**Abstract**

**Introduction**

Very synthetic technical part of pv and arguments for the need of the analysis

**Data**

Description of the data we had, the variables and the length

**Method**

The brief description of the method and its stages.

We can divide it into 3 stages: data exploration, anomaly detection with analysis methods and anomaly detection with ML models. Comparison of the outcomes of the methods.

The approach is to study the data of one plant and create the anomaly detection pipelines for that plant and then apply them to both plants in order to use the plant 2 data as test data and also see how generalisable are our methods and codes.

**Data exploration**

Exploration of weather data and how the 3 quantities are related to each other

* Their behaviour in time and their correlation (a plot of a section of data with the 3 curves?-maybe just 1 plot with also DC and AC)
* Their dependence
* Overplots of the curves for all the different days and considerations of the possibility to use them as a cloudiness measure

Exploration of power generation data

* DC power and its relation to irradiation. Overplots of DC power for a single inverter in time and multiple inverters in one day to show their differing behaviour.
* Considerations about the need to take into account the weather effects and the plot with respect to irradiation. Ideas about the inverter efficiency
* AC power and its relation to DC power. CD vs AC plot to show their straight relationship and the outlier points that will have to be found later

**Anomaly detection with simple data analysis methods**

1. Cloudiness measure
2. Anomalies in the DC power generation

* Method for their derivation and the types of anomalies we are aiming to derive (bullet points below)
* The inapplicability of the linear fit
* The alternative approach to avoid this problem and the derived efficiency values
  + Derivation of the efficiency trend
  + Derivation of the efficiency drops for one inverter
  + Derivation of long and short-term efficiency differences between inverters

1. Anomalies in the AC power generation from DC power generation

* Method for their derivation and the types of anomalies we are trying to derive (point anomalies and large sections of deviated efficiency)
* The analysis details and the conclusions about the AC conversion cut-off
* Derivation of DC to AC conversion efficiencies and their application to derive anomalies like in the above list
  + Derivation of the efficiency trend
  + Derivation of the efficiency drops for one inverter
  + Derivation of long and short-term efficiency differences between inverters

Note about the fact that we will not focus on this part in the following section

**Anomaly detection using ML models**

Describe the variables that will be used and why. Include only DC power to study the DC generation anomalies only.

Why we use the following two methods

* Use of simple regression models, the anomaly derivation
* Use of autoencoders. Description of autoencoders and the error derivation

How are the derived anomalies used to detect the 3 types of inefficiencies above. Basically using the same three function we created for trend, efficiency drops and inter/inverter differences.

**Method for performance comparison**

* Manual by comparing the points and anomaly origin
  + The points selected as outliers
  + The points selected as days with inverter low efficiency
  + The inverters selected as inefficient
* Calculate coefficients (Optional)
  + Recall and precision

**Results**

General insights that we got from the data

* The DC power what is it dependent on
* The AC power, what is it dependent on

Outcomes of anomalies and their comparison between the two methods

**Conclusions**

Should also contain some considerations of costs?