

Raport: Predicția Soldului Energetic Utilizând Algoritmii ID3 și Bayes Naiv

Descrierea problemei

Contextul acestui proiect este legat de analiza și predicția soldului energetic, care este diferența dintre consumul și producția de energie electrică:

$$\text{Sold[MW]} = \text{Consum[MW]} - \text{Producție[MW]}.$$

Scopul proiectului este de a compara două abordări pentru prezicerea soldului energetic folosind algoritmii ID3 și Bayes Naiv, și de a analiza performanțele acestora pentru a determina metoda optimă de predicție.

Justificarea abordării

Pentru rezolvarea problemei, am utilizat două metode distincte:

1. Predicția directă a soldului (ID3 și Bayes Naiv): Această metodă implică antrenarea modelelor direct pe datele aferente soldului energetic, utilizând un set de caracteristici relevante.
2. Predicția consumului și a producției pentru calcularea soldului: Deoarece soldul este definit ca diferența dintre consum și producție, această metodă implică antrenarea separată a modelelor pentru prezicerea consumului și a producției, iar apoi calcularea soldului ca diferență între aceste două valori.

Am ales aceste două metode pentru a testa atât o abordare directă, cât și o abordare indirectă care poate avea de relații mai puternice între datele inițiale și consum/producție. Algoritmii ID3 și Bayes Naiv au folosit pentru antrenament date din intervalul **ianuarie 2023 - noiembrie 2024**, iar pentru testare date din **decembrie 2024** și au fost adaptați astfel:

- ID3: S-a utilizat un arbore de decizie cu **adâncime maximă de 5** pentru a limita complexitatea modelului.
- Bayes Naiv: Modelul a fost antrenat pe **date normalizate** pentru a îmbunătăți performanța.

Prezentarea rezultatelor

Rezultatele pentru predicția directă a soldului

- ID3: RMSE: 248.09, MAE: 198.90, R-squared: 0.9606, Sold total prezis: 2,680,813.53 MW, Acuratețe: 94.56%.
- Bayes Naiv: RMSE: 998.14, MAE: 776.19, R-squared: 0.3617, Sold total prezis: 2,459,096 MW, Acuratețe: 86.74%.

```
Drive/Desktop/python/ML-API/prediction.py
ID3 - RMSE: 248.0917230361811, MAE: 198.90289257569108, R-squared: 0.960565809477226
Bayes naiv - RMSE: 998.1451085887196, MAE: 776.1872417355372, R-squared: 0.36168397828708787
Sold total: 2835061
Sold total prezis (ID3): 2680813.5319733284
Sold total prezis (Naive Bayes): 2459096
Acuratete ID3: 94.56%
Acuratete Bayes naiv: 86.74%
```

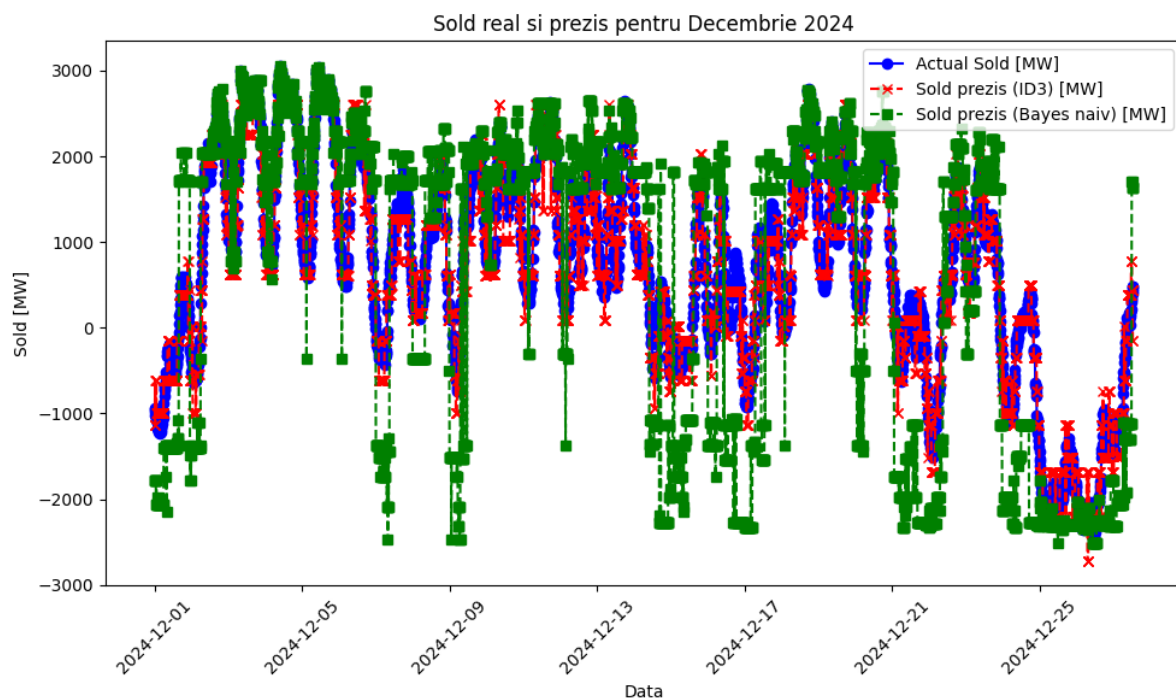


Figure 1: Predicție directă (ID3 și Bayes)

Rezultatele pentru predicția consumului și producției

- ID3:
 - Consum: RMSE: 48.41, MAE: 38.65, R-squared: 0.9979.
 - Producție: RMSE: 37.63, MAE: 32.30, R-squared: 0.9973.
 - Sold calculat: Acuratețe: 93.79%.

- Bayes Naiv:
 - Consum: RMSE: 0.0, MAE: 0.0, R-squared: 1.0.
 - Producție: RMSE: 0.0, MAE: 0.0, R-squared: 1.0.
 - Sold calculat: Acuratețe: 100.0%.

```

OneDrive/Desktop/python/ML-AP1/prediction_each.py
ID3 - Consum RMSE: 48.41454528135764, MAE: 38.65125361005721, R-squared: 0.9979189835695723
ID3 - Consum RMSE: 48.41454528135764, MAE: 38.65125361005721, R-squared: 0.9979189835695723
ID3 - Producție RMSE: 37.628770393700535, MAE: 32.29664284887943, R-squared: 0.997306611877755
Naive Bayes - Consum RMSE: 0.0, MAE: 0.0, R-squared: 1.0
Naive Bayes - Producție RMSE: 0.0, MAE: 0.0, R-squared: 1.0
Acuratete sold ID3: 93.79%
Acuratete sold Bayes naiv: 100.00%

```

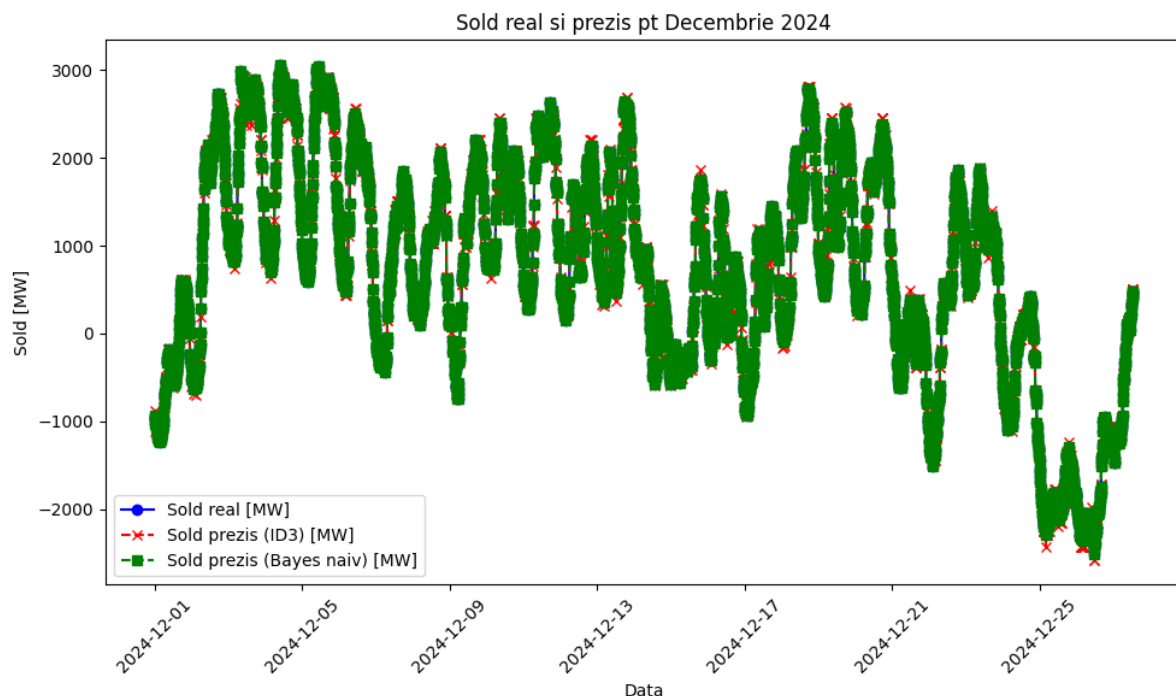


Figure 2: Predicție indirectă (ID3 și Bayes)

Analiza comparativă

Predicția directă a soldului a oferit rezultate bune pentru algoritmul ID3, cu un R-squared de 0.9606 și o acuratețe de 94.56%. Algoritmul Bayes Naiv, însă, a avut o performanță mai slabă pentru această metodă, cu un R-squared de 0.3617.

Predicția consumului și producției are o performanță mai bună în general, în special pentru Bayes Naiv, unde am obținut un R-squared de 1.0 și o acuratețe a soldului calculat de 100.0%, lucru ce indică o posibilă problemă de overfitting. Și ID3 a avut o acuratețe a soldului calculat bună, de 93.79%, dar mai mică față de cea din abordarea directă.

Concluzii

Rezultatele sugerează că metoda bazată pe prezicerea separată a consumului și a producției este mai bună, în special pentru Bayes Naiv. Algoritmul ID3 s-a descurcat bine în ambele metode, dar a fost mai puțin precis decât Bayes Naiv în metoda indirectă.

Îmbunătățiri posibile:

- Optimizarea hiperparametrilor pentru ID3 ar putea crește acuratețea.
- Utilizarea unui set de date mai mare pentru antrenare.
- Mărirea adâncimii arborelui de decizie pentru algoritmul ID3 în predicția directă.