

Predicția Soldului de Energie pentru Decembrie 2024

Boldeanu Razvan

January 5, 2025

1 Introducere

Scopul acestui proiect este de a prezice soldul pentru decembrie 2024 utilizând date istorice din Sistemul Energetic Național (SEN) al României. Setul de date conține informații despre consumul și producția de energie în funcție de diverse surse (de exemplu, hidro, solar, nuclear). Sarcina implică adaptarea algoritmilor ID3 (arbore de decizie) și clasificare Bayesiană pentru regresie, evaluarea performanței acestora și explorarea diferitelor abordări pentru rezolvarea problemei.

1.1 Context și Obiectiv

Obiectivul principal este de a prezice soldul, care reprezintă diferența dintre producție și consum, pe baza datelor din anii trecuți. Au fost utilizate două metode:

- **Metoda 1:** Predicția directă a soldului folosind caracteristici temporale.
- **Metoda 2:** Predicția componentelor (de exemplu, Consum[MW], Producție[MW]) și reconstrucția Sold[MW] ca diferență dintre acestea.

2 Preprocesarea Datelor

Setul de date a fost preprocesat pentru a gestiona valorile lipsă, a converti coloanele relevante în formate numerice și a extrage caracteristici temporale (oră, zi, lună, ziua săptămânii). Pașii aplicați au fost următorii:

1. Concatenarea mai multor fișiere de antrenament într-un singur set de date.
2. Conversia coloanei Data în format datetime și extragerea caracteristicilor temporale.
3. Conversia coloanelor legate de energie (de exemplu, Consum[MW]) în valori numerice.
4. Eliminarea rândurilor cu valori lipsă pentru a asigura integritatea datelor.

3 Metodologie

Au fost implementate două modele de învățare automată:

- **ID3:** Adaptat pentru regresie prin împărțirea variabilelor țintă continue în intervale discrete și prezicerea etichetelor de interval.
- **Clasificare Bayesiană:** Similar adaptată pentru regresie folosind valori împărțite în intervale.

Experimentele au explorat efectul scalării caracteristicilor asupra performanței modelelor. Predicțiile au fost evaluate utilizând următorii metrici:

- **RMSE (Eroarea Pătratică Medie):** Măsoară deviația standard a erorilor de predicție.
- **MAE (Eroarea Absolută Medie):** Măsoară magnitudinea medie a erorilor de predicție.
- **R^2 :** Măsoară cât de bine explică modelul variația variabilei țintă.

4 Rezultate

4.1 Metoda 1: Predicția directă a Sold[MW]

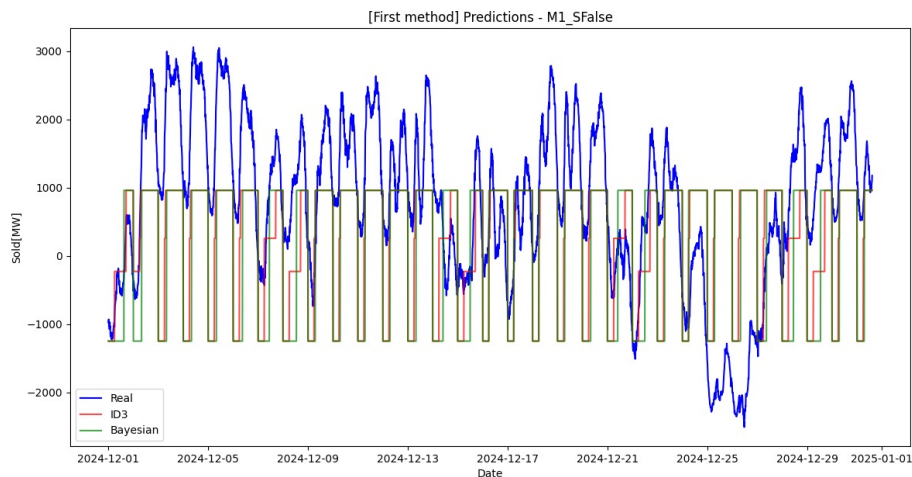


Figure 1: Predicții utilizând Metoda 1 fără scalare

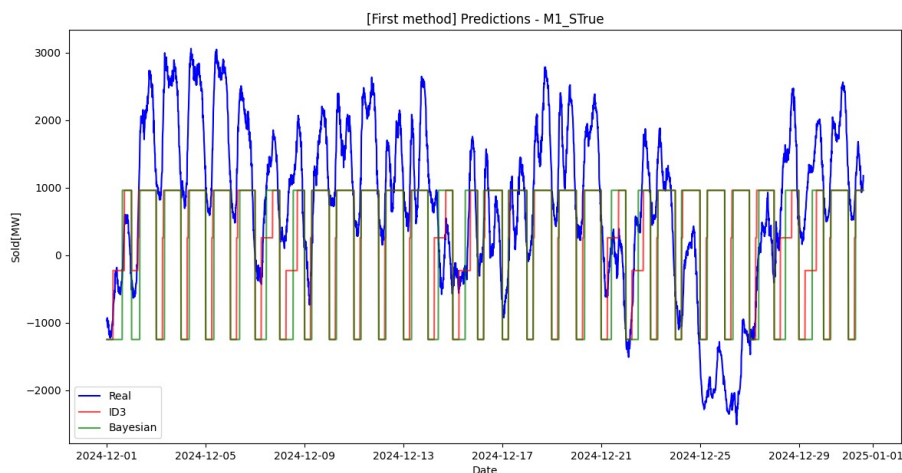


Figure 2: Predicții utilizând Metoda 1 cu scalare

În Metoda 1, predicțiile au fost foarte segmentate datorită utilizării regresiei bazate pe intervale. Scalarea a avut un impact minim asupra performanței, iar atât predicțiile ID3, cât și cele Bayesian au întâmpinat dificultăți în captarea fluctuațiilor reale din Sold[MW].

4.2 Metoda 2: Predicția Componentelor

Metoda 2 a produs predicții mai fluide, în special cu modelele Bayesian. Scalarea a îmbunătățit performanța prin alinierea predicțiilor mai aproape de valorile reale, dar a existat încă un decalaj notabil în cazul fluctuațiilor extreme.

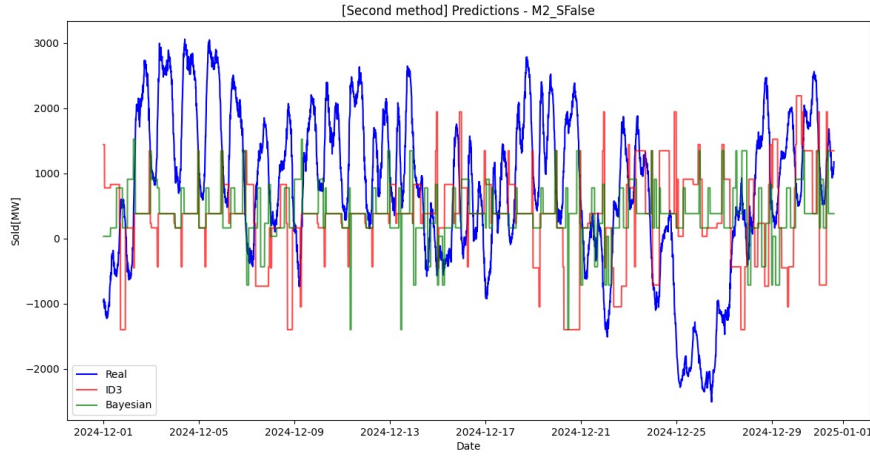


Figure 3: Predicții utilizând Metoda 2 fără scalare

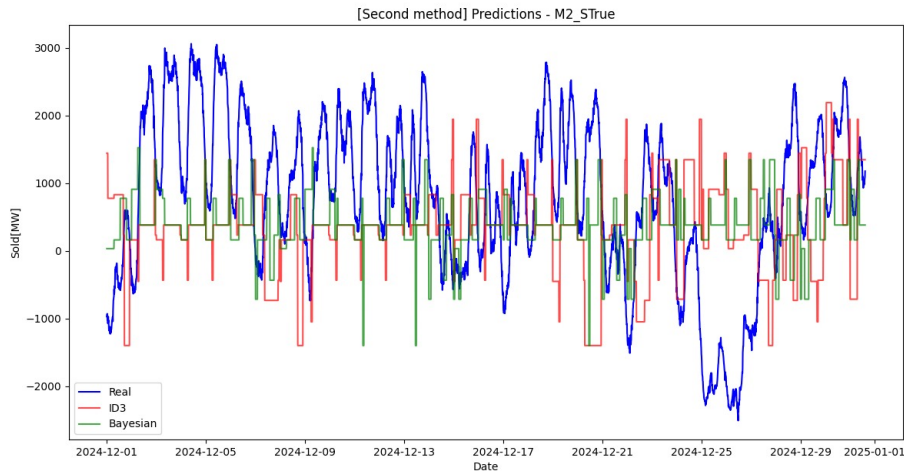


Figure 4: Predicții utilizând Metoda 2 cu scalare

5 Discuție și Comparație

5.1 Analiza Rezultatelor

- **Metoda 1:** Natura segmentată a predicțiilor indică faptul că regresia bazată pe intervale este insuficientă pentru captarea granularității din Sold[MW]. Atât modelele ID3, cât și cele Bayesian au avut rate de eroare ridicate.
- **Metoda 2:** Predicția componentelor a îmbunătățit rezultatele prin utilizarea informațiilor suplimentare despre consumul și producția de energie. Predicțiile Bayesian au fost mai fluide și mai precise decât cele ID3.
- **Scalarea Caracteristicilor:** Scalarea a îmbunătățit predicțiile în Metoda 2, evidențiind importanța sa în sarcinile de regresie ce implică date numerice.

6 Concluzii și Îmbunătățiri

6.1 Concluzii

- Metoda 2 a depășit Metoda 1 prin utilizarea informațiilor suplimentare despre componentele de energie.
- Modelele Bayesian au produs în mod constant predicții mai fluide și mai precise decât cele ID3.
- Scalarea caracteristicilor a îmbunătățit performanța în Metoda 2, dar a avut un impact limitat în Metoda 1.

6.2 Îmbunătățiri

- S-ar putea încerca modele bazate pe regresie (de exemplu, DecisionTreeRegressor, RandomForest) pentru a îmbunătăți granularitatea predicțiilor, precum și includerea caracteristicilor suplimentare (de exemplu, valori întârziate, medii mobile) pentru a surprinde dependențele temporale.
- De asemenea, o ajustare mai extinsă a hiperparametrilor pentru a optimiza performanța modelelor ar fi o idee bună.