

Sketch an analysis for Solar Generation Dataset

Introduction

With the increase of renewable energy generation and decentralized in power grids for instance wind farms and solar plants, dynamical stability of power grids becomes more important and challenging. In order to maintain stability of power grids, the data will be analyse and focus on the prediction of the values.

Description

- The data collected from two solar power plants in India over a 34 day period.
- Each of the plants consists of two datasets which are a power generation dataset recorded by inverters and a weather dataset gathered by sensors at solar panels.
- The task is to predict energy produced daily by solar power plants.

Data source

The dataset can be found and download at **Kaggle** via a topic called **Solar Power Generation Data**. (<https://www.kaggle.com/anikannal/solar-power-generation-data>)

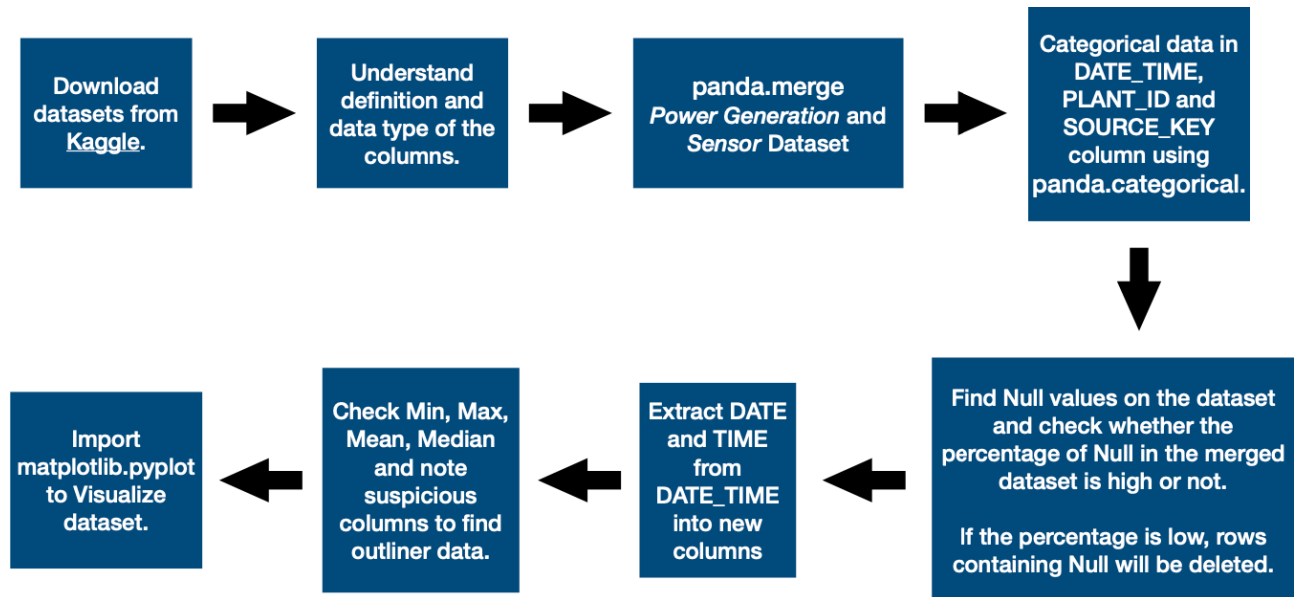
1. Power generation dataset contains seven columns.

Columns	DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
Data Type	String	String	String	Float	Float	Float	Float
Description	Date and time recorded every 15 minutes	Solar power plant ID	Inverter ID	DC power generated by the inverter	AC power generated by the inverter	Energy generated by the inverter on that day until DATE_TIME	Total Energy generated by the inverter

2. Weather Sensor dataset contains six columns.

Columns	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
Data Type	String	String	String	Float	Float	Float
Description	Date and time recorded every 15 minutes	Solar power plant ID	Inverter ID	Ambient temperature at the plant	Solar panels temperature	Amount of radiation which solar panels was exposed to.

Workflow



Visualization

- Scatterplot : Using DAILY_YIELD as X-axis and suspicious columns that we noted earlier as Y-axis to find outliers and delete that rows.
- Correlation Matrix : The graph will show correlation between variables in the data set. It will be used to analyse which data more or less relates to DAILY_YIELD by showing various numbers from -1.00 to 1.00. Then, some columns with high correlation will be picked and utilized as “features” for machine learning process.
- Histogram :
 - Compare data between AC_POWER and TIME of several solar panel to illustrate the trend of generated power within a day and know the time that solar panel generated power at peak.
 - Compare data between AC_POWER and weather variables (IRRADIATION, AMBIENT_TEMPERATURE, and MODULE_TEMPERATURE) makes us understand how the weather electrical generation. Therefore, we might manage to predict power accumulation by using irradiation. Moreover, the power generation can be improved by adjusting solar panel and make solar panel facing more or less radiation from the sun, for instance.