

Csalásfelderítés az energetikában adatbányászati módszerekkel

Kovács Levente

+36-70-430-9965

Konzulensek: Kertész Dávid, Dr. Vokony István

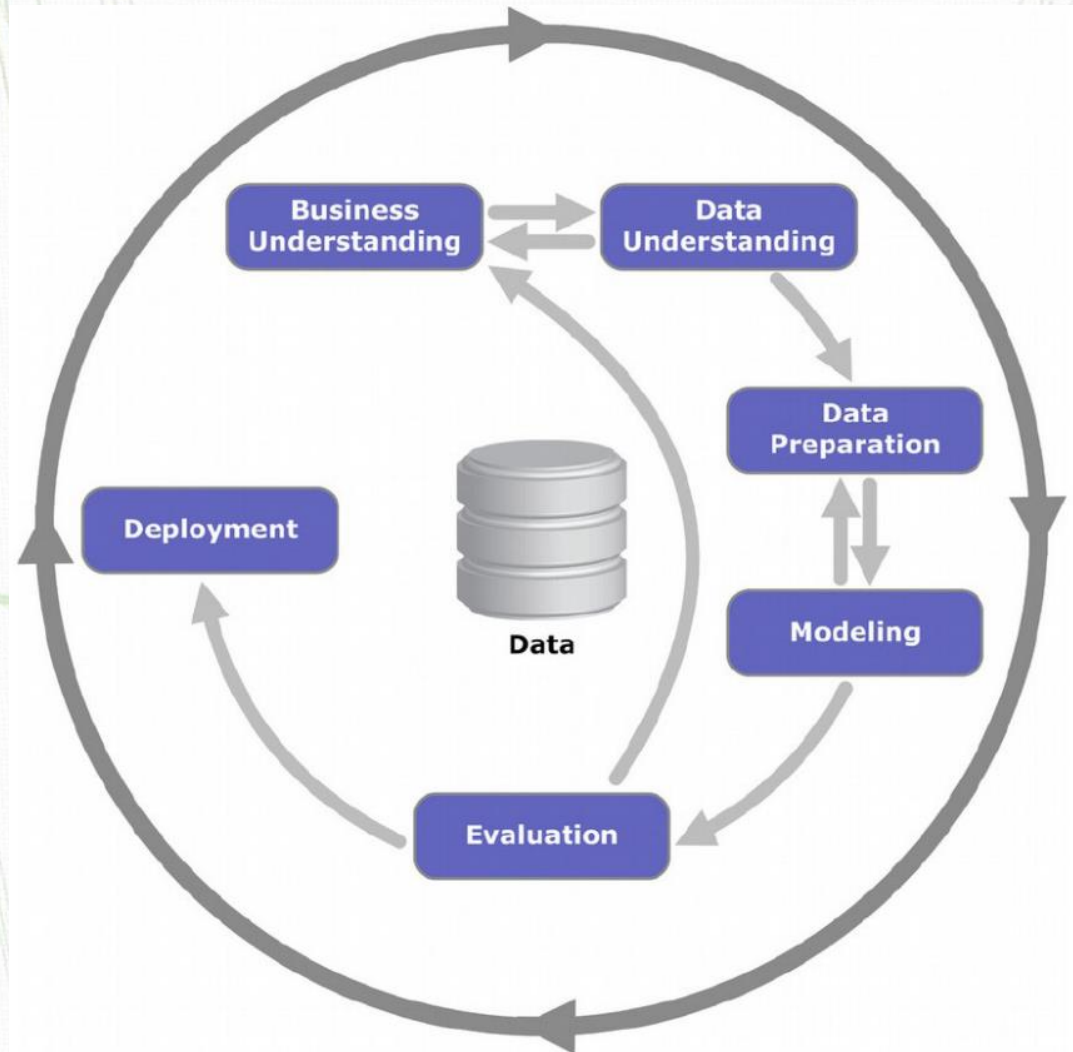
XI. Mechwart András Ifjúsági Találkozó – 67. Vándorgyűlés
2021. szeptember 21. Kedd



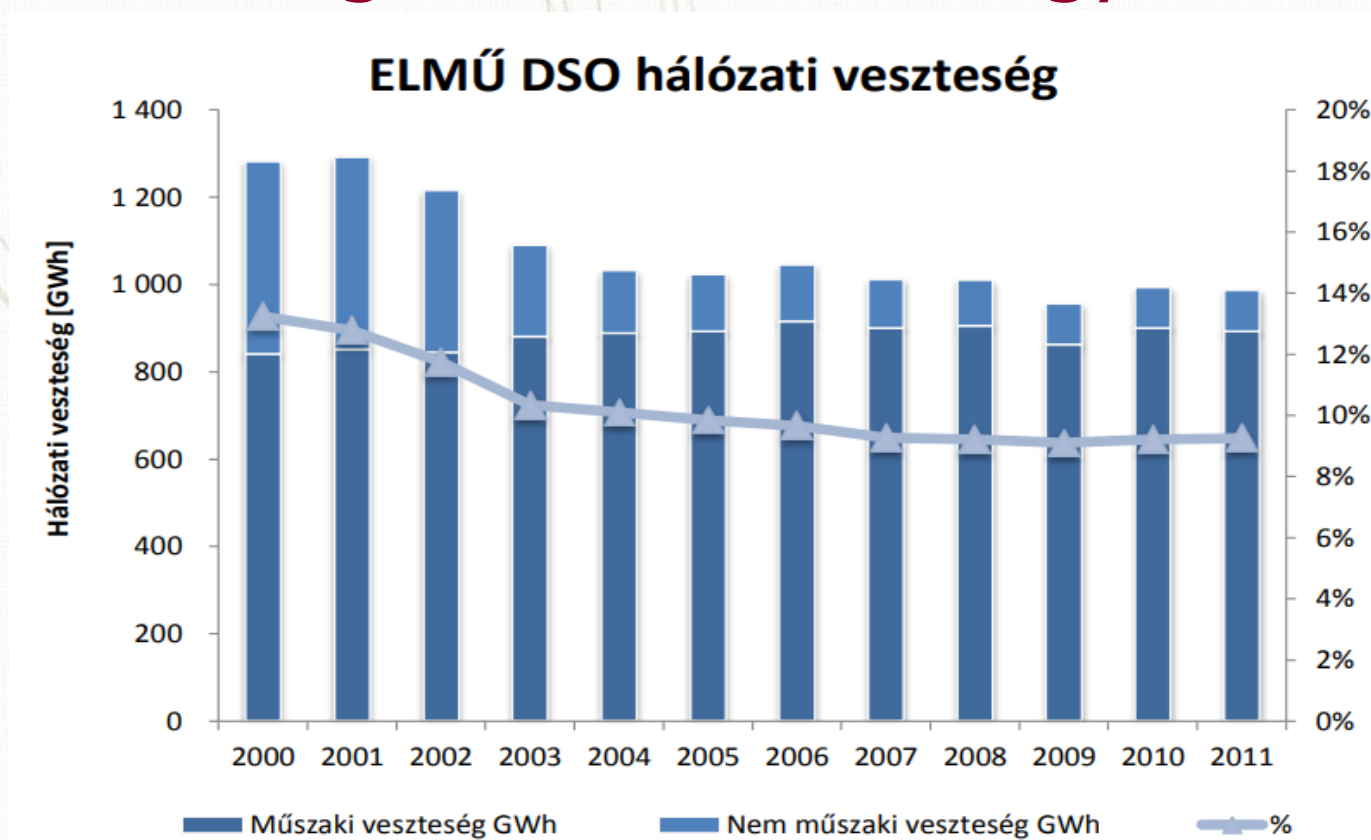
Villamos Energetika Tanszék
Villamos Művek és Környezet Csoport

CRISP-DM módszertan

- Üzleti célok megértése
- Adatok megértése
- Adatok előkészítése
- Modellezés
- Üzleti kiértékelés
- Telepítés
- Visszacsatolás



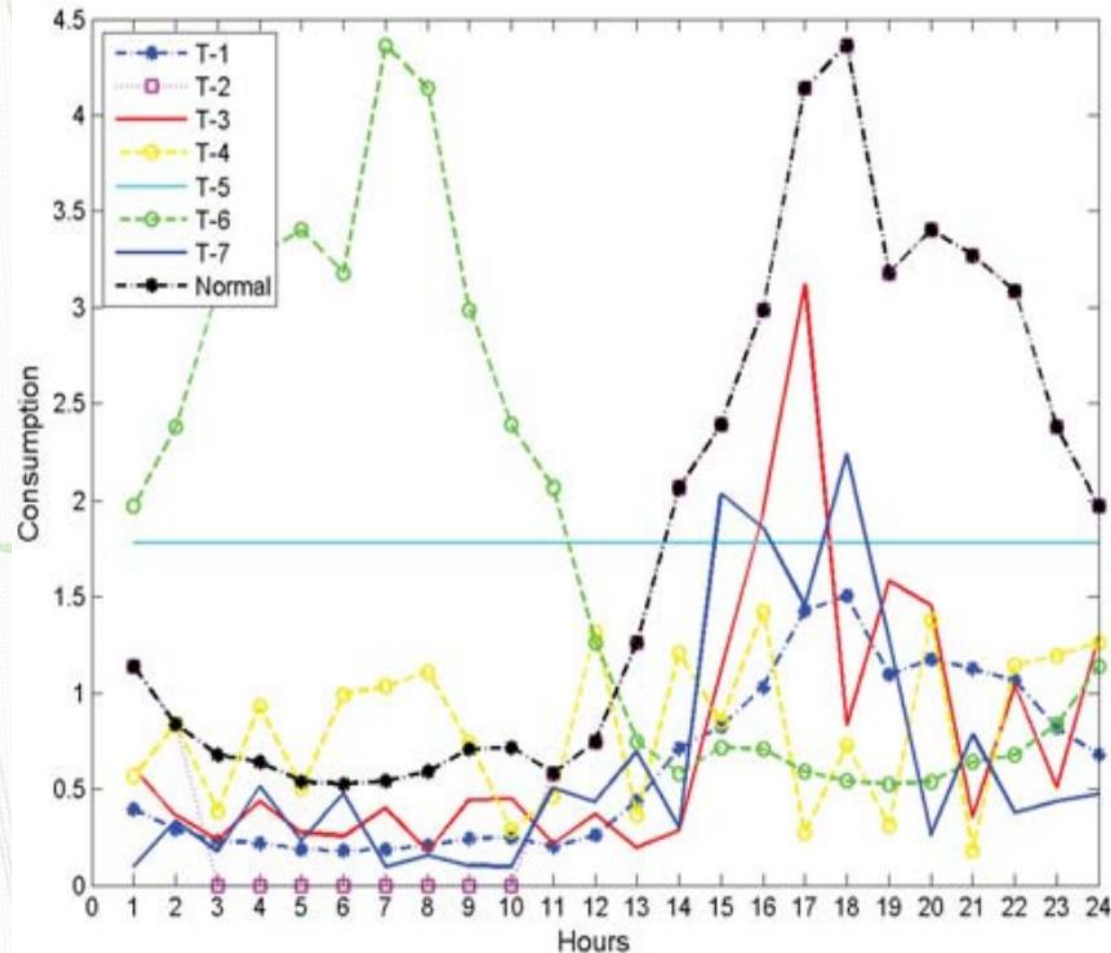
Villamosenergia-visszaélések Magyarországon



- Hálózati veszteség: betáplált villamosenergia 6-9%
- Áramlopás: betáplált villamosenergia **0,5-1%-a**

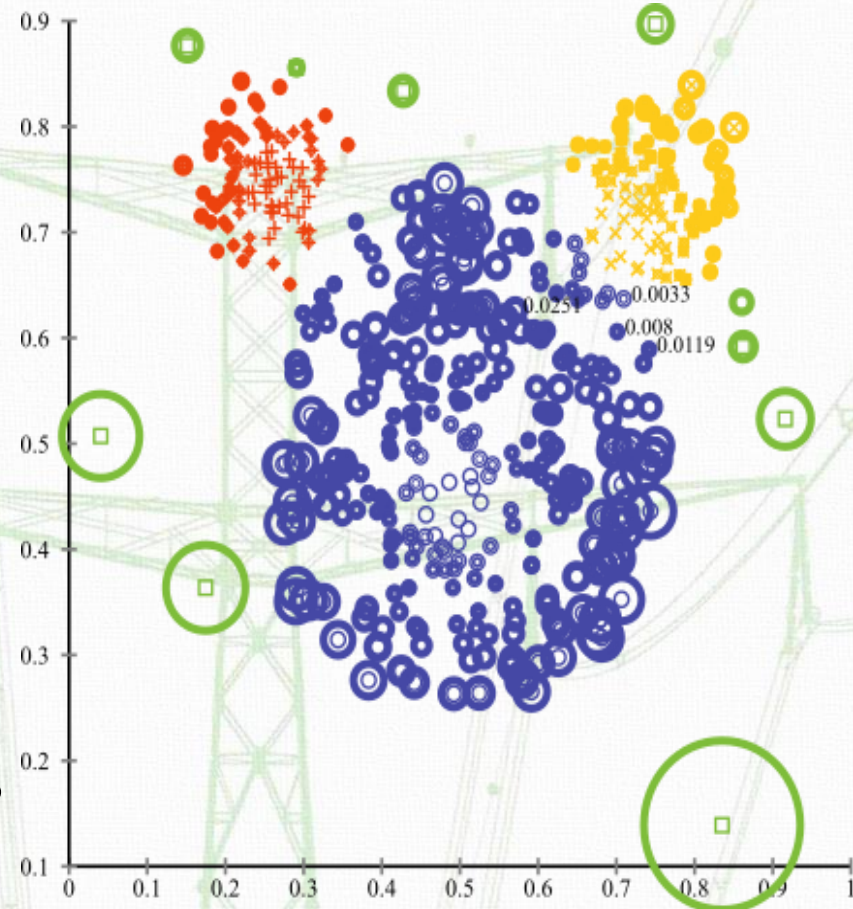
Visszaélési lehetőségek, hatásuk

- Több különböző módszer:
- Mérők manipulációja
- Mérőórán kívüli vételezés
- Számít a mérő típusa
- Identifikáció fogyasztói profilokból



Lehetséges módszerek

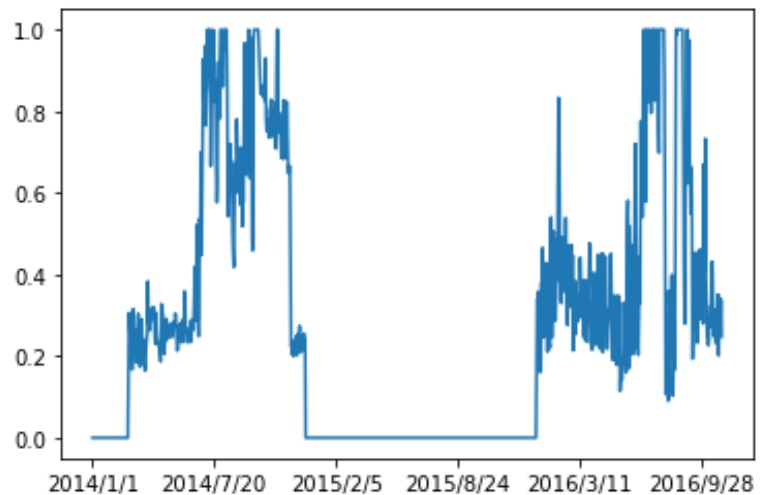
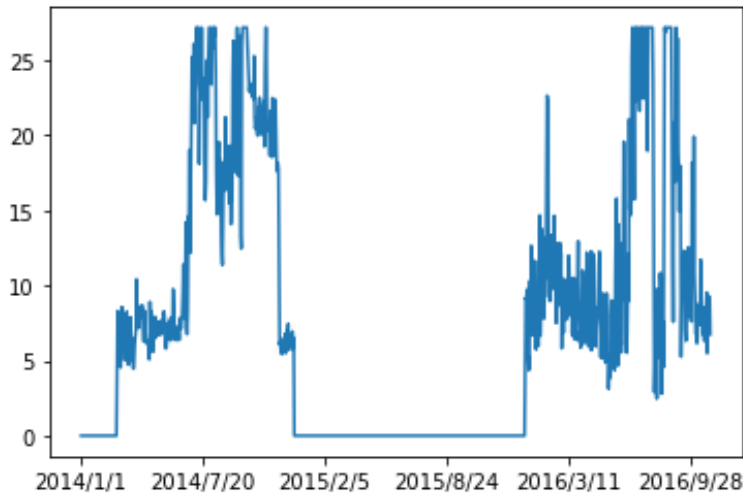
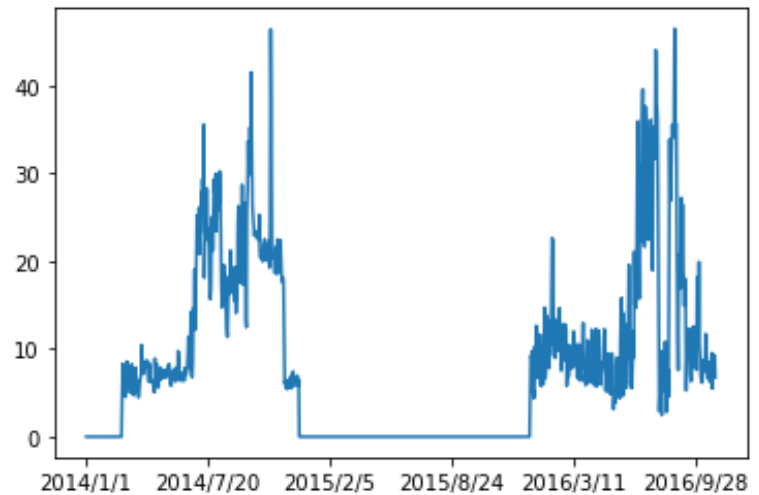
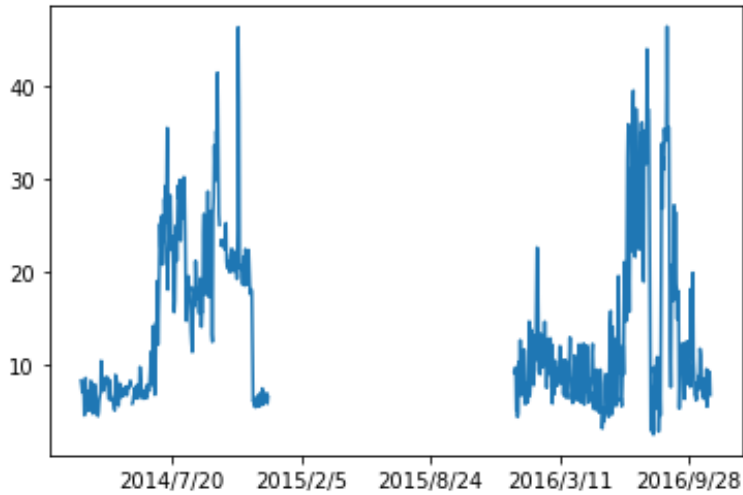
- Szükséges a modellezési eljárások tulajdonságait is vizsgálni adatkeresés előtt
- Felügyelt tanulás:
 - Döntés előzetes tudás alapján
 - Osztályozás, regresszió
- Nem felügyelt tanulás
 - Előzetes tudás nincs
 - Anomália detekció, klaszterezés



Felhasznált adatbázis

- State Grid Corporation of China (SGCC) által közzétett adathalmaz
- 42372 fogyasztó adatai, 2014.01-2016.10 között
- Csalási tevékenység jelölve (3615 eset)
- -> a visszaélések aránya 8,53%
- Egynapos felbontás

Adatelőkészítés

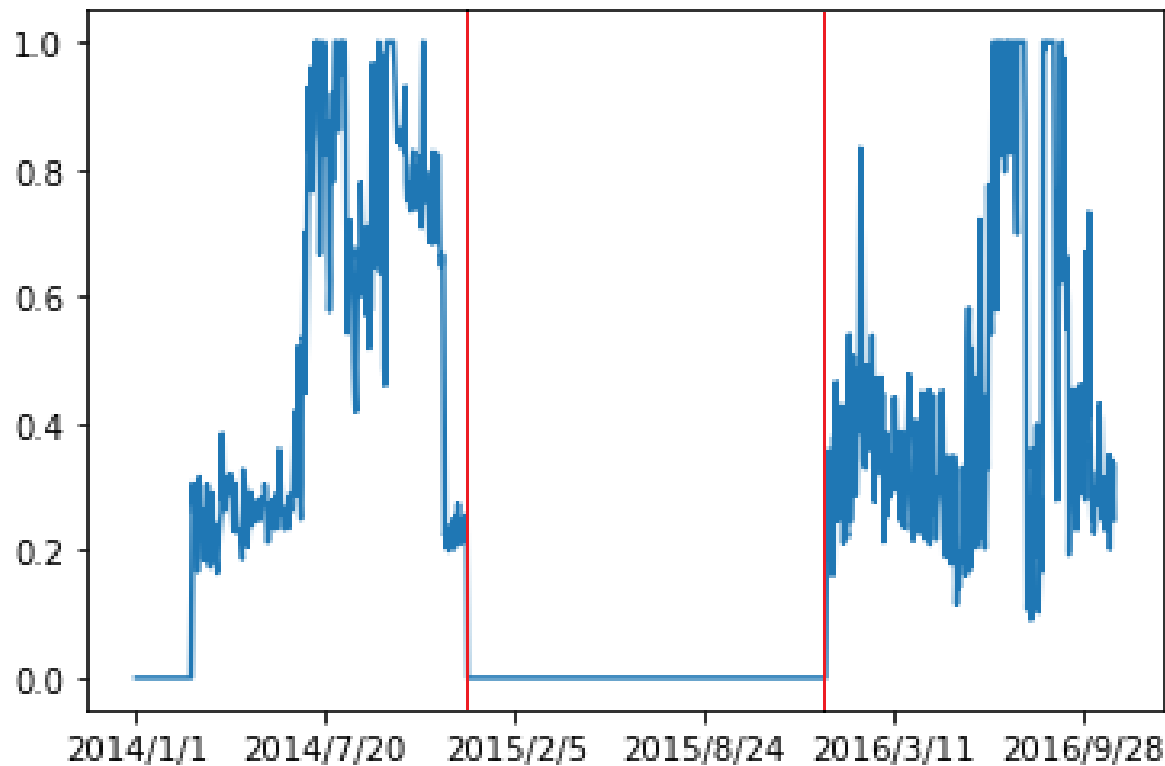


Modellezés

- Több modellező algoritmus:
 - - Logistic regression (LR)
 - - Support-vector machine (SVM)
 - - Random forest (RF)
- Train/test = 67%/33%
- Hiányos intervallumok
- Kiegyensúlyozatlan adathalmaz (~9%)

Modellezés

- Hiányos intervallumok
- Külön vizsgálat jobb eredményre juthat



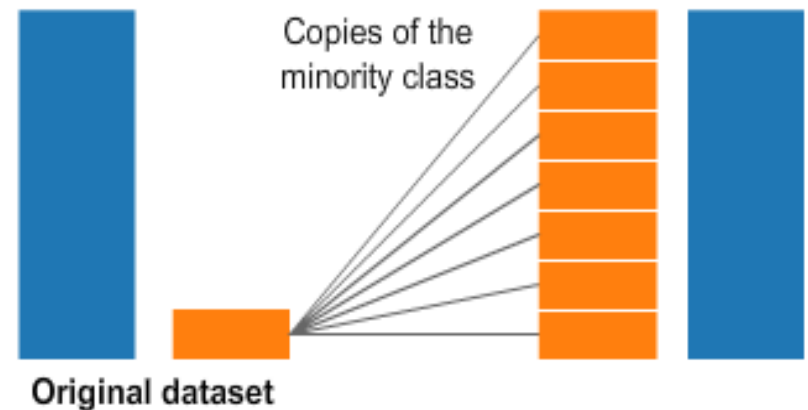
Modellezés

- Kiegyensúlyozatlan adathalmaz (~9%):
- -> Mintavételi eljárások
- - Undersampling
- - Oversampling

Undersampling



Oversampling

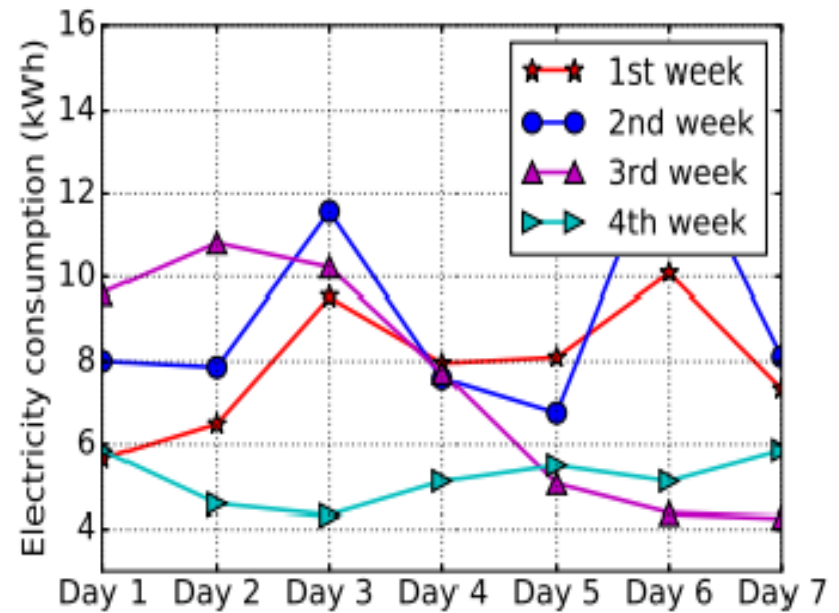
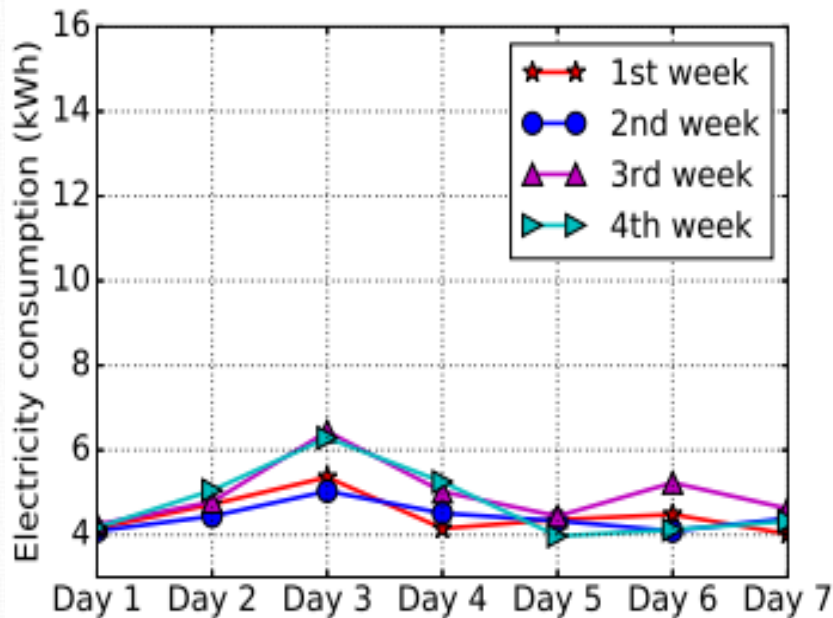


Modellezés - eredmények

	accuracy	f1-score	AUC-score
LR	0.9044	0.0864	0.5199
LR 1	0.9081	0.0333	0.5074
LR 2	0.9084	0.0225	0.5049
LR US	0.6419	0.6493	0.6424
LR OS	0.7011	0.7077	0.7012
SVM	0.9089	0.005	0.5012
SVM US	0.7031	0.7082	0.7037
RF	0.911	0.0787	0.5203
RF US	0.6812	0.6919	0.6819

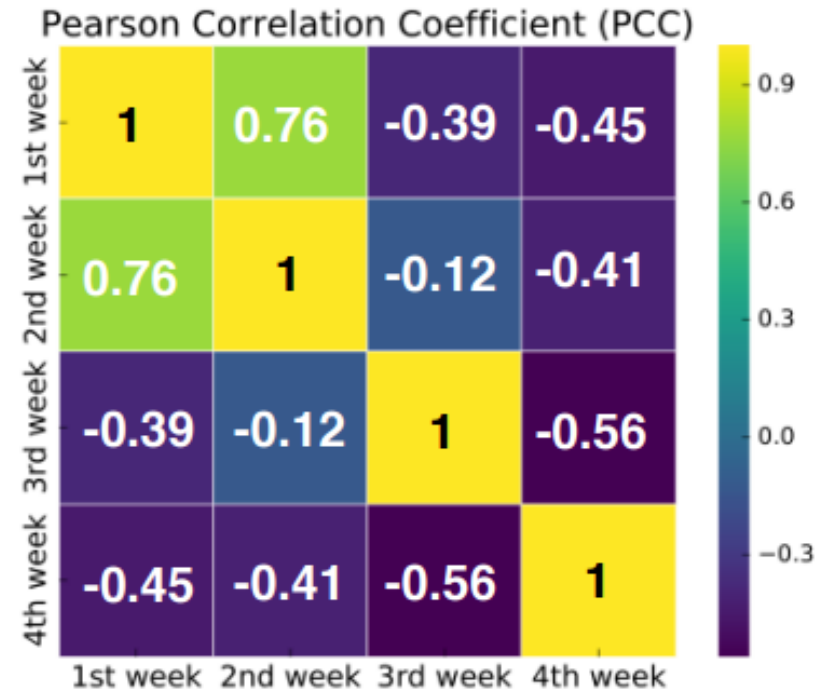
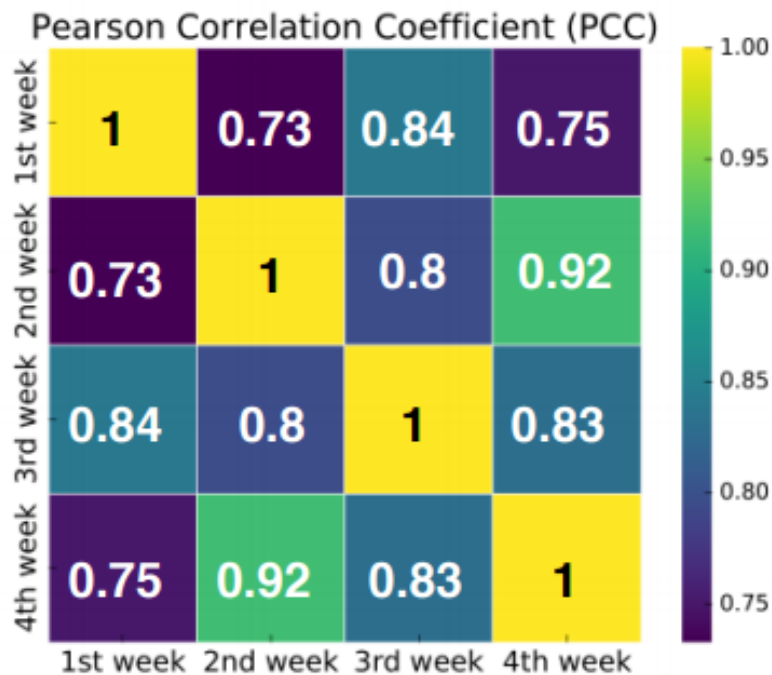
Egyéb lehetőségek – kétdimenziós modell

- Korreláció alapú analízis



Egyéb lehetőségek – kétdimenziós modell

- Korreláció alapú analízis



Köszönöm a figyelmet!



Villamos Energetika Tanszék
Villamos Művek és Környezet Csoport