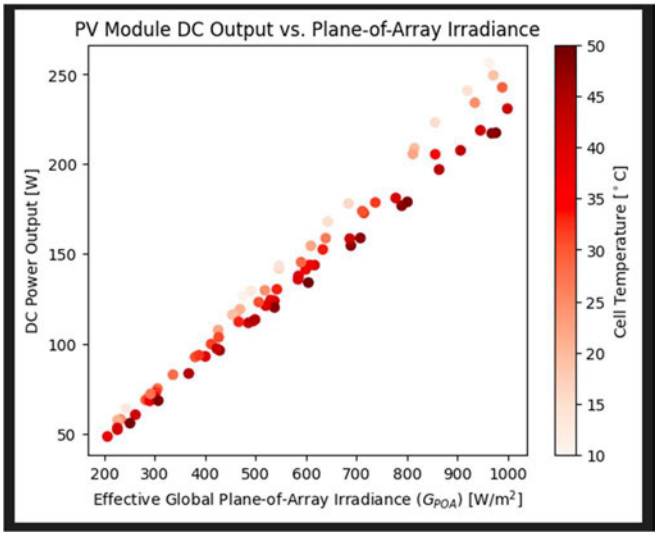
In closing, this study has been carried out, fully aware of the ethical precepts that guide moral scientific investigation. These ethical considerations are crucial for establishing the research's credibility, transparency, and dependability even though it does not directly involve human participants.

Chapter Four: Empirical analysis and results

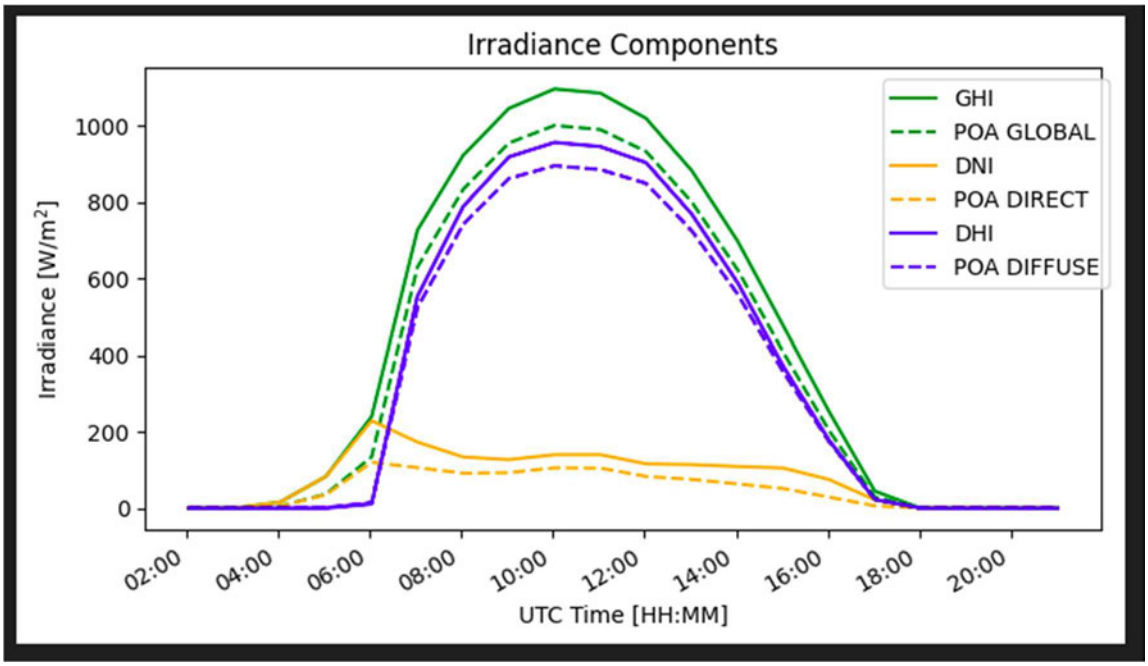
Results Of Data Analysis

In this section, the generated data and a detailed analysis of the results will be presented.

The study predicted a daily average plane of array (POA) irradiation of 50.38 kWh. This represents the amount of solar power on the surface of the solar panels. With different weather patterns, the irradiation will fluctuate and be higher in the summer and lower in the winter. In Figure 1 it can be seen that there is a linear relationship between the irradiance and the power that is generated. It can also be seen that the temperature has a negative impact on how the photovoltaic (PV) module performs. In Figure 2 the components of irradiation can be seen and how they affect the irradiation over a 18-hour period.

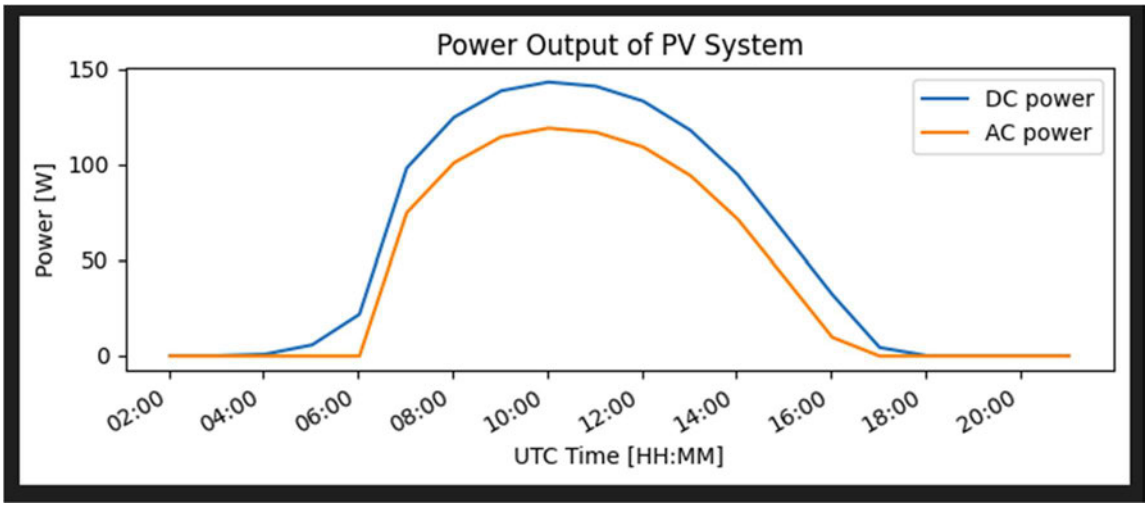


[Figure 1: PV Module DC Output vs. Plane-of-Array Irradiance]



[Figure 2: Irradiance Components over 18 Hours]

The PV capacity of the system used is 175.09 watts (W). Daily energy generation data for both direct current (DC) and alternating current (AC) outputs was generated. In figure 3 the daily generated power can be seen. The DC represents the actual energy produced by the panels while the AC is the usable energy that can be used by the household or fed back into the grid. It can be seen that the AC is lower than the DC which is normal as there will always be power lost in the conversion done by the installed inverter.



[Figure 3: Power Output of PV System over 18 Hours]

Chapter Five: Findings

Findings

The study's analysis has provided important new information about whether solar power systems are feasible for South African households. The findings show that an average household can generate enough energy with a modestly sized solar power system of 175.09 watts (W) to meet its electricity needs. The study's findings show that a solar power system with the required capacity can offer a practical and long-lasting energy source for individual homes. According to the analysis, this system is capable of producing a significant amount of energy that is suitable for household consumption when operating in ideal conditions.

A significant discovery concerns the ideal alignment of solar panels. The system produces an average of roughly 158.1 kilowatt-hours (kWh) of alternating current (AC) energy per month when the panels are facing south. However, this number rises to about 196.5 kWh when panels are facing northward. The latter considerably reduces the average household's electricity consumption, even though it might not entirely meet every household's monthly energy needs. This result emphasises the significance of panel orientation and the advantages of orienting solar panels northward for improved power output. By fully utilising the solar resources in the area, this orientation eventually increases the amount of energy produced for household needs.

Essential information about the system's sensitivity to sunlight levels can be obtained from the relationship between the direct current (DC) energy generated and the plane of array (POA) irradiation. The study shows that the solar system's power output and the amount of sunshine it receives are directly related. This sensitivity highlights the necessity of methodical optimization to guarantee optimal system performance. A number of variables are involved in optimization, one of which is modifying the tilt angle and cardinal direction of solar panels. In order to optimise energy generation, it is also essential to minimise shading effects on the panels. These results highlight how crucial it is to routinely observe and care for the solar system in order to guarantee optimal performance.

The data analysis concludes by highlighting the viability and possible advantages of solar power system adoption for South African households. Homeowners can

maximise their solar installations and increase energy generation and sustainability by making educated decisions about the impact of panel orientation and the system's sensitivity to sunlight levels.

These findings open up new avenues for investigation and useful applications to maximise solar energy harvesting, paving the way for a more sustainable energy future for South Africa.

Recommendations

A crucial area of further study involves conducting comprehensive comparative analyses across diverse South African regions to assess the viability and efficiency of solar power systems. These studies should explore the implications of geographical variations and topographical changes on the production of solar electricity, contributing to a deeper understanding of regional dynamics and how topographical alterations impact solar installations. This research can provide valuable insights for optimising solar power systems across the country.

To ensure the long-term sustainability and effectiveness of solar power systems, researchers should delve into the durability and dependability of specific solar power installations. Assessing their performance over extended periods is vital for the successful deployment of solar energy solutions. As the adoption of solar power systems continues to grow, it becomes essential to examine the effects of routine maintenance on the efficiency of solar panels. An in-depth analysis of the integration of energy storage devices like batteries with solar power systems is gaining prominence. Evaluating how this integration affects energy availability and reliability is crucial for maximising the benefits of solar energy.

Limitations

While this study has significantly advanced our knowledge of the viability of solar power systems for South African families, there are a number of limitations that should be taken into consideration when interpreting and extrapolating the results.

The use of simulated data produced by the pvlib library is one of the study's main drawbacks. Even though simulation has the benefit of producing huge datasets quickly, it might not accurately reflect the complexity of real-world situations. Certain assumptions made throughout the simulation process, like panel efficiency and weather patterns, might not accurately represent real-world circumstances. As a result, there is some doubt about how well the results translate to actual situations. To examine the data, the study used models from decision theory, optimization theory, and predictive analytics. To produce insights, these models reduce the complexity of real-world situations. Certain assumptions and simplifications made by these models may impair their accuracy, which could have an impact on the precision of the outcomes.

The study's assumptions about consistent weather patterns and trends in historical data might not always hold true in real-world situations. External factors that are difficult to fully account for, like climate change, have an impact on solar power generation in the real world.

The analysis of solar power generation in the short term is the main focus of the study. Extended periods of observation are necessary to evaluate the long-term durability, maintenance, and efficiency of solar systems. To gain a deeper understanding of system performance over time, future research should take long-term data into account.

References

Bal, H. (2020) What is Critical Realism?, Education studies. Available at: https://warwick.ac.uk/fac/soc/ces/research/current/socialtheory/maps/criticalrealism/#:~:text=Critical%20Realism%20(CR)%20is%20a,perceptions%2C%20theories%2C%20and%20constructions. (Accessed: 01 June 2023).

F. Holmgren, W., W. Hansen, C. and A. Mikofski, M. (2018) 'Pvlib python: A python package for Modeling Solar Energy Systems', Journal of Open Source Software, 3(29), p. 884. doi:10.21105/joss.00884.

Gordon, J. (2022) Decision theory - explained, The Business Professor, LLC. Available at: https://thebusinessprofessor.com/en_US/research-analysis-decision-science/decision-theory-definition (Accessed: 14 May 2023).

IBM. (2021) What is predictive analytics?, IBM. Available at: https://www.ibm.com/topics/predictive-analytics#:~:text=Predictive%20analytics%20is%20a%20branch,to%20identify%20risks%20and%20opportunities. (Accessed: 13 May 2023).

Labuschagne, H. (2022) Scary Load-shedding statistic revealed - mybroadband.co.za, Scary load-shedding statistic revealed. Available at: https://mybroadband.co.za/news/energy/470217-scary-load-shedding-statistic-revealed.html (Accessed: April 22, 2023).

Noble, H. and Mitchell, G. (2016) What is grounded theory?, Evidence-Based Nursing. Available at: https://ebn.bmj.com/content/19/2/34 (Accessed: 01 June 2023).

Morgan, B. (2021) Unit 1) optimization theory, Medium. Available at: https://towardsdatascience.com/unit-1-optimization-theory-e416dcf30ba8#:~:text=Optimization%20Theory%20is%20a%20branch,computer%20science%20and%20applied%20mathematics. (Accessed: 13 May 2023).

Overen, O.K. and Meyer, E.L. (2022) 'Solar Energy Resources and photovoltaic power potential of an underutilised region: A case of Alice, South Africa', Energies, 15(13), p. 4646. doi:10.3390/en15134646.

Ritchie, H., Roser, M. and Rosado, P. (2022) Fossil fuels, Our World in Data. Available at: https://ourworldindata.org/fossil-fuels#fossil-fuels-in-the-energy-and-electricity-mix (Accessed: April 22, 2023).

Ritchie, H., Roser, M. and Rosado, P. (2022) South Africa: Energy country profile, Our World in Data. Available at: https://ourworldindata.org/energy/country/south-africa (Accessed: April 22, 2023).

Sinha, G.R. (2020) 'Introduction and background to optimization theory', in Modern optimization methods for Science, Engineering and Technology. Bristol, UK: IOP Publishing, pp. 1–18.

Staub, N. (2021) What factors affect solar PV system efficiency, Eco Green Energy. Available at: https://www.eco-greenenergy.com/what-factors-affect-solar-pv-system-efficiency/ (Accessed: April 22, 2023).

Addenda

THE INDEPENDENT INSTITUTE OF EDUCATION

Vega

Varsity College

Boston College

MSA

SCHOOL OF HOSPITALITY & SERVICE MANAGEMENT

22 August 2023

Student name: Ashleigh Kingwill
Student number: [redacted]
Brand and campus: Varsity College Online

Approval of Postgraduate in Data Analytics Proposal and Ethics Clearance

Your research proposal and the ethical implications of your proposed research topic were reviewed by the School of IT Research Ethics Committee, a subcommittee of The Independent Institute of Education's Research and Postgraduate Studies Committee.

Your research proposal posed no significant ethical concerns and we hereby provide you with ethical clearance to proceed with your data collection.

There may be some aspects that you still need to address in your proposal. If this is the case, feedback will be provided to you in writing. You will need to address these aspects in consultation with your supervisor.

In the event that you decide to change your research topic or methodology in any way, kindly consult your supervisor to ensure that all ethical considerations are adhered to and pose no risk to any participant or party involved. A revised ethics clearance letter will be issued in such instances.

We wish you all the best with your research!

Yours sincerely,

[redacted]
Taryn Wilson
Programme Operations Manager

[redacted]
Mfowabo Maphosa
Research Supervisor Lead

ADVTECH GROUP

ADvTECH House: Inanda Greens, 54 Wierda Rd West, Wierda Valley 2196
P.O. Box 2338, Randburg 2125
Directors: RJ Douglas (UK), JDR Oasch, MD Aitken
Group Company Secretary: CB Grouse

BAC

The Independent Institute of Education (Pty) Ltd is registered with the Department of Higher Education and Training as a private education institution under the Higher Education Act, 1997 (reg. no. 2007/HE07/002). Company registration number: 1997/004764/07.