**Predictie cresterii temperaturii globale**

**Introducere**

Motivație: Studiul analizei predicției asupra datelor climatice reprezintă un domeniu de interes major datorită implicațiilor directe asupra mediului, economiei și vieții sociale. Identificarea unui algoritm eficient pentru predicții precise contribuie la luarea deciziilor bazate pe date.

Descriere generală temă: Lucrarea explorează utilizarea algoritmilor de învățare automată pentru predicția datelor climatice. Principalul obiectiv este compararea predicțiilor și evaluarea performanței unui algoritm ales.

Tip de algoritm: În acest proiect a fost utilizat un algoritm Gradient Boosting Regressor, ales pentru capacitatea sa de a modela relații complexe dintre variabile.

Algoritm ales: Gradient Boosting Regressor este o metodă puternică, care combină mai mulți arbori de decizie într-un mod secvențial, pentru a reduce eroarea de predicție și a crește performanța generală.

**Metodologie:**

1. **Datele inițiale și preprocesarea:**
   * **Setul de date**: Temperaturi istorice (1851-1975) și date reale (1976-2023) pentru validare ulterioară.
   * Crearea unui DataFrame pentru a organiza anii și temperaturile.
   * Preprocesare prin normalizarea caracteristicilor folosind StandardScaler pentru a îmbunătăți performanța modelului.
   * Această împărțire permite evaluarea performanței modelului pe date noi și evitarea supra-antrenării.
2. **Modelul de regresie:**
   * Utilizarea GradientBoostingRegressor cu parametri ajustați:
     + 100 de arbori de decizie.
     + Adâncimea maximă a arborilor este setată la 4 pentru a captura tendințele nelineare.
     + Rata de învățare este de 0.1 pentru un echilibru între viteză și acuratețe.
3. **Predicțiile și evaluarea:**
   * Predicțiile pentru setul de testare sunt evaluate folosind:
     + **MSE** (Mean Squared Error)
     + **R²** (coeficientul de determinare)
   * Performanța este măsurată pentru a evalua cât de bine se potrivesc datele modelului.
4. **Predicția pentru viitor (1976-2023):**
   * Se extinde predicția temperaturilor globale pentru următorii 50 de ani.
   * Ajustarea tendinței prin adăugarea unui factor liniar (trend\_adjustment) pentru a imita o creștere realistă.
5. **Validare cu date reale:**
   * Comparație între predicții și valori reale pentru 1976-2023:
     + Calcularea metricilor precum **MAE**, **RMSE** și **R²** pentru a evalua acuratețea extrapolării.
6. **Vizualizarea rezultatelor:**
   * Grafic cu trei componente:
     + Date istorice.
     + Predicții ajustate pentru viitor.
     + Valori reale din perioada de validare.
   * Marcatori colorați pentru a diferenția între tipurile de date.

**Particularități ale implementării algoritmului:**

1. **Modelarea tendințelor nelineare:** Gradient Boosting este potrivit pentru capturarea relațiilor complexe între ani și temperatură, evidențiate de creșterea graduală a temperaturilor.
2. **Normalizarea intrărilor:** Standardizarea X a îmbunătățit stabilitatea și convergența modelului, un pas esențial în prelucrarea datelor pentru multe algoritme bazate pe arbori.
3. **Ajustarea extrapolării:** Adăugarea unui trend liniar compensează limita algoritmului de a extrapola în afara domeniului de antrenare.
4. **Validarea cu date reale:** Compararea cu valori din lumea reală oferă o verificare practică a performanței predicției modelului, ceea ce este crucial pentru evaluarea aplicabilității.

**Rezultate obținute**

**Modelul a fost antrenat și testat utilizând datele istorice (1851-1975). Rezultatele arată că modelul are o performanță ridicată:**

* **Mean Squared Error (MSE): 0.0001  
  Aceasta indică o eroare foarte mică în predicțiile modelului comparativ cu valorile reale din setul de testare.**
* **R-squared (R²): 0.9992  
  Coeficientul de determinare foarte apropiat de 1 arată că modelul explică aproape în întregime variația temperaturilor în setul de testare.**

**Aceste valori confirmă că modelul este foarte precis pe datele folosite pentru antrenare și testare.**

**Performanța pe date reale (validare: 1976-2023)**

A black background with white numbers and symbols

Description automatically generated**Validarea modelului folosind date reale permite evaluarea capacității acestuia de a extrapola tendințele temperaturilor globale.**

* A black background with white text

  Description automatically generated**Mean Squared Error (MSE): 0.0025  
  Eroarea pătratică medie indică o potrivire bună între predicții și temperaturile reale, dar este mai mare comparativ cu setul de testare, din cauza extrapolării și a variațiilor reale.**
* **Root Mean Squared Error (RMSE): 0.0501  
  Oferă o măsură în unități de temperatură (°C) și indică o eroare medie de aproximativ 0.05°C pe perioada de validare.**
* A black background with white numbers and symbols

  Description automatically generated**Mean Absolute Error (MAE): 0.0410  
  Eroarea absolută medie arată cât de departe sunt predicțiile față de valorile reale în medie, confirmând o performanță solidă.**
* A black background with white text

  Description automatically generated**R-squared (R²): 0.8488  
  Aproximativ 84.88% din variația temperaturilor reale este explicată de modelul nostru, ceea ce este un rezultat bun.**

**Concluzii**

A graph with a blue line and orange dots

Description automatically generatedModelul Gradient Boosting Regressor s-a dovedit extrem de eficient în capturarea tendințelor istorice ale temperaturilor globale. Valoarea ridicată a R² (0.9992) pe setul de testare arată că datele istorice utilizate (1851-1975) au fost bine înțelese de model. Aceasta confirmă că tehnica Gradient Boosting este potrivită pentru a modela relații complexe, cum sunt cele între anii istorici și temperaturi..

**Analiza metricilor de eroare oferă o perspectivă asupra performanței reale a modelului:**

* **MAE (0.0410°C) și RMSE (0.0501°C) indică o precizie foarte bună, având în vedere intervalul mic de variație a temperaturilor medii globale.**
* **Diferențele între MSE din setul de testare (0.0001) și cel de validare pe date reale (0.0025) arată că extrapolarea are limitări naturale, dar performanța rămâne una satisfăcătoare.**

**Importanța datelor istorice și a complexității fenomenului climatic**

Modelul demonstrează că datele istorice sunt utile pentru identificarea unor tendințe generale, dar climatologia este un domeniu extrem de complex. Faptul că modelul nu a anticipat pe deplin toate variațiile temperaturilor reale subliniază:

* Necesitatea de a integra factori suplimentari (ex: emisiile de CO₂, defrișările, activitatea vulcanică).
* Limitările abordării bazate exclusiv pe statistici și tendințe istorice, fără considerarea fenomenelor neașteptate.

**Observații privind viitorul tendinței**

**Predicțiile ajustate pentru următorii ani sugerează o creștere constantă a temperaturii globale, ceea ce este în concordanță cu tendințele observate în ultimii 50 de ani. Totuși, aceste proiecții sunt doar estimări și pot varia în funcție de:**

* **Politicile globale legate de schimbările climatice.**
* **Evenimente naturale sau antropice majore.**