

# Predictia Valorilor PM2.5 pentru zona Ilfov

## Regresie Liniara

### Cuprins

1. Introducere
2. Metodologie
3. Rezultate obtinute
4. Concluzie

## **1. Introducere**

### **Motivatie**

Calitatea aerului afecteaza sanatatea publica prin cresterea riscului de afectiuni respiratorii, de aceea am realizat un studiu pentru monitorizarea valorilor acesteia, influentata de PM2.5 (Fine particulate matter) in zona Mogosoaia, cea mai afectata, fiind la periferia orasului Bucuresti. Studiul isi propune monitorizarea valorilor PM2.5 si utilizarea regresiei liniare pentru predictii relevante.

### **Algoritm ales**

Am folosit o regresie liniara care estimeaza dependenta valorilor PM2.5 in decursul unei saptamani. Acesta are scopul de a gasi natura relatiei dintre diferite variabile si de a facilita antrenarea rezultatului.

Pe langa acest algoritm, am mai folosit un program care va extrage datele de pe o pagina Chrome controlata automat de programul software realizat - `webdriver.Chrome()`.

## **2. Metodologie**

### **Implementare**

Datele au fost colectate prin intermediul bibliotecii Selenium, folosita pentru testarea web a framework-ului robot utilizat. Site-ul de pe care au fost preluate datele este: <https://www.iqair.com/romania/ilfov/mogosoia>, care mai apoi au fost salvate intr-un fisier CSV. Ca programul sa functioneze, a trebuit sa adaug o secventa de cod care sa-mi deruleze prin pagina pana sa gaseasca informatiile relevante, iar mai apo isa faca un screenshot.

De asemenea, modelul scikit-learn, a urmat pasii propriu-zisi:

- 1.Separarea datelor in seturi de antrenament
- 2.Antrenarea modelului pe acest set de testare
- 3.Evaluarea performantei, calculand MSE
4. Evaluarea unor predictii pentru PM2.5, adaugand valori viitoare

## Particularitati

Modelul presupune o evolutie in timp a valorilor PM2.5, insa isi propune sa ofere anumite predictii, raspunzand la intrebarea: “Care ar fi valorile normale pentru aceasta perioada, in functie de temperatura sau umiditatea?”, deoarece factori externi, precum poluarea industriala si gazele de esapament, afecteaza calitatea aerului.

## 3. Rezultate obtinute (si exemple de calcul)

- *Datele sunt preluate in ziua de 19.12.2024*
- *Alaturi avem screenshot-ul de pe care s-a bazat programul meu*

Daily forecast							3d	7d
Mogosoia air quality index (AQI*) forecast								
Today	112		10°	3°	▼ 10.8 km/h		61%	
Fri	86		9°	5°	▼ 10.8 km/h		55%	
Sat	60		7°	4°	▲ 25.2 km/h		64%	
Sun	62		6°	2°	▼ 14.4 km/h		69%	
Mon	67		5°	1°	▲ 14.4 km/h		57%	
Tue	64		5°	3°	▲ 14.4 km/h		97%	
Wed	72		7°	6°	▲ 39.6 km/h		82%	

## Performanța modelului

```
Eroarea Pătratică Medie (MSE): 5864.400000000001
Predictii pentru următoarele 7 zile: [72.8 75.4 78. 80.6 83.2 85.8 88.4]
```

Eroarea Patratica Medie, calculate de program, nu este in parametrii corespunzatori, aceasta initial fiind: **MSE = 828.98**, masurand diferenta medie patratica dintre valorile reale si cele prezise.

Am folosit ecuatia de regresie liniara( $Y = aX+b$ ) pentru a calcula:

- Coeficientul regresiei: -13.28 (a)
- Interceptul regresiei: 143.57 (b)

#### **Coeficientul de determinare**

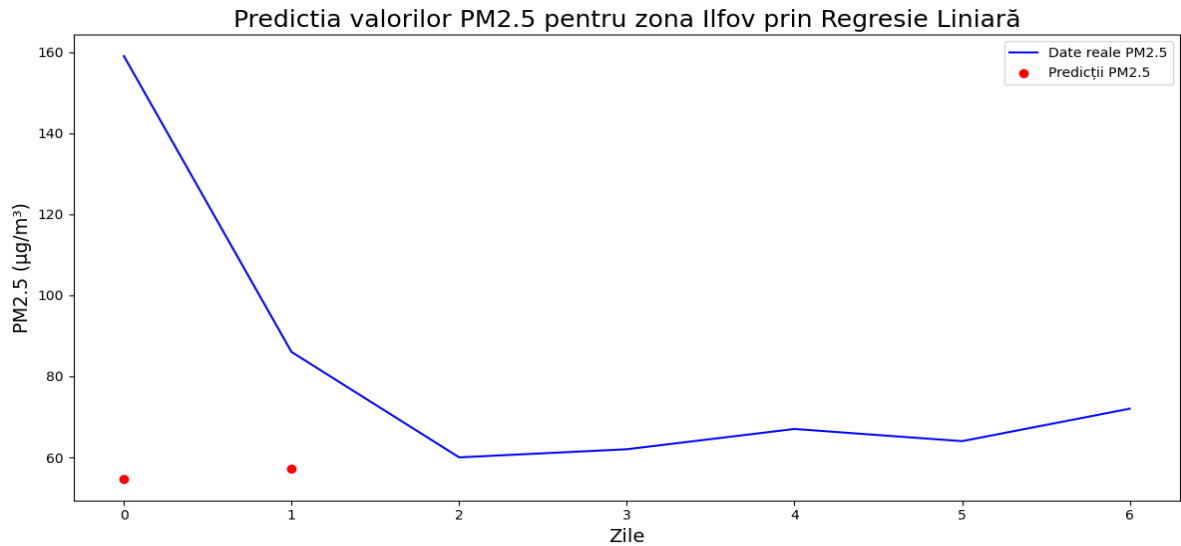
- Suma pătratelor reziduurilor
- Suma pătratelor deviațiilor
- Media valorilor reale

#### **Exemplu calcul:**

Modelul explica 89% din variatia valorilor PM2.5.

Predicții pentru următoarele 7 zile (pe baza datelor existente)

- Ziua 1: 143.57  $\mu\text{g}/\text{m}^3$
- Ziua 2: 130.29  $\mu\text{g}/\text{m}^3$
- Ziua 3: 117.01  $\mu\text{g}/\text{m}^3$
- Ziua 4: 103.73  $\mu\text{g}/\text{m}^3$
- Ziua 5: 90.45  $\mu\text{g}/\text{m}^3$
- Ziua 6: 77.17  $\mu\text{g}/\text{m}^3$
- Ziua 7: 63.89  $\mu\text{g}/\text{m}^3$



Graficul rezultat (punctele rosii sunt valorile prezise)

## 4. Concluzii

### Observatii

- ❖ Datele preluate justifica faptul ca PM2.5 depaseste limitele OMS, sugerand o poluare semnificativa
- ❖ Modelul estimeaza in regula valorile, insa prezinta anumite fluctuatii bruste, fiind imprevizibile, dat fiind situatia actuala, cu gradul ridicat de poluare
- ❖ In perioada urmatoare, predictiile arata o scadere a PM2.5, dat fiind temperaturile scazute si umiditatea ridicata
- ❖ Programul nu functioneaza eficient, din cauza unor valori aparute in tabelul de pe site (cantitatea de precipitatii care nu este specificata in fiecare zi), ceea ce a dus la perturbatia extragerii valorilor (de aceea am unit temperatura si viteza vantului intr-o singura coloana).

### Recomandari

- ❖ Extinderea datelor pentru o mai bună acuratețe.
- ❖ Utilizarea unor modele avansate (ex: regresie polinomială).
- ❖ Integrarea factorilor adiționali (ex: trafic, industrie).

### Surse

- [https://scikit-learn.org/1.5/auto\\_examples/index.html](https://scikit-learn.org/1.5/auto_examples/index.html)
- <https://robotframework.org/SeleniumLibrary/SeleniumLibrary.html>

- <https://www.nature.com/articles/s41598-023-49717-7>
- <https://chatgpt.com/>
- <https://www.iqair.com/romania/ilfov/mogosoia>