


Újfajta, automatikus, döntési fa alapú adatbányászati módszer idősorok osztályozására



Hidasi Balázs

bhidasi@t-online.hu

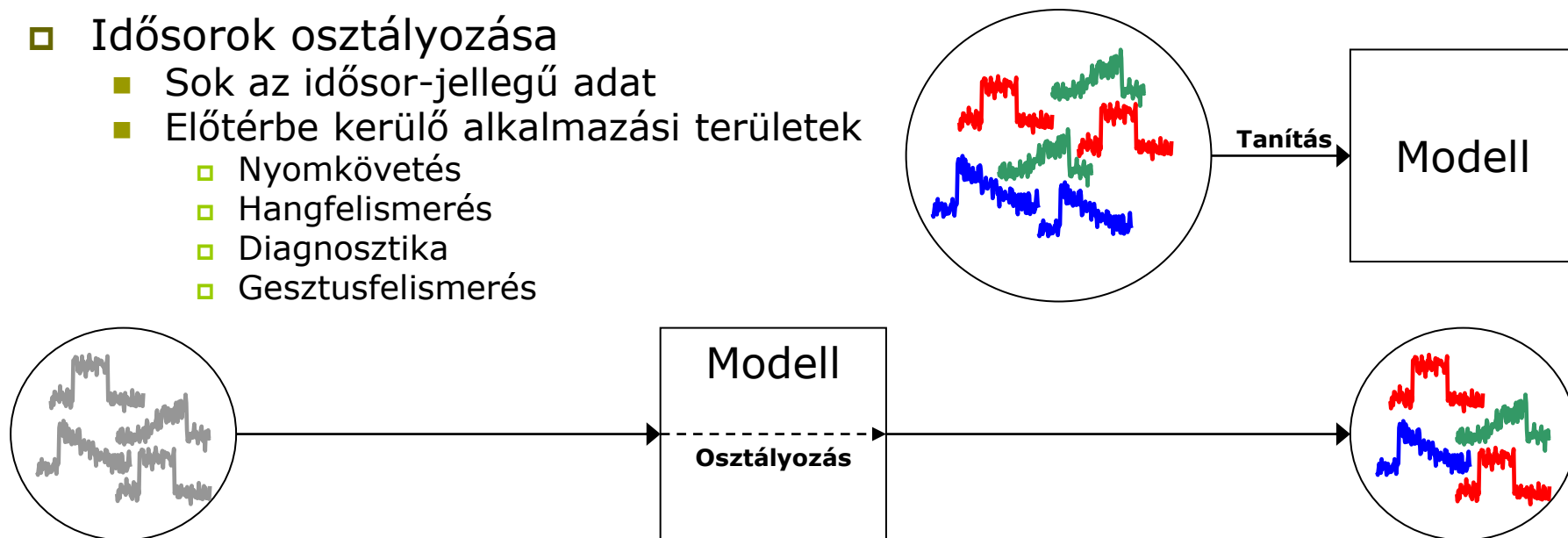
Konzulens: Gáspár-Papanek Csaba

Távközlési és Médiainformatikai Tanszék

2009. Január 8.

Motiváció

- Idősorok osztályozása
 - Sok az idősor-jellegű adat
 - Előtérbe kerülő alkalmazási területek
 - Nyomkövetés
 - Hangfelismerés
 - Diagnosztika
 - Gesztusfelismerés



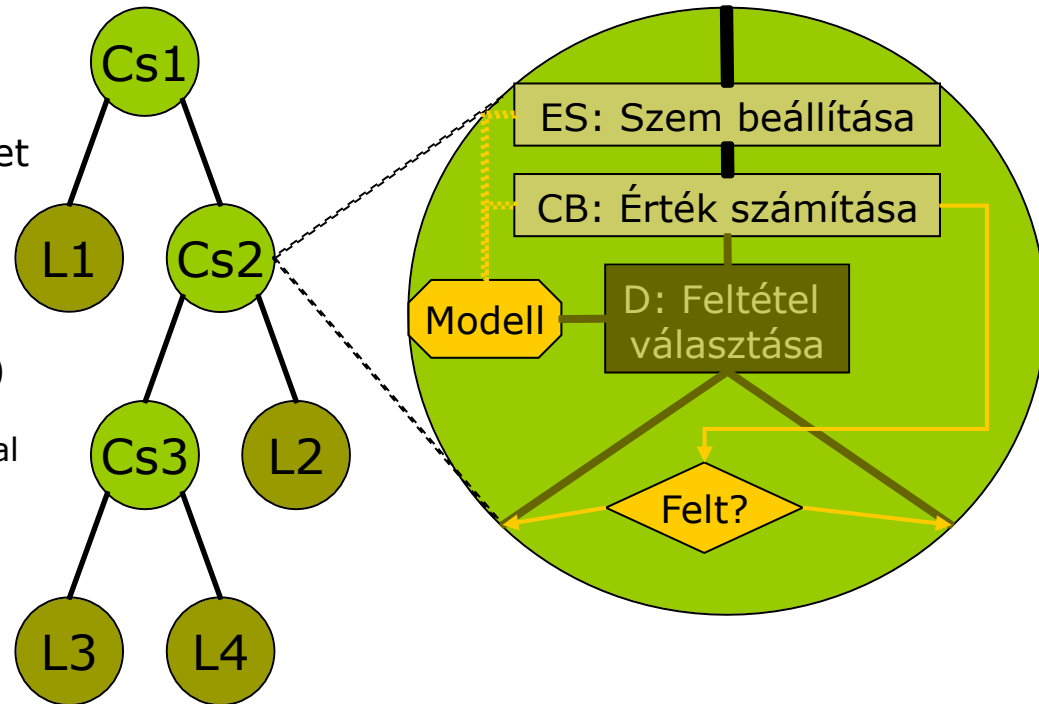
- Jelenlegi algoritmusok hátrányai
 - Klasszikus módszerek
 - Jelentős emberi munka (előkészítés)
 - Nem erre találták ki
 - Információvesztés (pontatlanság)
 - Általában nem magyaráz
 - Terület-specifikus algoritmusok
 - Más területen nem hatékony

Célok

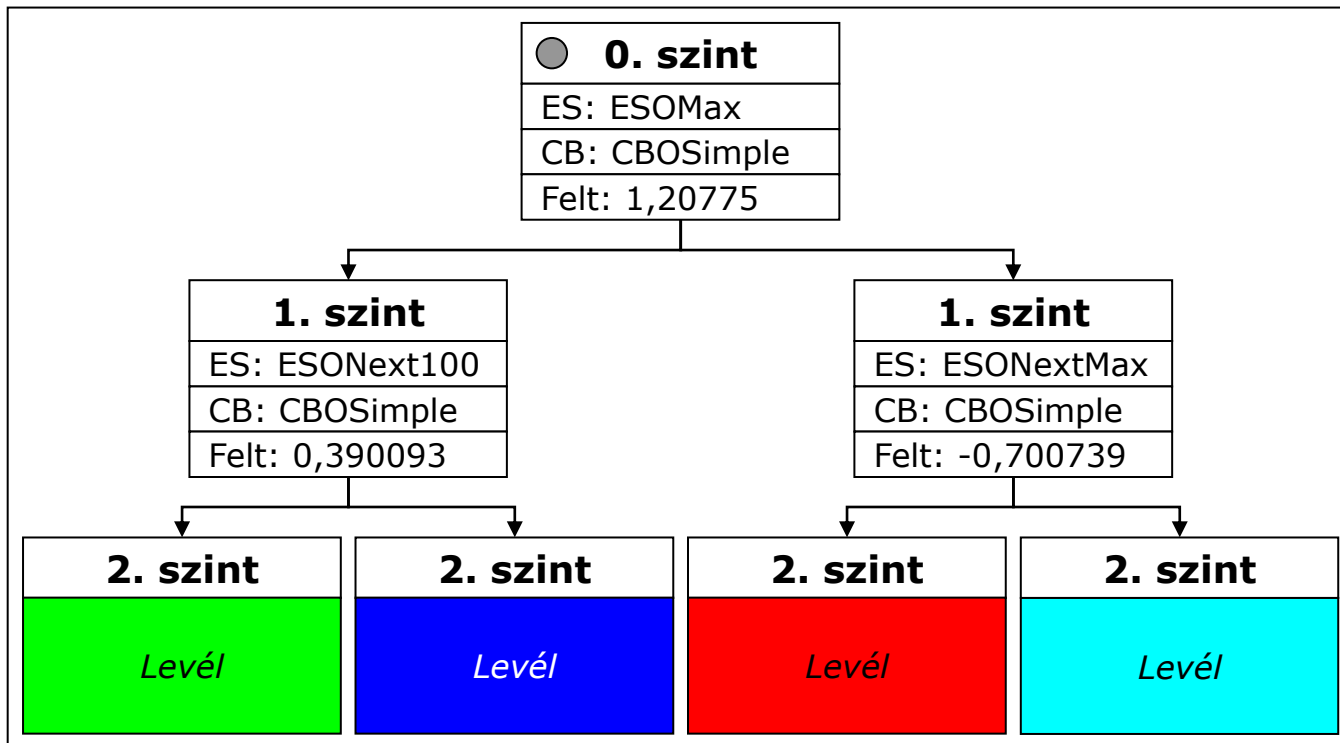
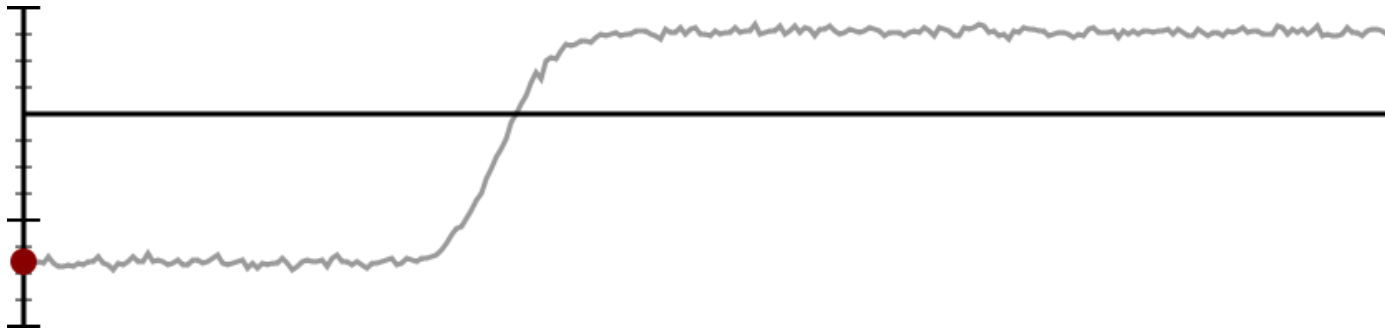
- Automatikus
 - Kevés emberi munka
 - Rövid előkészítési fázis
 - Minél több típus általános kezelése
 - Változók száma, osztályok száma, idősorok hossza, stb.
 - Több területen használható (általános)
- Pontos osztályozás
 - Magas találati arány
- Magyarázó
 - Könnyen értelmezhető modellt épít
 - Ellenőrizhető

ShiftTree – A módszer alapjai

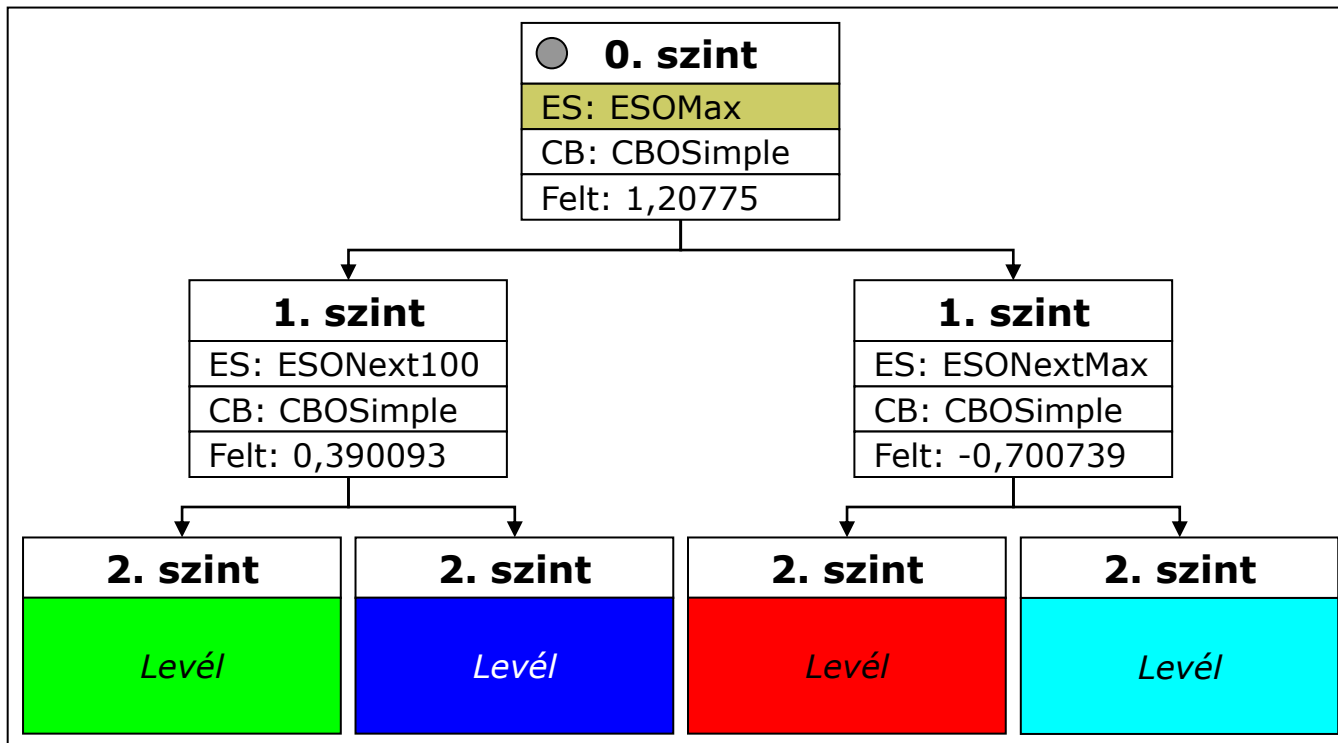
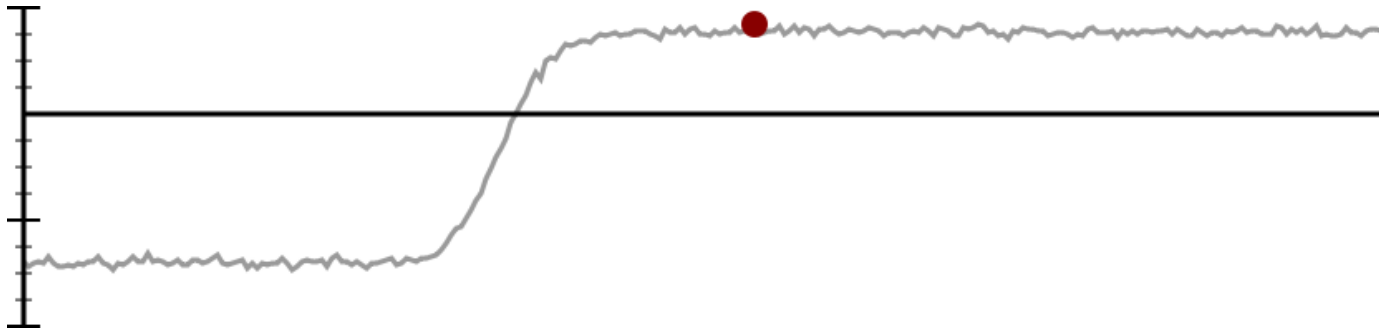
- Hibrid algoritmus
 - Döntési fa alap
 - Szerkezet
 - Vágások jóság értékei
 - Leállási feltételek
 - Módosított csomópont-szerkezet
- Moduláris felépítés
 - Szemtologató (EyeShifter)
 - ES-Operator (ESO)
 - Szem (pointer) mozgatás
 - Feltételállító (ConditionBuilder)
 - CB-Operator (CBO)
 - Érték származtatás a szem által mutatott értékből
 - (és környezetéből)
 - Döntő (Decider)
 - Vágási helyek vizsgálata
 - Jóságérték számítás
 - Optimális vágás választása a lehetőségekből
 - Feltétel kiszámítása



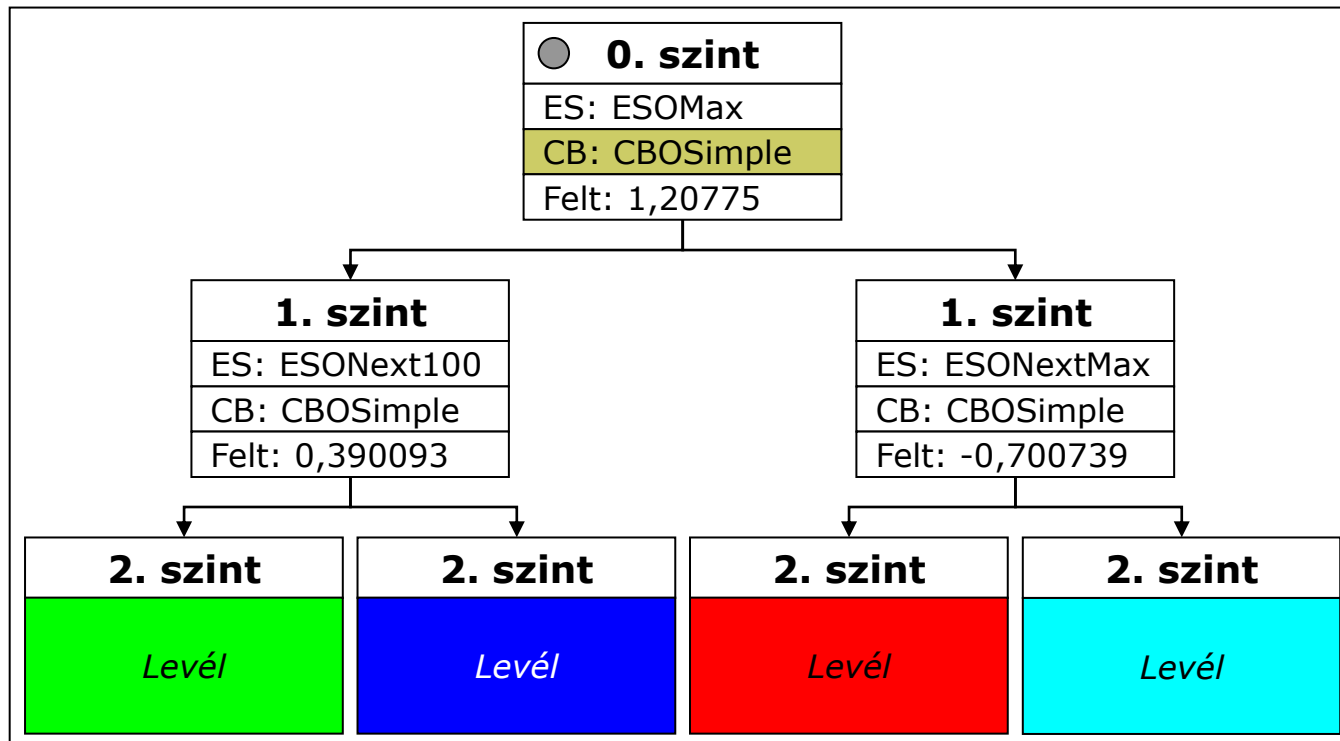
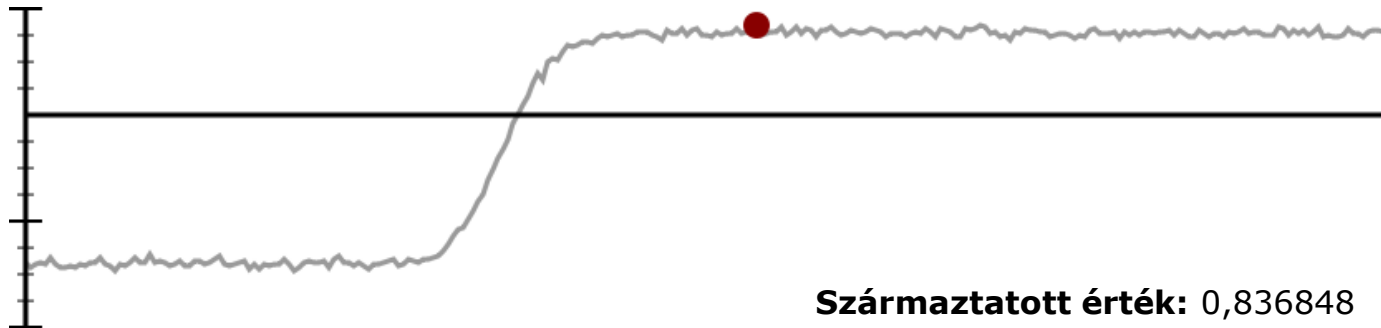
ShiftTree – Példa az osztályozásra



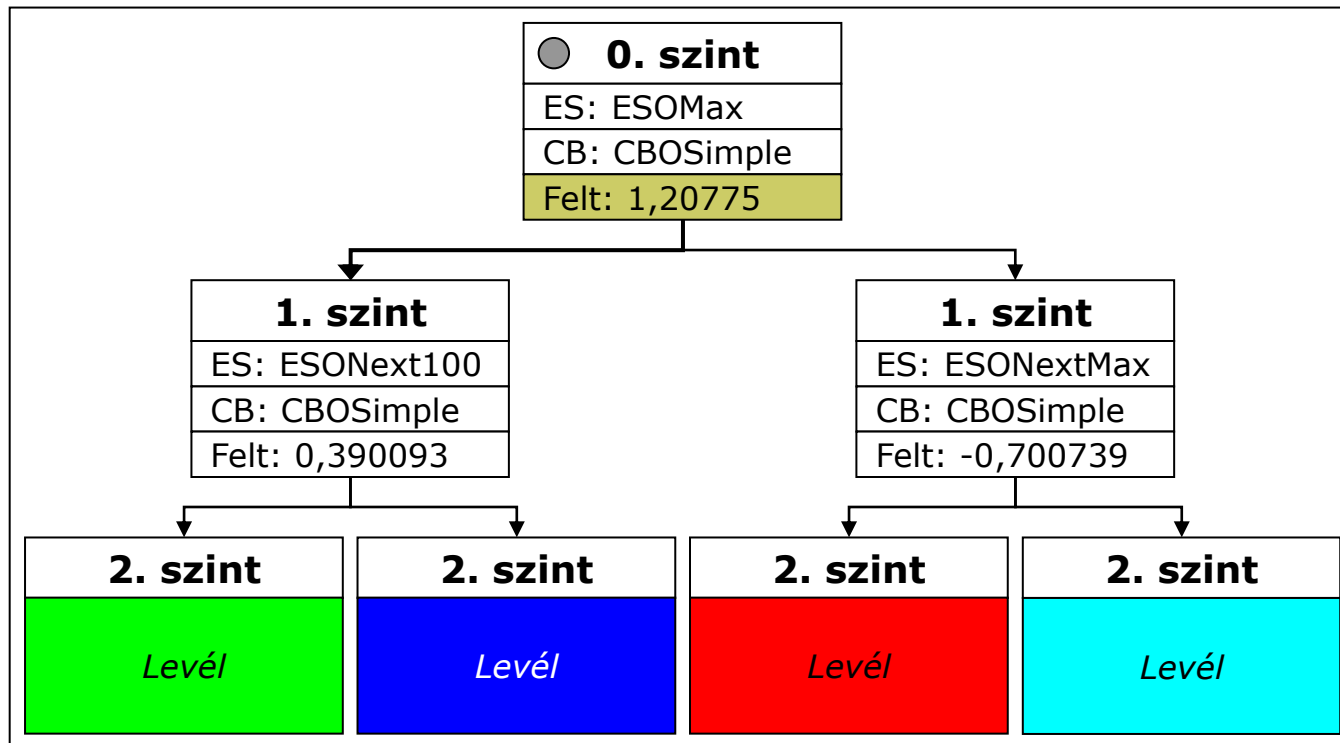
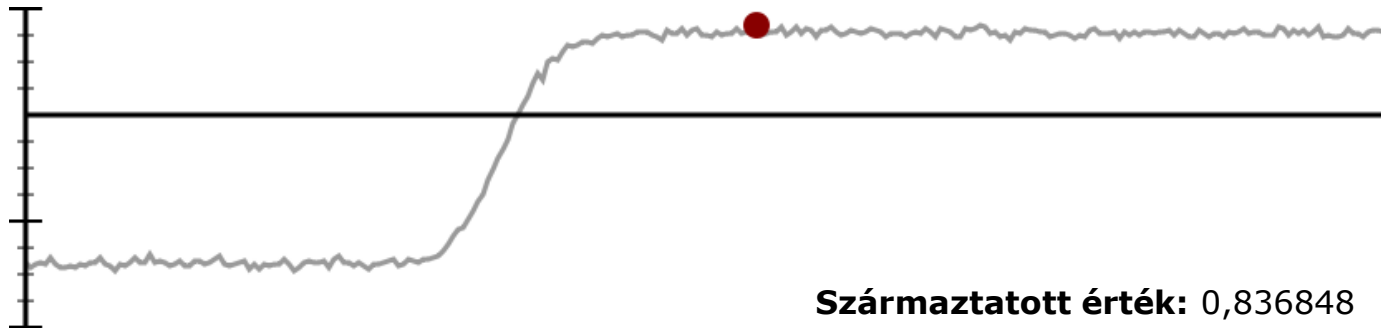
ShiftTree – Példa az osztályozásra



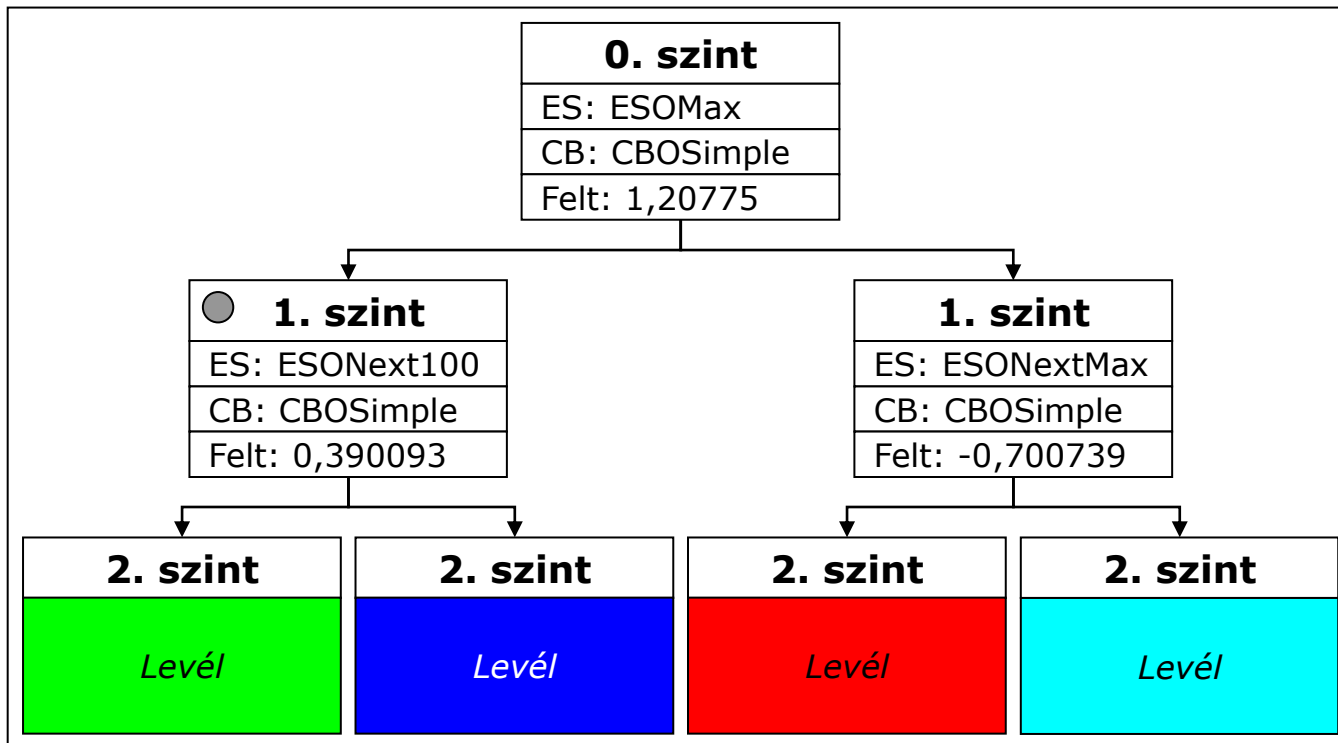
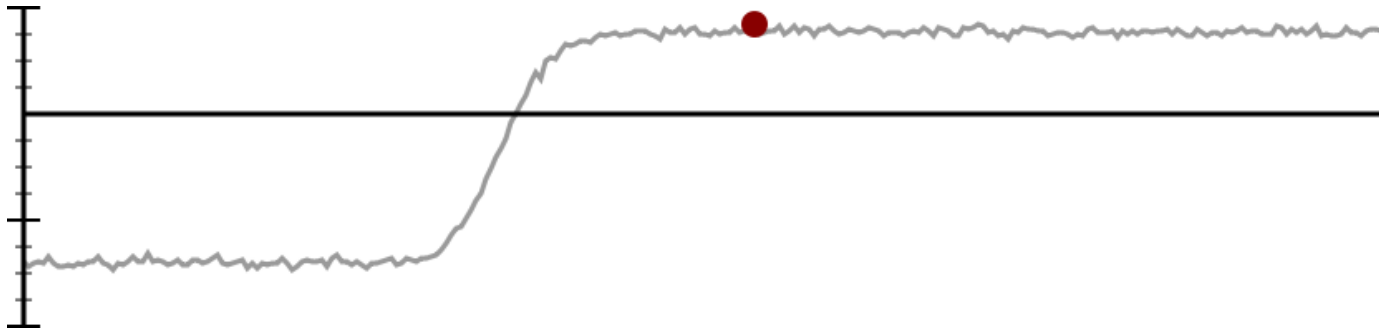
ShiftTree – Példa az osztályozásra



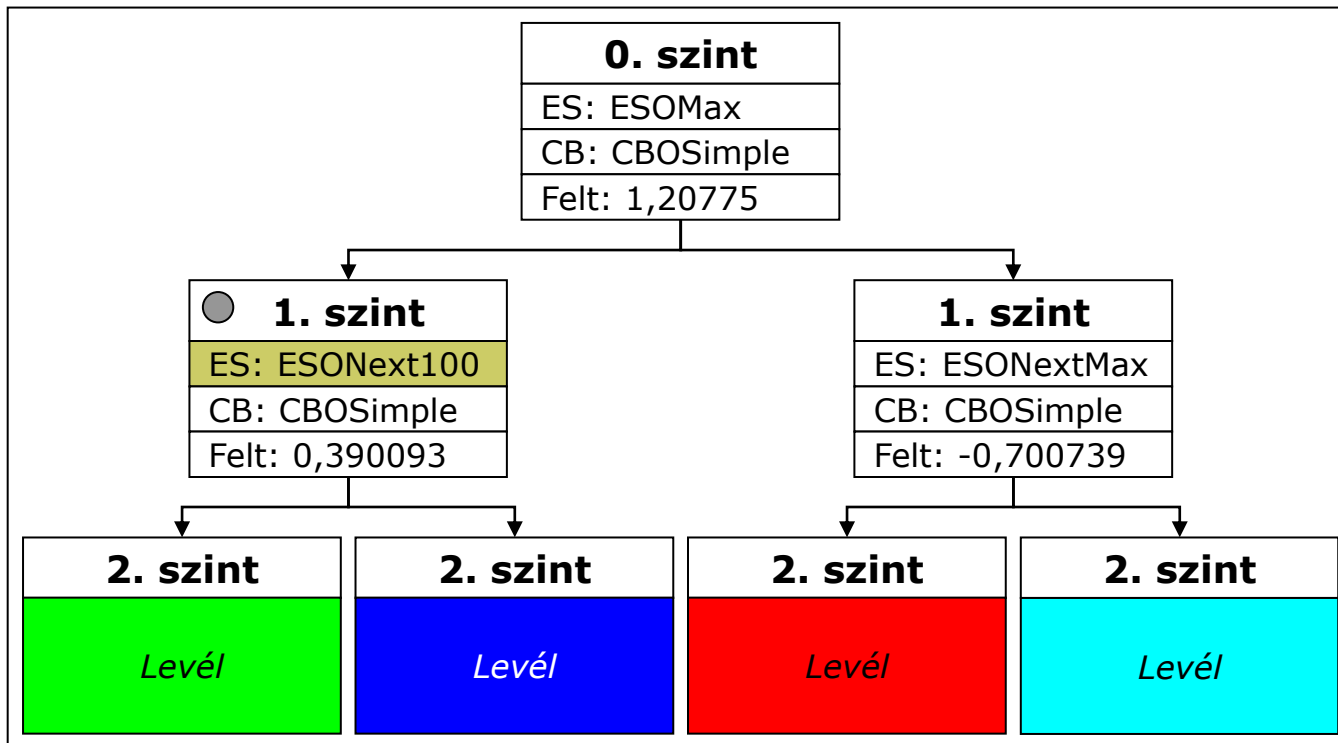
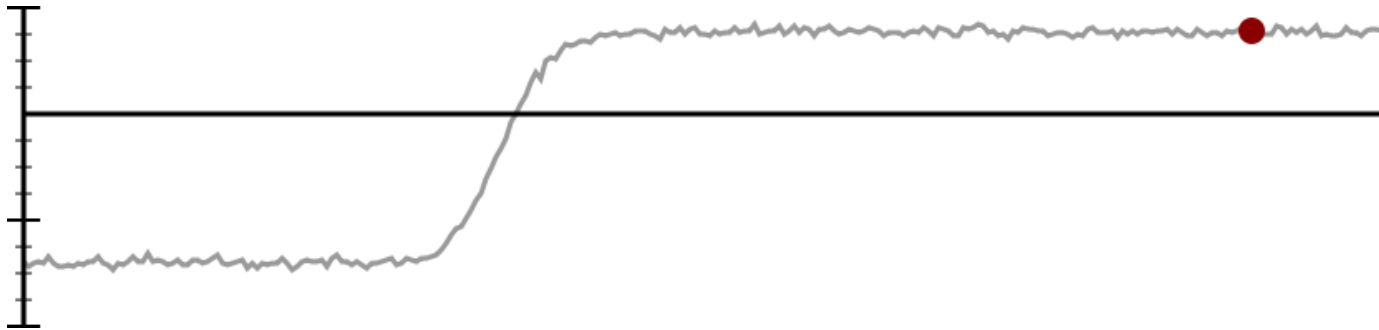
ShiftTree – Példa az osztályozásra



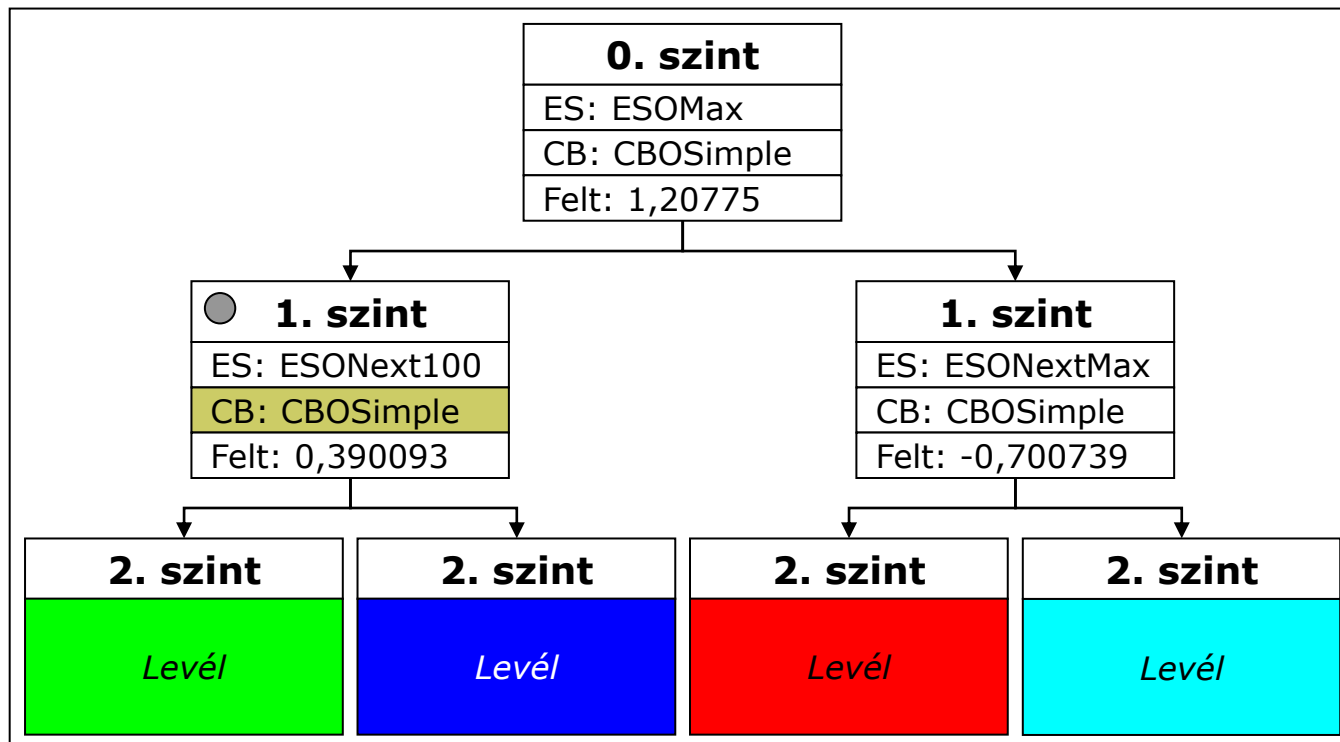
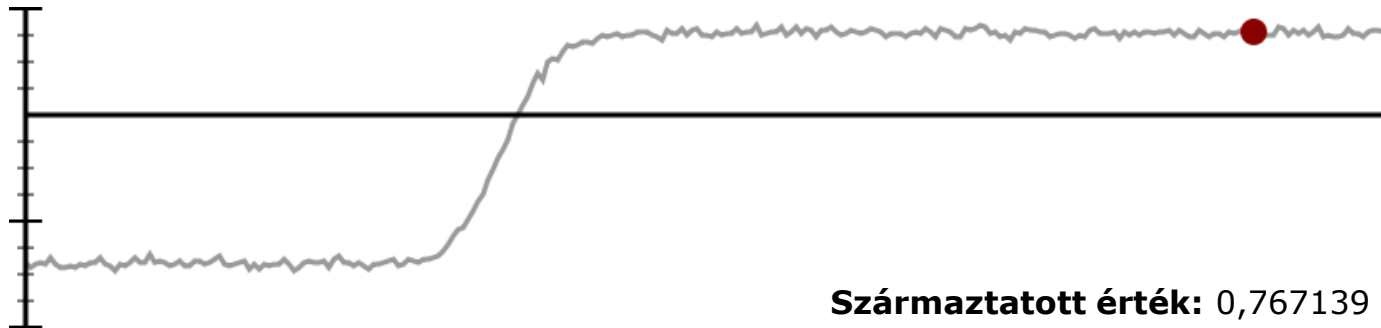
ShiftTree – Példa az osztályozásra



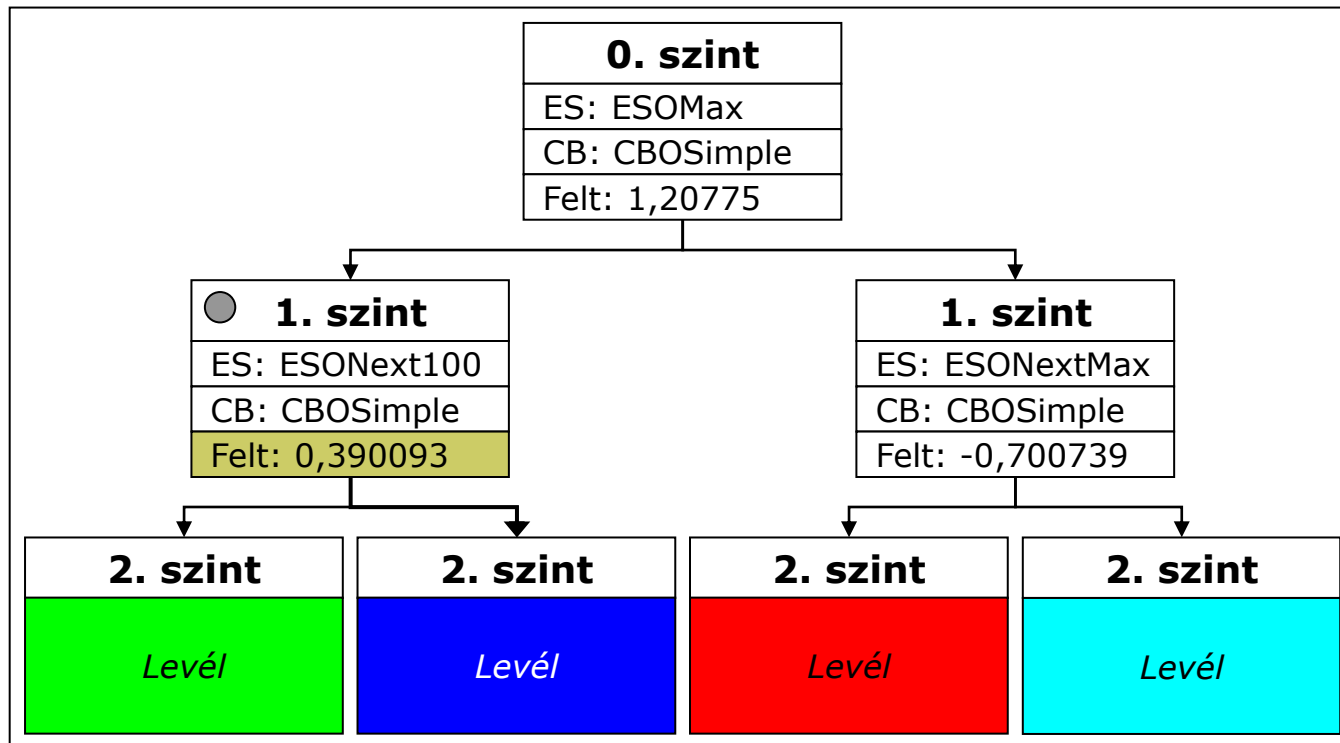
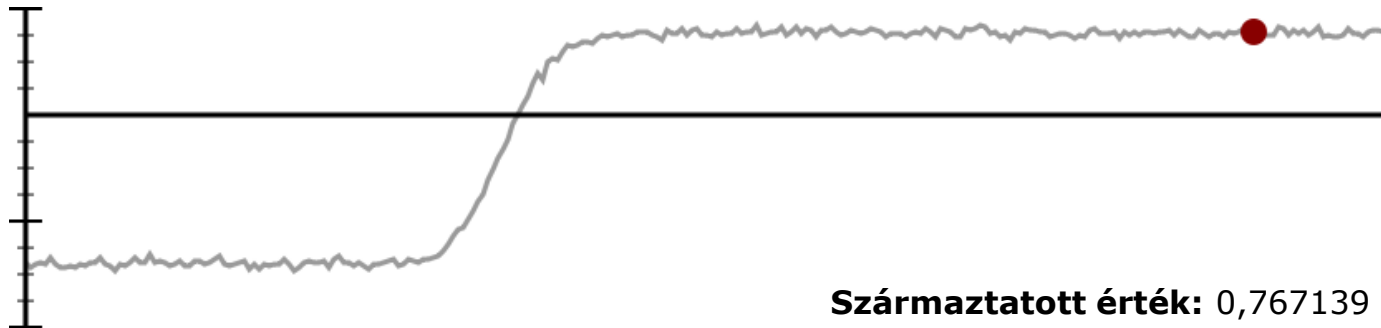
ShiftTree – Példa az osztályozásra



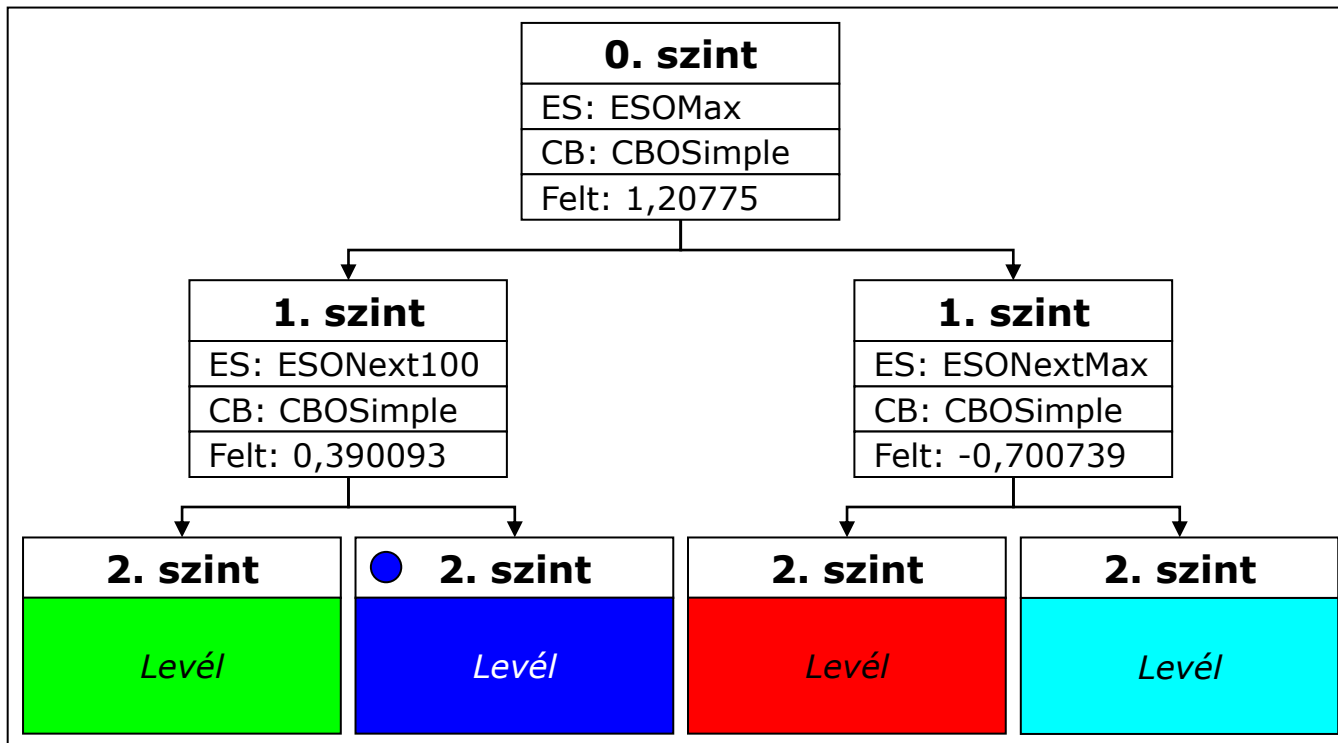
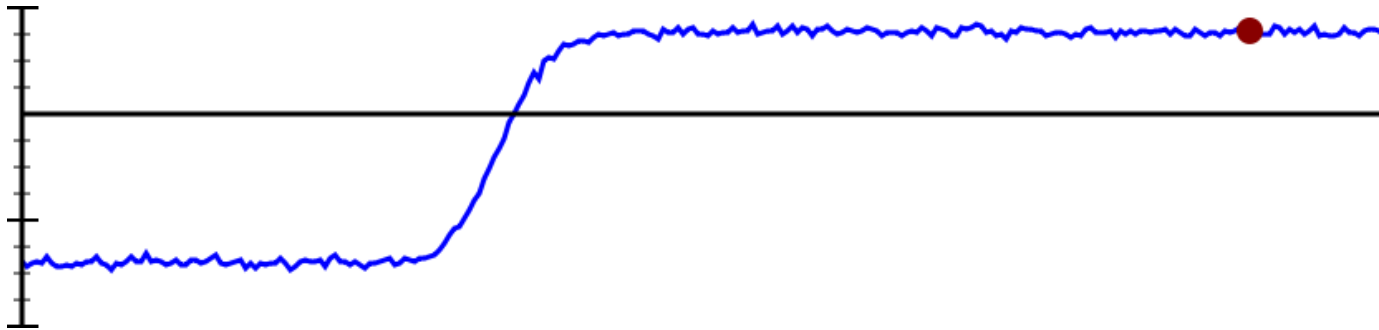
ShiftTree – Példa az osztályozásra



ShiftTree – Példa az osztályozásra

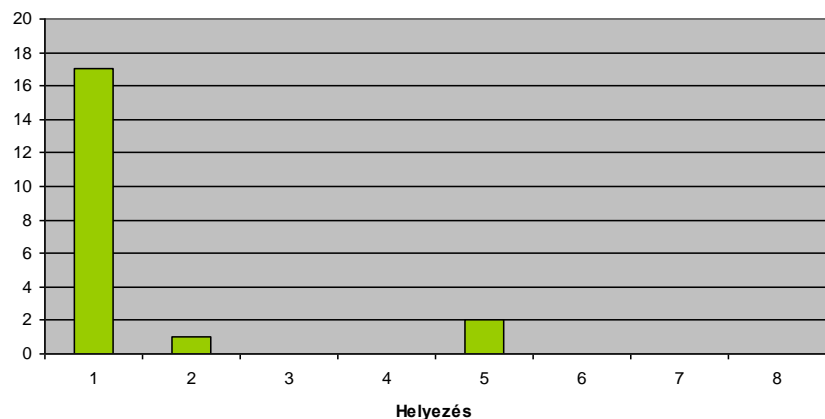


ShiftTree – Példa az osztályozásra

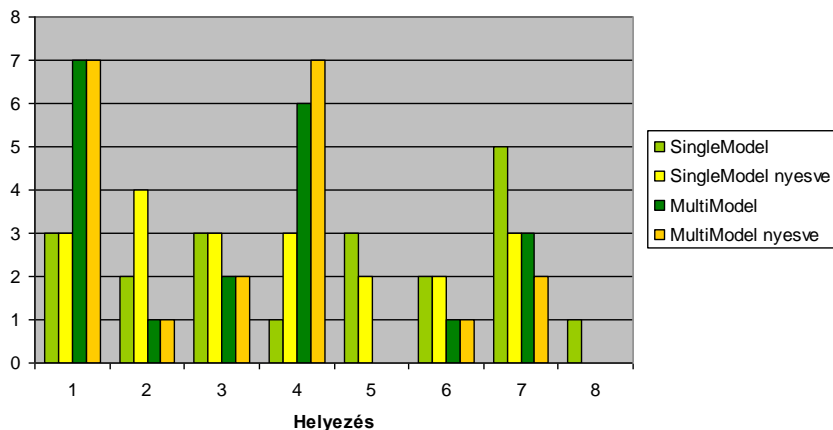


Tesztelés – benchmark eredmények

Az optimalizálatlan ShiftTree helyezései az optimalizálatlan benchmark eredmények között



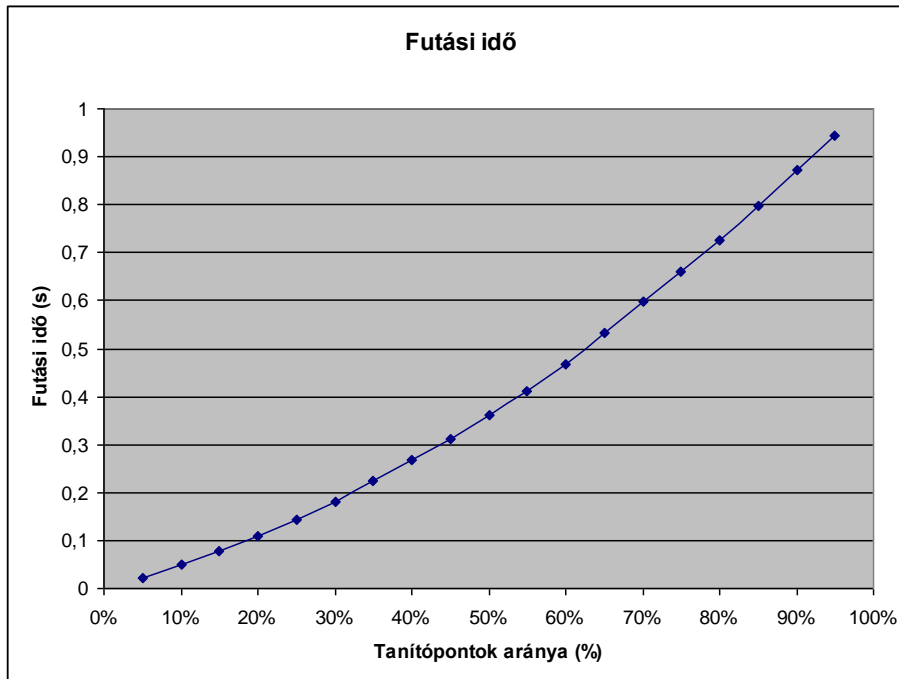
"Sima" és optimalizált ShiftTree helyezései az optimalizált benchmark eredmények között



- 20 különböző tématerületről származó adatsor
 - Különböző hossz/osztályok száma
 - Eltérő méretű tanító/teszt halmazok
- ShiftTree
 - Legegyszerűbb operátorok
- 7 osztályozó eredményei
 - KNN
 - C4.5 döntési fa
 - Logistic Model Tree
 - MLP
 - SVM
 - Naív Bayes háló
 - Random Forest
- Kétféle eredmény
 - Optimalizálatlan paraméterrel
 - Optimalizált paraméterrel
- Eredmények
 - Tanulás intenzív
 - Osztályonként kevés tanítópont esetén gyengébb eredmények
 - Jó eredmények
 - Élmezőny

Futási idő elméletben és gyakorlatban

- Futási idő elméletben (egyattribútumos esetben)
 - Csomópontonként
 - Függ a definiált operátorok számától és típusától
 - $\sim(\text{idősorok száma})^3 \cdot \log(\text{idősorok száma}) \cdot O(\text{idősorok hossza})$
 - Teljes fára
 - Csomópontok összege
 - Jelentősen függ:
 - Szerkezet
 - Idősorok eloszlása a csomópontokban
 - Nem becsülhető, a tanítóhalmaz határozza meg



- Futási idő gyakorlatban
 - Idősorok számától négyzetesen (ritkábban köbösen) függ
 - 24 adatsoron történő keresztvalidáció alapján
 - Adatsoronként 380 mérés történt

Összefoglalás

- Új idősor-osztályozó: ShiftTree
 - Automatikus
 - Minden egydimenziós idősorral működik
 - Operátorok definiálása a szakértő feladata
 - Nem automatikus
 - Pontos
 - Már egyszerű operátorokkal is kellően pontos
 - Magyarázó
 - Könnyen értelmezhető modellek, ellenőrizhető
- Legnagyobb előnye: általános
 - Tématerülettől függetlenül hatékonyan használható

Köszönöm a figyelmet!
