MACHINE LEARNING

Abstract:
The data comprises the responses from a gas multisensor device deployed in an urban area in Italy. Hourly averages of the sensor responses are documented alongside gas concentration readings obtained from a certified analyzer.

Problem Statement:
The air quality in urban environments has a significant impact on public health and the environment. In this project, we aim to identify the relative humidity (RH) levels in an Italian city, using data from a gas multisensor device deployed on the field. The dataset contains hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device, along with ground truth hourly averaged concentrations for CO, Non-Metanic Hydrocarbons, Benzene, Total Nitrogen Oxides (NOx), and Nitrogen Dioxide (NO2) provided by a co-located reference certified analyser. Our goal is to classify the RH levels into one of the five categories: "Dry", "Ideal", "Slightly Elevated", "Elevated", or "High".

Relative humidity plays a crucial role in indoor and outdoor air quality. High humidity levels can lead to the growth of mold and bacteria, while low humidity levels can cause dry skin and respiratory discomfort. Identifying the appropriate RH levels for the environment is essential for maintaining optimal air quality and reducing the risk of health problems associated with poor air quality.

Therefore, the ability to accurately classify RH levels based on sensor data can be a valuable tool for environmental monitoring and public health.

Attributes of the dataset are:

Sl No	Attribute	Description
1	Date	Date (DD/MM/YYYY)
2	Time	Time (HH.MM.SS)
3	CO(GT)	True hourly averaged concentration CO in mg/m^3 (reference analyzer)
4	PT08.S1(CO)	PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)
5	NMHC(GT)	True hourly averaged overall Non Metanic HydroCarbons concentration in microg/m^3 (reference analyzer)
6	C6H6(GT)	True hourly averaged Benzene concentration in microg/m^3 (reference analyzer)
7	PT08.S2(NMHC)	PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)

8	NOx(GT)	True hourly averaged NOx concentration in ppb (reference

		analyzer)
9	PT08.S3(NOx)	PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NOx targeted)
10	NO2(GT)	True hourly averaged NO2 concentration in microg/m^3 (reference analyzer)
11	PT08.S4(NO2)	PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO2 targeted)
12	PT08.S5(O3)	PT08.S5 (indium oxide) hourly averaged sensor response (nominally O3 targeted)
13	T	Temperature in °C
14	RH_type	RH category
15	AH	AH Absolute Humidity
16	ID	ID of the input