



Synthetic Time series Generation

Török Ákos

Önálló laboratórium

Konzulens: Ács Gergely

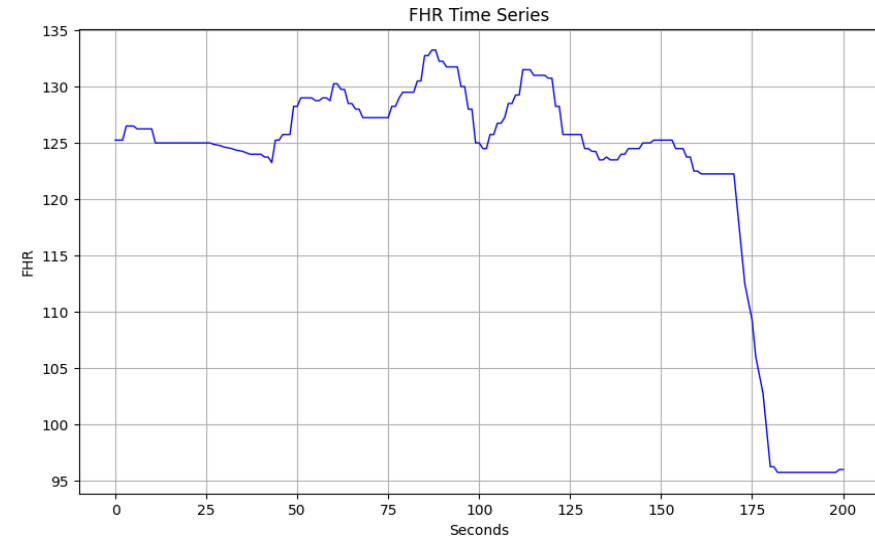
Background and motivation

- Adat generálás
 - Elengedhetetlen tényező
 - Adatok gyűjtése, rögzítése
 - Metódusok, modellek használata
 - Adatok elemzése
 - Döntések hozása
 - Kutatás
- Motiváció
 - Több adat az orvostan hallgatók számára
 - Tudás bővítése
 - Megfelelő minőség generálása



Problem statement and objectives

- CTG idősor (kardiotokográfia)
 - Fetal heart rate (FHR) - magzat szívverése
 - Uterine contractions (UC)
 - Méhösszehúzóadás
 - Nem vettük figyelembe
 - Nincs elég tanító adat az orvostanhallgatók számára
 - Úgy kell generálni többet, hogy a személyes jogok ne sérüljenek
- Mérföldkövek és elvárt eredmény
 - Adatok keresése, előkészítése
 - Generatív modellek tesztelése
 - Eredeti és generált adatok összehasonlítása
 - Minél realisztikusabb adatok



Results

- Adatok előkészítése
 - Megfelelő adatok keresése
 - Interpolálás, esetenként vágás (n/a)
 - Túl rövid idősorok eldobása
- Tanítás és generálás
 - Python csomagok használata (Y-data, Gretel-ai, deep echo)
 - Többnyire Generative adversarial network (GAN)
 - Neurális hálók versengése
 - Idősorra jellemző tulajdonságok
 - Generálás tanult modell alapján
- Y-data
 - Timegan - Time-series Generative Adversarial Network
 - Doppelganger model

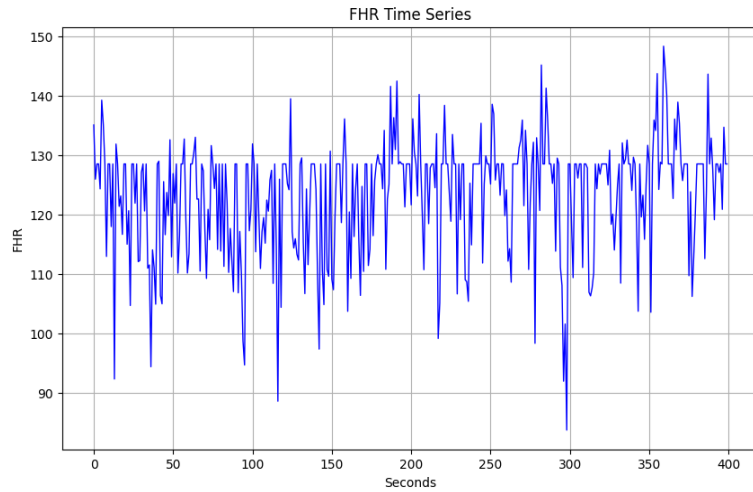
Results

- Gretel-ai
 - Szintén doppelganger
 - Limitált használat
- Deep echo
 - Synthetic Data Vault csomag része
 - ParModel - (Probabilistic AutoRegressive)
 - Deep learning alkalmazása

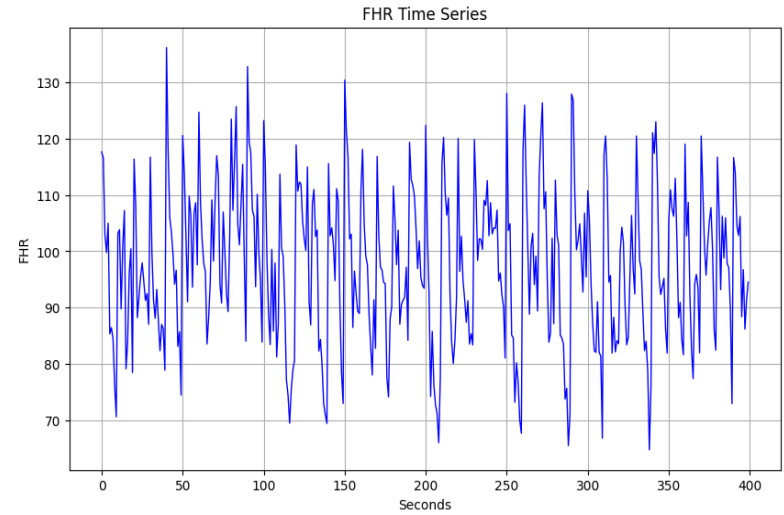
	Y-data	Gretel-ai	Deep echo
Fizetős	részben	igen	nem
Normalizálás	végeredmény manuális	auto	auto
Gyorsaság	átlagos	gyors (felhő)	átlagos

Evaluation

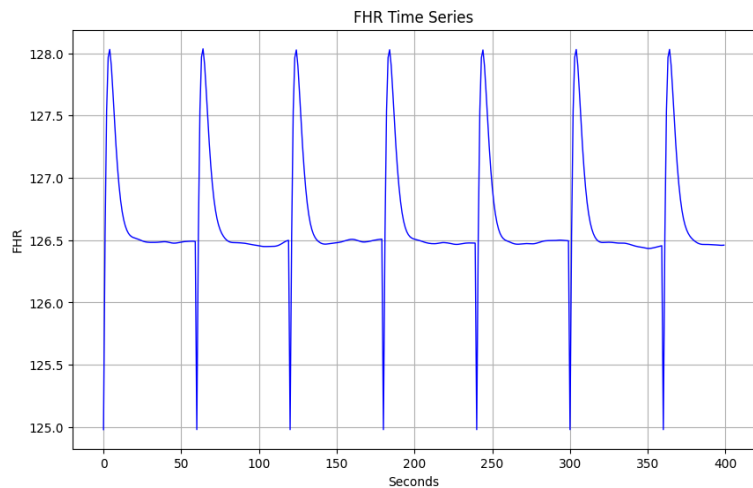
Deep echo



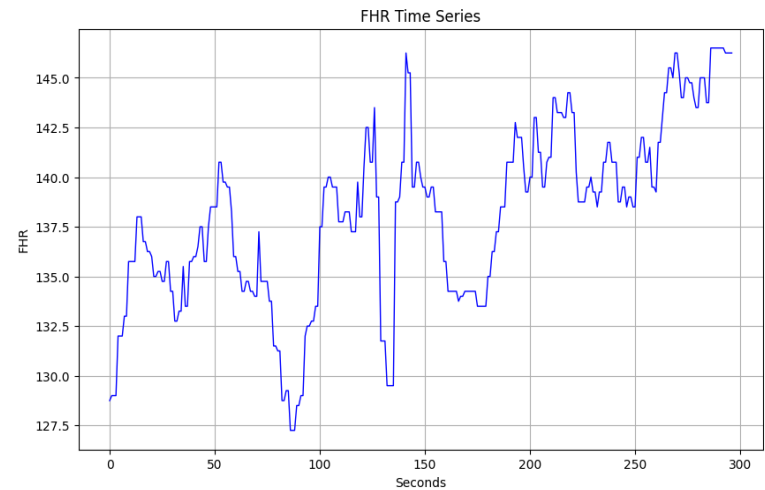
Gretel-ai



Y-data



Eredeti

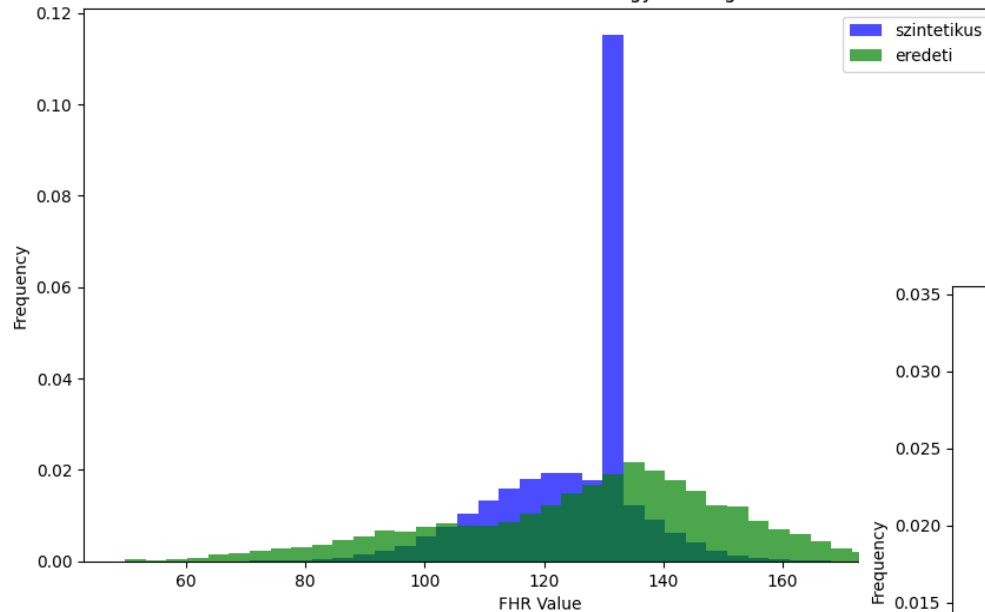


Evaluation

- Összevetés az eredeti adattal: értékek gyakorisága

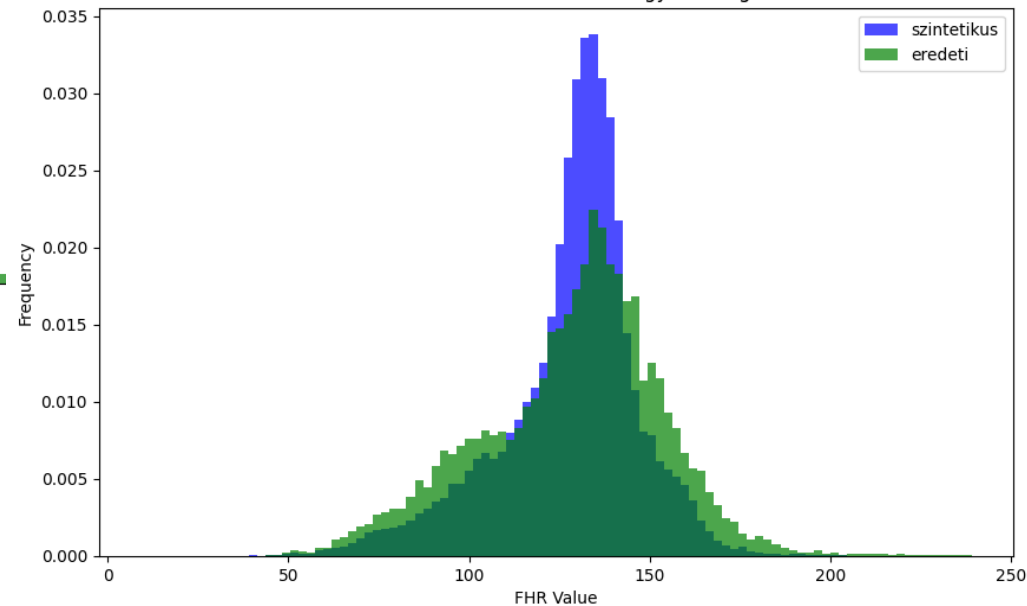
Deep echo

Eredeti és szintetikus értékek gyakorisága



Gretel-ai

Eredeti és szintetikus értékek gyakorisága

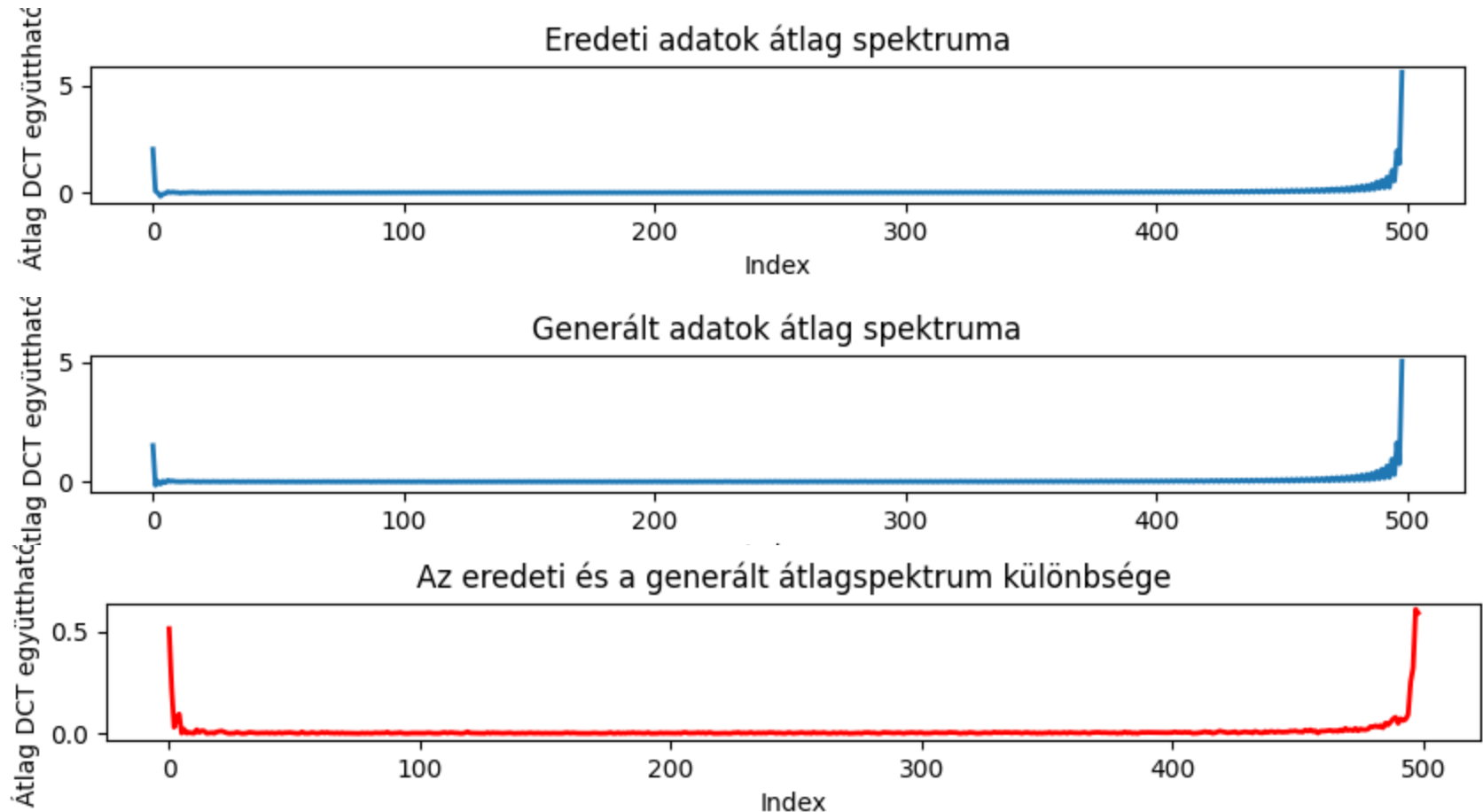


Evaluation

- Jensen-Shannon divergencia
 - Értékek valószínűségi eloszlásának hasonlósága
 - Értékek 100 részre csoportosítva
 - 0-hoz közelítve nagyobb a szimilaritás
 - Gretel ai
 - » 0.15914649623590504
 - Deep echo
 - » 0.43564005217707696
- DCT alkalmazása
 - Átlag spektrumok
 - Euklidészi távolság
 - » Szintén hasonlóság mérése
 - » Kisebb érték

Evaluation

- Átlag spektrumok és Euklideszi távolság - Deep echo
 - Euklidészi távolság: 2.586169973525308



Related work

- Y-data <https://github.com/ydataai/ydata-synthetic>
- Gretel-ai <https://github.com/gretelai/gretel-python-client>
- Deep echo <https://github.com/sdv-dev/DeepEcho/tree/main>



Conclusion and future work

- Összefoglalva
 - Megfelelő modell keresése
 - Ismerkedés a témával
 - Eltanult készségek
 - Python használata (numpy, plotters stb.)
 - Adatok megfelelő kezelése (szűrés, válogatási metódusok)
 - Metrika alkalmazása (divergenciák, gyakoriságok)
- Továbbhaladási lehetőségek
 - Modellek optimalizálása
 - Más modellek alkalmazása
 - Saját keretrendszer
 - Saját modellek fejlesztése