Monash Solar Farm Project Report
ADS2001

Batheendra, Byron, Nelson, Marco

24/05/2024

Executive Summary

Background and Context

Decarbonizing energy production with renewable sources such as wind and solar is crucial but challenging due to their dependence on weather conditions and the misalignment with peak electricity demand. Energy storage is costly and often inefficient. Accurate forecasting of energy demand and renewable production is essential for efficient grid management, including using on-demand sources, load shedding, and optimal battery usage. A common setup includes rooftop solar panels, batteries, and flexible demand schedules. Accurate forecasts help optimise battery charging and discharging and schedule energy use efficiently.

The IEEE Computational Intelligence Society (IEEE-CIS) has partnered with Monash University to improve solutions for battery and load scheduling in renewable energy systems. Monash aims to achieve Net Zero emissions by 2030 through its Monash Net Zero Initiative, which includes a microgrid with solar panels and battery storage. This challenge involves developing optimal schedules using machine learning to handle time series prediction and optimization under uncertainty. Successful solutions will enhance renewable energy reliability and affordability, with potential applications in various fields.

The historical 15-minute power production of six solar panels and energy consumption of six buildings was given by Monash University (Figure 2) along with the ERA 5 dataset (Figure 1) which contained recordings of many weather variables from 2010 to 2021. The goal of the project was to make predictions for the energy supply and demand for the first week of October in 2020 and determine whether the energy generated by solar panels on 6 buildings on campus meets the energy demand by 6 different buildings on campus. Exploratory data analysis and predictions were also made on the given data sets to identify trends, patterns, insights and relationships between variables.

Approach and Results

We were provided with two datasets, the ERA 5 data set (Figure 1) and the Phase 1 dataset (Figure 2), which contained 15-minute readings for the energy demand from 6 buildings on campus and the solar energy supplied by solar panels on 6 other buildings. With these data sets and the limited information that was provided to us, our initial steps were to first gain an understanding of the context of the data and understand the meaning behind the variables, as a lot of this information was not provided to us. We also expanded our knowledge on renewable energy, more specifically solar panels and how they work to prepare us for the project work ahead. We then later moved onto cleaning the datasets by looking for missing values, duplicate rows, and outliers and creating new variables by extracting information from pre-existing variables. Overall, the cleaning portion of our project wasn't too tedious. However, the phase 1 data was quite complex as the raw data was in the form of a tuple. This data had to be converted into a pandas data frame indexed by date time, which was done by using the starting date and the length of the series values array to determine an end date. For exploratory analysis, the findings for both solar supply and demand largely came from time series plots where patterns could be observed. Also, time series were plotted for a variety of variables from the ERA 5 data set and smoothed to examine more clear trends and patterns.

For our modelling, we set out a goal of assembling many models to predict and forecast the weather data and energy supply and demand. Many models were developed but only a few resulted in sufficient accuracy. The models that were not presented in our final presentation had low accuracy scores and due to our insufficient computing power, some of these models would not run on our devices. We evaluated a SARIMA (Seasonal Autoregressive Integrated Moving Average) model to forecast and analyse the time series data and a LSTM (Long Short-Term Memory) model which specialises in time series forecasting. Ultimately, we used a random forest regression model to predict the weather variables in the ERA 5 Data set. Features included year, month, day, and hour, while target variables were various weather metrics like temperature, dew point, wind speed, and cloud cover. The model achieved high accuracy with an r2 of 0.93 for temperature prediction and 0.98 for dew point temperature. The lowest r2 was 0.77 for total cloud cover. An interactive function was implemented into the code, "interactive_predict," which allowed users to input the variable they wanted to predict and date and time to get predictions. Random forest regressors were also trained taking feature variables from the era ERA 5 data and taking solar power production as a target variable, achieving r2 scores as high as 0.800.

Introduction

Decarbonizing energy production with renewable sources like wind and solar is crucial but challenging due to their weather dependency and misalignment with peak electricity demand. Efficient energy storage is costly and often inefficient, making accurate forecasting of energy demand and renewable production essential for effective grid management. This involves optimising battery usage, on-demand sources, and flexible demand schedules. The IEEE Computational Intelligence Society partnered with Monash University, aiming for Net Zero emissions by 2030 through the Monash Net Zero Initiative, which includes a solar-powered microgrid.

In this project, we were provided with two datasets, the ERA 5 data set and the Phase 1 data set which had 15-minute intervals of data from six solar panels and six buildings. The era five data set contains information on thirteen weather related variables every hour from 2010 to 2021. Some of these variables are temperature (degC), surface_solar_radiation (W/m^2), and surface_thermal_radiation (W/m^2). The Phase 1 data set contains historical data which describes the energy consumption of six buildings and the electricity generation of six solar panels at Monash University in Melbourne. During a period of fifteen minutes.

The objective of our project was to make predictions for the energy supply and demand for the first week of October in 2020 and determine whether the energy generated by solar panels on 6 buildings on campus meets the energy demand by 6 different buildings on campus. We were also tasked with performing exploratory data analysis on the datasets and creating predictions to identify trends, patterns, insights, and relationships between variables.

Data Quality and Cleaning

The initial stages of the project involved the cleaning and preprocessing of the two datasets so that they could be ready for exploratory data analysis, modelling and forecasting. For the ERA 5 data set, the cleaning process was quite simple. The ERA 5 data set had no missing values, no duplicate rows, no extreme outliers or unusual recordings and was in a pandas data frame. However, for future analysis and modelling, new columns were created by extracting data from pre-existing columns. For example the "coordinates (lat,lon)" variable had data in the form of (x,y), where x was the longitude and y was the latitude, therefore the coordinates column was removed from the data set and the data was extracted into two separate columns for longitude and latitude. From the “datetime (UTC)” we extracted the data and added “year,” “month,” “day” and “hour” columns for future modelling.

For the phase 1 data set, the cleaning is more complex than the ERA 5 data set. The phase 1 dataset contains values for energy consumption for the 6 buildings and solar supply for the 6 solar panels. The dataset contains missing values and the starting timestamp for each building and solar panel are not the same. First, the dataset was checked if there are duplicated timestamps by using the duplicate function. Secondly, each building and solar is split into individual data frames. Each building and solar data frame's starting timestamp are set to the date that is started to be recorded. Having filtered out the dates that are not used, missing or null values are replaced with “0”. This ensured that there will not be any missing or null value in between the starting timestamp and the ending timestamp due to closure for the building or solar panel. With a view to fitting in with the ERA 5 data set for future analysis and modelling, the phase 1 dataset is resampled hourly instead of every 15 minutes by calculating the mean for each hour. Also, it was important for the index of each solar supply and energy demand data frame to have a timestamp index so that exploration of time series could occur. To do this, as we knew the starting date and that each reading was taken 15 minutes apart, the end date and time could be found by taking the length of each supply and demand array and multiplying it by 15 minutes and adding that value to the start time. Once the end times for each array were calculated, a date time index could be created for each building which aligned with the provided energy readings.

Exploratory Data Analysis

ERA 5 Data Set

The first step of the exploratory data analysis involved reading the dataset. Since we were dealing with variables like “temperature (degC)”and “surface solar radiation (W/m^2)” with data connected to a specific time, we knew plotting time series for the variables in the data set would be important and crucial for our analysis. To plot the time series for “temperature (degC)”, “surface solar radiation (W/m^2)”, “Dew-Point Temperature (degC)” and “surface thermal radiation (W/m^2)” we defined a list of tuples, columns, specifying the variables to be plotted with their respective labels and colours. The loop iterates over the four variables to create subplots, each showing a time series for one variable. Figure 3 presents the raw time series for the four variables and clearly demonstrates how each variable behaves and changes over time. The plots show consistent seasonal patterns from 2010 to 2021. These raw time series plots are noisy and hide long-term trends, therefore, to reduce the noise and highlight the long-term trends better, we used a smoothing filter, uniform_filter1d function from the scipy.ndimage module to smooth the plots. This method smooths the data by averaging values within a specified window size, for our plots we used a window size of 17000. Another loop was then used to create subplots for the smoothed data, following the same format as the original plots. Figure 4 shows the smoothed plots, which give a clearer view of the long-term trends and remove the noise from the raw time series. The smoothed temperature time series possibly indicates a slight warming over the years. The smoothed dew point temperature demonstrates underlying humidity trends and shows a small decrease in dew point temperature from 2010 to 2020, and then a sudden increase in dew point temperature in 2020.

Solar Energy Supply

Due to the solar supply data being now indexed by date time with a column of energy readings as the values, it was clear that plotting time series for each solar supply would be a sensible start to EDA. It was predicted that there would be a seasonal, wave-like pattern to the time series data as the generation of solar power depends mainly on the surface solar radiation, which peaks in summer and decreases towards winter. As seen in figure 5, the solar power generated by the 6 buildings on campus fluctuate seasonally, validating the hypothesis of a wave-like pattern. Between the 6 solar readings provided, the starting dates varied. Solar 1 had some energy readings as early as January 2019 before a period of about 8 months with no readings. Whereas solar 5, 4, 2 and 3 didn’t have consistent readings until October 2019 with solar 0 only getting energy readings halfway through July 2020.

Something else that stood out in the time series was the gaps where readings were either missing or appeared to be incorrect or faulty. There are gaps where no readings were taken between February 2019 and July 2019, as well as in August/September 2019. There is a 6 week stretch from the end of April to the start of May 2020 where data readings appear to be faulty. This is suggested by the clear linear appearing energy readings, which vastly contrasts the sinusoidal, wave-like pattern present between October 2019 and October 2020. This 6 week stretch was investigated more closely by using the loc function in Pandas to print off the solar energy supply readings during this period. These readings were nearly constant for the solar 1 supply, maintaining a reading of 7.63 units, decreasing to 7.62 then eventually 7.61 over the 6 weeks, even during the night, where the solar supply should be 0. For solar 2 there was a gradual linear increase in this period, and solar 3 and 4 experienced gradual linear decreases. This period of anomaly was noted to be excluded from any modelling, as it didn’t contain accurate readings of solar power supply.

Another take away from this time series plot was the varying magnitudes in which solar power was generated between different buildings. There appeared to be 2 bins, an upper bin containing solar 5 and 0 and a lower bin containing solar 1, 2, 3 and 4. As all 6 buildings are on campus, weather factors such as the solar radiation won't vary enough from building to building to cause significant differences in the energy produced by each building's solar panels. Thus, the cause of the discrepancy in solar energy supply was theorised to be that the buildings in the upper bin must have a greater surface area of solar panels compared to those in the lower bin, but the actual reason is unknown. The wave-like pattern, missing values and faulty readings are accentuated when the lower and upper bin time series plots are plotted separately (figures 6 and 7), as the y-axis scale (solar energy supply) is fitted more appropriately.

Building Energy Demand

The building data is indexed by date time with a column of energy readings as the values, plotting a time series will be a perfect start for exploratory data analysis. The figure 8 time series graph shows a seasonal pattern for all the buildings. As shown in figure 8, the energy demand peaks at May and June, as it is getting close to exam period, while decreases in winter as summer holiday comes through, most all the building is closed during holiday. On top of what the 6 building values provided, the starting datetime for each building is not the same. Building 3 has the earliest record of energy consumption on the first of March in 2016, while building 0 has started being recorded four months later. Building 1, 4, 5 and 6 were being recorded in 2019, while building 1 starts having record on 9th January, 2019 as the earliest among the 4 buildings and building 5 and 6 were recorded at the latest on 25th July, 2019.

The first thing stood out from the time series plot was the gap between May 2018 to August 2019 for Building 0. There is no value in between these 15 months which seems to be faulty. This period was explored more deeply by using the loc function, which checks if the value for building 0 is missing in this 15 month. This reflected Building 0 may not be a suitable Building for prediction because as the missing values are replaced with 0, it will affect the accuracy of prediction.

The second thing that stood out was that two bins are observed from the time series model. Building 0 and Building 3 have a much higher energy consumption, while Building 1, 4, 5, 6 has a lower energy consumption. (figure 14). By grouping the buildings into two bins, we can raise the efficiency by only predicting the energy demand for one building in each bin and calculate the others by their ratio. This method works perfectly as all the buildings are located on campus, the weather factor will not affect significantly. The reason behind the bold difference between buildings will not be weather.

Modelling

Era 5

For the modelling portion of our group project, we developed an interactive model to predict all the weather variables in the ERA 5 data set using a Random Forest Regressor. The code began with defining the features and target variables, which included “year”, “month”, “day”, and “hour” for features, and all the weather-related variables such as “temperature (degC), “dew point temperature (degC)”, “wind speed (m/s)”, “mean sea level pressure (Pa)”, “relative humidity (0-1)”, “surface solar radiation (W/m^2)”, “surface thermal radiation(W/m^2)”, and “total cloud cover (0-1)” for the target variables. A dictionary named “models” was created to store the trained models. The Random Forest Regressor model was then trained for each target variable by splitting the data into training and testing sets using the train_test_split function from scikit-learn. Each model was then fitted to the training data and stored in the models dictionary. An interactive function called “interactive_predict” then prompts the user to choose a variable to predict, once the variable is chosen, it will ask the user to select a year, month, day then hour. Based on these inputs the model will output a prediction. Figure 10 shows the model when asked to predict the temperature on the 20/05/2024 at 5.00pm. The accuracy of the Random Forest Regressor model is quite high for predicting. The r2 value for temperature prediction is 0.93, showing a strong correlation between the predicted and actual values. The highest r2 score from all the weather variables is for the dewpoint “temperature (degC)”, with an r2 score of 0.98, representing near-perfect predictions. The lowest r2 value is for “total cloud cover (0-1)”, with an r2 score of 0.77, which, while lower compared to the other variables, still represents an accurate prediction. Other models were also evaluated. For example, we experimented with a SARIMA and LSTM model as well but due to insufficient computing power, we could not get these models to work. We did however complete a multivariate polynomial regression model to predict “temperature (degC)”, however due to time constraints we did not include it in the presentation.

Building Energy Demand

The goal for this prediction was to predict the energy demand for the 6 buildings for the first week of October 2020. Due to the nonlinear relationship between the weather data and building data, linear regression modelling wouldn’t be appropriate. Random forest regressor was chosen at last due to the fact that it can handle the wave shape or the energy demand and allow multiple variables to be used as features for input.

The Prediction model is using the columns in era 5 data set, it is essential to determine which features to be used for the modelling. Variables with larger influence to the energy demand should be taken into consideration first. Energy demand usually depends on the temperature, month and day. The energy demand model can be predicted by using the variables in the era 5 weather. A random forest regressor model was trained by using every model except coordinates, model and mean sea level pressure. The R2 score has achieved 0.57. It means that 57% of the total variability in the dependent variable is accounted for by the model. The remaining 43% of the variability is due to factors not included in the model or inherent randomness. With a view to knowing which features contributed the most in the prediction model(figure13), a feature importance model is created. The 5 most important features are surface solar, radiation, hour, relative humidity, temperature and surface thermal radiation respectively, which all of them has a higher value of 0.18. Provided there are data for era 5 weather data in October, 2020, no predictions are required for weather data.

For the prediction for building energy demand at the first week of October(figure 8), one model was trained using building 3 as the upper bin, while one model was trained using building 6 as the lower bin. The prediction for the rest of the building is calculated by the ratio. Building 0 is calculated by increasing 46.1% of the energy demand of building 3. Building 1 is calculated by increasing the energy demand of building 6 by 36.6%. Building 4 is calculated by decreasing the energy demand of building 6 by 61.9%. Building 5 is calculated by increasing the energy demand of building 6 by 49.6%.

Solar energy supply

The end goal for the project was to create predictions for the solar energy supply in the first week of October 2020, which was where the solar power generated was cut off. Due to the nonlinear nature of the solar supply data, linear regression of solar energy supply against time wouldn’t be appropriate. A polynomial regression was also considered, and due to there being a crest in solar supply around December 2019 and a trough in July 2020, a cubic polynomial could be fitted to the data. However, as time goes on this cubic would eventually tail off towards infinity as it only has a degree of 3, whereas the data would follow a wave pattern. A polynomial of any other degree would either overfit or underfit the data, and would only provide reasonably accurate results within the given time frame. Thus, a random forest regressor from sk learn was used to model solar energy supply, as it can handle the wave shape of the data and make predictions based on inputs from different variables.

The Era 5 data was used in conjunction with the solar energy supply data to create this model. It was important to determine which variables from the ERA 5 data had the greatest influence on the solar power generated at a particular time. As solar power solely depends on the weather and the sun, it was expected that a model for solar energy supply could be trained based on the weather data. Initially, a random forest regressor model was trained using every feature available from the ERA 5 data, achieving an R2 score of 0.784. This suggests 78.4% of the variation in the solar energy supply can be explained by variations in the time and weather, and 21.6% of the variation in solar power is explained by other factors. From this, the feature importance function of the random forest regressor was used to determine which features contributed the most in predicting solar energy supply (figure 11). The 5 most important features are the surface solar radiation in Watts per square metre, relative humidity, the hour of day, temperature, and month of the year, all having importances above 0.14. As we had access to the weather data for the first week of October 2020, we didn't have to worry about predicting any weather variables which would be needed to obtain a predicted solar energy supply value. However for future extrapolation, in order to predict solar energy supply, we would have to first predict values for all of the weather features used in making the model. If predicted values are used to predict another value, the accuracy of the final prediction will be further reduced.

For predicting the solar power values in the first week of October, one model was trained using solar 5 as the target variable for the upper bin and one was trained using solar 4 for the lower bin. The chosen feature variables were surface solar radiation, hour of day, relative humidity, temperature, month, wind speed, surface thermal radiation, dew point temperature and total cloud cover, each with a feature importance above 0.05. The model for the upper bin achieved an r2 score of 0.633 and the lower bin model achieved an r2 score of 0.800. A model for solar 0 was derived from the model for solar 5 by increasing all solar 5 predictions by 24.3% as the mean solar 0 value was 24.3% greater than the mean solar 5 predictions. The same was done with the lower bin predictions. The solar 4 predictions were increased by 57.3% for solar 2 predictions, 27.1% for solar 3 predictions and 41.7% for solar 1 predictions.

Predictions

To predict the solar energy supply in the first week of October 2020, a data frame containing all of the feature variables for that week was made using the loc function in pandas and dropping all of the columns not used in the model. The values for the predicted solar power generated were then found using the aforementioned models, and combined into their own data frame. An extra column was added for the average solar power generated each hour as well as the total power generated each hour. Finally, the total predicted solar power produced across all 6 buildings over the week was found by summing every element in the data frame, which produced an estimated solar energy supply of 3347.5269 units. When plotted (figure 11), the predicted solar energy supply from the 4th of October to the 5th was seen to have more than halved, which at first seemed like a big mistake in the model. After further research, it was found that this was likely an accurate prediction as the temperature dropped from a maximum of 24 degrees celsius on the 4th to only 11 degrees celsius on the 5th (figure 12, from The Global Historical Weather and Climate Data, 2020), which explains this drop. For the energy demand, it was predicted a total of 10665.351 units of energy would be used across the 6 buildings in the first week of October. A difference of 7317.824 units is calculated between energy demand and solar supply. The shortage of energy will need to be externally sourced in order to eliminate the shortage.

Conclusion

Through exploratory data analysis (EDA), insights into the behaviour of solar energy supply and building energy demand were gleaned. The solar energy supply exhibited a wave-like pattern, influenced by variables such as surface solar radiation, humidity, and temperature. Building energy consumption patterns were also observed as the Oscillating Pattern, influenced by variables such as surface solar radiation, temperature and hour.

In modelling, a Random Forest Regression model was utilised to predict weather variables from the ERA 5 dataset with high accuracy. Features such as temperature and surface solar radiation were found to be crucial in predicting solar energy supply and energy demand, with models achieving R2 scores of up to 0.800. Using these models, predictions for solar energy supply in the first week of October 2020 were made, totalling to 3347.5269 units of renewable solar energy. The total predicted energy demand across this week was 10665.351 units, meaning the shortfall of supply would have to be made up by sourcing energy externally, likely from non renewable sources. For Monash to achieve net zero emissions by 2030, other options must be considered for renewable energy supply, or alternatively more solar panels must be constructed on campus.

Figures

	datetime (UTC)	coordinates (lat,lon)	model (name)	model elevation (surface)	utc_offset (hrs)	temperature (degC)	dewpoint_temperature (degC)	wind_speed (m/s)	mean_sea_level_pressure (Pa)	relative_humidity ((0-1))	surface_solar_radiation (W/m^2)	surface_thermal_radiation (W/m^2)	total_cloud_cover (0-1)
0	2010-01-01 00:00:00	(-37.91, 145.13)	era5	69.59	10.0	18.26	16.39	2.60	101046.38	0.89	287.01	408.35	1.00
1	2010-01-01 01:00:00	(-37.91, 145.13)	era5	69.59	10.0	18.67	16.29	2.91	101037.96	0.86	360.79	411.02	1.00
2	2010-01-01 02:00:00	(-37.91, 145.13)	era5	69.59	10.0	18.16	15.89	3.26	101017.26	0.87	291.54	410.67	1.00
3	2010-01-01 03:00:00	(-37.91, 145.13)	era5	69.59	10.0	18.46	15.33	3.17	101022.56	0.82	357.11	410.95	1.00
4	2010-01-01 04:00:00	(-37.91, 145.13)	era5	69.59	10.0	18.53	15.11	2.95	100940.03	0.80	459.91	410.00	0.90
...	...	...	...	...	...	...	...	...	...	...	...	...	...
100052	2021-05-31 20:00:00	(-37.91, 145.13)	era5	69.59	10.0	8.31	3.28	4.91	102033.73	0.71	0.00	278.11	0.93
100053	2021-05-31 21:00:00	(-37.91, 145.13)	era5	69.59	10.0	8.15	3.16	4.63	102033.49	0.71	0.00	277.98	1.00
100054	2021-05-31 22:00:00	(-37.91, 145.13)	era5	69.59	10.0	8.72	3.31	5.29	102069.84	0.69	8.99	281.65	1.00
100055	2021-05-31 23:00:00	(-37.91, 145.13)	era5	69.59	10.0	9.83	3.88	5.60	102077.12	0.66	80.46	296.20	1.00
100056	2021-06-01 00:00:00	(-37.91, 145.13)	era5	69.59	10.0	10.73	4.25	5.43	102109.62	0.64	149.84	305.43	0.99

Figure 1 - ERA 5 data frame

```
(   series_name     start_timestamp  \
 0    Building0 2016-07-03 21:30:00
 1    Building1 2019-01-09 23:15:00
 2    Building3 2016-03-01 04:15:00
 3    Building4 2019-07-03 04:45:00
 4    Building5 2019-07-25 23:00:00
 5    Building6 2019-07-25 01:45:00
 6       Solar0 2020-04-25 14:00:00
 7       Solar1 2018-12-31 13:00:00
 8       Solar2 2019-06-05 14:00:00
 9       Solar3 2019-06-05 14:00:00
 10      Solar4 2019-06-05 14:00:00
 11      Solar5 2019-01-15 13:00:00

                                          series_value
 0   [283.8, 283.8, 283.8, 606.0, 606.0, 606.0, 606...
 1   [8.1, 15.7, 22.8, 32.7, 8.1, 16.5, 24.7, 34.5,...
 2   [1321.0, 1321.0, 1321.0, 1321.0, 1293.0, 1293....
 3   [2.0, NaN, 1.0, 2.0, NaN, 2.0, NaN, NaN, 2.0, ...
 4   [30.0, 31.0, 24.0, 34.0, 30.0, 31.0, 26.0, 33....
 5   [36.8, 34.6, 34.6, 36.2, 36.2, 35.2, 35.2, 35....
 6   [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
 7   [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
 8   [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
 9   [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
 10  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
 11  [0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...  ,
 '15_minutes',
 None,
 True,
 False)
```

Figure 2 - Phase 1 data as tuple

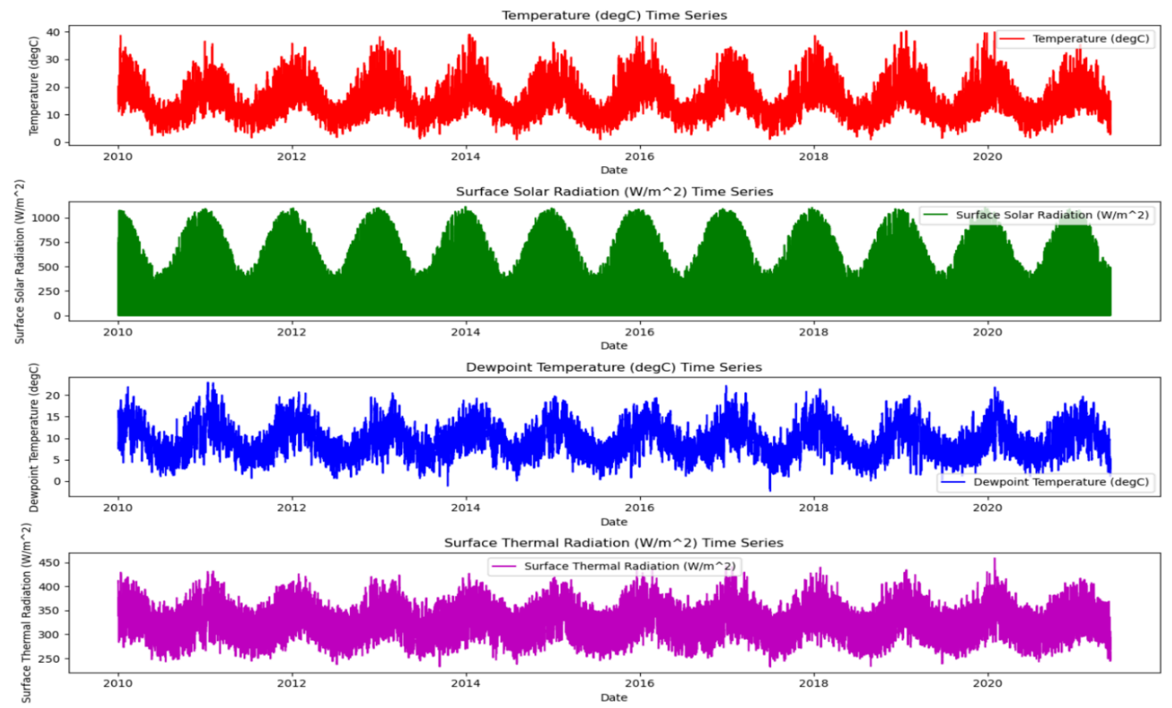

Figure 3 - raw plots of weather data

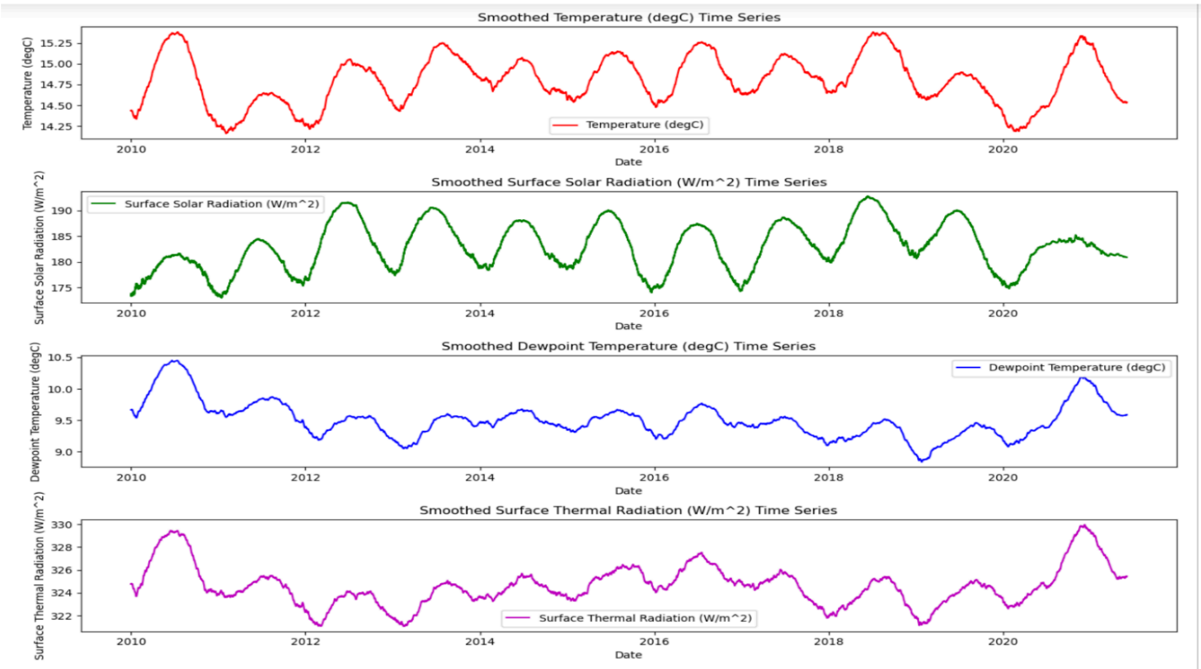

Figure 4 - smoother plots of weather data

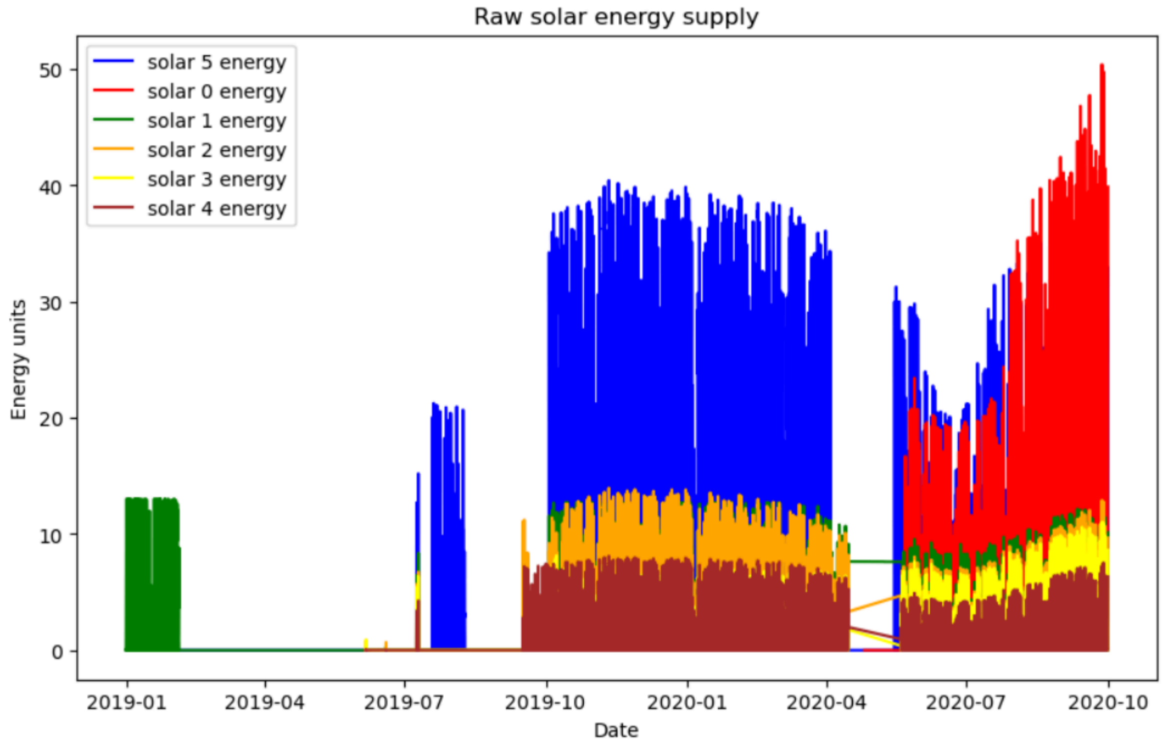

Figure 5 - raw solar energy supply time series

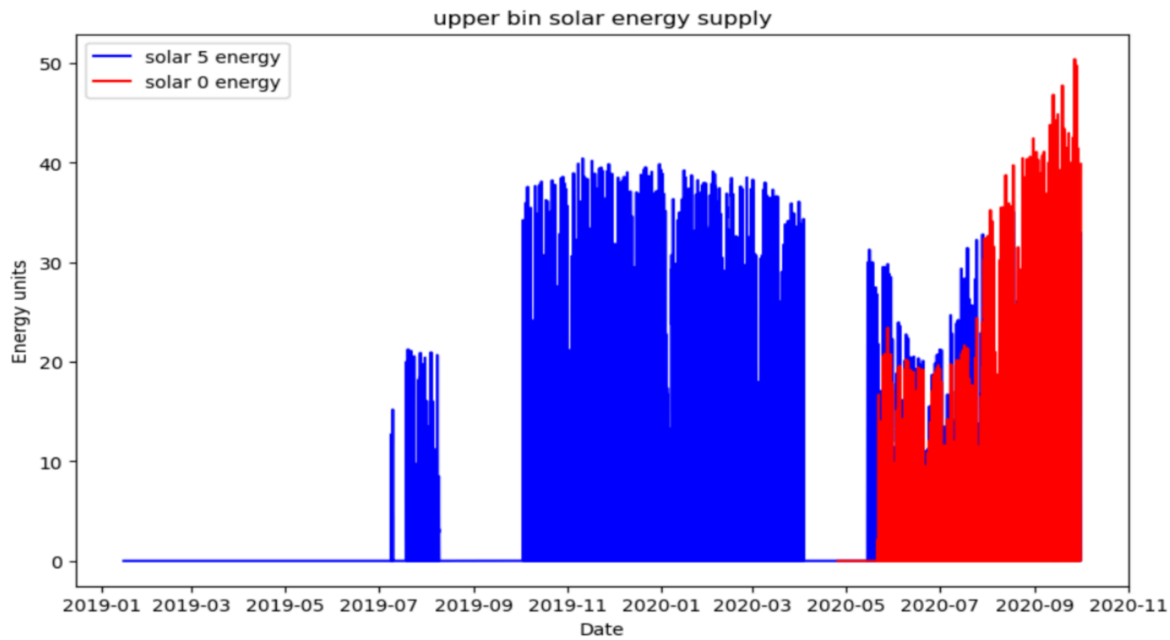

Figure 6 - upper bin solar supply time series

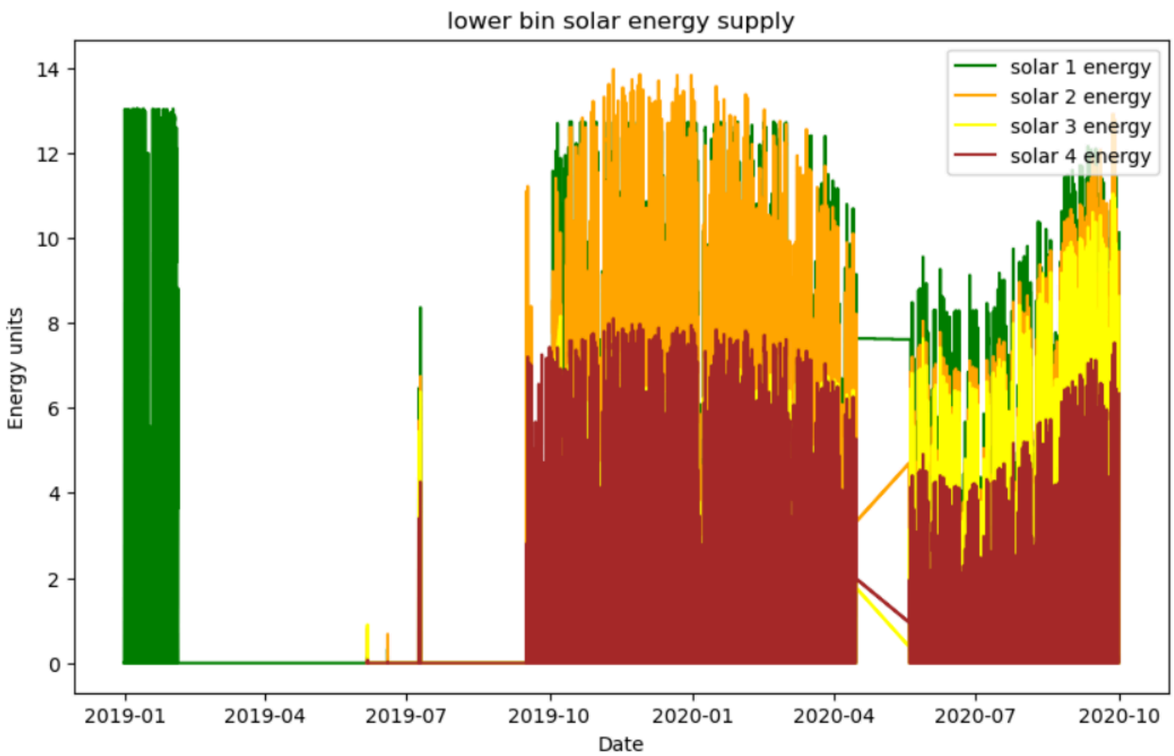

Figure 7 - lower bin solar supply time series

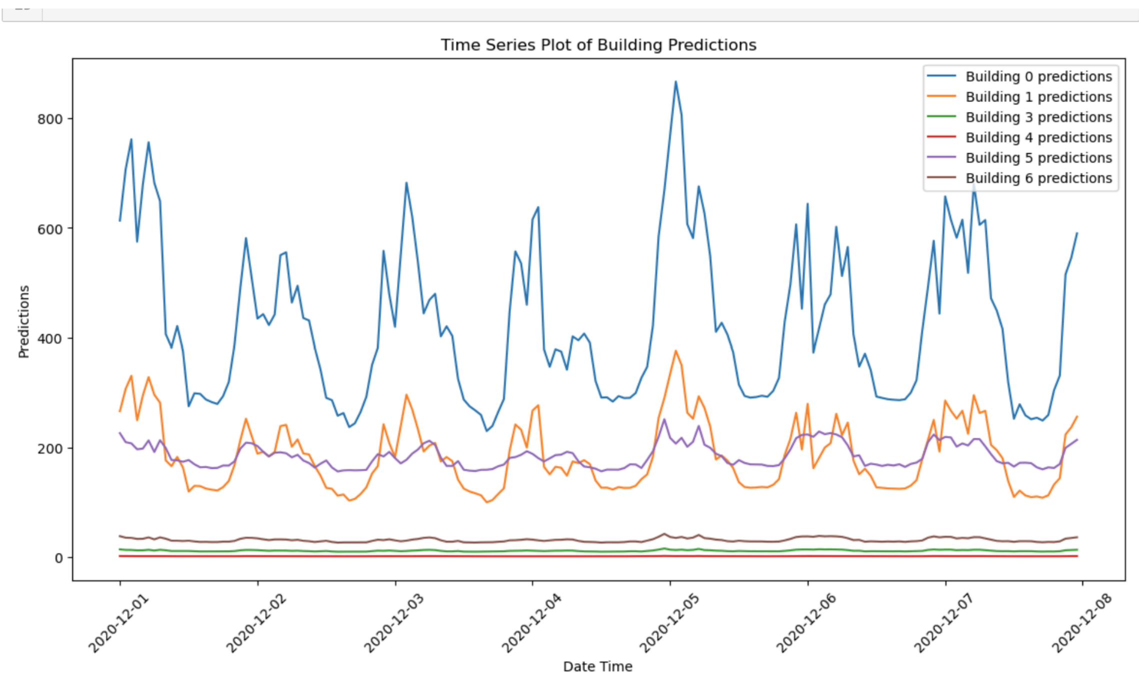

Figure 8 - time series for predicted energy demand

Feature	Importance
surface_solar_radiation (W/m^2)	0.463708
month	0.127764
day	0.064572
mean_sea_level_pressure (Pa)	0.055604
wind_speed (m/s)	0.048377
surface_thermal_radiation (W/m^2)	0.045235
dewpoint_temperature (degC)	0.045024
relative_humidity ((0-1))	0.039606
temperature (degC)	0.038883
hour	0.037516
total_cloud_cover (0-1)	0.033712

Figure 9 - feature importance for solar supply random forest regressor model

```
Choose a variable to predict from the following: temperature (degC), dewpoint_temperature (degC), wind_speed (m/s), mean_sea
_level_pressure (Pa), relative_humidity ((0-1)), surface_solar_radiation (W/m^2), surface_thermal_radiation (W/m^2), total_c
loud_cover (0-1)
temperature (degC)
Enter year: 2024
Enter month: 5
Enter day: 20
Enter hour: 7

C:\Users\bathe\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but Random
ForestRegressor was fitted with feature names
  warnings.warn(

The predicted value for temperature (degC) is 14.075000000000005

1  interactive_predict()

Choose a variable to predict from the following: temperature (degC), dewpoint_temperature (degC), wind_speed (m/s), mean_sea
_level_pressure (Pa), relative_humidity ((0-1)), surface_solar_radiation (W/m^2), surface_thermal_radiation (W/m^2), total_c
loud_cover (0-1)
temperature (degC)
Enter year: 2024
Enter month: 5
Enter day: 20
Enter hour: 9
```

Figure 10 - interactive predictor model

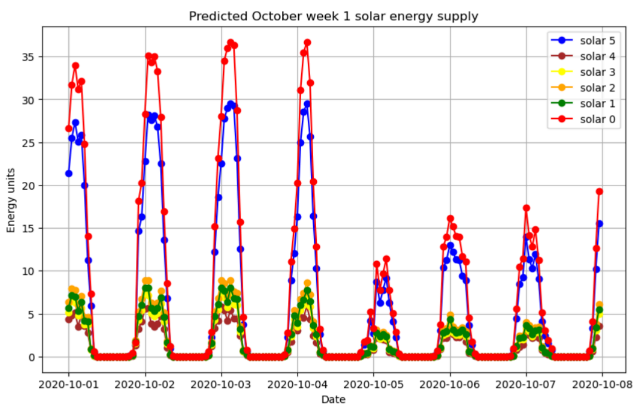

Figure 11 - predicted solar energy supply in week 1 of October 2020

Time	Temperature
October	° C | ° F
2020-10-01	14 | 57.2
2020-10-02	19 | 66.2
2020-10-03	23 | 73.4
2020-10-04	24 | 75.2
2020-10-05	11 | 51.8
2020-10-06	11 | 51.8
2020-10-07	15 | 59.0

Figure 12 - temperature readings in first week of October 2020

```
Mean Squared Error: 30182.178800119244
R^2 Score: 0.5731103219098874
Building 3                        1.000000
surface_solar_radiation (W/m^2)   0.537142
hour                              0.436351
relative_humidity ((0-1))         0.422487
temperature (degC)                0.355528
surface_thermal_radiation (W/m^2) 0.179674
wind_speed (m/s)                  0.095461
dewpoint_temperature (degC)       0.080004
month                             0.078161
day                               0.013116
total_cloud_cover (0-1)           0.000629
model elevation (surface)              NaN
utc_offset (hrs)                       NaN
Name: Building 3, dtype: float64
[0.12044513 0.04759418 0.12346739 0.09111873 0.44759316 0.09847201
 0.07130941]
                             Feature  Importance
4    surface_solar_radiation (W/m^2)    0.447593
2                 temperature (degC)    0.123467
0                               hour    0.120445
5                   wind_speed (m/s)    0.098472
3  surface_thermal_radiation (W/m^2)    0.091119
6          relative_humidity ((0-1))    0.071309
1                              month    0.047594
```

Figure 13 - Feature importance

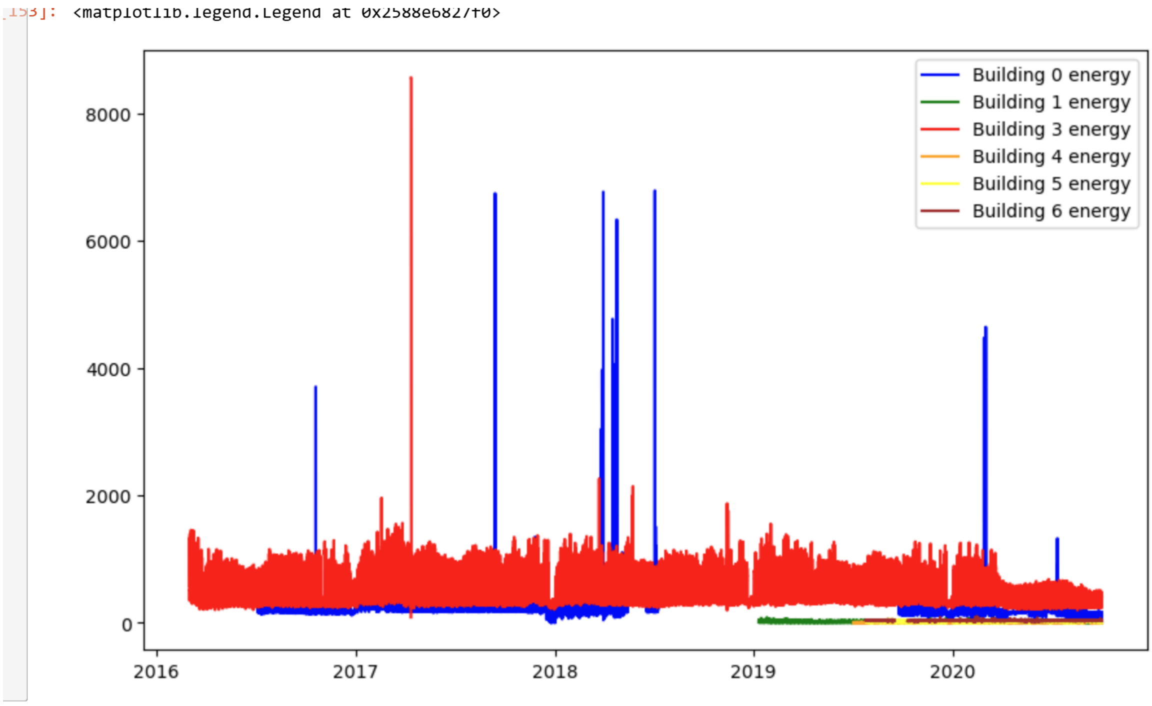

Figure 14 time series for actual energy demand

References

Australian Government, (n.d.). *Solar Panels*. https://www.energy.gov.au/solar/get-know-solar-technology/solar-panels#:~:text=Solar%20panels%20respond%20to%20both,direct%20sunlight%20hits%20the%20panels.

Bergmeir, C. (2022, March 1). *IEEE-CIS technical challenge on Predict+Optimize for renewable energy*. IEEE DataPort. https://ieee-dataport.org/competitions/ieee-cis-technical-challenge-predictoptimize-renewable-energy-scheduling

The Global Historical Weather and Climate Data, (2020). *Melbourne Weather In October 2020.* https://weatherandclimate.com/australia/victoria/melbourne/october-2020

Kolkowska, N (June 23 2023). *Factors affecting solar panel output.* https://sustainablereview.com/factors-affecting-solar-panel-output/