

Raport: Analiza Predictiilor cu Algoritmii ID3 si Bayes

Pantaziu Andrei Rares

January 3, 2025

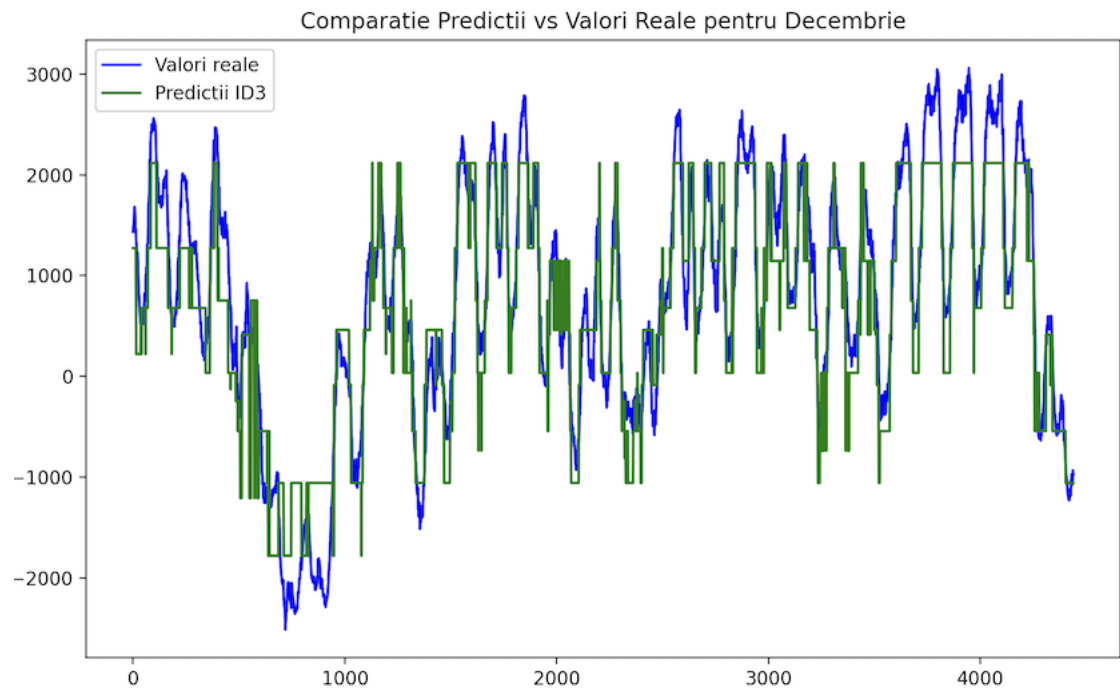
1 Descrierea problemei

Scopul acestui proiect este de a analiza performanta algoritmilor de predictie ID3 și Bayes in estimarea valorilor de sold energetic pentru luna decembrie 2024, utilizand un set de date istorice. Contextul proiectului se bazeaza pe necesitatea de a anticipa valori energetice pentru o mai buna planificare a resurselor în sistemul energetic national. Setul de date utilizat contine informatii despre productia si consumul de energie, segmentate pe diferite surse (carbune, nuclear, eolian etc.).

2 Justificarea abordarii

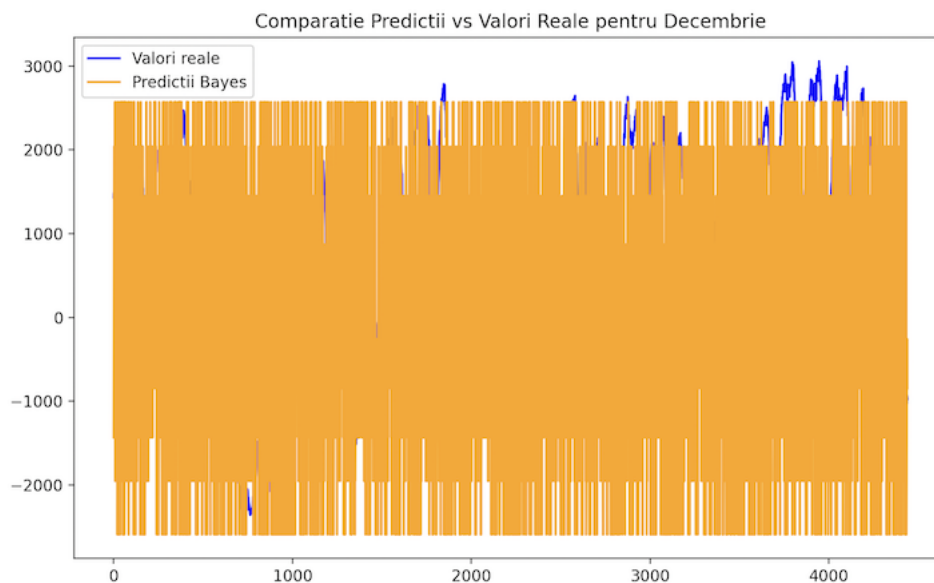
2.1 Algoritmul ID3

Arborele de decizie ID3 a fost ales datorita capacitatii sale de a gasi relatii complexe si non-liniare intre caracteristicile de intrare (ex. consum, surse de energie) si tinta de predictie (soldul energetic). Modelul a fost configurat pentru regresie, cu o adancime maxima de 5, pentru a evita suprainvatarea.



2.2 Algoritmul Bayes

Modelul bayesian, desi simplificat, a fost ales datorita abordarii sale probabilistice care poate oferi informatii despre incertitudine. Algoritmul a fost adaptat pentru regresie prin discretizarea variabilei tinta în intervale (*buckets*). Acest proces a fost implementat folosind metode statistice de mediere a valorilor în cadrul fiecarui interval.



3 Prezentarea rezultatelor

3.1 Rezultate numerice

- Media valorilor reale (decembrie): 786.6436 MW
- Media predictiilor ID3: 666.5598 MW
- Media predictiilor Bayes: 55.5882 MW

```
silence this warning.
data['Data'] = pd.to_datetime(data['Data'])
Media valorilor pentru luna decembrie:
Valori reale: 786.6436134264474
Predictii ID3: 666.5598930625738
Predictii Bayes: 51.58822517761045
```

Modelul ID3 produce rezultate mult mai apropiate de valorile reale comparativ cu Bayes, care esuează în estimare

4 Concluzii

În urma analizei, am constatat că algoritmul ID3 este mult mai potrivit pentru acest set de date decât abordarea bayesiană. Acest lucru se datorează capacității ID3 de a modela relații non-liniare complexe.

Pentru a imbunatati metoda:

- Implementarea unui model bayesian mai avansat.
- Optimizarea hiperparametrilor ID3 pentru performanta maxima.
- Extinderea setului de date pentru a include mai multe caracteristici relevante.