# Analiza calității apei

Călugăru Olga

18.12.2023

**Abstract**

# Introducere

Într-o lume unde resursele de apă dulce devin din ce în ce mai prețioase și vulnerabile la poluare, monitorizarea și asigurarea calității apei potabile devine o necesitate urgentă. Calitatea apei influențează sănătatea populației, ecosistemele acvatice și eficiența proceselor industriale. Setul de date "waterQuality1.csv" oferă o perspectivă valoroasă asupra acestei teme, prin analiza a diferiți parametri ce indică prezența de metale grele, compuși organici și alte substanțe nocive în apa de consum.

Datasetul cuprinde 7999 de înregistrări, fiecare reflectând concentrația diferitelor substanțe precum aluminiu, amoniac, arsen, bariu, cadmiu, cloramini, crom, cupru, fluorură și altele în mostre de apă. De asemenea, sunt monitorizați indicatori cum ar fi bacteriile, pH-ul, conductivitatea, turbiditatea și alții, care sunt esențiali în evaluarea calității apei.

Analiza inițială relevă un interval larg de variație pentru fiecare parametru. De exemplu, concentrațiile medii observate sunt de 0.666 mg/L pentru aluminiu și 9.819 mg/L pentru nitrați, indicând o posibilă variație în sursele de apă sau în procesele de purificare.

Mai mult, setul de date include și o coloană 'is\_safe', care indică dacă apa este considerată sigură pentru consum, bazându-se pe standardele stabilite. Aproximativ 88.6% din înregistrări indică apă nesigură pentru consum, ceea ce subliniază importanța continuării și aprofundării acestei analize.

Această analiză oferă o bază pentru înțelegerea mai profundă a stării calității apei și a factorilor care influențează această calitate. Prin aplicarea unor tehnici statistice și de data mining, putem identifica modele și tendințe care ne ajută să înțelegem mai bine impactul activităților umane asupra resurselor de apă și să dezvoltăm strategii eficiente pentru gestionarea și îmbunătățirea calității apei.

# Expunerea Problemei

**Nitratii în Apa Potabilă: O Problemă Invizibilă cu Impact Vizibil**

Atunci când vorbim despre poluarea apei, ne gândim adesea la imagini dramatice ale deversărilor industriale sau la pete de ulei pe suprafața râurilor. Cu toate acestea, unele dintre cele mai periculoase forme de poluare sunt invizibile ochiului uman. Printre acestea se numără nitrații, compuși chimici care, deși esențiali pentru viața pe Pământ, în concentrații ridicate pot deveni un pericol major pentru sănătatea umană.

Setul de date "waterQuality1.csv" aruncă lumină asupra acestei probleme prin evidențierea nivelurilor de nitrați în diferite mostre de apă potabilă. Cu o medie de 9.819 mg/L, aceste cifre nu sunt doar niște numere abstracte. Ele reprezintă o amenințare reală, mai ales în comunitățile care depind de surse de apă cu risc de contaminare.

Nitrații ajung în apa potabilă din diverse surse: fertilizanți agricoli, deșeuri animale, sisteme de canalizare defecte și chiar precipitații. Odată consumați, aceștia pot transforma hemoglobina din sânge într-o formă care nu mai poate transporta oxigen eficient, cauzând metahemoglobinemie sau „sindromul bebelușului albastru” la sugari.

Dar problema nitraților nu se oprește aici. Pe lângă riscurile directe pentru sănătate, prezența lor în concentrații mari poate indica și alte probleme de mediu, cum ar fi poluarea agricolă nereglementată sau sistemele de tratare a apei ineficiente. Aceasta pune în lumina reflectoarelor nevoia urgentă de acțiune și inovare în domeniul tratării și purificării apei, precum și în practicile agricole sustenabile.

Analiza datelor din "waterQuality1.csv" nu este doar un exercițiu academic; este un apel la acțiune. Trebuie să ne întrebăm: cum putem reduce nivelurile de nitrați în sursele noastre de apă? Ce strategii putem adopta pentru a preveni contaminarea inițială? Și cel mai important, cum putem lucra împreună pentru a asigura că apa pe care o bem nu doar că ne păstrează în viață, dar ne protejează sănătatea?

Aceste întrebări nu sunt ușor de răspuns, dar înțelegerea profundă a problemei nitraților în apa potabilă este un pas crucial în direcția corectă. Este timpul să privim dincolo de ceea ce este vizibil și să abordăm aceste amenințări invizibile care pândesc în apa noastră.

## Obiectivele Analizei

Analiza datelor privind calitatea apei, cu un accent particular pe problema nitraților, își propune să ofere o înțelegere profundă și cuprinzătoare a provocărilor și soluțiilor în domeniul gestionării calității apei potabile. Prin intermediul acestui studiu, urmărim să abordăm problema nitraților din diferite perspective, integrând analiza datelor cu înțelegerea contextului ecologic, agricol și urban.

**1.Predicția Concentrațiilor de Nitrați**

Obiectivul principal este de a dezvolta modele predictive pentru concentrațiile de nitrați în diferite surse de apă. Aceasta implică analizarea datelor istorice și identificarea modelelor și factorilor care contribuie la variații în concentrațiile de nitrați. Înțelegerea acestor factori va îmbunătăți capacitatea noastră de a anticipa și de a gestiona riscurile de poluare a apei cu nitrați.

**2.Evaluarea Sursei și Impactului Nitraților**

Un alt obiectiv crucial este de a evalua sursa și impactul nitraților asupra calității apei potabile. Acest lucru implică investigarea relației dintre nivelurile de nitrați și diferite surse de poluare, cum ar fi activitățile agricole, industriale și urbane, pentru a înțelege mai bine cum poluarea afectează sursele de apă.

**3.Analiza Impactului Factorilor Extern**

Vom analiza, de asemenea, impactul factorilor externi, cum ar fi condițiile climatice, practicile agricole și schimbările în infrastructura urbană, asupra nivelurilor de nitrați în apă. Înțelegerea acestor factori va ajuta la dezvoltarea de strategii pentru reducerea poluării cu nitrați.

**4.Îmbunătățirea Strategiilor de Purificare a Apei**

Un obiectiv important este îmbunătățirea și optimizarea proceselor și strategiilor de purificare a apei, pentru a reduce eficient nivelurile de nitrați. Prin analiza datelor, ne propunem să identificăm cele mai eficiente metode de tratament și să sugerăm îmbunătățiri ale tehnologiilor existente.

**5.Elaborarea Strategiilor de Prevenire a Poluării**

Obiectivul este de a folosi înțelegerile obținute pentru a elabora strategii de prevenire a poluării cu nitrați. Acest lucru include dezvoltarea de recomandări pentru practici agricole sustenabile, gestionarea deșeurilor și planificarea urbană.

**6.Contribuția la Cunoașterea Științifică și Practică**

În final, studiul își propune să contribuie la literatura existentă în domeniul calității apei, oferind noi perspective și înțelegeri despre managementul nitraților. Publicarea rezultatelor va oferi informații valoroase pentru comunitatea științifică, precum și pentru practicienii în domeniul managementului resurselor de apă.

**Provocări și Ipoteze pentru Analiza Calității Apei cu Accent pe Nitrați**

**Provocări:**

**Calitatea și Consistența Datelor:**

Provocarea principală în analiza calității apei este asigurarea calității și consistenței datelor. Datele privind nivelurile de nitrați pot fi afectate de variații în metodele de colectare, erori de laborator, sau lipsuri de date. Acestea pot duce la interpretări greșite sau concluzii inexacte. Este crucial să se verifice și să se prelucreze datele cu atenție pentru a asigura fiabilitatea analizei.

**Variabilitatea Geografică și Temporală:**

Variabilitatea nivelurilor de nitrați în diferite regiuni și perioade poate constitui o provocare în înțelegerea și generalizarea rezultatelor. Factorii climatici, geografici și sursa de poluare pot influența semnificativ concentrațiile de nitrați, necesitând o abordare atentă și contextualizată în analiză.

**Complexitatea Sursei de Poluare:**

Identificarea surselor precise ale poluării cu nitrați poate fi complicată datorită multiplelor posibile surse, inclusiv agricultura, deșeurile industriale și urbane. Această complexitate poate îngreuna formularea de strategii eficiente de reducere a poluării.

**Ipoteze:**

**Corelație între Activități Agricole și Niveluri de Nitrați:**

Se presupune că există o corelație directă între intensitatea activităților agricole într-o regiune și nivelurile crescute de nitrați în sursele de apă. Această ipoteză se bazează pe utilizarea extensivă a fertilizanților care conțin nitrați în agricultură.

**Influența Condițiilor Climatice asupra Concentrațiilor de Nitrați:**

Se ipotezează că variațiile climatice, cum ar fi precipitațiile abundente sau seceta, pot influența semnificativ concentrațiile de nitrați în apă. Ploaia abundentă poate spăla nitrații din sol în sursele de apă, în timp ce seceta poate concentra nivelurile de nitrați.

**Eficiența Sistemelor de Tratare a Apei:**

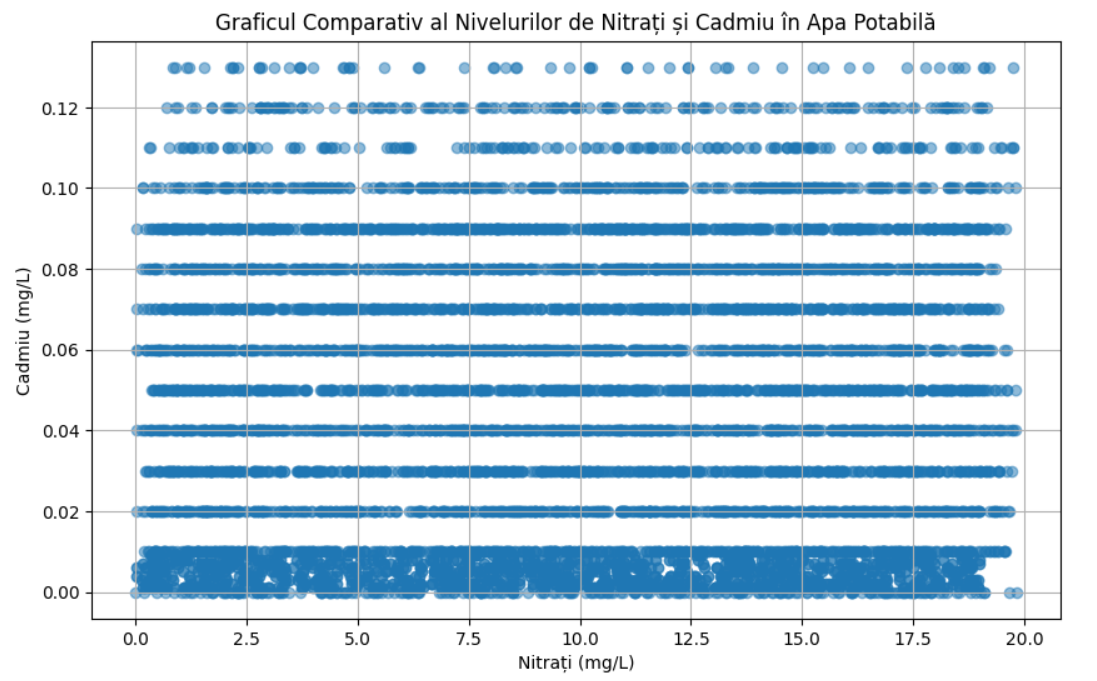
Se presupune că eficiența sistemelor de tratare a apei influențează semnificativ nivelurile de nitrați în apa potabilă. Regiunile cu sisteme de purificare mai avansate ar trebui să prezinte niveluri mai scăzute de nitrați în apa potabilă.

Prin identificarea și abordarea acestor provocări și prin testarea ipotezelor formulate, studiul poate oferi o perspectivă valoroasă asupra gestionării calității apei și a impactului nitraților, contribuind astfel la dezvoltarea de strategii mai eficiente pentru asigurarea sănătății publice și a sustenabilității mediului.

## Anexe

### Anexa A: Graficul Comparativ al Nivelurilor de Nitrați și Cadium în Apa Potabilă

Graficul compară concentrațiile de nitrați cu nivelurile de cadmiu, oferind o reprezentare vizuală a modului în care aceste două contaminanți coexistă în diferite probe de apă potabilă. Graficul ilustrează relația dintre concentrațiile de nitrați și cadmiu, evidențiind orice tendințe sau corelații dintre aceste două elemente importante în contextul calității apei. Prin acest grafic, se poate observa dacă există o tendință comună în creșterea sau scăderea nivelurilor acestor metale și cum această dinamică variază în funcție de diferitele surse de apă analizate.



Observații:

**Distribuția Nitraților:** Nivelurile de nitrați par să fie distribuite pe o gamă largă, care se întinde de la aproape 0 până la aproximativ 20 mg/L. Există o concentrare notabilă de puncte la niveluri mai scăzute, ceea ce sugerează că majoritatea probelor de apă au concentrații de nitrați în partea inferioară a gamei.

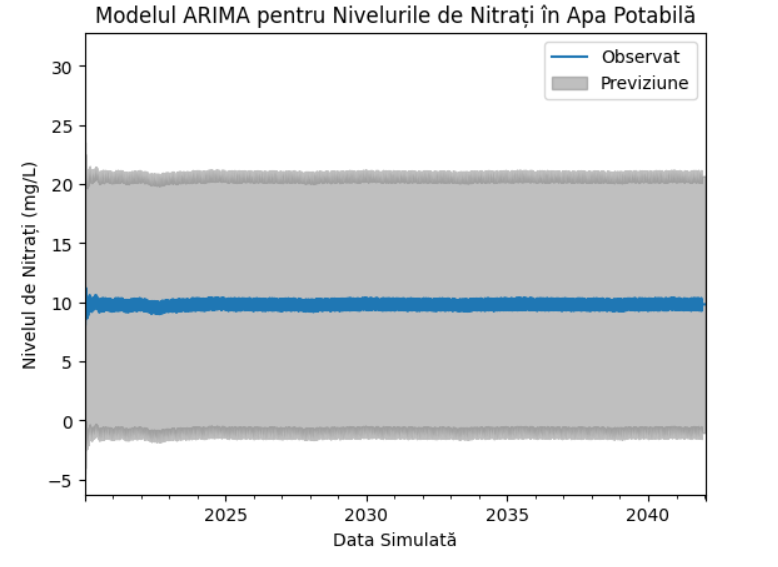
**Distribuția Cadmiului:** Nivelurile de cadmiu par să fie în general mai scăzute comparativ cu cele ale nitraților, concentrându-se în principal sub 0.02 mg/L. Se observă și o distribuție destul de uniformă pe verticală a punctelor de cadmiu în grafic, indicând o variație relativ constantă a nivelurilor de cadmiu în probele de apă.

**Concentrații Ridicate:** Există câteva puncte care indică niveluri mai ridicate atât pentru nitrați, cât și pentru cadmiu. Acestea ar putea reprezenta outlieri sau cazuri specifice în care apa este semnificativ mai poluată.

**Lipsa unei Corelații Aparente:** La o privire generală, nu pare să existe o corelație directă între nivelurile de nitrați și cele de cadmiu. Punctele sunt dispersate pe grafic fără un model clar sau o tendință care să indice o relație directă între cele două variabile.

### Anexa B: Modelul ARIMA pentru Nivelurile de Nitrați în Apa Potabilă

Modelul ARIMA (AutoRegressive Integrated Moving Average) este o abordare de modelare statistică utilizată pentru a analiza și prognoza seriile de timp, care poate fi aplicată pentru a înțelege dinamica nivelurilor de nitrați în apă potabilă. Utilizând date istorice, modelul ARIMA ia în considerare atât tendințele, cât și ciclicitatea potențială pentru a face predicții precise despre nivelurile de nitrați în viitor. Acest tip de model este deosebit de valoros în gestionarea resurselor de apă, deoarece permite autorităților și companiilor de utilități să anticipeze și să planifice pentru posibile schimbări ale calității apei. Prin identificarea modelului potrivit și ajustarea parametrilor, modelul ARIMA poate contribui la asigurarea calității apei potabile și la protejarea sănătății publice.



Observații:

**Stabilitatea Nivelurilor de Nitrați:** Linia albastră, reprezentând observațiile istorice, pare să fie relativ stabilă de-a lungul timpului, sugerând că nivelurile de nitrați în probele de apă potabilă nu au variat semnificativ în perioada analizată.

**Previziunile Modelului ARIMA:** Umbrirea gri indică previziunile modelului ARIMA. Lățimea benzii gri reflectă incertitudinea sau eroarea de predicție, care pare să fie mai mare pe măsură ce ne îndepărtăm de datele istorice.

**Intervalul de Încredere:** Benzile mai deschise de gri în jurul previziunilor sugerează un interval de încredere pentru predicțiile modelului. Acest interval este relativ larg, indicând un grad mare de incertitudine în previziunile pe termen lung.

**Lipsa Tendinței:** Graficul nu arată o tendință clară ascendentă sau descendentă în datele istorice sau în previziuni, ceea ce sugerează că nivelurile de nitrați au fost relativ constante sau că modelul nu a detectat nicio tendință persistentă.

**Potențial pentru Fluctuații:** În timp ce previziunile pe termen scurt par să fie relativ apropiate de ultimele observații istorice, există potențial pentru fluctuații semnificative pe măsură ce ne îndepărtăm în timp, după cum sugerează lărgirea benzii de previziune.

**Necesitatea Monitorizării Continue:** Observațiile stabile și previziunile incerte subliniază importanța monitorizării continue a nivelurilor de nitrați în apă pentru a detecta orice schimbări semnificative care ar putea să apară în viitor.

### Anexa C: Rezultatele Modelului de Regresie Liniară

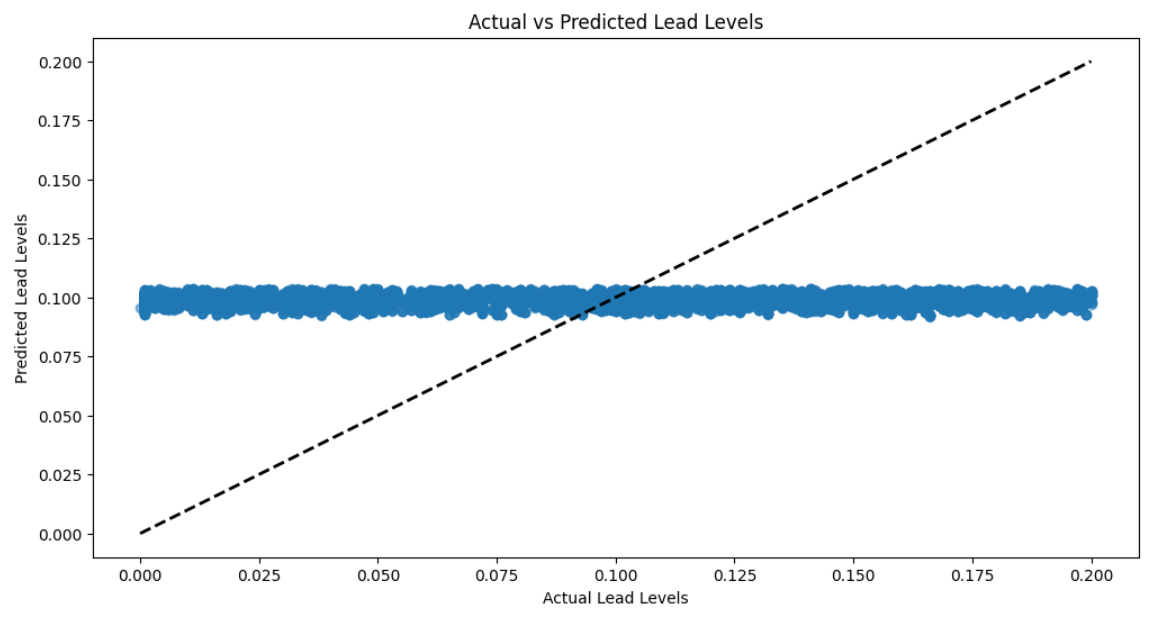
Am aplicat modelul de regresie liniară folosind concentrațiile de nitrați și cloramini pentru a prezice nivelul de plumb. Iată rezultatele:

Mean Squared Error (MSE): 0.003290.00329

Coeficientul de determinare (R²): 0.00170.0017

Aceste rezultate indică faptul că modelul nu are o acuratețe foarte mare în predicția nivelului de plumb, bazându-se pe concentrațiile de nitrați și cloramini. Un R² aproape de zero sugerează că modelul explică o proporție foarte mică a variabilității variabilei dependente (plumbul) pe baza variabilelor independente (nitrații și cloraminii).





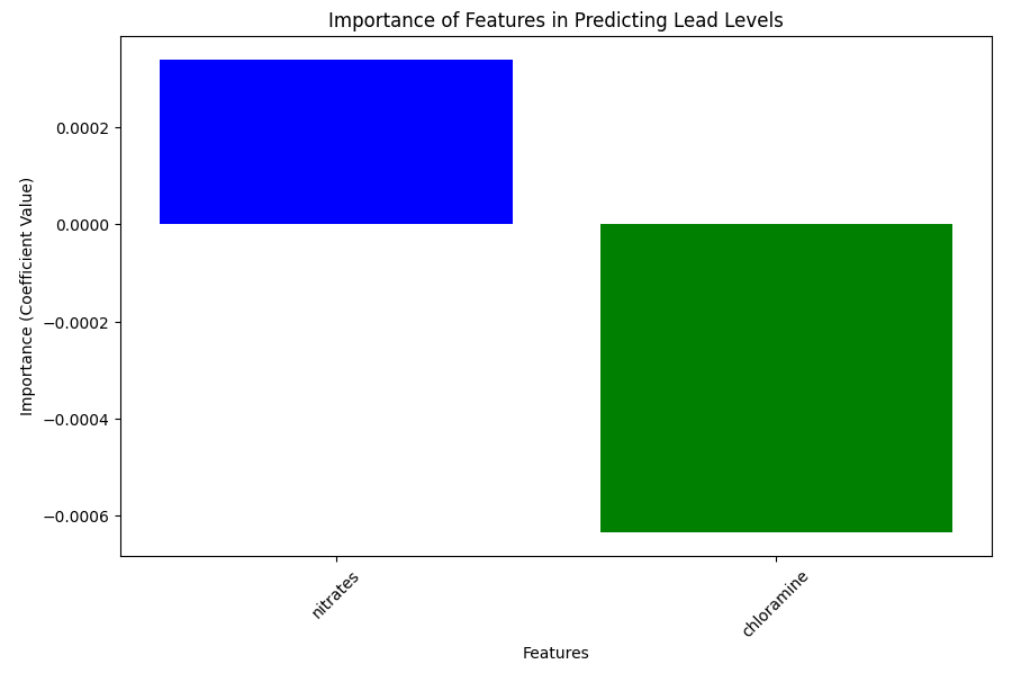
### Anexa D: Importanta caracteristicilor

Importanța caracteristicilor în modelul de regresie liniară este după cum urmează:

**Nitrați:** Coeficientul este 0.000340. Acest lucru indică faptul că o creștere a concentrației de nitrați este asociată cu o creștere ușoară a nivelului de plumb, menținându-se constante celelalte variabile.

**Cloramini:** Coeficientul este - 0.000634. Aceasta sugerează că o creștere a concentrației de cloramini este asociată cu o scădere ușoară a nivelului de plumb, tot menținându-se constante celelalte variabile.

Trebuie să rețineți că, deși acești coeficienți indică o anumită relație, coeficientul scăzut de determinare (R²) al modelului sugerează că aceste variabile explică doar o mică parte din variația nivelului de plumb. Prin urmare, ar putea exista și alți factori influenți care nu sunt incluși în acest model. ​



În graficul prezentat, avem două bare care reprezintă importanța a două caracteristici (nitrați și cloramini) în predicția nivelului de plumb, în cadrul modelului de regresie liniară:

**Bara albastră (Nitrați):** Aceasta are o valoare pozitivă, indicând că o creștere a concentrației de nitrați este asociată cu o creștere a nivelului de plumb prezis de model. Cu alte cuvinte, nitrații au un efect pozitiv asupra nivelului de plumb, conform modelului.

**Bara verde (Cloramini):** Aceasta are o valoare negativă, semnificând că o creștere a concentrației de cloramini este asociată cu o scădere a nivelului de plumb prezis de model. Prin urmare, cloraminii au un efect negativ asupra nivelului de plumb, conform modelului.

Dimensiunea barelor reflectă mărimea coeficienților din modelul de regresie liniară. Coeficienții mai mari (fie pozitivi, fie negativi) indică o influență mai puternică a caracteristicii respective asupra variabilei țintă (nivelul de plumb). Totuși, este important de menționat că, deși aceste caracteristici influențează nivelul de plumb conform modelului, acuratețea modelului în sine (indicată de coeficientul R²) este destul de scăzută, ceea ce sugerează că alte variabile sau factori pot juca un rol semnificativ în determinarea nivelului de plumb.