ABSTRACT

This project explores the use of machine learning techniques for predicting solar irradiance, a critical factor in optimizing solar energy systems and integrating renewable energy into power grids. By leveraging data from NASA's POWER project, key machine learning models such as Polynomial Regression, K-Nearest Neighbors (KNN), Random Forest, and Long Short-Term Memory (LSTM) networks were employed to analyze historical meteorological and solar radiation data. The study highlights the effectiveness of these models in capturing non-linear patterns and temporal dependencies, ensuring improved prediction accuracy. Performance metrics like Mean Squared Error (MSE) and R-Squared (R^2) were used to evaluate model performance, providing insights into their reliability. This work underscores the potential of advanced predictive analytics in enhancing energy management, grid integration, and the efficiency of renewable energy systems. Future efforts will focus on integrating adaptive learning systems and hybrid energy solutions to further optimize energy sustainability.

CHAPTER 1

1.1 INTRODUCTION

Solar irradiance, which refers to forces of solar energy received over a given unit area, is one of the basic parameters in systems of renewable energies. Highly accurate prediction of solar irradiance is important to maximise solar panel performance, to efficiently manage energy storage systems, and to introduce solar power delivery into the energy grids. It is however difficult to forecast solar irradiance because of its dependence to a number of meteorological factors such as temperature, humidity, cloud cover, and time of day, the relationships between which are complex and nonlinear.

Developments of machine learning (ML) have provided new ways of solving these challenges. There are various techniques of ML such as regression models, ensemble methods and deep learning networks that are able to find complex patterns in data and build accurate predictions. This project utilizes these techniques to forecast solar irradiance through the use of meteorological data history and real-time data. Based on data on NASA's Prediction of Worldwide Energy Resources (POWER) project, this study develops and tests such models as Polynomial Regression, K-Nearest Neighbors (KNN), Random Forest, and Long Short-Term Memory (LSTM) networks.

The details of the interdisciplinary approach undertaken by the team have been discussed in the report, which includes incorporation of ideas of electrical engineering and data science. It underlines the capability of machine learning in solving real world energy issues towards sustainable energy management and climate change mitigation.

1.2 OBJECTIVES

- To give accurate forecasts on solar irradiance levels.
- To make use of the NASA POWER information effectively
- To apply and compare various machine learning procedures
- To examine the effects of meteorological factors.
- To adopt an interdisciplinary approach
- In order to contribute to sustainable energy management

CHAPTER 2

2.1 UNDERSTANDING SOLAR RADIATION

Definition: Solar radiation is the energy emitted by the sun, reaching Earth as sunlight, measured in irradiance (W/m^2).

Types of Solar Radiation:

- Direct Solar Radiation: Sunlight reaching Earth without scattering.
- Diffuse Solar Radiation: Scattered sunlight due to atmospheric particles.
- Global Solar Radiation: Combination of direct and diffuse radiation.

2.2 BENEFITS OF SOLAR IRRADIANCE PREDICTION

- Reliable Energy Generation: Neutralizes fluctuation of solar power supply.
- Grid Management: Enables its integration into the grid through the balancing supply and demand.
- Optimizing Solar Panel Operation: Increases efficiency by directing orientation of the panels as well as storage of energy.
- Smart Grid Applications: Helps in managing load and distributing energy.

2.3 DATA COLLECTION FROM NASA POWER

Source: NASA's Prediction of Worldwide Energy Resources (POWER) project provides historical solar and meteorological data.

Data Parameters:

- Solar radiation
- Year
- Month
- Day
- Hour

	YEAR	MONTH	DAY	HOUR	ALLSKY_SFC_SW_DWN	T2M	QV2M	WS10M
0	2021	1	1	5	0.00	15.23	11.90	1.93
1	2021	1	1	6	20.94	15.57	11.84	2.09
2	2021	1	1	7	146.68	17.51	12.15	3.26
3	2021	1	1	8	323.63	19.92	12.45	3.78
4	2021	1	1	9	500.41	22.02	12.45	3.95

Fig 1: Data taken from NASA website

2.4 INSIGHTS:

- Yearly patterns in average solar irradiance are similar, making yearly prediction less effective.
- Monthly patterns show little variation; hence monthly predictions are not very useful for grid operations.
- Daily predictions offer significant potential for grid efficiency due to substantial day-wise variations.

2.5 ANALYSIS OF NASA DATA:

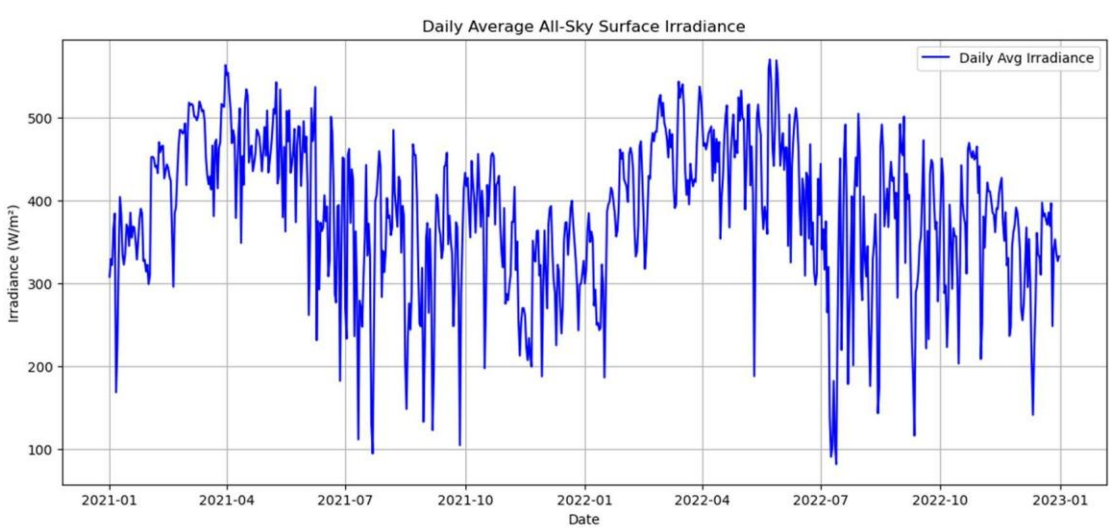

Fig 2: Daily Average of 2 years

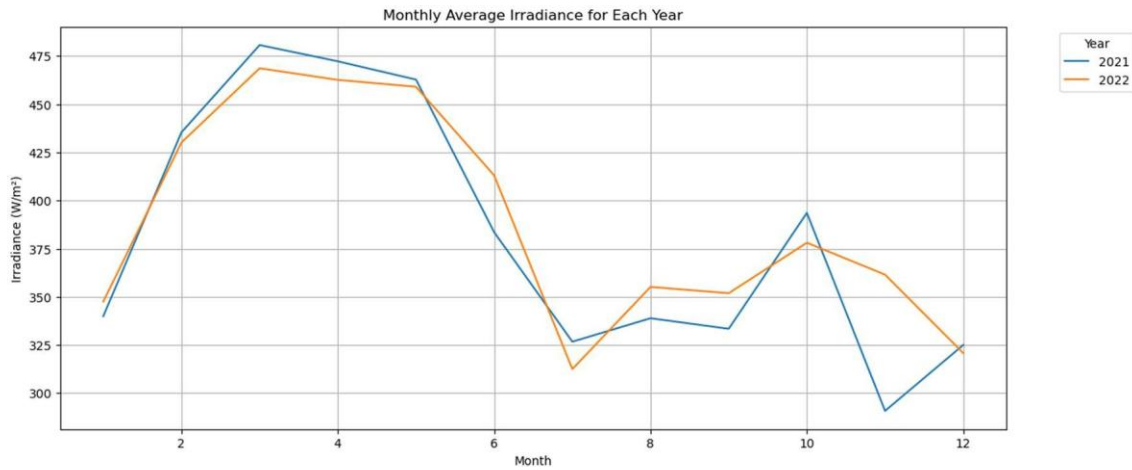

Fig 3: Monthly average of given data

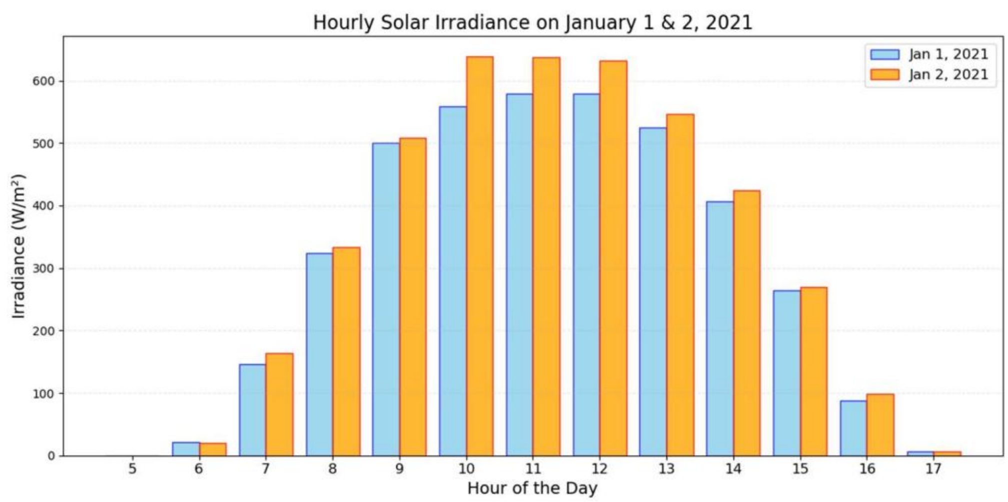

Fig 4: Hourly Solar Irradiance on 1 & 2 January

CHAPTER 3
MACHINE LEARNING TECHNIQUES FOR PREDICTION

3.1 TECHNIQUES OVERVIEW:

Polynomial Regression: An extension of linear regression that fits a polynomial curve, allowing for better modeling of nonlinear relationships.

- Models complex nonlinear relationships.
- Flexible and captures intricate patterns.
- Risk of overfitting with higher degrees.

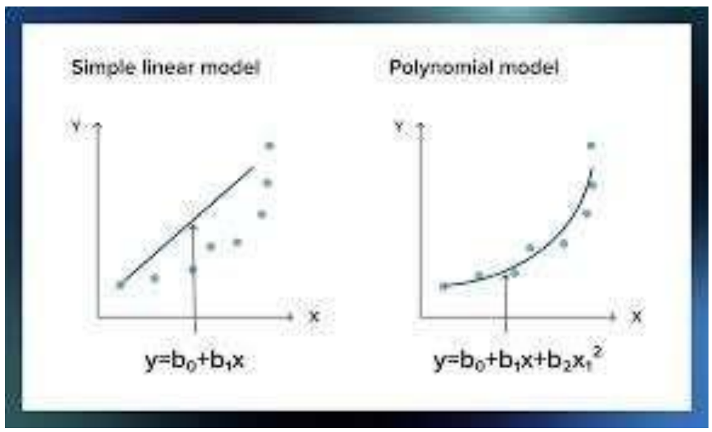

Fig 5: linear and polynomial model

K-Nearest Neighbors (KNN):

- A non-parametric method that predicts the values based on average of the nearest datum, local trends in the data are captured.
- Deals with clusters and local trends in solar irradiance data.
- Non-parametric and flexible.
- Computationally expensive on large datasets

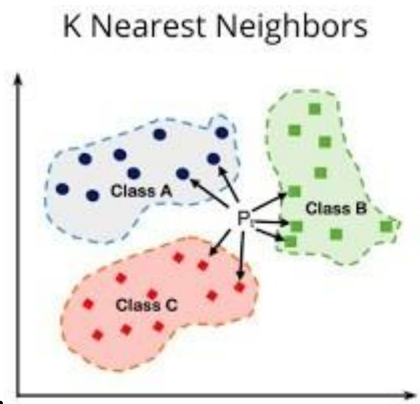

Fig 6: KNN Representation

Random Forest:

It is the most popular machine learning algorithm that bring together the theory of both ensemble learning and decision trees. It bags a number of decision trees (“forest”) and makes compilation of their predictions in order to generate a more accurate and stable outcome.

Workflow:

- Data Sampling: Bootstrapped subsets train decision trees. Feature Selection: Random selection of features eliminates over fitting. Aggregation: Averaging is done for regression tasks in terms of outputs.
- It is strong and works well on complicated, non-linear relations.

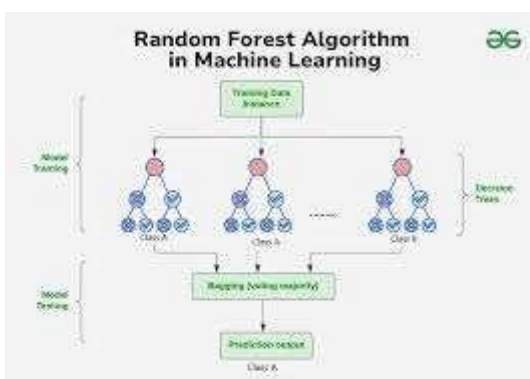

Fig 7: Random Forest Representation

LSTM (Long Short-Term Memory): It is an example of some variant of artificial recurrent neural network (RNN) architecture used in the frame of deep learning. It is very good at processing and making predictions from sequential data, such as time series, natural language, or video frames.

Workflow:

- LSTM units use gates to manage flow of information in processed sequences. Stores the temporal dependences of time-series information.
- Time-series data of long-term dependencies.

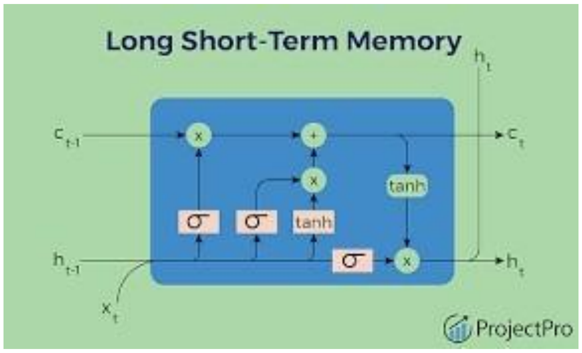

Fig 8: Workflow of Long short term memory

3.2 POLYNOMIAL REGRESSION

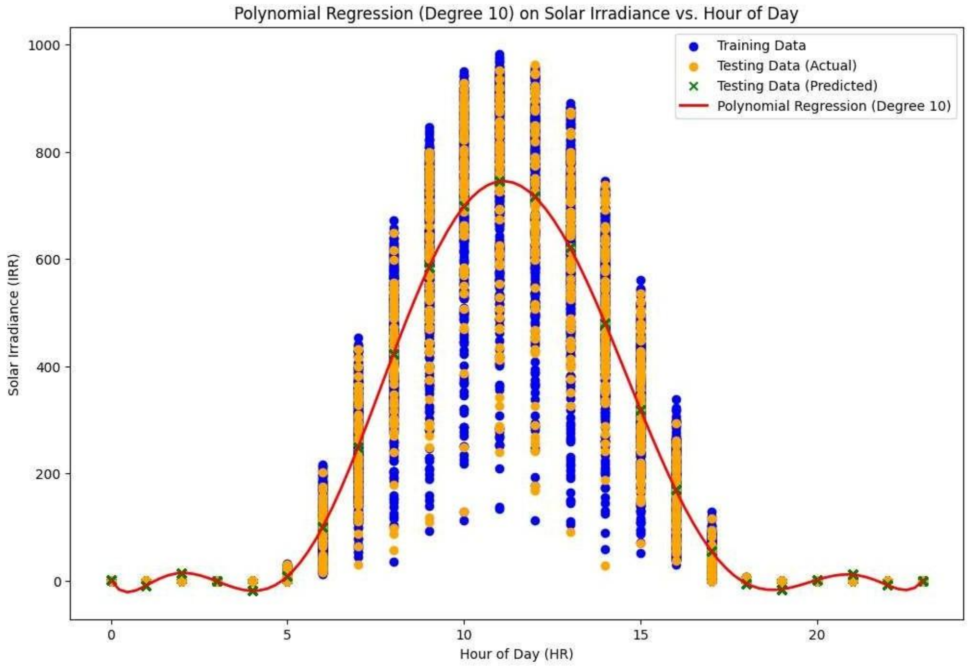

Fig 9: Prediction using polynomial regression

```
Mean Squared Error: 8747.557847564723
R-squared: 0.8903780291557282

Sample of Actual vs Predicted Irradiance:
   Hour (HR)  Actual Irradiance (IRR)  Predicted Irradiance (IRR)
0          1                     0.00                   -8.641186
1          0                     0.00                    1.492969
2         11                   770.45                  744.624814
3          9                   744.38                  583.549932
4         16                   156.02                  169.407000
5         14                   558.72                  480.318561
6         23                     0.00                    0.621116
7         14                   485.70                  480.318561
8         14                   507.65                  480.318561
9          6                    18.58                  100.789812
```

Fig 10: Performance Metrics

3.3 K-Nearest Neighbors (KNN)

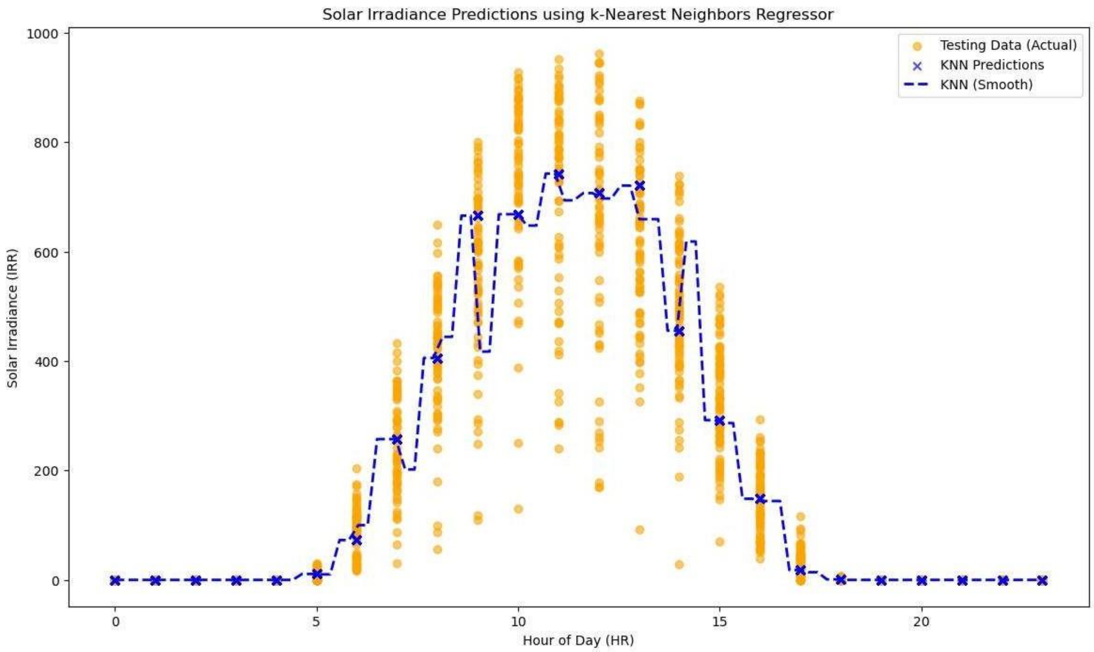

Fig 11: Prediction Using KNN

```
Sample of Actual vs Predicted Irradiance (First 10 rows):
   Hour (HR)  Actual Irradiance (IRR)  KNN Predicted (IRR)
0          1                     0.00                0.000
1          0                     0.00                0.000
2         11                   770.45              742.400
3          9                   744.38              665.524
4         16                   156.02              148.292
5         14                   558.72              455.420
6         23                     0.00                0.000
7         14                   485.70              455.420
8         14                   507.65              455.420
9          6                    18.58               72.774
k-Nearest Neighbors Regressor - Mean Squared Error: 9693.03288997382
k-Nearest Neighbors Regressor - R-squared: 0.8785295979319426
```

Fig 12: Performance Metrics

3.6 PERFORMANCE METRICS:

1. **Mean Squared Error (MSE):** Measures are average squared differences between the observed and expected value. Lower values indicate better performance.
2. **R-Squared (R^2):** Indicates the percentage of variance that is explained by the model. The closer the values are to 1, the better fit.

CHAPTER 4

4.1 CONCLUSION

Solar irradiance prediction has undergone a great leap with the use of the machine learning techniques. Key outcomes include:

- Increased ML models based prediction accuracy.
- Understanding of effectiveness of different models through metrics of performance.
- Opportunities to further integrate real-time as well as adaptive learning systems in order to improve energy management.
- Using cutting-edge predictive tools, this project illustrates how data-oriented solutions can be used to solve problems of renewable energy, preparing the ground for more sustainable energy systems.