

Global Power Plant Database

Project Description

The Global Power Plant Database is a comprehensive, open source database of power plants around the world. It centralizes power plant data to make it easier to navigate, compare and draw insights for one's own analysis. The database covers approximately 35,000 power plants from 167 countries and includes thermal plants (e.g. coal, gas, oil, nuclear, biomass, waste, geothermal) and renewables (e.g. hydro, wind, solar). Each power plant is geolocated and entries contain information on plant capacity, generation, ownership, and fuel type. It will be continuously updated as data becomes available.

Key attributes of the database

The database includes the following indicators:

-
- ``country`` (text): 3 character country code corresponding to the ISO 3166-1 alpha-3 specification [5]
- ``country_long`` (text): longer form of the country designation
- ``name`` (text): name or title of the power plant, generally in Romanized form
- ``gppd_idnr`` (text): 10 or 12 character identifier for the power plant
- ``capacity_mw`` (number): electrical generating capacity in megawatts
- ``latitude`` (number): geolocation in decimal degrees; WGS84 (EPSG:4326)
- ``longitude`` (number): geolocation in decimal degrees; WGS84 (EPSG:4326)
- ``primary_fuel`` (text): energy source used in primary electricity generation or export
- ``other_fuel1`` (text): energy source used in electricity generation or export
- ``other_fuel2`` (text): energy source used in electricity generation or export
- ``other_fuel3`` (text): energy source used in electricity generation or export
- ``commissioning_year`` (number): year of plant operation, weighted by unit-capacity when data is available
- ``owner`` (text): majority shareholder of the power plant, generally in Romanized form
- ``source`` (text): entity reporting the data; could be an organization, report, or document, generally in Romanized form
- ``url`` (text): web document corresponding to the ``source`` field
- ``geolocation_source`` (text): attribution for geolocation information
- ``wepp_id`` (text): a reference to a unique plant identifier in the widely-used PLATTS-WEPP database.
- ``year_of_capacity_data`` (number): year the capacity information was reported
- ``generation_gwh_2013`` (number): electricity generation in gigawatt-hours reported for the year 2013
- ``generation_gwh_2014`` (number): electricity generation in gigawatt-hours reported for the year 2014
- ``generation_gwh_2015`` (number): electricity generation in gigawatt-hours reported for the year 2015
- ``generation_gwh_2016`` (number): electricity generation in gigawatt-hours reported for the year 2016
- ``generation_gwh_2017`` (number): electricity generation in gigawatt-hours reported for the year 2017

- `generation_gwh_2018` (number): electricity generation in gigawatt-hours reported for the year 2018
- `generation_gwh_2019` (number): electricity generation in gigawatt-hours reported for the year 2019
- `generation_data_source` (text): attribution for the reported generation information
- `estimated_generation_gwh_2013` (number): estimated electricity generation in gigawatt-hours for the year 2013
- `estimated_generation_gwh_2014` (number): estimated electricity generation in gigawatt-hours for the year 2014
- `estimated_generation_gwh_2015` (number): estimated electricity generation in gigawatt-hours for the year 2015
- `estimated_generation_gwh_2016` (number): estimated electricity generation in gigawatt-hours for the year 2016
- `estimated_generation_gwh_2017` (number): estimated electricity generation in gigawatt-hours for the year 2017
- `estimated_generation_note_2013` (text): label of the model/method used to estimate generation for the year 2013
- `estimated_generation_note_2014` (text): label of the model/method used to estimate generation for the year 2014
- `estimated_generation_note_2015` (text): label of the model/method used to estimate generation for the year 2015
- `estimated_generation_note_2016` (text): label of the model/method used to estimate generation for the year 2016
- `estimated_generation_note_2017` (text): label of the model/method used to estimate generation for the year 2017

Fuel Type Aggregation

We define the "Fuel Type" attribute of our database based on common fuel categories.

Prediction : Make two prediction 1) **Primary Fuel** 2) **capacity_mw**

Dataset Link-

- https://github.com/wri/global-power-plant-database/blob/master/source_databases_csv/database_IND.csv

Temperature Forecast Project using ML

Project Description

This data is for the purpose of bias correction of next-day maximum and minimum air temperatures forecast of the LDAPS model operated by the Korea Meteorological Administration over Seoul, South Korea. This data consists of summer data from 2013 to 2017. The input data is largely composed of the LDAPS model's next-day forecast data, in-situ maximum and minimum temperatures of present-day, and geographic auxiliary variables. There are two outputs (i.e. next-day maximum and minimum air temperatures) in this data. Hindcast validation was conducted for the period from 2015 to 2017.

Attribute Information:

For more information, read [Cho et al, 2020].

1. station - used weather station number: 1 to 25
2. Date - Present day: yyyy-mm-dd ('2013-06-30' to '2017-08-30')
3. Present_Tmax - Maximum air temperature between 0 and 21 h on the present day ($^{\circ}\text{C}$): 20 to 37.6
4. Present_Tmin - Minimum air temperature between 0 and 21 h on the present day ($^{\circ}\text{C}$): 11.3 to 29.9
5. LDAPS_RHmin - LDAPS model forecast of next-day minimum relative humidity (%): 19.8 to 98.5
6. LDAPS_RHmax - LDAPS model forecast of next-day maximum relative humidity (%): 58.9 to 100
7. LDAPS_Tmax_lapse - LDAPS model forecast of next-day maximum air temperature applied lapse rate ($^{\circ}\text{C}$): 17.6 to 38.5
8. LDAPS_Tmin_lapse - LDAPS model forecast of next-day minimum air temperature applied lapse rate ($^{\circ}\text{C}$): 14.3 to 29.6
9. LDAPS_WS - LDAPS model forecast of next-day average wind speed (m/s): 2.9 to 21.9
10. LDAPS_LH - LDAPS model forecast of next-day average latent heat flux (W/m^2): -13.6 to 213.4
11. LDAPS_CC1 - LDAPS model forecast of next-day 1st 6-hour split average cloud cover (0-5 h) (%): 0 to 0.97
12. LDAPS_CC2 - LDAPS model forecast of next-day 2nd 6-hour split average cloud cover (6-11 h) (%): 0 to 0.97
13. LDAPS_CC3 - LDAPS model forecast of next-day 3rd 6-hour split average cloud cover (12-17 h) (%): 0 to 0.98

14. LDAPS_CC4 - LDAPS model forecast of next-day 4th 6-hour split average cloud cover (18-23 h) (%): 0 to 0.97
15. LDAPS_PPT1 - LDAPS model forecast of next-day 1st 6-hour split average precipitation (0-5 h) (%): 0 to 23.7
16. LDAPS_PPT2 - LDAPS model forecast of next-day 2nd 6-hour split average precipitation (6-11 h) (%): 0 to 21.6
17. LDAPS_PPT3 - LDAPS model forecast of next-day 3rd 6-hour split average precipitation (12-17 h) (%): 0 to 15.8
18. LDAPS_PPT4 - LDAPS model forecast of next-day 4th 6-hour split average precipitation (18-23 h) (%): 0 to 16.7
19. lat - Latitude ($^{\circ}$): 37.456 to 37.645
20. lon - Longitude ($^{\circ}$): 126.826 to 127.135
21. DEM - Elevation (m): 12.4 to 212.3
22. Slope - Slope ($^{\circ}$): 0.1 to 5.2
23. Solar radiation - Daily incoming solar radiation (wh/m²): 4329.5 to 5992.9
24. Next_Tmax - The next-day maximum air temperature ($^{\circ}$ C): 17.4 to 38.9
25. Next_Tmin - The next-day minimum air temperature ($^{\circ}$ C): 11.3 to 29.8T

You have to build separate models that can predict the minimum temperature for the next day and the maximum temperature for the next day based on the details provided in the dataset.

Dataset Link-

<https://github.com/dsrscientist/Dataset2/blob/main/temperature.csv>

Loan Application Status Prediction

Project Description

This dataset includes details of applicants who have applied for loan. The dataset includes details like credit history, loan amount, their income, dependents etc.

Independent Variables:

1. Loan_ID - This refers to the unique identifier of the applicant's affirmed purchases
2. Gender - This refers to either of the two main categories (male and female) into which applicants are divided on the basis of their reproductive functions
3. Married - This refers to applicant being in a state of matrimony
4. Dependents - This refers to persons who depend on the applicants for survival
5. Education - This refers to number of years in which applicant received systematic instruction, especially at a school or university
6. Self_Employed - This refers to applicant working for oneself as a freelancer or the owner of a business rather than for an employer
7. Applicant Income - This refers to disposable income available for the applicant's use under State law.
8. CoapplicantIncome - This refers to disposable income available for the people that participate in the loan application process alongside the main applicant use under State law.
9. Loan_Amount - This refers to the amount of money an applicant owes at any given time.
10. Loan_Amount_Term - This refers to the duration in which the loan is availed to the applicant
11. Credit History - This refers to a record of applicant's ability to repay debts and demonstrated responsibility in repaying them.
12. Property_Area - This refers to the total area within the boundaries of the property as set out in Schedule.
13. Loan_Status - This refers to whether applicant is eligible to be availed the Loan requested.

You have to build a model that can predict whether the loan of the applicant will be approved(Loan_status) or not on the basis of the details provided in the dataset.

Dataset Link- https://github.com/dsrs Scientist/DSDData/blob/master/loan_prediction.csv