

Analiza și Predicția Soldului Energetic

Enache Samira

Decembrie 2024

1. Analiza problemei

Setul de date utilizat conține informații despre consumul energetic, producția de energie (pe diferite surse, inclusiv nuclear, eolian, fotovoltaic, etc.) și soldul energetic calculat ca diferența dintre producție și consum. Scopul acestui proiect este de a construi modele predictive pentru soldul energetic ($Sold[MW]$) folosind trei metode diferite:

- Predicția folosind doar coloana $Sold[MW]$ (Metoda 1).
- Predicția fiecărei componente (consum și producție pe surse) și calcularea $Sold[MW]$ ulterior (Metoda 2).
- Agregarea producției în categorii (intermitentă și constantă) și utilizarea acestora pentru predicție (Metoda 3).

Datele au fost preprocesate pentru a elimina valorile lipsă, iar caracteristicile temporale (ora, ziua, luna, ziua săptămânii) au fost extrase din coloana $Data$ pentru a fi utilizate în modelele predictive.

2. Implementarea algoritmilor

Pentru predicție, am folosit două modele:

- Arborele de decizie ID3 adaptat
- Modelul Naive Bayes

Pentru evaluarea performanței, au fost utilizate două metrici: **RMSE** (Root Mean Square Error) și **MAE** (Mean Absolute Error).

Metoda 1: Predicția folosind doar $Sold[MW]$

Această metodă utilizează doar coloana $Sold[MW]$ pentru a antrena modelele predictive. Datele sunt împărțite în seturi de antrenare (lunile până la decembrie) și testare (luna = 12).

Metoda 2: Predicția fiecărei componente și calcularea $Sold[MW]$

În această metodă, fiecare componentă individuală (consum, producție pe surse) este prezisă separat, iar $Sold[MW]$ este calculat ulterior ca diferență între producția totală și consum.

Metoda 3: Agregarea producției pe categorii

Producția este agregată în două categorii:

- **Producție intermitentă:** suma producției din surse eoliene și fotovoltaice.
- **Producție constantă:** suma producției din surse nucleare, hidrocarburi, ape, cărbune și biomasă.

Modelele sunt antrenate pentru a prezice aceste categorii, împreună cu consumul total, iar $Sold[MW]$ este calculat ca diferență.

3. Rezultate și Analize Comparative

Performanțele celor trei metode au fost evaluate folosind RMSE și MAE. Rezultatele sunt prezentate în Tabelul 1.

Table 1: Rezultatele modelelor ID3 și Bayesian pentru cele trei metode

Metodă	Model	RMSE	MAE
Metoda 1	ID3	940.64	803.16
	Bayesian	1289.49	1058.39
Metoda 2	ID3	2636.53	2245.72
	Bayesian	3438.19	2772.29
Metoda 3	ID3	2944.46	2592.92
	Bayesian	2548.35	1985.50

Graficul comparativ pentru Metoda 1 este ilustrat în Figura 1.

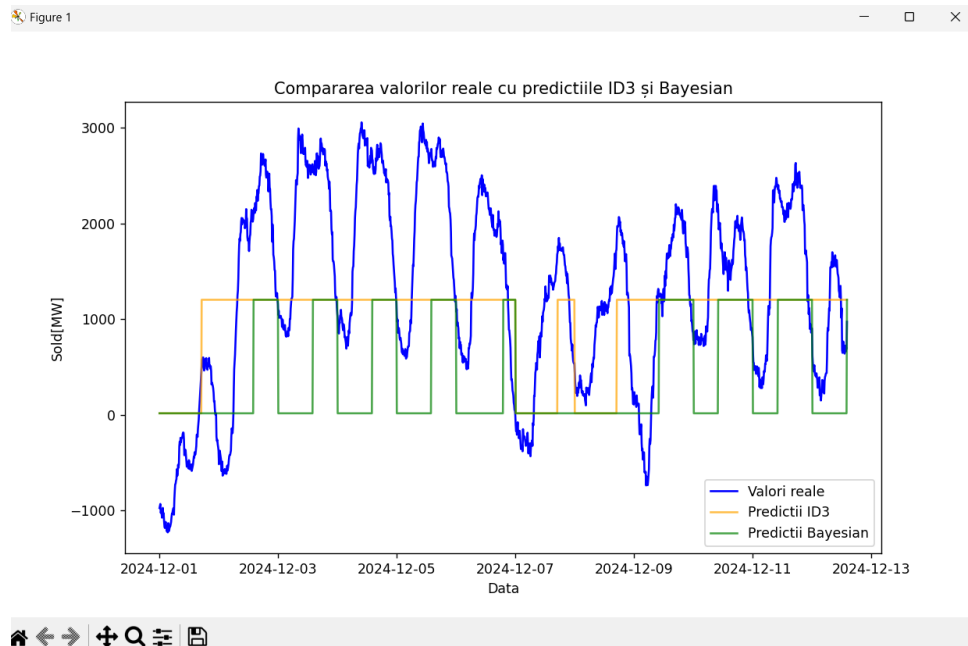


Figure 1: Comparația valorilor reale cu predicțiile (Metoda 1)

4. Concluzii

- **Metoda 1** a oferit cele mai bune rezultate, cu un RMSE de 940.64 pentru ID3 și 1289.49 pentru Bayesian. Acest lucru sugerează că predicția directă a $Sold[MW]$ este mai eficientă decât predicția componentelor individuale.
- Metodele 2 și 3 au performat semnificativ mai slab, probabil din cauza acumulării erorilor de predicție pentru componentele individuale.
- Păstrarea a 5 intervale (pentru discretizare) a fost justificată prin faptul că o complexitate mai mare (mai multe intervale) a dus la performanțe mai slabe, crescând erorile RMSE și MAE.
- Pentru a îmbunătăți metodele 2 și 3, ar fi utilă utilizarea unor modele mai complexe, cum ar fi Random Forest sau Gradient Boosting.

5. Codul sursă

Codul complet al proiectului este disponibil în repozitoriul GitHub, împreună cu raportul în format PDF și explicațiile markdown:

- Link către repozitoriu: https://github.com/SamiiraEnache/ML_AP1