**Forecasting renewable energy assets**

**Abstract**: UTIP21  wind turbine farm contains 3 wind turbines. The sample dataset has power values generated by each turbine captured at one-minute granularity. The columns in the dataset are given below.

1. time
2. power\_turbine1
3. power\_turbine2
4. power\_turbine3

Windfarm (farm level) power measurements are typically taken in at the substation and are a summation of the power produced by all individual turbines of the farm.

**Problem Statement**: UTIP21 is a wind turbine farm situated north of Primrosk city. The power values generated by each turbine are collected at one-minute intervals. Explore the dataset and build a machine learning model to forecast the power of the UTIP21 farm for the next 6 hour from the last available measurement time.

**Requirements:** The dataset is given that contains the power generated by turbines and their capacity is given.

**Softwares and Libraries :** I used the Anaconda navigator Jupiter Notebook . Numpy Scipy Matplot libraries are used.

**Algorithim :** ARIMA(Auto Regressive Integrated Moving Average):

**Steps Followed in project:** I have followed the following steps while doing the project:

1. Setting up the environment
2. Data Preprocessing and Cleaning
3. Data Transformation
4. Forecasting
5. Data Visualization

**1.Setting up Environment:** In this Step I used Anaconda Navigator Jupiter Notebook. In this libraries are already installed. Just b writing import statements we can use in our code.The libraries I used are Numpy, Scipy, Pandas and Matplotlib .

**2.Data Preprocessing and Cleaning:** In every datascience project this is one of the important stage. The data is huge and contains null and corrupt values. I used basic python code to remove null values and corrupt values. Based on requirements I convert the time column to timestamp. Time stamp is already given in data itself and power column to float as there is no double datatype in python.

**3.Data Transformation:** The data that is obtained is not convenient for forecasting so based on requirement I caluculated the sum ofPower of three turbines and reshaped the data in to required format timestamp and Farm power. The file I attached in task 3.

**4.Data Forecasting:** I forecasted the power generated by turbine for next 6 hours by using a machine learning model that is ARIMA(Auto Regressive Integrated Moving Average) algorithim.

**Explaination of Algorithim:**

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses [time series data](https://www.investopedia.com/terms/t/timeseries.asp) to either better understand the data set or to predict future trends.

A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a turbine future power based on its past performance.

* Autoregressive integrated moving average (ARIMA) models predict future values based on past values.
* ARIMA makes use of lagged moving averages to smooth time series data.
* They are widely used in technical analysis to forecast future security prices.
* Autoregressive models implicitly assume that the future will resemble the past.
* Therefore, they can prove inaccurate under certain market conditions, such as financial crises or periods of rapid technological change.

**Understanding Autoregressive Integrated Moving Average (ARIMA):**

An autoregressive integrated moving average model is a form of [regression analysis](https://www.investopedia.com/terms/r/regression.asp) that gauges the strength of one dependent variable relative to other changing variables. The model's goal is to predict future securities or financial market moves by examining the differences between values in the series instead of through actual values.

* [*Autoregression (AR)*](https://www.investopedia.com/terms/a/autoregressive.asp): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
* *Integrated (I):*represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
* [*Moving average (MA)*](https://www.investopedia.com/terms/m/movingaverage.asp): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

## **ARIMA Parameters**

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

* p: the number of lag observations in the model; also known as the lag order.
* d: the number of times that the raw observations are differenced; also known as the degree of differencing.
* q: the size of the moving average window; also known as the order of the moving average.

In a [linear regression](https://www.investopedia.com/terms/n/nonlinear-regression.asp) model, for example, the number and type of terms are included. A 0 value, which can be used as a parameter, would mean that particular component should not be used in the model. This way, the ARIMA model can be constructed to perform the function of an ARMA model, or even simple AR, I, or MA models.

## **Autoregressive Integrated Moving Average (ARIMA) and Stationarity**

In an autoregressive integrated moving average model, the data are differenced in order to make it stationary. A model that shows stationarity is one that shows there is constancy to the data over time. Most economic and market data show trends, so the purpose of differencing is to remove any trends or seasonal structures.

[Seasonality](https://www.investopedia.com/terms/s/seasonality.asp), or when data show regular and predictable patterns that repeat over a calendar year, could negatively affect the regression model. If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made with great efficacy.

## **Special Considerations**

ARIMA models are based on the assumption that past values have some residual effect on current or future values. For example, an investor using an ARIMA model to forecast stock prices would assume that new buyers and sellers of that stock are influenced by recent market transactions when deciding how much to offer or accept for the security.

Although this assumption will hold under many circumstances, this is not always the case. For example, in the years prior to the [2008 Financial Crisis](https://www.investopedia.com/articles/economics/09/financial-crisis-review.asp), most investors were not aware of the risks posed by the large portfolios of [mortgage-backed securities](https://www.investopedia.com/terms/m/mbs.asp) (MBS) held by many financial firms.

During those times, an investor using an autoregressive model to predict the performance of U.S. financial stocks would have had good reason to predict an ongoing trend of stable or rising stock prices in that sector.  However, once it became public knowledge that many financial institutions were at risk of imminent collapse, investors suddenly became less concerned with these stocks' recent prices and far more concerned with their underlying risk exposure. Therefore, the market rapidly revalued financial stocks to a much lower level, a move that would have utterly confounded an autoregressive model.

**5.Data Visualization:** I have visualized the data and drawn the graph for time and forecast for next 6 hours how much turbine will generate. The graph of screenshots I attached in Final deliverables as Google drive link.

**Conclusion:** In this project time series analysis are studied and based on this data I forecasted the future data of 6 hrs at 10min granularity.