AI Hackathon ISEP X GarageISEP

29 - 30 - 31 March 2019

Technical Challenge

Challenge presentation

Agriculture is a field that is very much affected by climatic events: heat, cold, frost, rain, hail, wind. These events will:
Impact crops by causing crop losses, damage to fruits, sometimes for several years (impact of hail on vines, can be felt the following year, with fewer buds)

Desolation scene after a hail storm on the vine.

Impact farming operations: the wind limits the possibilities of treatment, rain can make the soil impassable for agricultural machinery.

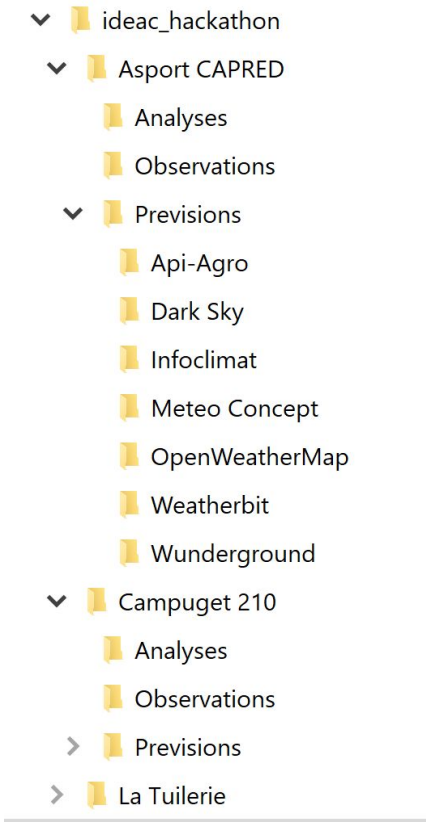
Tractor seriously stuck

The farmer is therefore a very assiduous consumer of weather forecasts. Aware that the latter have limited reliability, he will consult 3, 4, sometimes more, sources of weather forecasts and arbitrate between sources in a subjective way. These sources do not necessarily always perform in the same way. Validations may show that some weather forecast data providers will sometimes outperform (i.e. more accurate forecasts when compared to measurements made by weather stations), others underperform, depending on the period, type of climate, area.
In addition, many weather forecasting providers now offer APIs to make their data interoperable with agricultural decision-making tools. It is therefore relevant to use different weather forecast sources to improve decision-making, but how? The aim is to improve decision-making within a few hours, so the focus will be on immediate forecasting. The most interesting parameters in agriculture are temperature, especially for extreme cold events (frost at a sensitive stage of the crop which can destroy flowers and therefore future fruits in arboriculture or vines), rainfall, which makes it possible to control crop irrigation, wind which constrains cultural interventions and can also cause crop lodging.

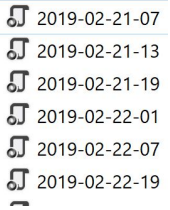
The dataset

You will have weather forecast data from different weather forecasting providers. These data were collected using NIFI (Apache) during the months of February and March 2019. We focus in a small area located in the Gard department, where we have 4 weather stations collecting measured data (observations) and on which we will also collect forecast data from 7 different suppliers.

This data is available in a tree structure as described below, in an easily usable JSON format.



There are stations : Asport CAPRED, Campuget, La Tuilerie et MASDEMIAN.
The Forecast directory contains the successive forecasts retrieved from the station point for the 7 forecasters.



We retrieve the forecasts four times a day at 1:00, 7:00, 1:00, 7:00, 7:00. They may or may not have changed depending on the forecast provider.


```
{
  "time": 1550883600,
  "temperature": 11,
  "precipitation": 0,
  "wind": 5,
  "wind_dir": 26,
  "pressure": 1028,
  "humidity": 73,
  "cloud_cover": 0
},
{
  "time": 1550887200,
  "temperature": 10,
  "precipitation": 0,
  "wind": 5,
  "wind_dir": 26,
  "pressure": 1028,
  "humidity": 74,
  "cloud_cover": 0
},
{
  "time": 1550890800,
  "temperature": 9,
  "precipitation": 0
```


Example of a forecast made on February 22, 2019 at 7 p.m. for the supplier Wunderground: for a given time stamp, corresponding to a forecast time for temperatures (°C), precipitation (mm), wind speed (m/s), wind direction (hourly degrees with 0 to the north), pressure (hPa), relative humidity (%) and cloud cover (%).

Be careful, not all suppliers necessarily provide the same things:
-the scope of the forecast (number of hours) is different, but we have limited it to 2 days so that it is homogeneous and because we are primarily interested in forecasts for the next 12 to 24 hours.
-the time step of the forecast may differ, we have suppliers who provide information for each hour of the forecast, others every three hours.
-the parameters recorded may differ: not all suppliers produce the same parameters.

Analysis files are also available.

The latter is an example of the very simple kind of processing that can be done to exploit forecasts from multiple suppliers. They summarize the differences observed between suppliers for a given time step and a given forecast:


```
{
  "time": "2019-02-20-18",
  "temperature": [
    11.13,
    11,
    10.420923,
    10,
    10.9
  ],
  "source": [
    "dark_sky",
    "wunderground",
    "api-agro",
    "meteo_concept",
    "weatherbit"
  ],
  "min": 10,
  "moy": 10.690185,
  "max": 11.13
}
```



```
{
  "time": "2019-02-20-18",
  "temperature": [
    11.13,
    11,
    10.420923,
    10,
    10.9
  ],
  "source": [
    "dark_sky",
    "wunderground",
    "api-agro",
    "meteo_concept",
    "weatherbit"
  ],
  "min": 10,
  "moy": 10.690185,
  "max": 11.13
}
```


Content of the analysis file : 2019-02-20-13_2019-02-20-18

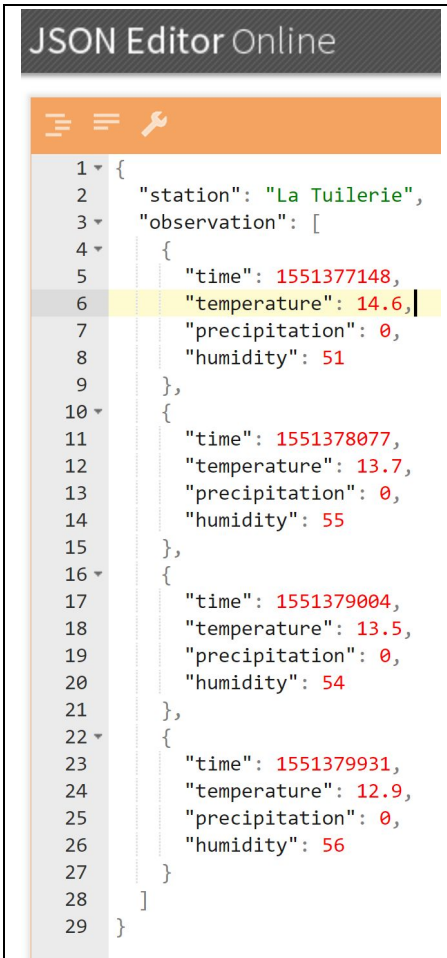
Here we see the temperatures forecast by the various suppliers for 20/02/2018 at 6 pm. The forecast is dated 20/02/2019 at 1:00 p. m.
Suppliers here, were in relative agreement with each other. This is not always the case! Identifying significant differences between suppliers may be a sign of a particular weather event.

what forcast

Finally, we have observation files. From the four weather stations in the area, these files make it possible to verify how a forecaster performed during a forecast by comparing the forecasted values with the measured values. Obviously, this requires waiting until the observations are available for the planned time step.

Since we are working on past forecast data, these observations can easily be used to validate and evaluate a forecaster's performance at a given time.

Warning: weather stations may have failures, so the observation data sets are more or less complete depending on the station considered.



Here we have the data of the 28/02/2019 at 8pm. Each hour is the statement of a JSON file and contains 1 to 4 readings depending on data availability (one statement every ¼ hour maximum)

<u>Potential subjects:</u>

1. Can we find signs of the arrival of a given event (e.g. low temperature, below 2°C, rain above 5 mm, other...) by analyzing forecasts and observations.

2. Can we determine with good certainty the most reliable forecast based on past forecasts and observations?
3. Can we integrate observation data into predictive models, but also forecast errors?
4. Propose AI models that can be applied to this data and that are of interest to farmers?
5. …