GDG Algiers

Fine Tune Model

DevFest 2020

Description

Given a data set, the aim of this challenge is to obtain **a high accuracy model** with decision trees algorithm to predict **solar radiation** using the dataset provided by NASA.The expected output of this challenge is a well presented notebook that illustrates the process of the models' creations

To pass the Challenge:

1. Create a notebook.
2. Load the data.
3. Preprocess the loaded data.
4. Create and tune the specified model in order to achieve a high accuracy model.
5. Give a **detailed** models' comparison based on their performance on **the same test set.**

Data guide

These datasets are meteorological data provided by NASA from the HI-SEAS weather station from four months (September through December 2016) between Mission IV and Mission V. The dataset's size is 2.82 MB.

❑ *For each dataset, the fields are:*

1. A row number (1-n) useful in sorting this export's results
2. The UNIX time_t date (seconds since Jan 1, 1970). Useful in sorting this export's results with other export's results
3. The date in yyyy-mm-dd format
4. The local time of day in hh:mm:ss 24-hour format
5. The numeric data, if any (may be an empty string)
6. The text data, if any (may be an empty string)

❑ *The units of each dataset are:*

1. Solar radiation: watts per meter^2
2. Temperature: degrees Fahrenheit
3. Humidity: percent
4. Barometric pressure: Hg
5. Wind direction: degrees
6. Wind speed: miles per hour
7. Sunrise/sunset: Hawaii time

Technical Requirements:

1. Use the given dataset.
2. You should split your dataset into a **train**, **dev** and **test** with a ratio of **60/20/20** after shuffling using the random_state = 42 parameter so we all get to work with the same subsets.
3. The result of each treatment should appear in the notebook.
4. The submitted file should have the **.html** extension.
5. The notebook **must be well commented and presented**.