1-napos időjárás-előrejelzések pontossága

Három különböző időjárást előrejelző weboldal összehasonlítása

1-napos előrejelzések esetén különféle meteorológiai jelenségek esetén

Budapesten

Kodolányi János Egyetem – Gazdálkodási és menedzsment alapképzési szak

Információs rendszerek

Deak Zsombor – SV1LDA

2022.

Tartalomjegyzék

[Bevezetés 3](#_Toc124790406)

[Feladat 3](#_Toc124790407)

[Motiváció 3](#_Toc124790408)

[Cél 3](#_Toc124790409)

[Adatok és Módszerek 3](#_Toc124790410)

[Saját adatvagyon 3](#_Toc124790411)

[Saját módszertan 4](#_Toc124790412)

[Eredmények 5](#_Toc124790413)

[Elvárások 5](#_Toc124790414)

[Válaszok 5](#_Toc124790415)

[Vita 7](#_Toc124790416)

[Következtetések 8](#_Toc124790417)

[Háttér dokumentumok: 8](#_Toc124790418)

# Bevezetés

## Feladat

A feladatom az volt, hogy összehasonlítsam különböző időjárás előrejelző oldalak jövőt érintő adatait egy adott helyen (Budapest) és időszakra (2022\_09\_19-2022\_10\_19) vonatkozóan. Ezeket az adatokat az utólag begyűjtött tényadatokhoz viszonyítva megállapításra jutottam arra vonatkozóan, mennyire pontosak valójában az előrejelzések?

## Motiváció

A feladatot azért is választottam, mert én és a családom több tagja is mindig is figyelemmel kísértük a különböző időjárás előrejelzéseket. Szeretünk kültéri sportokat (pl. biciklizés, síelés stb., illetve egyéb olyan tevékenységeket) végezni, amelyekre befolyással bír az időjárás.

Ez előtt is megbízhatatlannak gondoltam az előrejelzéseket, de tételes statisztikát ez idáig még nem készítettem arról, hogy valójában mennyire azok vagy sem? Emellett az alábbiak szerinti pontosság-kimutatásokat sehol nem találtam – a szolgálató oldalán önértékelésként sem…

## Cél

Azért is gyűjtöttem össze több ezzel a témával foglalkozó website előrejelzési adatait, hogy a jövőben majd melyiket ajánljam mások számára is, vagyis ne csak a megérzéseimre, hanem tényadatokra alapozva értékelhessem az időjárás előrejelző szolgáltatásokat.

# **Szakirodalmi előzmények**

Az alábbi web-oldal egy korábbi, önkéntes pontosság-értékelés eredményeit tartalmazza: <https://miau.my-x.hu/myx-free/olap/olap3/4_olap_m.php3>

# Adatok és Módszerek

## Saját adatvagyon

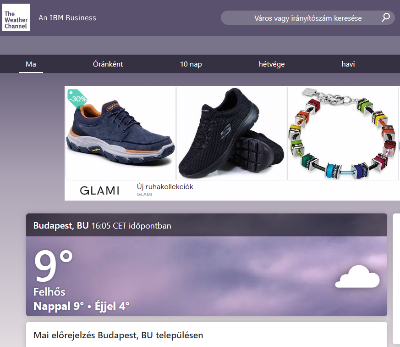
Az általam választott 3 időjárás előrejelző websiteok a következőek:

Időkép - <https://www.idokep.hu/idojaras/Budapest>

Köpönyeg - <https://koponyeg.hu/elorejelzes/Budapest%20IV.%20ker%C3%BClet>

The Weather - <https://weather.com/hu-HU/idojaras/ma/l/HUXX0002:1:HU?Goto=Redirected>

Azért is ezekre esett a választásom mivel ezeket használom a legrendszeresebben és ezek a mások által legtöbbet használt oldalak (vö. 0. ábra):



1. Ábra: A felhasznált weboldalak (forrás: saját ábrázolás)

Az általam vizsgált időszak 2022. szeptember. 19. és 2022. október 19. közötti idősávot öleli fel.

Az előrejelzés adatokat úgy gyűjtöttem, hogy minden nap készítettem egy képernyőfotót a mindenkori **következő napra** jósolt időjárásról, tehát a 09.20-ára előjelzett adat 09.19-én készült (került publikálásra). Az általam vizsgált adatok a hőmérséklet napi minimuma, átlaga, illetve maximuma, ezen felül a csapadék átlagos mennyisége és a szél átlagos sebessége.

A tény adatok gyűjtése ennél sokkal nehezebbnek bizonyult. Számomra is meglepő módon az OMSZ honlapján és a KSH holnapján sem tudtam megfelelő adatokat találni, mivel ezeken a site-okon, vagy csak 2022 első féléves adatok, vagy 2021-es adatok álltak rendelkezésre, és ezek többsége pedig nem napi, hanem heti vagy havi bontásban.

Végül sok kutatómunka eredményeképpen az AMSZ.hu website adatait tudtam felhasználni az adatvagyonom létrehozására.

## Saját módszertan

Létrehoztam egy Excel adatbázist, amelybe a képernyőképekről kinyerhető adatokat összegyűjtöttem (vö. háttér-dokumentumok fejezet).

Ezt megtettem külön mind a három website fellelhető adataival, valamint a tény adatokkal is.

Egy statisztikai táblába illesztettem mindent. A vizsgált adatok nem mindegyik honlapon voltak megadva, így a végső következtetéseket, csak olyan információk alapján határoztam meg, amelyet minden website vizsgál.

Azt vizsgáltam, hogy a tényleges adathoz viszonyítva mekkora százalékos eltérés volt az egy napra előrejelzett adathoz képest. Tehát pl. a szeptember 19-én, pl. 20-ára vonatkozóan előrejelzett adat, mekkora százalékban teljesült valójában 20-án. Ezt a gondolatmenetet végig számoltam minden meteorológiai jelenségre és website-ra nézve. Ahol valamelyik jelenséget az adott oldal nem tette közzé, ott természetesen nem került be adat az adatbázisba.

A további számítások elvégzése előtt, konzultáltam a tantárgy oktatójával, mert nem voltam benne biztos, hogy az irány, amelyet megadtam az valóban helyesnek mondható e.

Az első verziójú háttérdokumentumban a hőmérséklet változásának irányának pontosságát vizsgáltuk, illetve a csapadák mennyiségének változását és a szélerő változásának irányát.

A további egyeztetés után a statisztikai mutató számok alapján szerettem volna vizuálisan könnyebben érthetővé tenni a vizsgált adatokat, ezért csoportosított sávdiagrammot készítettem azokból.

Az eltérések arányosításánál felmerült egy olyan probléma, hogy ha a tényleges adatoktól pozitívan tért el az előrejelzett adat, akkor a százalékos arány meghaladta a 100%-ot (pl. a tényleges maximum hőmérséklet 19-én 14,3C° az Időképen előjelzett hőmérséklet pedig 16C°, ebből a maximum hőmérséklet a ténylegeshez viszonyítva, 112 % lett).

A végleges verzióban korrigáltam a statisztikai mutatókat, hogy azok százalékos aránya ne haladja meg a 100-at (pl. az előző alapján, ami 112%-ot ért el az 88% pontosságnak felel meg).

# Eredmények

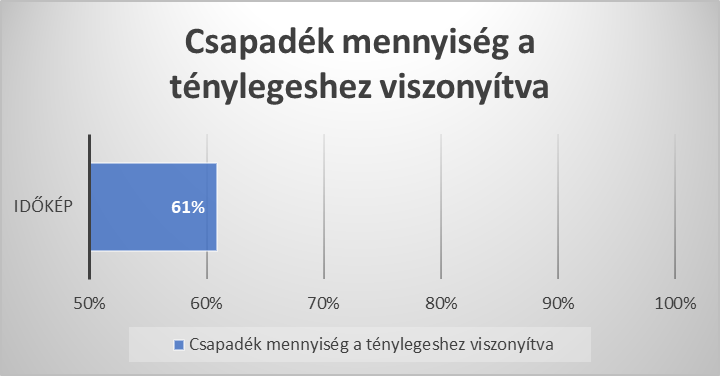
## Elvárások

Az adatok kiértékelése előtt, korábbi tapasztalataimból kiindulva az volt a várakozásom, hogy nem lesz nagy eltérés a site-ok sikerességi aránya között, viszont én abszolút győztesként az Időképet prognosztizáltam. Adat szempontból az volt a megérzésem, hogy az átlag hőmérséklet lesz a legpontosabb adat, valamint, hogy a csapadék a legpontatlanabb.

## Válaszok

A maximum hőmérséklet változását lehetett a szolgáltatók által a legpontosabban előre jelezni, ez 94,7% aránnyal volt (numerikusan) pontos. A második legpontosabb az átlag hőmérséklet előrejelzése lett, ami 91,3%-os pontossággal sikerült a szolgáltatóknak. A minimum hőmérséklet előrejelzés sikerült a legkevésbe, itt 79%-os pontossággal tudták meghatározni a napi előrejelzést (vö. 1. ábra).

1. Ábra: Abszolút pontosságok (forrás: háttér XLS)

Csapadék összehasonlítás csak az Időképen volt mérhető, ami átlagosan 91% pontossággal tudta meghatározni a mennyiséget (vö. 2. ábra):

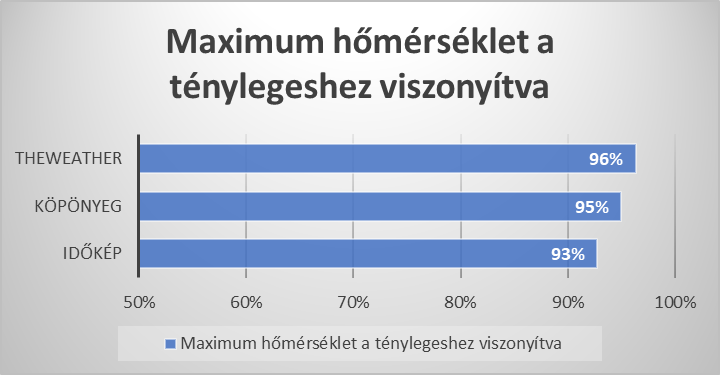
1. Ábra: A csapadék előrejelzés numerikus pontossága (forrás: háttér XLS)

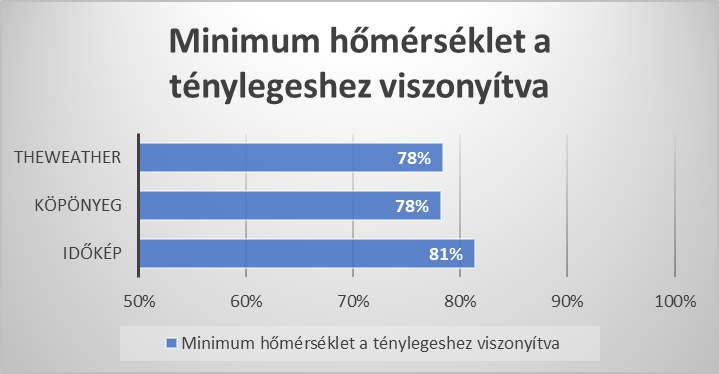
A szélsebesség összehasonlításában a Köpönyeg és a The Weather adatait tudtam összehasonlítani a tényleges adatokkal, ami 68 és 82%-ban sikerült csak a két időjárás előre jelző honlapnak (vö. 3. ábra):

1. Ábra: A szélsebesség előrejelzésének numerikus pontossága (forrás: háttér XLS)

A hőmérsékletet vizsgálva az derült ki, hogy az Időkép tudja a legpontosabban meghatározni átlagosan az időjárást a következő napra vizsgálva.

A maximum hőmérséklet előrejelzésében, a TheWeather 96%-os pontossági aránya a legmagasabb (vö. 4. ábra):

1. Ábra: A maximum hőmérséklet előrejelzésének numerikus pontossága (forrás: háttér XLS)

A minimum hőmérséklet előrejelzésében, az Időkép 81%-os pontossági aránya a legmagasabb (vö. 5. ábra):

1. Ábra: A minimum hőmérséklet előrejelzésének numerikus pontossága (forrás: háttér XLS)

Az átlag hőmérsékletnél pedig, a Köpönyeg 92%-os pontossági aránya a legmagasabb (vö. 6. ábra).

1. Ábra: Az átlaghőmérséklet előrejelzésének numerikus pontossága (forrás: háttér XLS)

# Vita

A statisztikai hibák/eltérések felvázolására két külön metódust vázoltam fel:

A numerikus pontosság esetében elmondható, hogy nincs hiba, így a tökéletes pontosság 100%-os. Mivel az előjeles változat egy speciális információs réteggel rendelkezik, nevezetesen azzal, hogy a becslés nagyobb valószínűséggel túl magas (>100) vagy túl alacsony (<100). Itt is látszólag a 100=100 a hibátlanság, de a hibák átlaga az előjeles nézetben sajnos egymást kioltó is lehet.

Az egyik megközelítés szerint a tényleg adatokhoz viszonyított előre jelzett adatok, mindegy, hogy pluszban vagy mínuszban tévesztenek, ugyan olyan arányban rontják le a jóslat pontossági arányát a 100%-hoz képest.

A másik megközelítés pedig ezt nem veszi figyelembe tehát következésképp lehet magasabb a hiba százaléka mint 100.

# Következtetések

A TheWeather, mint legjobb szolgáltató sem éri el a 60%-os átlagos iránytartást érintő sikerarányt. A maximumhőmérséklet változásának irányát lehetett szolgáltatók által a legjobban előrejelezni az adatok alapján. Furcsa, hogy nem a minimum, hanem az átlag a második legpontosabb (átlag feletti, ahol az átlag is csak 54.2%). A csapadék mennyiségének növekedése/csökkenés+nem-változása a találgatás (50%) szintjét sem éri el, ahogy a minimum-változás iránya és a szélerő változásának iránya esetén sem érhető el a vagy-vagy aránya.



1. ábra: Az előrejelzések utolsó ismerthez képesti eltérés-irányának találati arányai (forrás: háttér XLS)

# Háttér dokumentumok:

* <https://miau.my-x.hu/miau/293/idojaras%20tenyek_vs_elorejelzesek_2022_09_19-10_19_V3.xlsx>
* <https://miau.my-x.hu/miau/293/idojaras%20tenyek_vs_elorejelzesek_2022_09_19-10_19_V2.xlsx>
* https://miau.my-x.hu/miau/293/idojaras%20tenyek\_vs\_elorejelzesek\_2022\_09\_19-10\_19.xlsx