Solar Radiation Prediction

Battery: Solar storage
Vasileios Papadopoulos

Motivation

- Estimate levels of solar radiation
- Why:
    - Solar energy fluctuations
    - Predictability, easier integration to conventional production
- How:
    - Analyze samples from 4 months period (1-9-2016 - 31/12-2016)
    - Machine learning model(s)

The Data

- HI-SEAS weather station, Hawaii
- Collected data/features (32686 rows):
  - Solar radiation [W/m^2]
  - Temperature [F]
  - Atmospheric pressure [Hg]
  - Humidity [%]
  - Wind speed [miles/h]
  - Wind direction [degrees]
  - Time sun rise
  - TIme sun set
- Response variable:
  - Solar radiation [W/m^2]

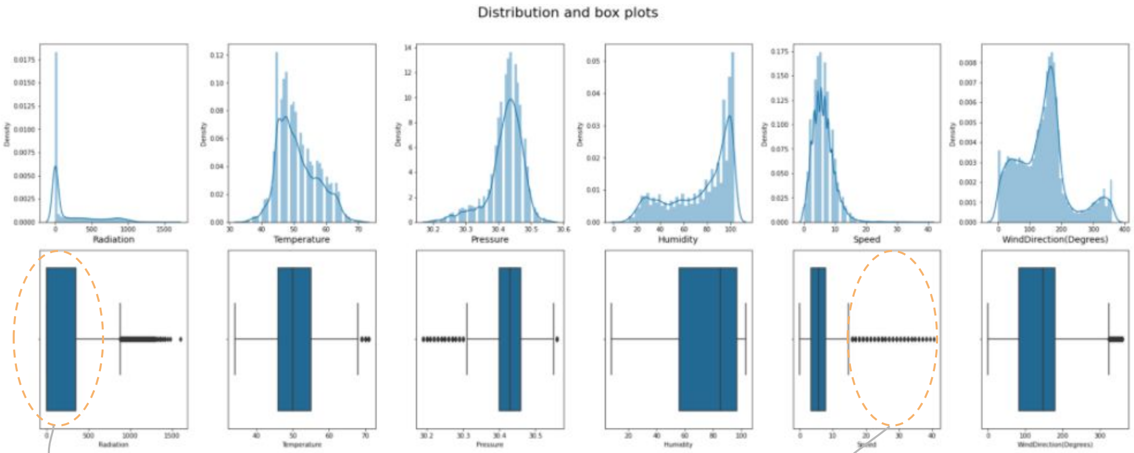
Approach

- Challenge:
  - Time Series
    - Cyclicity
    - Seasonality
- Feature Engineering
  - Deal with Time
- Models
  - Simple Linear Regression
  - Random Forest
  - Gradient Boosting
- Metrics:
  - R-Squared, Mean-Squared Error, CV

Train/Test Split

- Begin Sep. - End Nov.
  - As train set (70%)
- Begin Dec. - End Dec.
  - As test set (30%)
- Training Set, TimeSeriesSplit:
  - 3 Folds to ‘simulate’ seasonality

Explore - Distribution



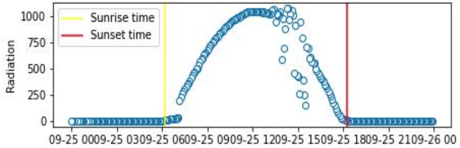
Roughly 50% of values
located between 0 and 250
W/m^2

Extreme outliers

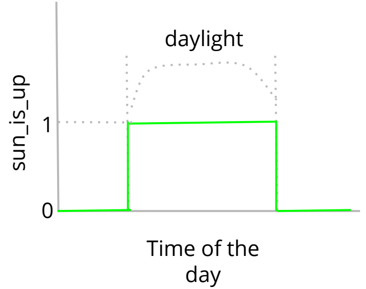
Feature Engineering: Time

2 cases

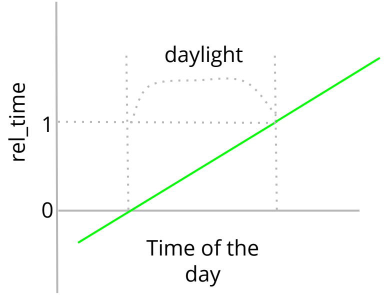
Typical solar radiation dist.



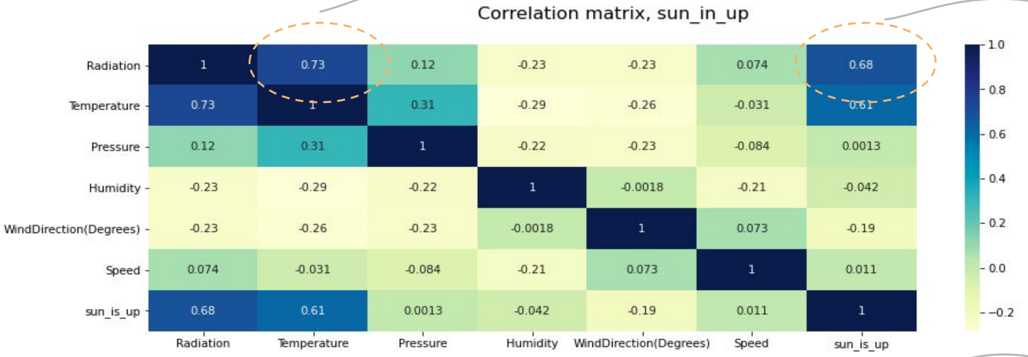
- Sun_is_up feature:
  - 0 - when outside of [SunRise - SunSet]
  - 1 - when during [SunRise - SunSet]
  - Band pass filter, hard cut-off



- Rel_time feature:
  - (Current_time - SunRise_time) / Daylight_duration
    - Daylight_duration = SunRise_time - SunSet_time
      - Rel_time < 0, if before sunrise
      - Rel_time = 0 at exact sunrise time
      - Rel_time (0,1) between sunrise and sunset, ***linear***
      - Rel_time = 1 at exact sunset time
      - Rel_time > 1 if after sunset

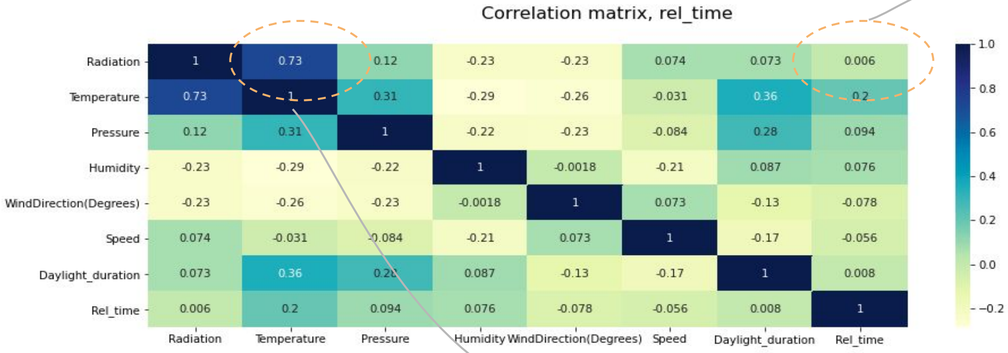


Explore - Correlation



Correlation matrix, sun_in_up

	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	sun_is_up
Radiation	1	0.73	0.12	-0.23	-0.23	0.074	0.68
Temperature	0.73	1	0.31	-0.29	-0.26	-0.031	0.61
Pressure	0.12	0.31	1	-0.22	-0.23	-0.084	0.0013
Humidity	-0.23	-0.29	-0.22	1	-0.0018	-0.21	-0.042
WindDirection(Degrees)	-0.23	-0.26	-0.23	-0.0018	1	0.073	-0.19
Speed	0.074	-0.031	-0.084	-0.21	0.073	1	0.011
sun_is_up	0.68	0.61	0.0013	-0.042	-0.19	0.011	1



Correlation matrix, rel_time

	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	Daylight_duration	Rel_time
Radiation	1	0.73	0.12	-0.23	-0.23	0.074	0.073	0.006
Temperature	0.73	1	0.31	-0.29	-0.26	-0.031	0.36	0.2
Pressure	0.12	0.31	1	-0.22	-0.23	-0.084	0.28	0.094
Humidity	-0.23	-0.29	-0.22	1	-0.0018	-0.21	0.087	0.076
WindDirection(Degrees)	-0.23	-0.26	-0.23	-0.0018	1	0.073	-0.13	-0.078
Speed	0.074	-0.031	-0.084	-0.21	0.073	1	-0.17	-0.056
Daylight_duration	0.073	0.36	0.28	0.087	-0.13	-0.17	1	0.008
Rel_time	0.006	0.2	0.094	0.076	-0.078	-0.056	0.008	1

High positive corr with radiation

Relatively high positive corr with radiation

Practically no corr with radiation, **BUT... see later**

High positive corr with radiation

Models

- Multivariate Linear Regression
- Random Forest
  - 3 folds - Cross Validation (sklearn.timeseriessplit)
  - Randomized best hyper-parameters search
- Gradient Boosting
  - 3 folds - Cross Validation (sklearn.timeseriessplit)
  - Randomized best hyper-parameters search

nr_estimators **=** [100, 300, 500, 800]

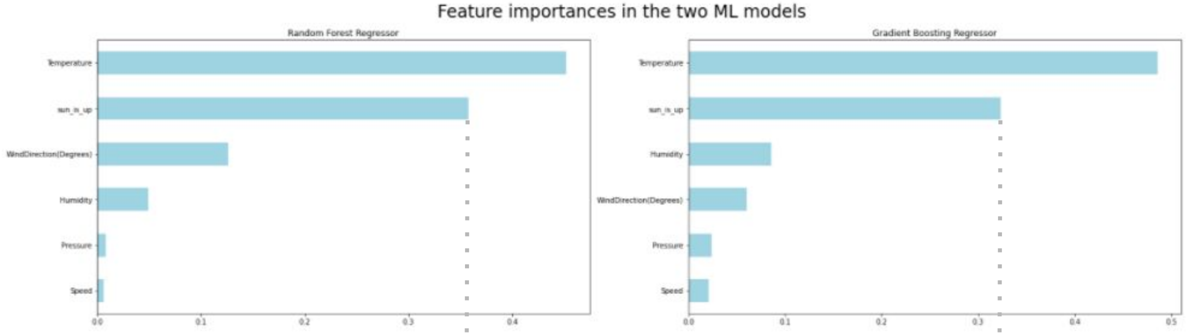
Average to control overfit

Boosting stages, control overfit

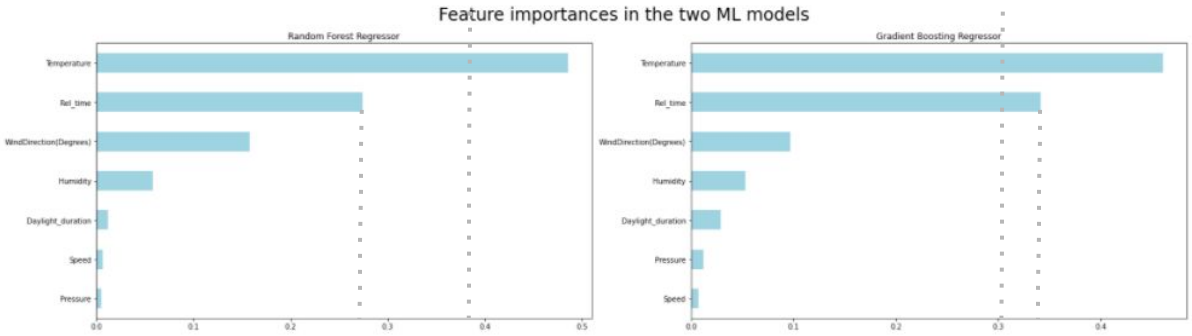
Feature Importance

Gini/Information Gain

sun_is_up feature

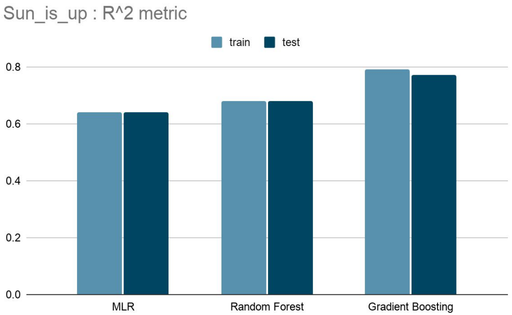


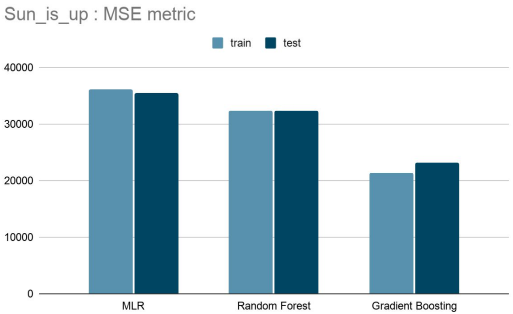
rel_time feature



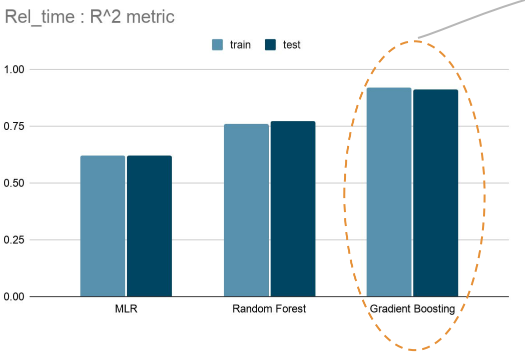
Results - Metrics

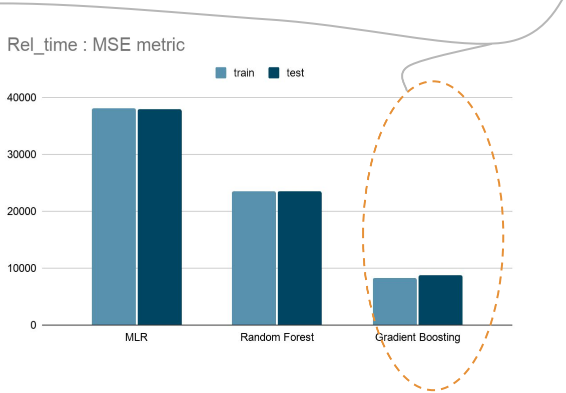
sun_is_up feature





rel_time feature



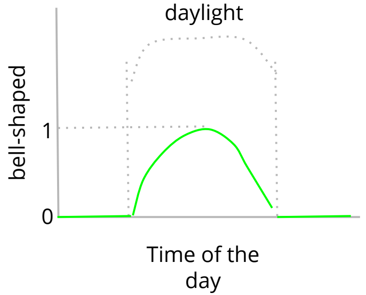


Best combination of model/feature

{
"model": "GradientBoosting",
"feature_case": "rel_time",
"cv_mse": "9792.25",
"train_mse": "8253.57",
"test_mse": "8690.75",
"train_r2": "0.92",
"test_r2": "0.91"
}

Future work

- Feature Engineering
  - Model time within daylight period as bell-shaped
  - Exclude samples outside daylight period
    - Since solar radiation in practically zero during night hours
- Models
  - Try more models
    - **XGBoost**...
    - **SVR**



Questions