Meie meeskond:

* …

Mis on meie ülesanne:

* Anda jaama omanikule ülevaade, millised inverterid vajavad hooldust/parandust
* *Võiks olla ka DC POWERi ehk tootlikkuse ennustamine*

Mis on meie blocker:

* Meil on kaks erinevat jaama ja me ei tea, mis põhjustab seda erinevust
* See mõjutab mudelite täpsust
* *Lihtne regressioonimudel ei tule jaama ega inverteri eripärade arvestamisega toime*

Mis oleme sellega siiamaani teinud:

* Oleme loonud arusaama andmetest
  + Suhe kiirguse ja Alalisvoolu vahel
  + Sama suhe ajajoonel - mis juhtub ‘kehvade tulemustega’ jaamas
* Linear Regression baseline model
* Lõikame arusaamatud andmepunktid välja
  + Linear Regression Threshold
  + KMeans clustering
* *Jaama eripärade arvestamisega võivad paremini toime tulla keerukamad mudelid nagu XGBoost - lisaks võivad LSTM-id kasutada ära andmestiku ajalist (time series) iseloomu*

Mis meil tulevikuks plaanis on?

* Kui saame andmed soovitud kujule, rakendame
  + XGBoost
  + LSTM
* Ilus mudel
  + Iga inverteri kohta annaks hinnangu - töötab/ei tööta

In your presentation:

* introduce **your team** and project owner (if applicable);
* briefly **describe the problem** you are trying to solve (say why it needs to be solved);
* mention **progress** you have managed so far;
* tell about your **blockers**/**problems**;
* Please, specify who in your team is responsible for which part of the work;
* lastly, say a few words about **future steps** (what you will accomplish for the final presentation).

**Who are we?**

* …

**What problem are we solving?**

* The panels are in a remote area
  + Difficult to check for the need for maintenance
  + Difficult to check for faulty panels/inverters
* Difficult to predict the power generation

**What progress have we made?**

* We have gone through the data (finding and fixing the flaws, figuring out the data correspond to reality (written != reality)).
* We have visualised what it means for solar panels to be faulty
* We have successfully implemented K-means clustering to divide inverters into “Good” and “Faulty.”
* We have implemented linear regression to predict power generation from incoming sun radiation

**What has been keeping us back?**

* Understanding the data and faults in the data
* Not having specific goal or result to aim (how should we implement the models we make in real life?)

**What are our future steps?**

* If the data is in the correct shape to implement XGBoost and other more complex algorithms
* To implement more techniques we have covered in lectures.

|  | Binary classification task - Lauri and Martin | Continuous value prediction task - Gustav | Overall - Work with Solar Power Plant Dataset |
| --- | --- | --- | --- |
| Problem | Identify faulty equipment | Predict power generation |  |
| Challenges | No labels for equipment health in the dataset | Dataset contains data for faulty equipment | Dataset contains flaws,  Understanding each other’s tasks and goals |
| Progress | Use supervised learning,  Visualize data | Exclude faulty data  Train LR model | Correct data, Interpret data, Visualize data |
| Future | Label data manually,  Train supervised classification model | Add equipment health labels to dataset  Train a more sophisticated model | Come to an understanding of what each member of the team is doing and what are our common goals |

**PRESENTATION MUSTAND**

**TASK 1**

**Speech**: One of our tasks is to identify faulty equipment. This is a classification task that Lauri and Martin have been working on. We now understand what faulty equipment looks like by looking at the data.

**Text on slide**: What does faulty equipment look like?

***Slide images****: 1) IR and DC on sunny day,*

*2) cloudy day and*

*3) for faulty inverter*

**Speech:** This is what a piece of healthy equipment looks like on a sunny day. You can see a strong correlation between the amount of radiation coming in and power going out. This is healthy equipment on a cloudy day. You can see the clouds here, and here the clouds go away. Now this is unhealthy - radiation is going up, and power generation plummets, and we don’t know why. Clearly, this piece of equipment needs to be looked at.

**Text on slide**: Challenge 1 - No labels for equipment health

**Speech:** The challenge we face with this task is that the dataset does not contain labels for equipment health. So there are two possible solutions.

**Text on slide**: Solution 1 - Cluster equipment by **radiation** and **power**

***Slide image:*** *Lauri klasterdus-algoritmi tulemus*

**Speech:** For the first solution, Lauri has successfully clustered the equipment into healthy and faulty groups using K-Means clustering.

**Text on slide**: Solution 2 - Label the data manually

***Slide image:*** *Martini IRRADIATION-DC\_POWER grid-joonised*

**Speech:** As a second solution, we could use these beautiful graphs that Martin generated to pick out equipment that we deem unhealthy and label the dataset by hand. This approach might give us more certainty in the labels and, since the dataset is not too big, it will not be very time consuming.

**TASK 1**

**Speech**: Our second task, that Gustav is working on, is predicting power output from incoming radiation. Thus far we have a Linear Regression model that works pretty well, but only if the equipment is healthy and operational. This brings us to the main challenge of the prediction task which is faulty equipment.

**Text on slide**: Dataset contains faulty equipment data

***Slide images****:*