**Cerinte Laborator 2**

**15 puncte**

* Identificarea setului de date: [2019 - OneDrive (sharepoint.com)](https://ubbcluj-my.sharepoint.com/personal/adriana_coroiu_ubbcluj_ro/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fadriana%5Fcoroiu%5Fubbcluj%5Fro%2FDocuments%2FArhiva%202020%2D2022%2FdiverseUBB%2Fmediu%2Fdate%2F2019&ga=1)
* Colectarea setului de date: Nu e cazul
* Intelegerea setului de date: Momentul masurarii, UID-ul aparatului, ce masoara (Radon/CO2, dar se afla si in nume), valoarea masurata si date pe care nu le vom folosi (prag pornire, prag oprire si alarma)
* **Analiza exploratorie a datelor folosite ca input pentru proiect**

(mai multe detalii aici: [2023\_\_Analiza Exploratorie a datelor.pdf](https://ubbcluj.sharepoint.com/:b:/s/ModeledeInteligenaArtificialnSchimbareaClimatic2022-2023/EazhPcxvrMpOpQPp5guGmAEBmG9GAwbv4VsfwxqsUBYZkA?e=GIkjUq) )

* + **Description of data** 
    - results from the *describe()* function from python (this)
  + **Handling missing data**
    - drop NULL or missing values
    - fill Missing Values (this for now)
    - predict Missing values with an ML Algorithm
  + **Handling outliers**
    - BoxPlot
    - Scatterplot (this for now)
    - Z-score
    - IQR (Inter-Quartile Range)
  + **Understanding relationships and new insights through plots**
    - Histogram (this for now)
    - Heatmap
* **Preprocesarea setului de date**

(mai multe detalii aici: [2023\_Preprocesarea datelor.pdf](https://ubbcluj.sharepoint.com/:b:/s/ModeledeInteligenaArtificialnSchimbareaClimatic2022-2023/EdIMwjNuAThAv0Bt_CNTgF0BxdGJU12QQ2JmRIDCBlzlGg?e=7yass7) )

* + **Standardization**
    - StandardScaler (this)
    - MinMaxScaler
    - RobustScaler
  + **Normalization** (L1 or L2) (we used L1)
  + **Discretization (yes)**
  + **Encoding categorical features**
  + **Features engineering**
    - Features selection
    - Features extraction

We analysed what we considered would be useful for future modelling.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

A screen shot of a graph

Description automatically generated

A screenshot of a table

Description automatically generated

In the last image we have the final table (form) that we are going to use for modelling. Because of the dataset that we have little information about, we are still in discussion with the teacher to find out more about the columns.