

Aplicație Practică 1 - Electricity Predictor

Nume: Pricope Marius-Andrei

Grupă: A2

29 decembrie 2024

Cuprins

1	Descrierea problemei	2
1.1	Contextul și scopul proiectului	2
2	Analiza problemei	2
2.1	Înțelegerea setului de date	2
2.2	Preprocesarea datelor	2
3	Implementarea algoritmilor	3
3.1	Adaptarea tabelului de valori pentru regresie	3
3.2	Evaluarea performanței modelelor	5
4	Prezentarea rezultatelor	5
5	Concluzii	5
5.1	Ce am învățat?	5
5.2	Cum poate fi îmbunătățită metoda?	6
6	Referințe	6

1 Descrierea problemei

1.1 Contextul și scopul proiectului

În cadrul acestei teme, obiectivul este dezvoltarea unei soluții pentru predicția soldului total al Sistemului Energetic Național (SEN) pentru luna decembrie 2024. Aceasta implică utilizarea și adaptarea algoritmilor ID3 și clasificării bayesiene pentru o problemă de regresie.

Setul de date: Datele includ consumul și producția de energie electrică, defalcate pe diverse surse, și sunt furnizate de Transelectrica SEN Grafic.

2 Analiza problemei

2.1 Înțelegerea setului de date

Setul de date conține următoarele coloane principale:

- **Data:** Timpul specific al înregistrării.
- **Consum[MW]:** Consumul total de energie electrică.
- **Medie Consum[MW]:** Media consumului.
- **Producție[MW]:** Producția totală de energie.
- **Carbune[MW]:** Producție pe bază de cărbune.
- **Hidrocarburi[MW]:** Producție pe bază de hidrocarburi.
- **Ape[MW]:** Producție hidro.
- **Nuclear[MW]:** Producție nucleară.
- **Eolian[MW]:** Producție eoliană.
- **Foto[MW]:** Producție solară.
- **Biomasă[MW]:** Producție din biomasă.
- **Sold[MW]:** Diferența dintre producție și consum.

2.2 Preprocesarea datelor

Inițial am realizat un tabel, cu media valorile pentru fiecare coloana pentru fiecare lună. Coloanele pe care le-am păstrat, au fost doar cele de consum și producție, și soldul era coloana rezultat. Motivul pentru care am ales să încerc așa prima dată a fost că producția este suma tuturor surselor de producție, și m-am gândit că e suficient dacă mă raportează doar la producția totală, astfel rezultatul algoritmului devenind mai general.

Date	Consum	Productie	Sold
Nov. 2023	5402	5384	18
Dec. 2023	6753	7059	-306
Jan. 2024	8223	7574	649
Feb. 2024	6281	6690	-409
Mar. 2024	4808	6337	-1529
Apr. 2024	5334	5045	289
May 2024	6011	5126	885
Jun. 2024	5847	5449	398
Jul. 2024	7622	6073	1549
Aug. 2024	5659	5515	144
Sept. 2024	5374	5156	218
Oct. 2024	7055	6308	747
Nov. 2024	5784	5086	698

Am ales pe parcurs să mai adaug și o coloana numită ”Intermitentă” pentru productia de Hidrocarburi, Eolian, Foto si una numită ”Constantă” pentru productia de Carbune, Nuclear, Ape, Biomasa. Am făcut suma coloanelor corespunzătoare și le-am pus in tabel.

Date	Consum	Productie	Intermitent	Constant	Sold
Nov. 2023	5402	5384	1913	3471	18
Dec. 2023	6753	7059	3012	4047	-306
Jan. 2024	8223	7574	2014	5560	649
Feb. 2024	6281	6690	1709	4981	-409
Mar. 2024	4808	6337	2301	4036	-1529
Apr. 2024	5334	5045	1833	3212	289
May 2024	6011	5126	840	4286	885
Jun. 2024	5847	5449	1951	3498	398
Jul. 2024	7622	6073	2123	3950	1549
Aug. 2024	5659	5515	2057	3458	144
Sept. 2024	5374	5156	2367	2789	218
Oct. 2024	7055	6308	3903	2405	747
Nov. 2024	5784	5086	2182	2904	698

3 Implementarea algoritmilor

3.1 Adaptarea tabelului de valori pentru regresie

- Am utilizat *bucketing-ul* pentru a transforma soldul într-o variabilă discretă.
- Am ales sa impartim intervalele uniform. Regulile de separare sunt in tabelul de mai jos.

Interval Sold	Etichetă
< -500	Negativ Mic
$[-500, 500]$	Echilibrat
$[500, 1500]$	Pozitiv Mic
> 1500	Pozitiv Mare

Date	Consum	Producție	Intermitent	Constant	Sold
Nov.2023	5402	5384	1913	3471	Echilibrat
Dec.2023	6753	7059	3012	4047	Echilibrat
Jan.2024	8223	7574	2014	5560	Pozitiv Mic
Feb.2024	6281	6690	1709	4981	Echilibrat
Mar.2024	4808	6337	2301	4036	Negativ Mic
Apr.2024	5334	5045	1833	3212	Echilibrat
May.2024	6011	5126	840	4286	Pozitiv Mic
Jun.2024	5847	5449	1951	3498	Echilibrat
Jul.2024	7622	6073	2123	3950	Pozitiv Mare
Aug.2024	5659	5515	2057	3458	Echilibrat
Sep.2024	5374	5156	2367	2789	Echilibrat
Oct.2024	7055	6308	3903	2405	Pozitiv Mic
Nov.2024	5784	5086	2182	2904	Pozitiv Mic

- Mai departe am discretizat intervalele pentru variabile continue. Am luat pentru fiecare coloana valoare cea mai mica si cea mai mare, am rotunjit in sus, pentru valoarea maxima si in jos pentru valoarea minima, si am impartit intervalul mare in intervale mai mici si egale.

Interval Consum	Etichetă
$[4500-5500]$	Mic
$[5501-6500]$	Mediu
$[6501-7500]$	Mare
$[7501-8500]$	Foarte Mare

Tabela 1: Reguli pentru etichetarea Consumului.

Interval Producție	Etichetă
$[5000-6000]$	Mic
$[6001-7000]$	Mediu
$[7001-8000]$	Mare

Tabela 2: Reguli pentru etichetarea Producției.

Interval Intermitent	Etichetă
$[800-2000]$	Mic
$[2001-3200]$	Mediu
$[3201-4400]$	Mare

Tabela 3: Reguli pentru etichetarea Intermitentului.

3.2 Evaluarea performantei modelelor

Mai jos avem rezultatele evaluarii performanței, folosit metrici precum RMSE sau MAE. Pentru antrenare am folosit 80% din datele din tabelul construi la inceput si 20% pentru teste.

Model	RMSE	MAE
ID3	0.0	0.0
Bayes	0.75	1.12

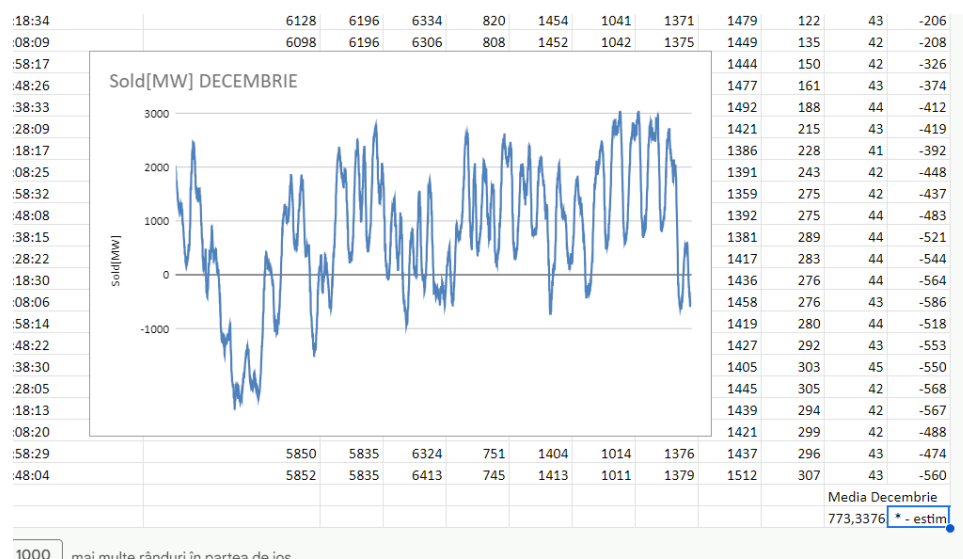
Tabela 4: Performanta modelelor

4 Prezentarea rezultatelor

În urma rularii ambilor algoritmi, am obținut următoarele rezultate:

- Bayes: Echilibrat
- ID3: Pozitiv Mic

Am calculat separat media soldului pentru luna Decembrie 2024. Si conform graficului de mai jos, se poate vedea ca media este de 773MW. Care conform regulilor setate la inceput, se incadreaza la eticheta Pozitiv Mic. Deci ID3 este cel care prezice corect rezultatul.



5 Concluzii

5.1 Ce am învățat?

Lucrând la această temă, am citit mai multe informatii despre metrice de performanță, si mai multe tipuri de bucketing. Am încercat sa aleg cele mai bune variante pentru tematica primită. Am observat ca pe cazul meu, ID3 si-a facut mai bine treaba decat Bayes.

5.2 Cum poate fi îmbunătățită metoda?

În soluția mea, am folosit foarte puține date, făcând media pe fiecare lună. Cred că dacă aș folosi ceva mai multe date, cel mai probabil și Bayes ar avea o performanță mai bună.

6 Referințe

- Transelectrica SEN Grafic: <https://www.transelectrica.ro>