

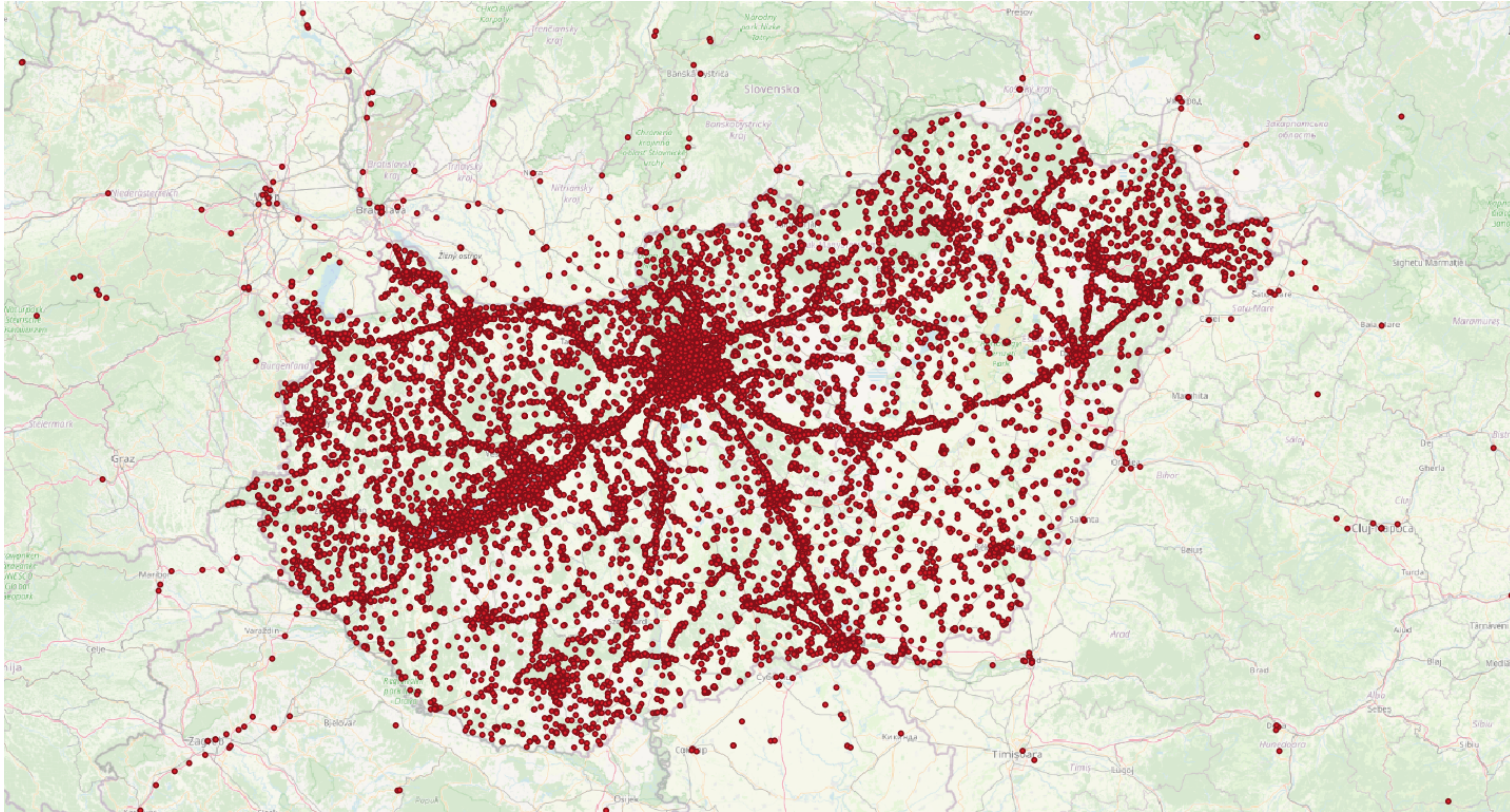
MI és a

M E T  É S Z

PWS adatfeldolgozás és tanuló algoritmus  
fejlesztések bemutatása

# MET-ÉSZ adatok, 2024 október

- Észlelők száma: 1 904
- Helyszínek száma: 21 150



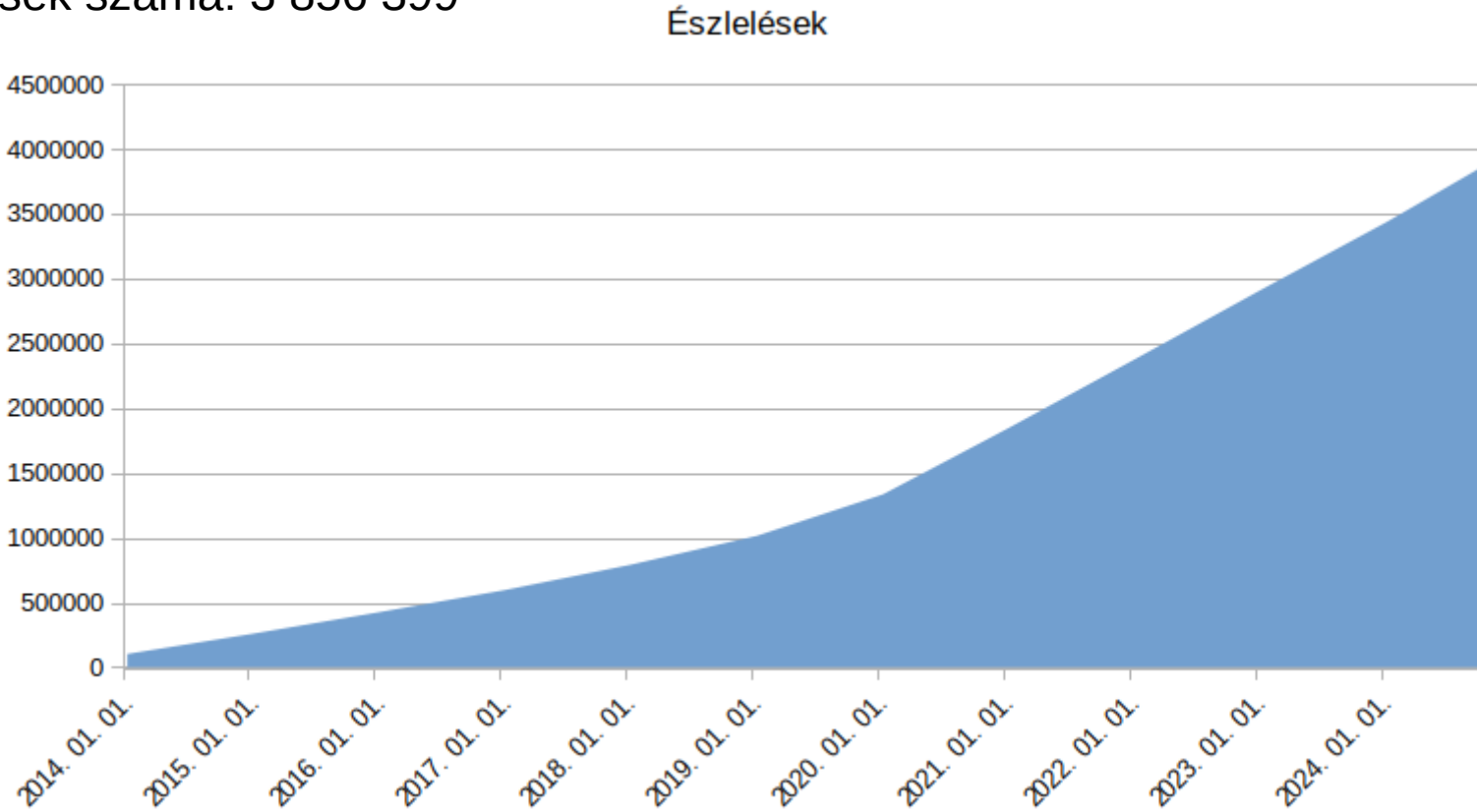
## MET-ÉSZ adatok, 2024 október

- Észlelők száma: 1 904
- Helyszínek száma: 21 150



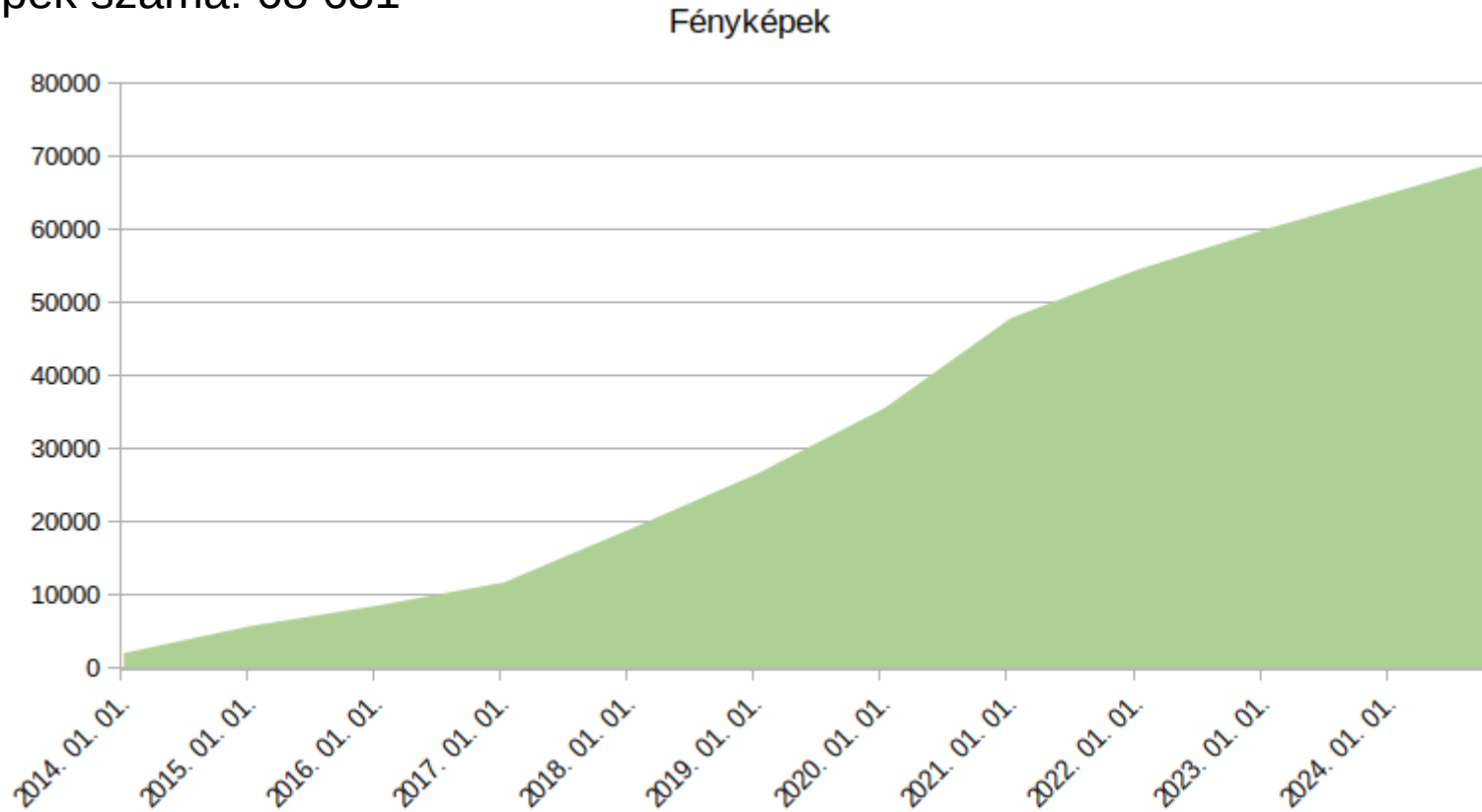
# MET-ÉSZ adatok, 2024 október

- Észlelések száma: 3 856 399



# MET-ÉSZ adatok, 2024 október

- Fényképek száma: 68 631

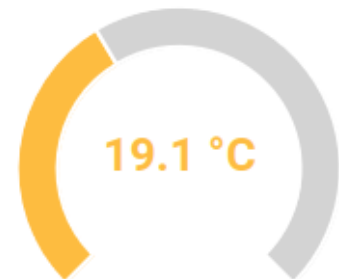
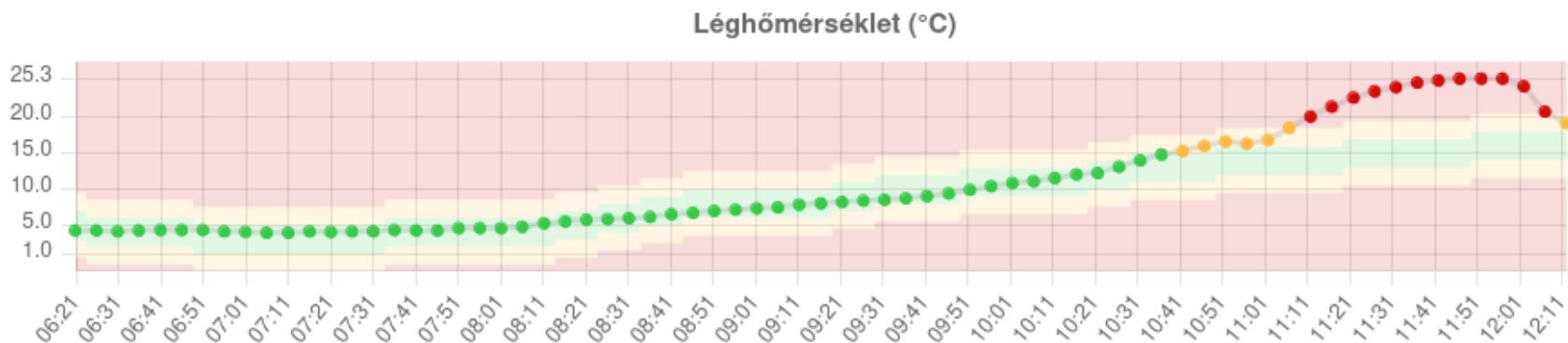


## Közelmúlt MET-ÉSZ fejlesztései

PWS (“mini meteorológiai állomás”) adatfeldolgozás tovább fejlődött:

- adatfájl feltöltés
- új grafikonok automata minősítést megjelenítő sávokkal

Mi ez az automata minősítés?



- Minden bejövő adat **minősítést** kap (DADER projekt)
- A minősítést adatellenőrök által felügyelt algoritmusok végzik
- A MET-ÉSZ azért jeleníti meg, hogy segítse a műszer gazdáját

## A minősítés jövője: tanuló algoritmusok

A most minősítésre használt küszöbök

- fix mérőhelyek adatai alapján lettek kiszámolva
- ideális mérési körülményeket feltételeznek

Túl merev, olyan helyről származó adatot is fel szeretnénk használni,

- ahonnan nincs korábbi adatunk
- ahol nem ideálisak a mérési körülmények
- ahol ismeretlenek a mérési körülmények
- ahol változhatnak a mérési körülmények

Mérési körülmény minden, ami hatással lehet a mérésre (és így annak minőségére).

Vagyis mindent menet közben kell megtanulni (és folyamatosan újratanulni).

# A DADER projektben fejlesztendő tanuló algoritmusról

Kiindulás alap: **a rendszeres dolgokat lehet megtanulni**

Cél a rendszeresen együtt változó körülmények és mérési eredmények felismerése

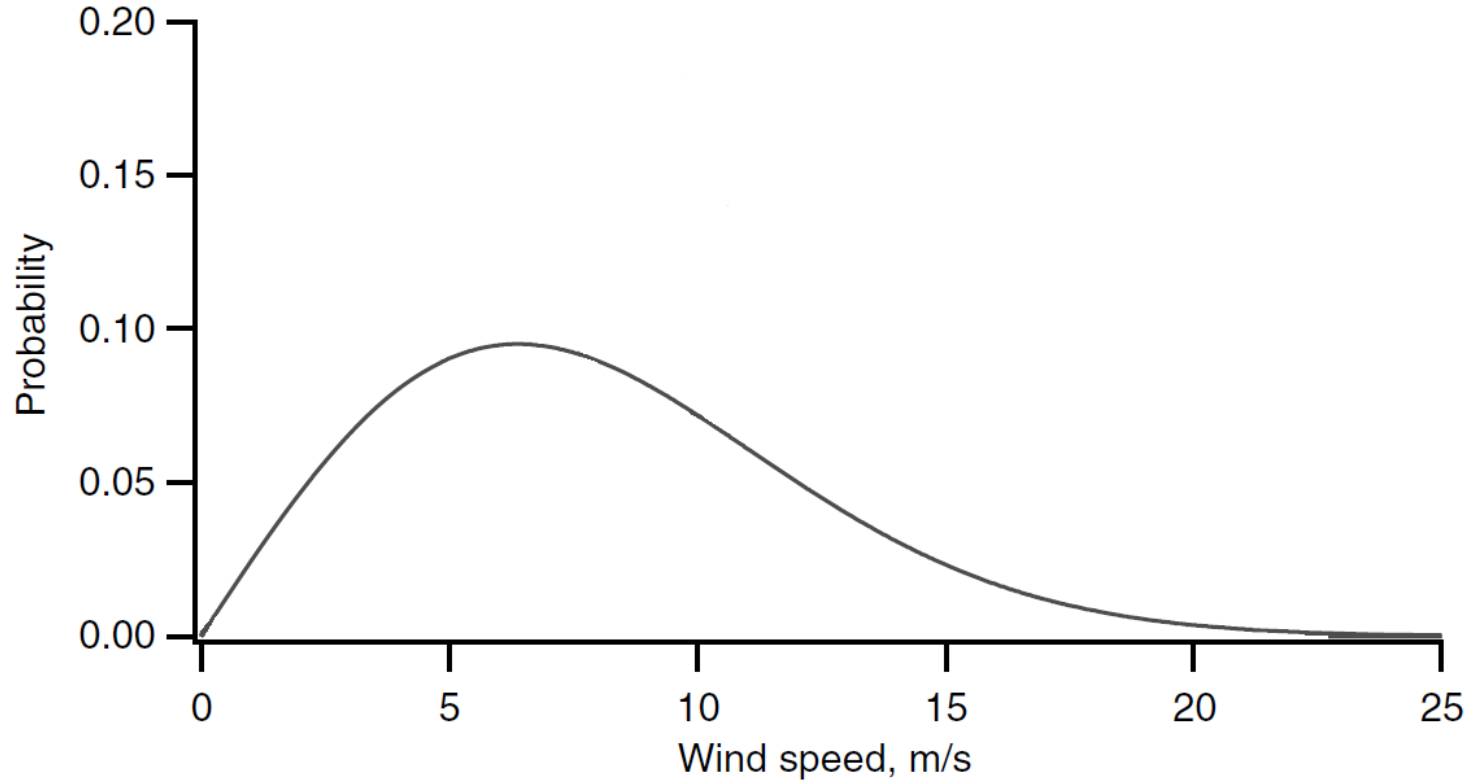
Mit használhatunk?

- (Humán input)
- Ha a **korábbi mérések**hez van valamilyen **viszonyítási alap**, pl. interpolált hivatalos adat (nem kell pontos), akkor föl lehet deríteni a rendszeres eltéréseket
- **Korábbi mérésekből** referencia nélkül is lehet tanulni: a természetes folyamatokra jellemző **eloszlásoktól való eltéréseket** is rendszeres jelenségek okozzák



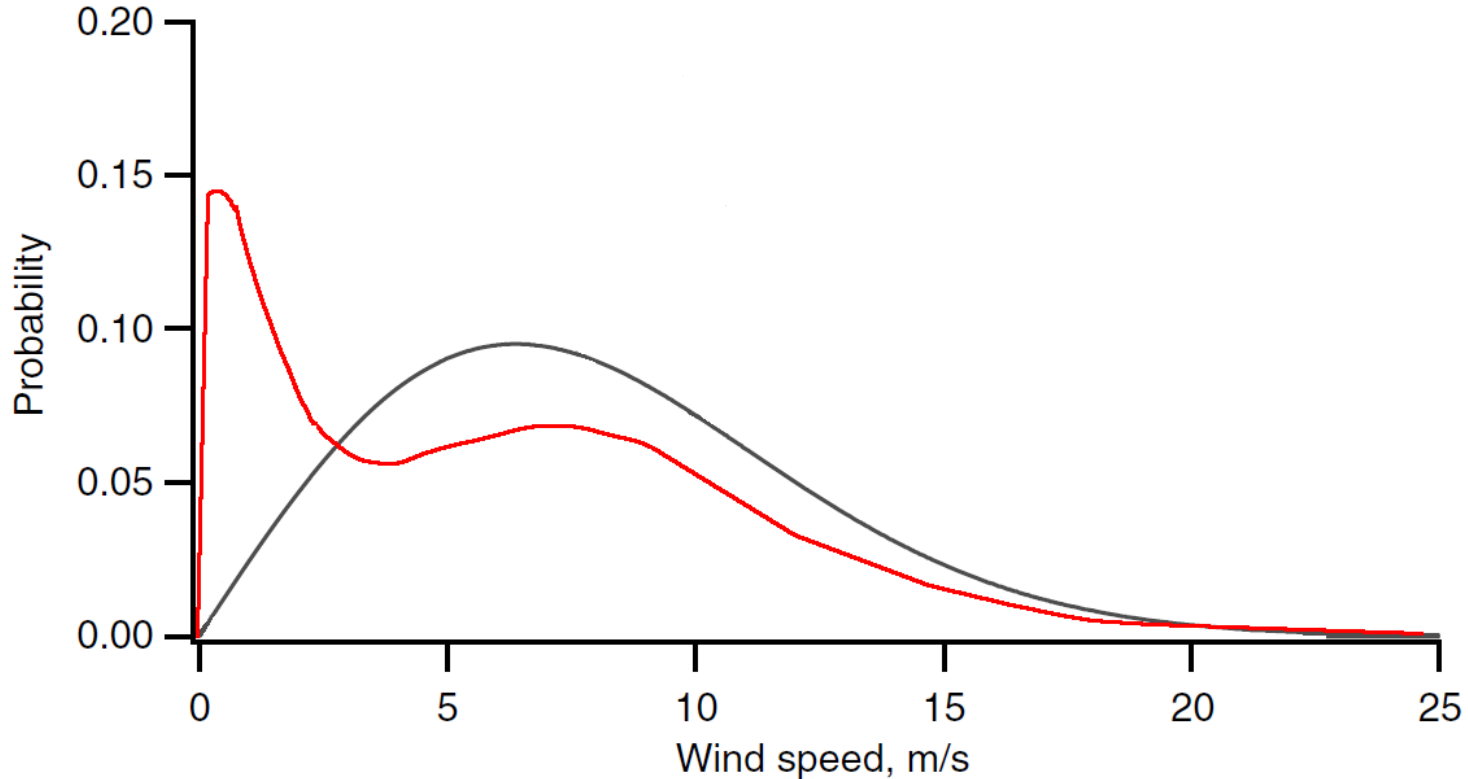
# Tanulás az eloszlások alapján

Szélesebbég természetes (idealizált) eloszlásának egy görbéje:



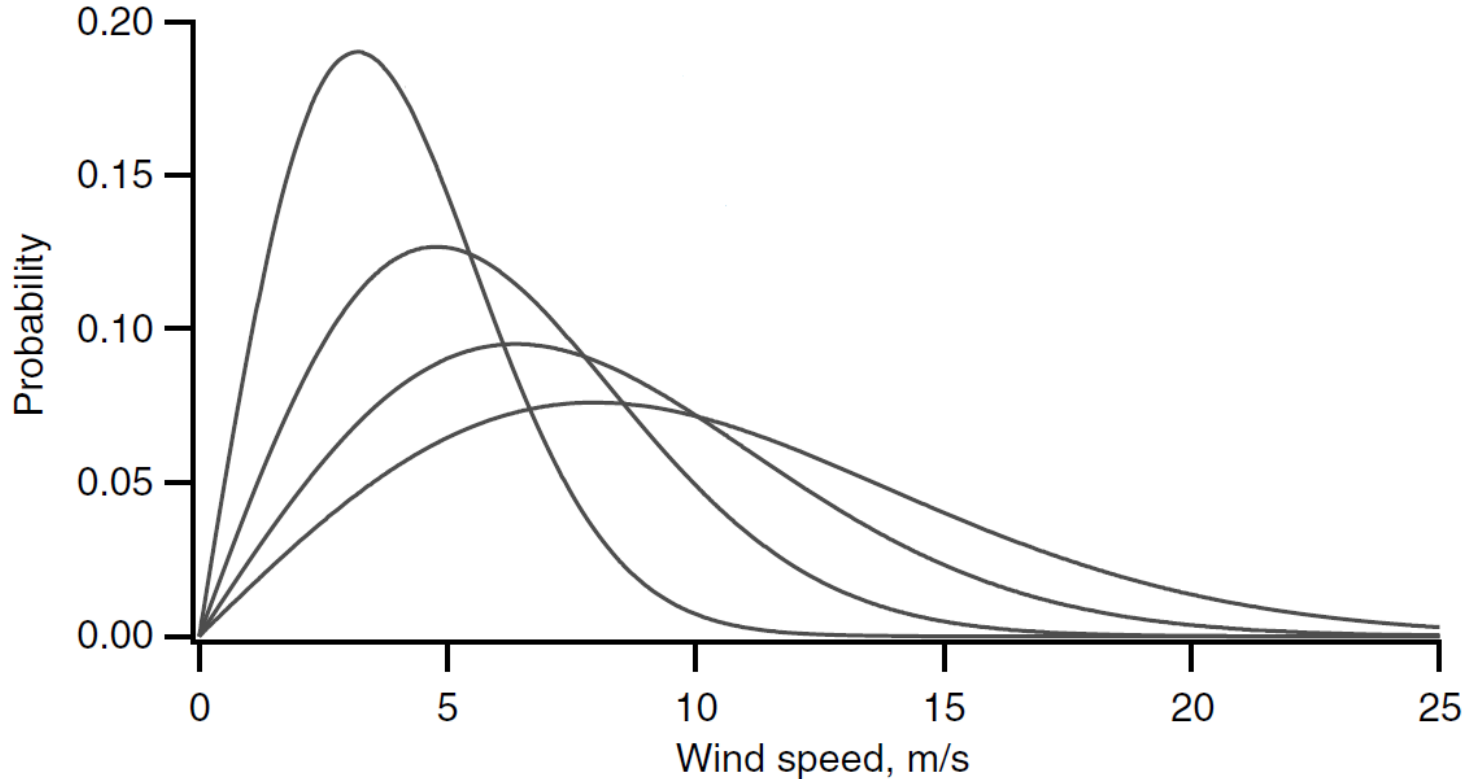
## Tanulás az eloszlások alapján

Ha a mérési eredmények eloszlása **torzul**, eltér az idealizált formától, akkor valamilyen zavaró **körülmény befolyásolja** a mérést



## Tanulás az eloszlások alapján

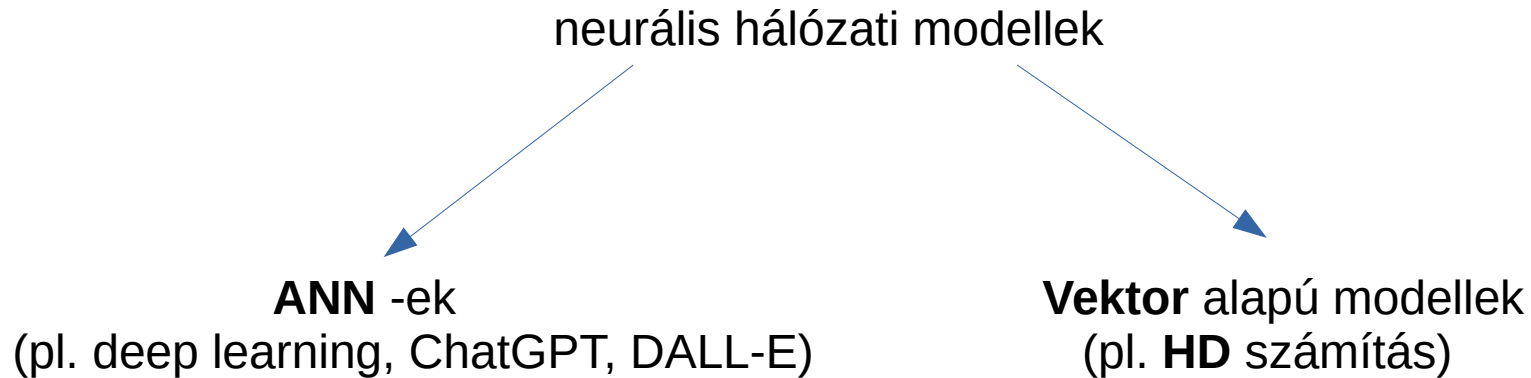
Adott helyre jellemző természetes eloszlás is lehet más, mint az átlagos → megtanulhatók a paraméterei → ez alapján be le állítani az **anomáliák** figyelését



## Tanuló rendszer ismertetése

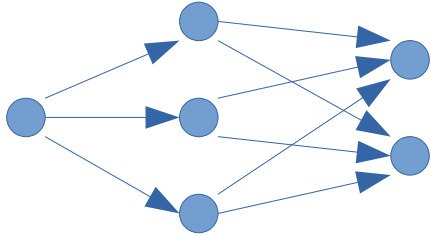
70 -es évek: **neuronhálózatok** működésének **modellezésével** szerettek volna megoldani olyan feladatokat, amik a hagyományos számítógépes architektúrával nehéznek bizonyultak (mert zajos az input, nehéz a számítás, stb):

- pl. nyelvtanulás, mintafelismerés, stb...



## Tanuló rendszer ismertetése

- Mesterséges neurális hálózatok (**ANN**): az idegrendszer hálózatos **felépítésének** és az elemek működésének lemásolása



- **Vektoros** számítási modellek: az idegrendszer legfontosabb **működési jellemzőinek** lemásolása. Mik ezek?
  - **Statisztikusan jó működés** ( $\leftrightarrow$  mindig jó működés), ami gyors  $\rightarrow$  véletlen választáson alapuló gyors folyamatok használata, a sokdimenziós tér teszi lehetővé
  - **Hibatűrés**  $\rightarrow$  a statisztikus működés során óhatatlanul sérülnek itt-ott az elemeket kódoló (bit)mintázatok, de a helyfüggetlen (holografikus) tárolás miatt a reprezentáció egyértelmű marad

## Tanuló rendszer ismertetése

**HD számítás:** a hagyományos számítási modellt terjeszti ki **sokdimenziós vektorokra:**

- 10 000 dimenziós bitvektor (“HD vektor”) minden
- A **magas dimenziószám** miatt könnyen létre tudunk hozni a különböző elemeinkhez (statisztikusan) **különböző**, vagyis merőleges **vektorokat** (random generált vektor ilyen)
- Vektorok:
  - összerendelhetők (“hőmérséklet=10”)
  - összevonhatók, szuperpozícióban kezelhetők (“hőmérséklet=10 és páratartalom=50”)
  - rákövetkezés kódolható (“hőmérséklet=10 majd hőmérséklet=12”)
- a fenti a műveletekkel ill. megfelelő tulajdonságaiknak hála (invertálhatóság, disztributivitás, távolságtartás, stb.) **számításokat tudunk végrehajtani**

# Tanuló rendszer ismertetése

Mi jellemzi a HD számításokat?

- **Hibatűrés**

Holografikus reprezentáció, távolság alapján megtalálható a legközelebbi "korrekt" érték

- Adatok kezelhetők **szuperpozícióban**, ezt a hagyományos rendszerek nem tudják

- **Valószínűségek** kódolhatók (szuperpozícióban is) → statisztikus tanulás

Egymenetes tanulás is lehetséges, alacsony erőforrásigény

- Adatstruktúrák **explicit** módon kódolhatók, manipulálhatók, dekódolhatók, szemben az ANN -ekkel

# Tanuló rendszer ismertetése

Eddig vizsgált tanulási módszerek:

- Klasszifikáció
- Referencia nélküli tanulás eloszlás torzulása alapján

További lehetőségek:

- Jellemzők helyi eloszlásának megtanulása, ez alapján anomália-detekció



# Tanuló rendszer ismertetése

Klasszifikáció lépései HD számítással:

## 1) Előkészítés:

- Vizsgált **körülmények** kiválasztása: mi okozhat rendszeres eltérést a mérés során
- **Osztályok** definiálása: hány osztályba soroljuk be a körülményeket? Mi alapján?

## 2) Tanítás: az egyforma osztályú eltérést előidéző **körülmények összesítése osztályonként** korábbi adatok alapján

## 3) Lekérdezés:

- Osztály: az aktuális **mérés körülményei alapján az osztályoktól való távolság mérése** (a közeliak közül a legközelebbi a jellemző osztály)
- Diagnosztika (melyik körülmény okozza az adott osztályba tartozást leginkább?): a **körülményektől való távolság mérése**

## Egyszerű példa klasszifikációra

Cél: megmutatni, hogy ha a 11 és 12 óra közötti mérések idejére megemelem a hőmérsékletet, akkor meg tudja tanulni a rendszer, hogy a 11-12 közötti mérések nem jók.

Kivitelezés:

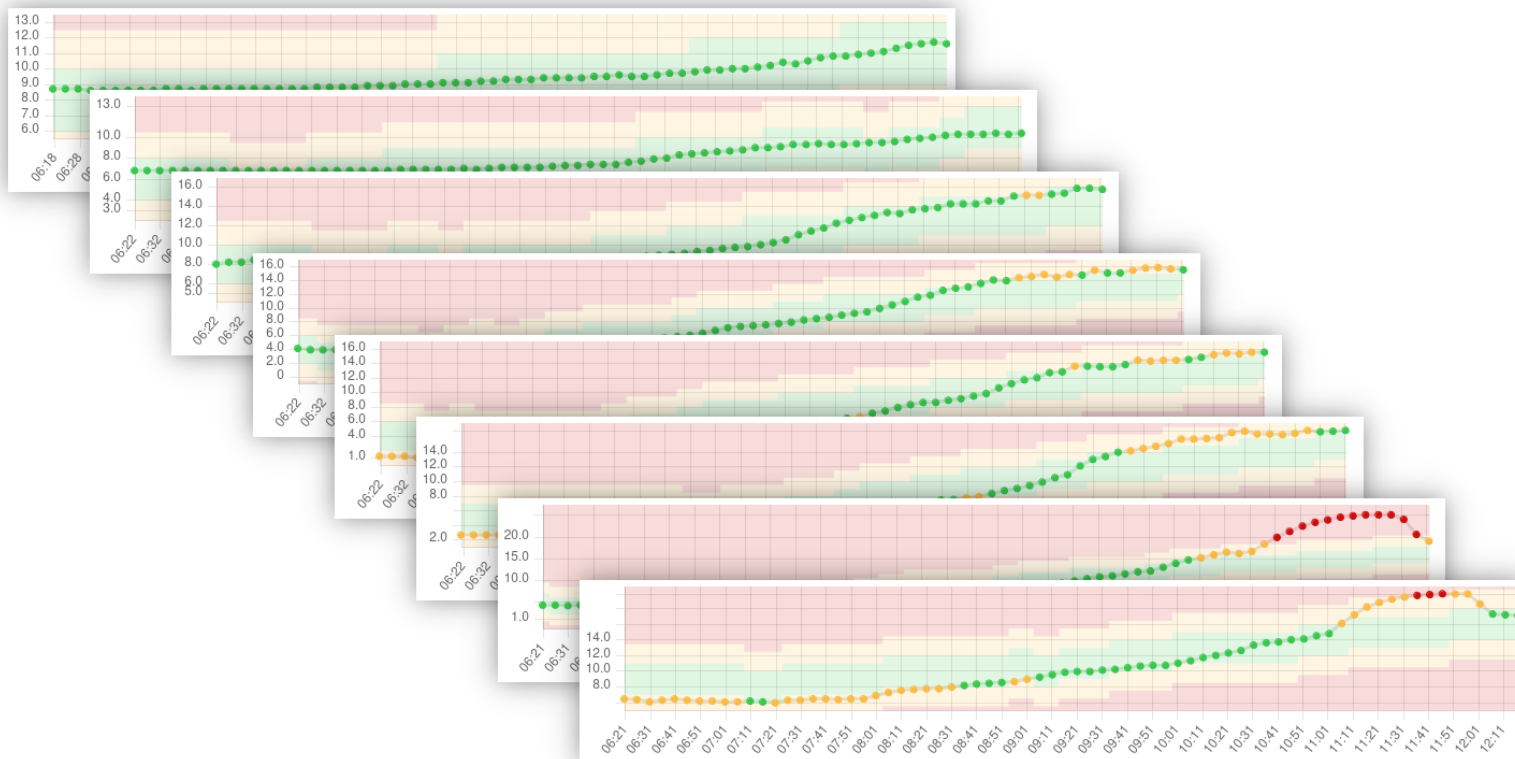


# Egyszerű példa klasszifikációra

Vizsgált mérési körülmény: időpont (nap órája)

Referencia: a korábbi mérésekhez a meander modellből vett becslés

Osztályok: a PWS sávokhoz hasonlóan 3 osztály



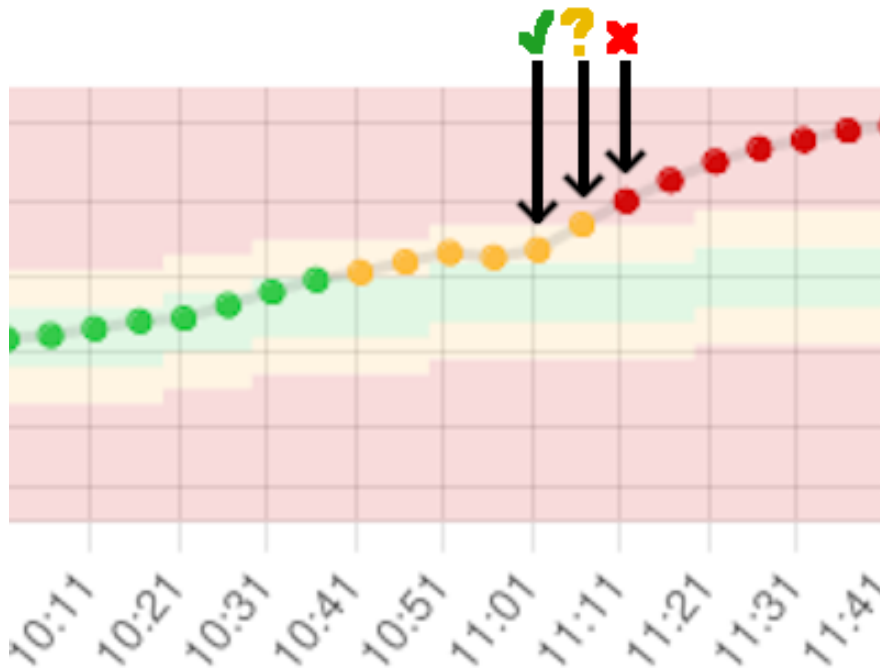
## Egyszerű példa klasszifikációra

Teszt: 1 hetes csúszó ablakot használva egyre több “rossz” 11 órás adattal tanítani a rendszert, majd minden tanítási ablak után lekérdezni, hogy mit tanult meg a mérési körülményekről

## Eredmények

Az első “piros” 11 órás adat után megtanulta, hogy a 11 órás adat rossz, az órája miatt → ???

Amire számítottam: tovább kell tanítani  
Amire nem gondoltam: nem volt más piros minősítésű adat, az első ilyen dominálta a piros kategóriát



## Egyszerű példa klasszifikációra

Nem úgy sikerült, ahogy terveztem, a képességeit a rendszernek nem tudtam demonstrálni.

A rendszer működése így viszont így még jobban látszik:

- egyszerű dolgot csinál egy ilyen klasszifikáló rendszer: megtanulja, hogy milyen jellemzők dominálnak egyes kategóriákban, csak egyszerre sok kategóriával és gyorsan dolgozik

Ráadásul még triviálisnak is tűnik a megoldás, hiszen az első rossz után azt mondta, hogy a következő is rossz lesz, de ez csak a választott mérési körülmény (nap órája) miatt látszik ilyennek, hiszen a bejövő adatok is aszerint vannak rendezve.



## Egyszerű példa klasszifikációra

A másik ötletem: szelesebb időben  
a mérő mellett gyújtok egy jó nagy máglyát!

A magasabb hőmérséklet itt a széliránytól függ.

Ugyanúgy megtanulta volna-e a rendszer  
az összefüggést?

Gondolatkísérlet maradt,  
de szimulált adatokon jól működött a rendszer.

Alapvetően szimulált adatokat  
használunk az algoritmus-fejlesztéshez.



## **MET-ÉSZ kapcsolódása a minősítési folyamathoz, várható fejlesztések**

Az automaták (HWS, PWS) adatait jelenleg a MET-ÉSZ bevivő Automata oldalán lehet elérni.

Tervek:

- PWS dashboard a Meteora app
- Tanuló algoritmus automatikus minősítésének megjelenése a PWS piros/sárga/zöld minősítésekben

A PWS a tanuló pálya, ez a rendszer része a fejlesztés alatt álló DADER rendszernek, ami majd az összes HungaroMet mérés automatikus minősítését végzi.

# MET-ÉSZ kapcsolódása a minősítési folyamathoz, várható fejlesztések

Terv:

- MET-ÉSZ bevivő Automata oldalán MI felügyelet funkció saját PWS -re

Felügyelt tanulás, mint utolsó feldolgozási lépés:

- rosszul kategorizált régiók bejelölése a grafikonokon
- ahol van rá lehetőség, ott fell kell használni az emberi észlelők tudását az MI betanítására
- amiért érdemes lesz csinálni: **jobb visszajelzés a mérési körülmények romlásáról** (most grafikon, később, ha már robusztusabban működik a rendszer, akkor pl. email), **megnevezve a lehetséges okot** is, amit elhárítva javul a mérés minősége



# Köszönöm a figyelmet, kérdések?



DALL-E gépi tanuló MI: “hyperdimensional computing”, ceruzarajz fehér papíron, 2023  
Egyéb felhasznált ábrák: pngtree.com, windroseexcel.com