

# Ensemble szegmentáló módszerek idősoros adatokon

Önálló laboratórium szóbeli beszámoló

---

Készítette: Czotter Benedek (TFB4FY)

Konzulensek: Dr. Szűcs Gábor, Németh Marcell

2024/25/2

# Feladat leírása

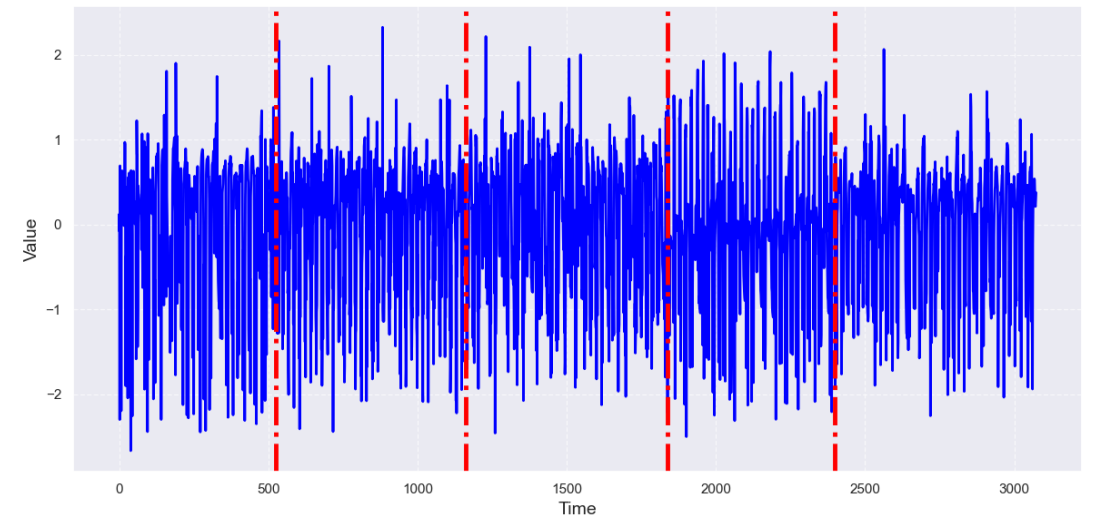
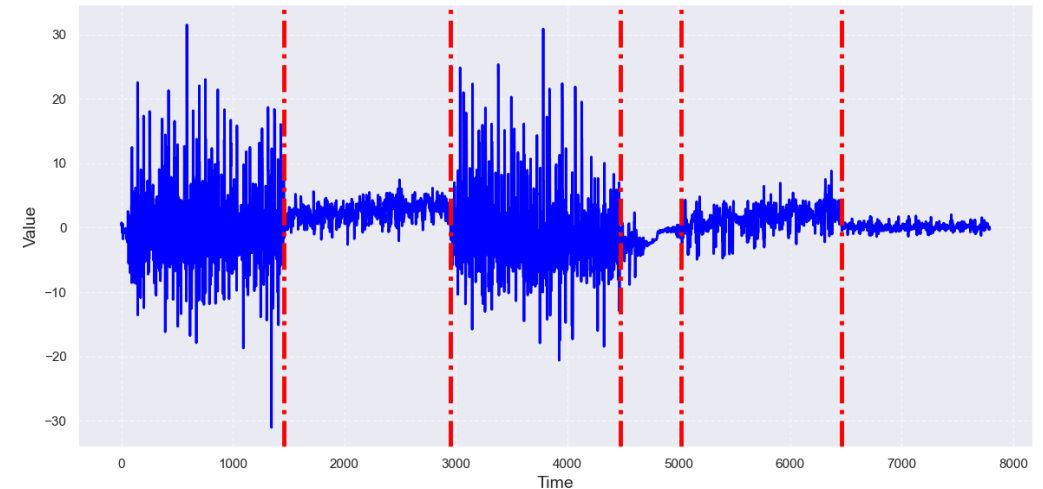
---

- Van több szegmentáló módszer
- Hogyan lehet kiválasztani  $k$  darabot, hogy pontosabb eredményt érjünk el?
- Ensemble módszer hogyan tud dönteni

# Idősor szegmentálás

---

- kisebb, egymástól elkülönülő szakaszokra bontás
- egyes szakaszok statisztikai vagy szerkezeti tulajdonságai hasonlóak legyenek



# Idősor szegmentálás területei

---

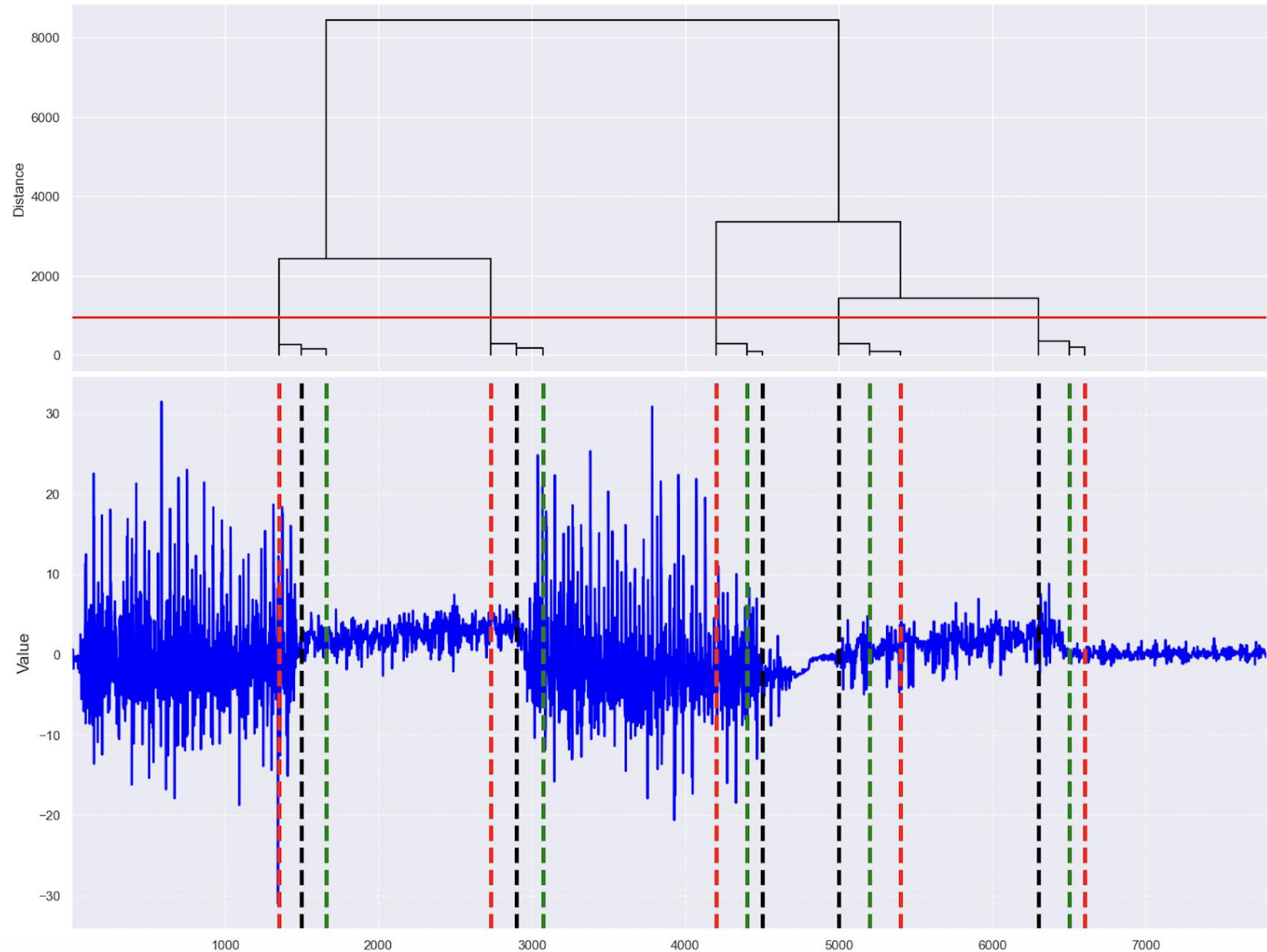
- Pénzügy
- Egészségügy és orvosi diagnosztika
- Ipari gépmonitorozás
- Beszédfelismerés és jelfeldolgozás
- Közlekedés és mobilitás

# State-of-the-art algoritmusok

<b>Statisztikai és szabályalapú módszerek</b>	<b>Klaszterezési és tanulási alapú módszerek</b>	<b>Mélytanulás alapú módszerek</b>
<p>Pruned Exact Linear Time (PELT)</p> <p>Binary Segmentation (BinSeg)</p> <p>Bottom-Up</p> <p>Window-Based</p>	<p>Dynamic Time Warping KMeans (DTW KMeans)</p> <p>Classification Score Profile (CLaSP)</p>	<p>Temporal Convolutional Network (TCN) Autoencoder</p> <p>Long Short-term Memory (LSTM)</p> <p>Konvolúciós neurális háló (CNN)</p>

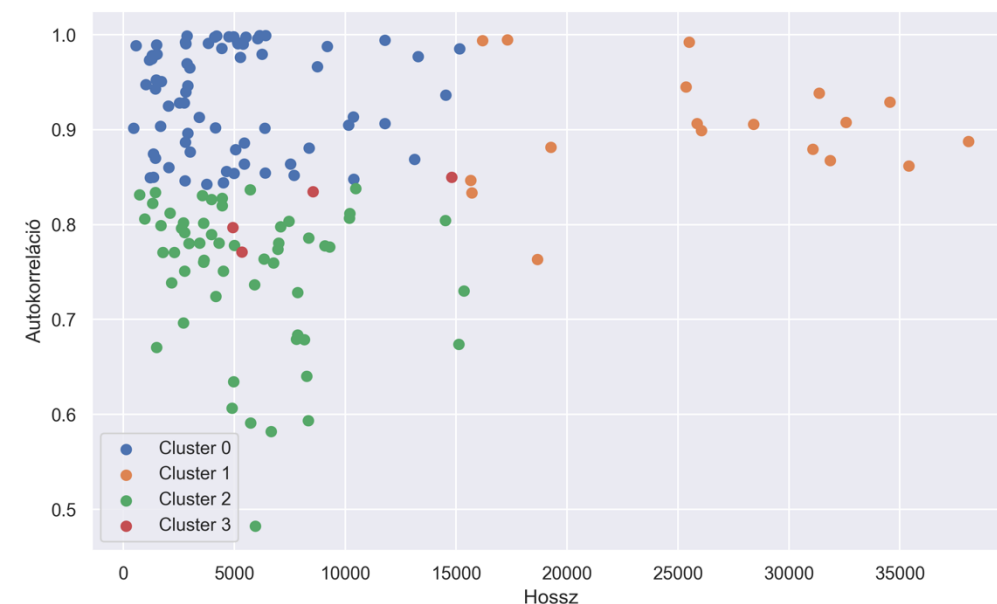
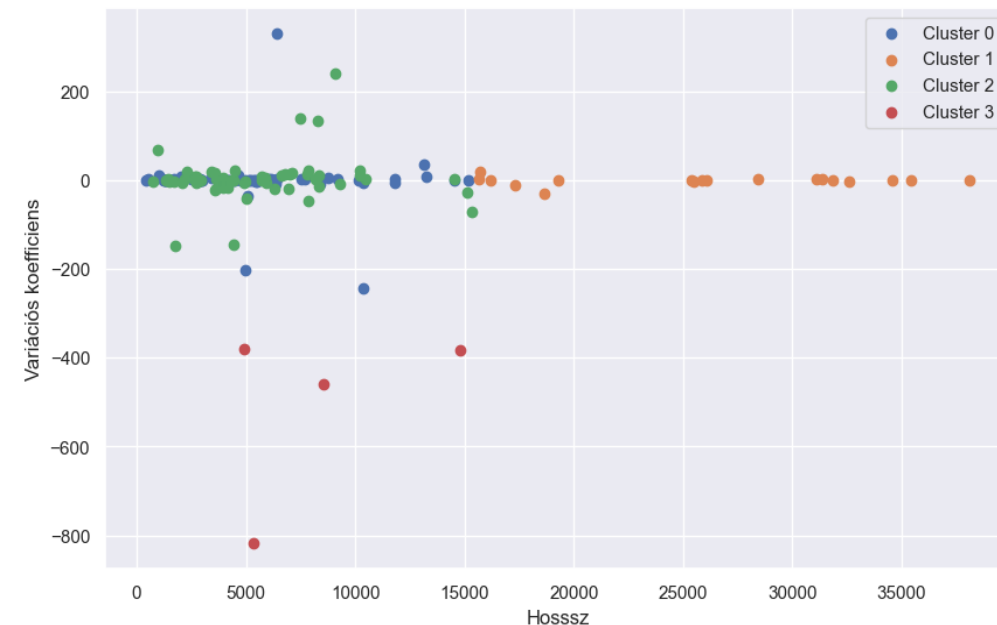
# Többségi szavazás idősor szegmentáláshoz

---

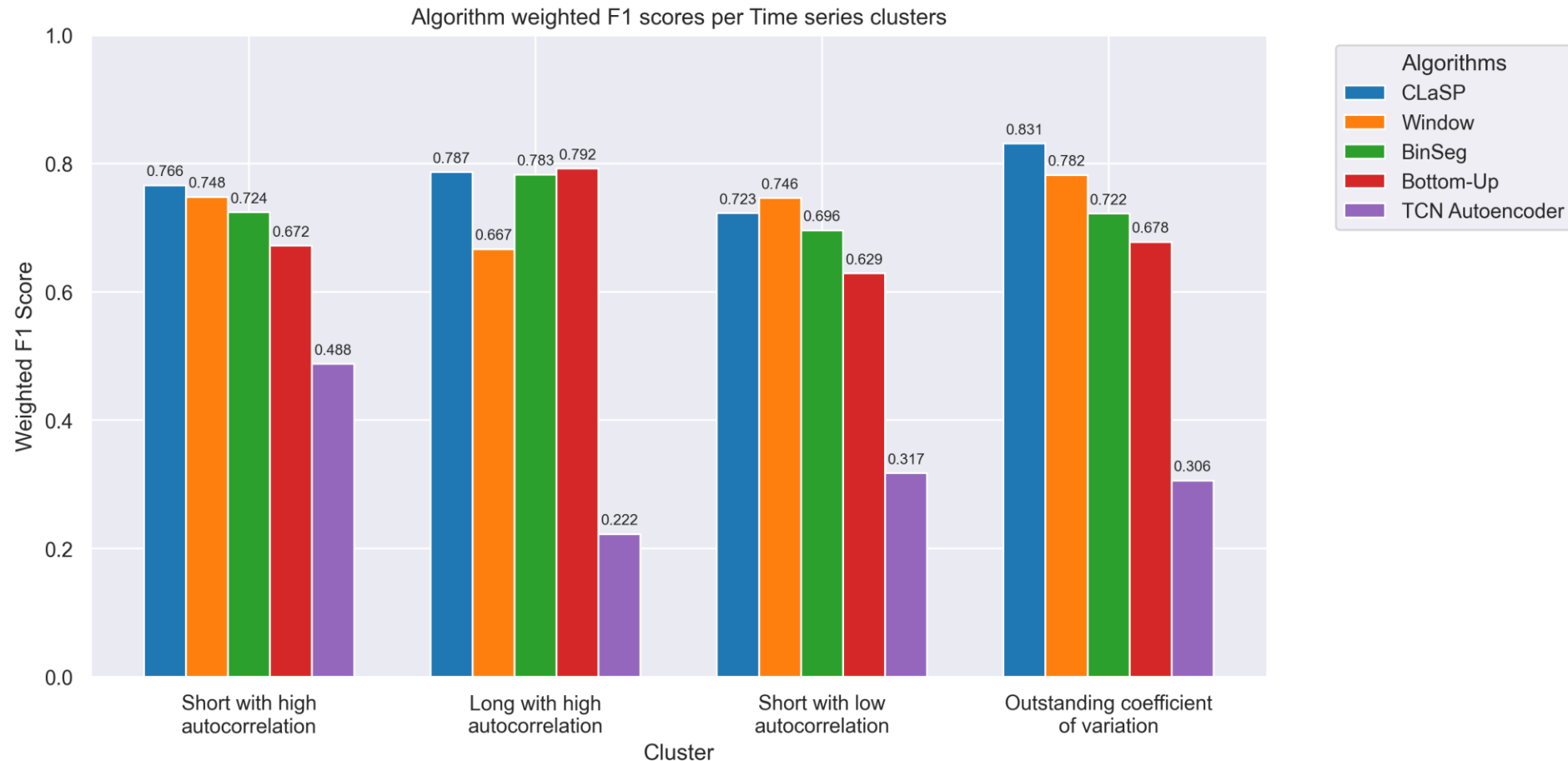


# Adott klaszteren legjobban teljesítő algoritmus választása

- Idősorok klaszterezése KMeans-sel
- Vizsgált paraméterek:
  - Hossz
  - Autokorreláció
  - Variációs koefficiens
  - Trend



# Adott klaszteren legjobban teljesítő algoritmus választása





# Eredmények


- Használt adathalmazok:
  - TSSB – 75 idősor
  - HAS – 250 idősor
- Teszthalmaz nagysága: 40%

Algoritmus	HAS adathalmaz		TSSB adathalmaz	
	F1 score	Covering	F1 score	Covering
CLaSP	0.7235	<b>0.6459</b>	<b>0.8890</b>	<b>0.8385</b>
Window	0.7462	0.6100	0.7705	0.5962
BinSeg	0.7385	0.6412	0.6560	0.3663
BottomUp	0.6940	0.6175	0.6376	0.5225
TCN autoencoder	0.3173	0.4592	0.4233	0.4472
PELT	<b>0.7835</b>	0.5759	0.6401	0.4215
CLaSP_Window_BinSeg	0.5538	0.6164	0.4395	0.6764
PELT_BinSeg_BottomUp	0.5335	0.5364	0.4260	0.4670
Window_BinSeg_BottomUp	0.5246	0.5955	0.5492	0.5492
Pre-clustering the time series	<b>0.7504</b>	<b>0.6368</b>	<b>0.8890</b>	<b>0.8385</b>

# Összegzés és kitekintés

---

- Idősorok megkülönböztetése statisztikai jellemzők alapján
- Egyes algoritmusok erősségeinek feltérképezése
- Robusztusabb algoritmus
- Jövőbeli feladatok:
  - Hierarchikus klaszterezés más módszerrel történő optimalizálása
  - TCN autoencoder fejlesztése
  - Idősorok megkülönböztetése további paraméterek alapján
  - Több előre szegmentált adathalmaz gyűjtése



# Köszönöm szépen a figyelmet!

- Elvégzett munka:
  - Szegmentáló algoritmusok optimalizálása
  - Ensemble modellek:
    - Hierarchikus klaszterezéssel
    - Kmeans és statisztikai jellemzők alapján
  - Robusztusabb, pontosabb eredmények elérése