



M Ű E G Y E T E M 1 7 8 2

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM

TERMÉSZETTUDOMÁNYI KAR

ALGEBRA ÉS GEOMETRIA TANSZÉK

**AZ INFARKTUSOK KIALAKULÁSÁT BEFOLYÁSOLÓ IDŐJÁRÁSI ÉS EGYÉNI
PARAMÉTEREK FELTÁRÁSA ADATELEMZÉSI MÓDSZEREKKEL**

TUDOMÁNYOS DIÁKKÖRIONFERENCIA

PESTI PATRIK

KONZULENSEK:

DR. SZILÁGYI BRIGITTA

EGYETEMI DOCENS

SIPOS BENCE

PHD HALLGATÓ

BUDAPEST, 2024.



NEMZETI KUTATÁSI, FEJLESZTÉSI
ÉS INNOVÁCIÓS HIVATAL

Abstract

In Hungary, cardiovascular diseases are responsible for nearly 50% of deaths and a significant portion of illnesses associated with severe complications. Infarctions are one of the most characteristic subgroups of these diseases. The formation of infarctions is influenced by a number of individual (e.g., age, medical history, gender) and global parameters (weather, place of residence). Our research aims to uncover the relationships among these datasets using various data analysis methods.

In this paper, we examine the connections between infarctions and their possible influencing parameters using databases provided by the National Myocardial Infarction Registry and the National Meteorological Services, with particular attention to STEMI/NSTEMI type infarctions. To achieve this, we identify days with unusually high case numbers and investigate their similarities. Additionally, we study the impact of specific weather parameters on case numbers using regression methods. Furthermore, we create more robust predictive models employing XGBoost and neural network-based techniques to examine the properties of more complex systems. Finally, we assess the causal relationships between the variables.

Our goal is to assist in identifying higher-risk groups and high-risk parameters, as well as in forecasting the load on the healthcare system through our findings.

Kivonat

Hazánkban a halálozások közel 50%-áért, valamint a súlyos szövődményekkel járó kórok jelentős részéért a kardiovaszkuláris betegségek felelősek. Ezen betegségek egyik legjellemzőbb alcsoportját az infarktusok alkotják. Az infarktusok kialakulását számos egyéni (például életkor, kórelőzmények, nem) illetve globális paraméter (időjárás, szociokulturális háttér) befolyásolja. Kutatásunk célja, hogy feltárjuk ezen adatok közti összefüggéseket különböző adatelemzési módszerek segítségével.

A dolgozatban a Nemzeti Szívinfarktus Regiszter és az Országos Meteorológiai Szolgálat által biztosított időjárási adatbázisokon keresztül vizsgáljuk az infarktusok és az azokat befolyásoló paraméterek közti összefüggéseket, különös tekintettel a STEMI/NSTEMI típusú infarktusokra. Ehhez megkeressük a kiugró esetszámmal rendelkező napokat, és megvizsgáljuk az ezeket övező hasonlóságokat. Emellett megvizsgáljuk az adott időjárási paraméterek esetszámmra gyakorolt hatását regressziós eljárásokkal, továbbá robusztusabb XGBoost és neurális háló alapú eljárások segítségével prediktív modelleket is felállítunk a komplexebb rendszerek tulajdonságainak vizsgálatára. Végezetül a változók közti ok-okozati kapcsolatokat is értékeljük.

Célunk, hogy eredményeinkkel segítsünk a magasabb kockázatú csoportok és a rizikót okozó paraméterek azonosításában, valamint az egészségügyi rendszer terhelésének előrejelzésében.

Tartalom

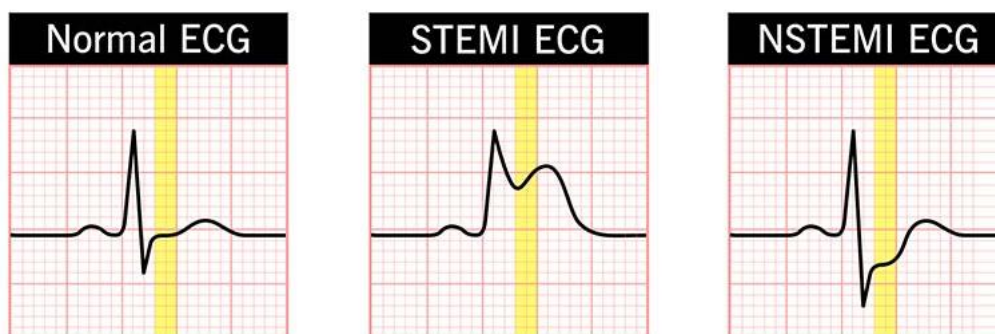
1.	Bevezetés.....	5
1.1	Az infarktus, mint jelenség hazánkban, STEMI / NSTEMI	5
1.2	Irodalom az infarktus befolyásoló időjárási tényezők vizsgálatáról	7
1.3	A dolgozat felépítése, módszertana és célja	9
1.4	A felhasznált adatbázisok jellemzői	10
1.4.1	Anonim Nemzeti Szívinfarktus Regiszter	10
1.4.2	Meteorológiai adatbázis	11
2.	Kiugró esetszámmal rendelkező napok	11
2.1	Kiugró napok meghatározása	12
2.2	Eloszlásvizsgálat a paraméterek függvényében	13
3.	Az esetszám becslése (regresszió).....	14
3.1	Poisson regresszió	14
3.2	Robusztusabb regressziós modell, XGBoost	17
4.	STEMI / NSTEMI-t meghatározó paraméterek	20
4.1	Klasszifikáló algoritmus.....	20
5.	Ok-okozatiság.....	21
5.1	Ok-okozati vizsgálatok	21
5.1.1	Az ok-okozati elemzés alapjai és módszertani kérdései.....	22
5.1.2	Kísérleti és kvázi-kísérleti megközelítések	22
5.1.3	Instrumentális változók és fordított valószínűség súlyozás	23
5.1.4	Következmény-elemzés (Counterfactual Analysis)	23
5.1.5	A DoWhy előnyei a hagyományos modellekkel szemben	24
5.2	Oksági vizsgálat eredménye adatbázisainkon	25
6.	Összegzés	25
7.	Köszönetnyilvánítás	26
8.	Irodalomjegyzék.....	28

1. Bevezetés

1.1 Az infarktus, mint jelenség hazánkban, STEMI / NSTEMI

Hazánkban a halálozások közel 50%-áért, valamint a súlyos szövődményekkel járó kórok jelentős részéért a kardiovaszkuláris betegségek felelősek [14]. Ezen betegségek egyik legjellemzőbb alcsoportját az infarktusok alkotják. Az infarktusok kialakulását számos egyéni (például életkor, kórelőzmények, nem) illetve globális paraméter (időjárás, szociokulturális háttér) befolyásolja. Kutatásunk elsődleges célja, hogy feltárjuk ezen adatok közötti összefüggéseket különböző adatelemzési módszerek segítségével, különös hangsúlyt fektetve a STEMI és NSTEMI infarktusokra.

Az akut miokardiális infarktus (AMI), közismert nevén szívroham, világszerte a morbiditás és mortalitás egyik vezető oka. Az AMI-t elektrokardiogram (EKG) leletek alapján két fő típusba sorolják: ST-szakasz elevációs miokardiális infarktus (STEMI) és nem ST-szakasz elevációs miokardiális infarktus (NSTEMI). [1.1-es ábra]]



1.1 ábra

Egészséges, STEMI infarktusú és NSTEMI infarktusú páciens EKG-jának illusztrációja [7]

A STEMI-t az EKG-n megjelenő ST-szakasz eleváció jellemzi, ami egy koronária artéria teljes elzáródását jelzi. Ez a teljes elzáródás kiterjedt miokardiális ischémiához és nekrozishoz vezet, ha nem kezelik gyorsan. A betegek általában súlyos mellkasi fájdalommal jelentkeznek, amely szorító vagy nyomó érzésként írható le és sugározhat a nyakba, az állkapocsba vagy a bal karba. Egyéb tünetek lehetnek a légszomj, izzadás, hányinger és a szédülés.

Az NSTEMI esetében nincs ST-szakasz eleváció az EKG-n, de más jelei vannak a miokardiális károsodásnak, például emelkedett szívenzimszintek (Tropin T és I, CKMB, LDH,

Mioglobin). Az ok általában egy koronária artéria részleges vagy intermittáló elzáródása. A betegek hasonló tüneteket mutathatnak, mint az STEMI esetén, de a tünetek lehetnek mérsékeltebbek vagy atipikusak, különösen időseknél vagy cukorbetegyeknél. Az NSTEMI-s betegek általában idősebbek és gyakrabban nők, mint a STEMI-s betegek. Az öregedés több krónikus betegség és komorbiditások előfordulásával jár, ami hozzájárulhat az NSTEMI kialakulásához. Gyakoribb a hipertónia, a cukorbetegség és a diszlipidémia ezen betegek körében, míg az STEMI-s betegek nagyobb valószínűséggel fiatalabb férfiak és aktív dohányosok.

Az STEMI kezelése a véráramlás gyors helyreállítására összpontosít az érintett artériában, általában primer perkután koronária intervenció (PPCI) vagy fibrinolitikus terápia révén. Az idő kritikus tényező az STEMI esetén a miokardiális károsodás minimalizálása érdekében. Az NSTEMI kezelése gyakran korai invazív stratégiát foglal magában koronária angiográfiával és esetleges revaszkularizációval, de általában kevésbé sürgős, mint az STEMI esetén. A komorbiditások magasabb előfordulása miatt az NSTEMI-s betegek esetében gyakrabban alkalmaznak gyógyszeres kezelést, és a kezelési tervet az egyéni kockázati profilhoz igazítják.

Világszerte az NSTEMI gyakoribb, mint az STEMI, az összes miokardiális infarktus körülbelül 60–70%-át teszi ki. Az Egyesült Királyságban például az NSTEMI az akut miokardiális infarktusok 59,5%-át teszi ki, míg a STEMI 40,5%-át. Bár az in-hospital mortalitási arányok hasonlóak a STEMI és az NSTEMI betegek között, a hosszú távú kimenetelek eltérhetnek. Az NSTEMI-s betegek hosszú távú prognózisa rosszabb lehet, magasabb egyéves halálozási arányokkal és nagyobb gyakoriságú kombinált kardiovaszkuláris eseményekkel, mint például miokardiális infarktus, stroke és kardiovaszkuláris halál.

A környezeti tényezők, mint például az időjárás és a légszennyezés befolyásolják a STEMI és az NSTEMI előfordulását. Mind a magas, mind az alacsony hőmérséklet összefügg a miokardiális infarktus kórházi felvételeinek növekedésével. A hideg hőmérséklet általában magasabb akut koronária események kockázatával jár, különösen az NSTEMI esetén, míg a magas hőmérséklet növelheti a STEMI felvételek kockázatát [9, 13, 20]. Egyéb időjárási tényezők, mint a nagy havazás és a légköri nyomás változásai, szintén összefüggésbe hozhatók a miokardiális infarktus előfordulásának növekedésével. A légszennyezés, beleértve a finom részecskék magas szintjét, hozzájárul a mind a STEMI, mind az NSTEMI fokozott kockázatához, mivel a szennyező anyagok súlyosbíthatják az alapvető kardiovaszkuláris állapotokat és akut eseményeket válthatnak ki.

Az életkor és a nem befolyásolja a környezeti tényezők miokardiális infarktus kockázatára gyakorolt hatását. Fiatalabb nők (≤ 55 év) esetében a magasabb hőmérséklet és a hosszabb napsütéses órák növelik az STEMI felvételek számát. Ezzel szemben a fiatalabb férfiak jobban érintettek lehetnek a hideg hőmérsékletek által, bár az eredmények kevésbé következetesek. Az idősebb felnőttek, különösen a nők, hajlamosabbak a hideg által kiváltott miokardiális infarktus kockázatára.

Összefoglalva, a STEMI és az NSTEMI különböznek klinikai megjelenésükben, betegdemográfiájukban, kezelési stratégiáikban és kimenetelükben. Ezeknek a különbségeknek a megértése elengedhetetlen a klinikusok számára a megfelelő betegellátás biztosításához. A környezeti tényezők miokardiális infarktus előfordulására gyakorolt hatásának felismerése pedig segíthet a megelőzési stratégiák és közegészségügyi beavatkozások kidolgozásában, csökkentve az akut koronária események terhét.

1.2 Irodalom az infarktus befolyásoló időjárás tényezők vizsgálatáról

Az időjárás kardiovaszkuláris történésekre gyakorolt hatását több tanulmány elemzi. Tekintettel arra, hogy az időjárás mindenkire hat és a populáció, különösen a magyarországi populáció jelentős része érintett a betegségben, ezért még ha gyenge hatásról van is szó, komolyan számolnunk kell vele. Egy-egy – az AMI szempontjából jelentős – időjárás esemény komoly leterhelést okozhat az ellátórendszerben. A légkörfizikai és légkörkémi paraméterek hatásainak hazai elemzése nemcsak azért fontos, mert Magyarországon igen magas az infarktusok száma, hanem azért is, mert az eltérő időjárás viszonyok hatásait elemző cikkek eredményei nem szükségképpen relevánsak. Az időjárás paraméterek utóbbi évtizedekben bekövetkező gyors és jelentős változása ugyancsak azt teszi szükségessé, hogy a régebbi eredmények ellenére ismételt feltegyük és megválaszoljuk ugyanazokat a kérdéseket.

A fent leírtak vezetnek ahhoz is, hogy az időjárás infarktusra gyakorolt hatását elemző publikációk között nem egyszer egymásnak ellentmondókat találunk. A továbbiakban néhány korábbi eredményt gyűjtöttünk össze. (A források felkutatásában a MI segítségét vettük igénybe.)

- Mind az extrém magas, mind az alacsony hőmérséklet növeli a STEMI- és NSTEMI-felvételek kockázatát [12].

- "V", "U", "J" vagy lineáris alakú korrelációt figyeltek meg az ischémiás szívesemények és a külső környezet hőmérséklete között [12].

- Az alacsony hőmérséklet általában a STEMI és az NSTEMI fokozott kockázatával jár [10], [12].

- A napi 2 cm-t meghaladó havazás emelheti a STEMI arányát, különösen a férfiaknál [11].

- Egyes tanulmányokban a nagyobb szélsébség a szívinfarktus megnövekedett gyakoriságával járt [12].

- Bizonyos tanulmányok szerint a STEMI esetszámok növekedése a magas hőmérséklettel, míg az NSTEMI esetszámok az alacsony hőmérséklettel hozhatók kapcsolatba [13].

- Egy izraeli tanulmány a téli hónapokban szignifikáns növekedést mutatott ki a STEMI esetek számában, de nem mutatott statisztikailag szignifikáns szezonális eltérést az NSTEMI-ben [13].

Vannak szezonális mintázatok a STEMI és az NSTEMI előfordulásában, bár a mintázatok olyan tényezőktől függően változhatnak, mint az életkor, a nem vagy a földrajzi elhelyezkedés. A legfontosabb megállapítások összefoglalása:

- Számos tanulmány az akut koronária szindrómák (ACS) általános növekedését mutatja - beleértve a STEMI-t és az NSTEMI-t is - a téli hónapokban.

- A SWEDEHEART regiszter (a svédországi kardiovaszkuláris regiszter) negatív összefüggést talált a levegő hőmérséklete és a szívinfarktus kockázata között, a legmagasabb kockázatot a hidegebb hőmérsékletek idején figyelték meg [10].

- Egyes tanulmányok a STEMI megnövekedett arányát találták a nyári hónapokban, különösen bizonyos alcsoportokban.

- Egy izraeli tanulmány szerint a téli hónapokban jelentősen megnőtt a STEMI esetek száma [13].

- A fiatal férfiak (≤ 55 év) esetén a téli hónapokban a STEMI-k magasabb aránya volt megfigyelhető [11].

- Nyáron a STEMI arány jelentős növekedését észlelték fiatalabb, 55 év alatti nőknél [11].

- A fiatalabb nők esetén a STEMI aránya 2010 augusztusában érte el a csúcst (az éves esetszám 17,5%-a) a legalacsonyabb februárban és márciusban volt (az éves felvételek 3,9%-a) [11].

- Az izraeli vizsgálat nem talált statisztikailag szignifikáns különbséget az NSTEMI szezonális változásai között [13].
- A SWEDHEART regiszter megállapította, hogy a leghidegebb évszakban (január-március) pozitív kapcsolat volt a hőmérséklet növekedése és az NSTEMI aránya között [10].
- Egy tanulmány szerint jelentős növekedést mutatott a STEMI esetszáma a nyári hónapokban és a magasabb hőmérsékletű napokon [11].
- A hideghatás kifejezettebb volt a nőbetegeknél [13].
- Svédország északi egészségügyi régióiban a szél sebessége és a hócsapadék az NSTEMI és a STEMI nagyobb kockázatával járt [10].

Ezek a szezonális minták rávilágítanak az időjárási viszonyok és a STEMI / NSTEMI előfordulása közötti összetett kapcsolatra. A különböző alcsoportok és régiók között megfigyelt eltérések arra utalnak, hogy számos tényező, köztük az életkor, a nem és a helyi éghajlat fontos szerepet játszik az akut koszorúér események szezonális kockázatának meghatározásában.

Ezért a következő fejezetekben bemutatott kutatásunkat nem csak az alkalmazott módszerek újszerűsége, hanem a jelenségek regionalitása miatt, Magyarországspecifikussága is kiemelten relevánssá teszi.

1.3 A dolgozat felépítése, módszertana és célja

A dolgozatban a Nemzeti Szívinfarktus Regiszter és az Országos Meteorológiai Szolgálat által biztosított időjárási adatbázisokon keresztül vizsgáljuk az infarktusok és az azokat befolyásoló paraméterek közti összefüggéseket, különös tekintettel a STEMI/NSTEMI típusú infarktusokra. A kidolgozott módszerek azonban általánosak, a jövőben az eljárásokat orvosokkal együttműködve, szakmai irányításukkal a SE Érsebészeti Klinikáján működő Éghajlatváltozás Multidiszciplináris Nemzeti Laborban alkalmazzuk majd.

Először megkeressük a kiugró esetszámmal rendelkező napokat és elemezzük az őket jellemző légkörfizikai és légkörkémi paramétereket. Ezt követően megvizsgáljuk az adott időjárási paraméterek esetszáma gyakorolt hatását regressziós eljárásokkal, továbbá robusztusabb XGBoost és neurális háló alapú eljárások segítségével prediktív modelleket is felállítunk a komplexebb rendszerek tulajdonságainak vizsgálatára. Emellett a STEMI/NSTEMI-t

befolyásoló kórelőzményeket is megvizsgáljuk klasszifikáló algoritmusokkal. Végezetül a változók közti ok-okozati kapcsolatokat is értékeljük.

Célunk, hogy eredményeinkkel segítsünk a magasabb kockázatú csoportok és a rizikót okozó paraméterek azonosítását, valamint az egészségügyi rendszer terhelésének előrejelzését.

1.4 A felhasznált adatbázisok jellemzői

1.4.1 Anonim Nemzeti Szívinfarktus Regiszter

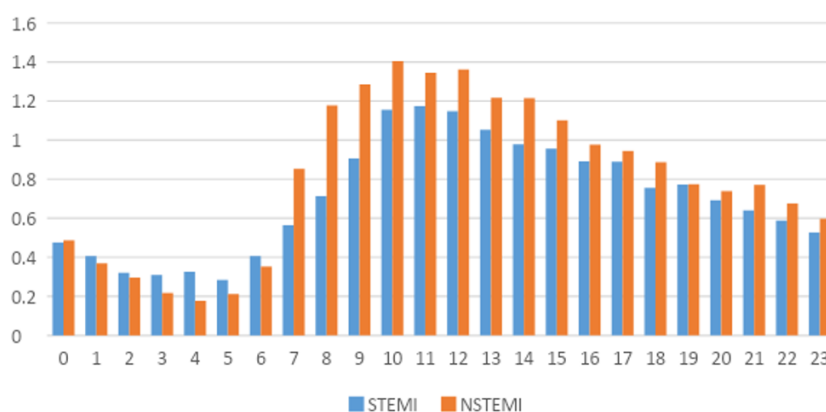
Ez az adatbázis 2014.01.01-2023.12.31 között a magyar egészségügyi ellátás által elkönyvelt infarktus eseteket és számos azokkal kapcsolatos adatot tartalmaz. A NSZR értékes része a bejelentkezés nélkül elérhető interaktív lekérdező felület, amely sok hasznos információt tartalmaz területi és időbeli bontásban. Ezen adatok közül az általunk felhasználtak a következők:

A Beteg életkora; neme; az esemény időpontja (kórházba érkezést követő regisztráció); típusa (STEMI/NSTEMI); kórelőzmények [korábbi infarktus, PCI, cukorbetegség, COPD, stroke, szívkatéter, BMI index].

Már szimplán ezen adatbázis segítségével is le tudunk vonni érdekes következtetéseket az infarktusok jellegéről hazánkban.

Először is 2021.01.01-2023.12.31-ig az esetek 47% volt STEMI míg 53%-a NSTEMI. Ez az arány 2014.01.01 és 2020.03.03 közt, azaz a kovidot megelőző időszakban 42:58 volt. Azt láthatjuk, hogy még egy adott régió belül is mint Magyarország ez az arány változik, szükségessé téve ezen adatok időszakos vizsgálatát.

Emellett még az infarktusok és típusai napi eloszlására érdemes ránézni, ami a következő [1.2-es ábra]:



1.2-es ábra

Ezen az ábrán a késődélelőtti tetőzési órák mellett érdemes észrevenni, hogy a hajnali órákban megnő a STEMI aránya az NSTEMI-hez képest. Ez utalhat arra, hogy bekövetkezése kevésbé érzékeny az időpontokra vagy adottesetben arra, hogy a betegek kevésbé képesek megvárni a reggelt egy ilyen infarktus esetén ezért mentőt hívnak. Az viszont biztosan állíthatjuk, hogy a STEMI esetszám paraméterekre vonatkozó érzéketlensége végig fog kísérni minket dolgozatunk során.

1.4.2 Meteorológiai adatbázis

Ez az adatbázis 2015.01.01. és 2023.12.31. között tartalmaz légkörfizikai és légkörkémiail adatokat, különböző származtatott paramétereket (AQI, WEI, ...) az ország területére kiátlagolva, napi bontásban.

Ezen paraméterek közül azokat, amelyeket módszereink a kutatási kérdések megválaszolása szempontjából relevánsnak találtak, a 2.fejezetben listázzuk. Az időjárást leíró indexek közül sok, speciális empirikus képlettel számolható. Ezek pontos meghatározását – elsősorban terjedelmi korlátok miatt - nem tárgyalnánk, de hatásukat figyelembe vesszük tekintve, hogy a meteorológiai adatbázis fontos adatként kezeli őket. Itt érdemes még megjegyezni, hogy néhány paraméter hiányosan szerepel az adatbázisban. Ezeket természetesen később szűrjük, a hiányos adatokkal rendelkező napokat figyelmen kívül hagyjuk. Ezt megtehetjük, mert releváns paraméterink értékét a napok több mint 90%-ában ismerjük.

2. Kiugró esetszámmal rendelkező napok

Az infarktusok előfordulásának alapos vizsgálatához elengedhetetlen az olyan napok elemzése, amelyek esetszáma jelentősen eltér az átlagostól. Ennek érdekében nemcsak a kiugróan magas (felül kiugró) esetszámú napokat vettük figyelembe, hanem azokat is, amikor a szívinfarktusok száma a vártnál alacsonyabb volt (alul kiugró napok). Az alul és felül kiugró napok összehasonlító elemzése lehetővé teszi számunkra, hogy mélyebb következtetéseket vonjunk le az infarktusok előfordulását befolyásoló tényezőkről.

Az ilyen napok vizsgálata azért különösen fontos, mert ezek tükrözik az egészségügyi rendszer terhelésének szélsőségeit. A kiugróan magas esetszámú napok során az egészségügy komoly kihívásokkal szembesül, mivel a hirtelen megnövekedett betegszám fokozott kapacitásokat igényel. Ezzel szemben az alacsony esetszámú napok elemzése segít megérteni, milyen körülmények között csökken az infarktusok előfordulása, ami értékes információkat nyújthat a preventív intézkedések és az erőforrások optimális elosztása szempontjából.

Az elemzés során megvizsgáltuk a különböző időjárási és környezeti paraméterek eloszlását az alul-, normál és felül kiugró napok függvényében. Ez lehetővé tette számunkra, hogy azonosítsuk, hogyan térnek el ezek a napok az olyan tényezők tekintetében, mint például a hőmérséklet, a légköri nyomás vagy a páratartalom. Az ilyen összehasonlítások segítenek feltárni azokat a paramétereket, amelyek jelentős hatással vannak az infarktusok előfordulására.

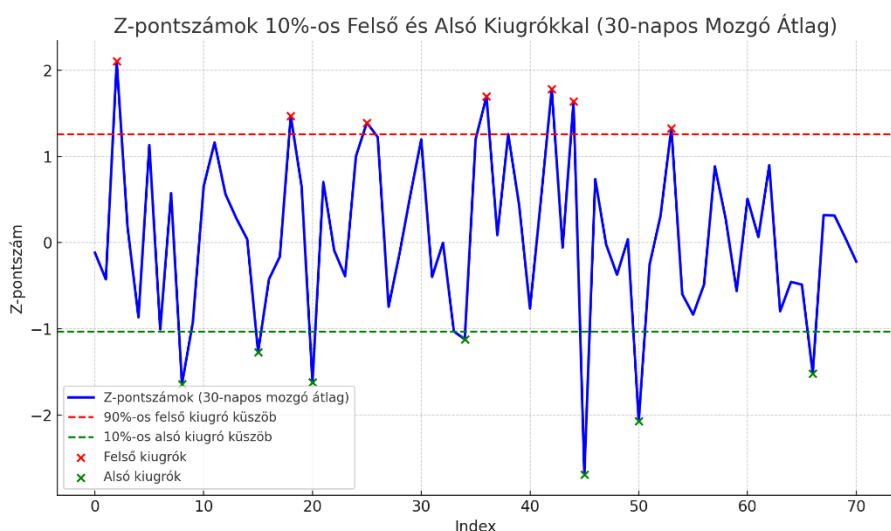
Fontos szempont volt az is, hogy az időjárási paraméterek hatása évszakonként eltérő lehet. Ezért az évet két részre bontottuk: téli és nyári időszakra. A téli és nyári hónapokban a különböző időjárási tényezők más-más módon befolyásolják az emberi szervezetet, így az infarktusok előfordulását is. Például a hideg időjárás és a hőmérsékletingadozások télen más fiziológiai stresszt jelentenek, mint a nyári hőhullámok. Ezen évszakok külön kezelésével pontosabb és relevánsabb eredményeket érhetünk el.

2.1 Kiugró napok meghatározása

A kiugró napok meghatározásához először minden naphoz rendelünk egy Z pontszámot, melyet a következő módon számolunk:

$$Z_i = \frac{x_i - \mu_i}{\sigma_i},$$

ahol Z_i az i -edik nap pontszáma, μ_i az adott nap 15 napos sugarába eső 30 nap átlagos esetszáma ($[i-15, i+15]$), σ_i ezen napok esetszámának szórása.



2.1-es ábra

Ezzel standardizáljuk adatainkat és azok az adatok fognak nagyobb, illetve kisebb Z pontszámmal rendelkezni, melyek eltérnek a körülöttük lévőktől [2.1-es ábra]. Ezután a napoknak azt a 10%-át melyek, melyek a legmagasabb Z pontszámmal rendelkeznek „felül kiugróknak”, míg azt a 10%-ot, melyek legalacsonyabbal rendelkeznek „alul kiugróknak” neveztük el.

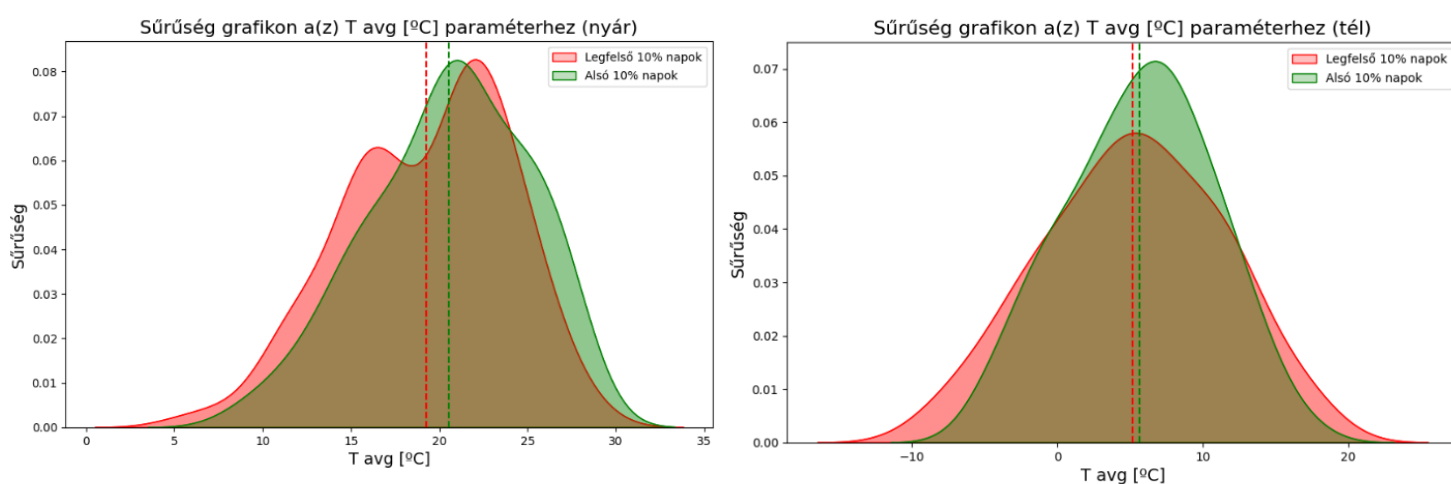
2.2 Eloszlásvizsgálat a paraméterek függvényében

Felmerül, hogy hogyan is vizsgáljuk a paraméterek hatását a kiugró napokra. Elképzelhető, hogy egyszerű módszerekkel nem mutatható ki a paraméterek jelentősége, mert hiába nagyon más az alsó/felső kiugró napok eloszlása az adott paraméter függvényében, a kapcsolat nem írható le alapvető függvények használatával.

Ennek okán magukat az eloszlásokat hasonlítjuk össze [2.2-es ábra]. Megvizsgáljuk őket a súlypont eltolódása és a Kolmogorov–Szmirnov statisztika [8] segítségével, amely megmutatja, mennyire hasonlók az eloszlások. Ez a módszer érzékeny az eloszlások eltolódására is, ami számunkra előnyös. A [8]-as forrás szerint kiszámoljuk a p-értéket, amely megadja annak valószínűségét, hogy a megfigyelt vagy annál nagyobb különbség előfordul, ha a két minta ugyanabból az eloszlásból származik. Röviden, a kisebb KS-érték és a nagyobb p-érték hasonló eloszlásra utalhat. Az eloszlásokat még egy súlypont differencia szerint is megvizsgáljuk amit a következőképpen számolunk:

$$\text{súlypont differencia \%} = \left| \frac{\mu_{\text{also}} - \mu_{\text{felso}}}{P_5 - P_{95}} \right| * 100,$$

ahol, $\mu_{\text{also}}, \mu_{\text{felso}}$ a napok alsó, felső 10%-ának súlypontja a paraméter szerinti eloszlásban, P_5, P_{95} az alsó, illetve felső napok átlagos 5 és 95 percentilise.



2.2-es ábra

Ezek alapján a tél, nyár lebontásban a következő légkörfizikai paramétereket és a levegőminőség indexet vizsgáltuk. A statisztikai vizsgálatok értékei paraméterenként:

Időjárási paraméterek	Alsó / felső súlypont differencia % NYÁR; TÉL	K-s statisztika értéke NYÁR; TÉL	P-érték NYÁR; TÉL	Megjegyzés
ciklon_c	2.6%; 1.7%	0.09; 0.07	0.87; 0.99	diszkrét értékei miatt kevésbé releváns itt
AQI (air quality index)	5.8 %; 1.9%	0.10; 0.07	0.78; 0.97	-
T avg [C°]	11%; 2.8%	0.18; 0.08	0.11; 0.95	-
WS[m/s]	1.8%; 7.6%	0.08; 0.12	0.95; 0.51	-
ΔT napok közti [C°]	7.6%; 6.8%	0.10; 0.22	0.76; 0.02	-
ΔT nap folyamán [C°]	6.3%; 0.2%	0.15; 0.08	0.30; 0.95	nincs jelentős különbség az eloszlásokban télen
Nyomás [hPa]	1.0%; 2.8%	0.13; 0.1	0.4; 0.76	-

3. Az esetszám becslése (regresszió)

3.1 Poisson regresszió

Az előző fejezetben megvizsgáltuk a kiugró esetszámmal rendelkező napok időjárását, de ez nekünk nem elég arra, hogy az egyes paraméterek pontos hatását megértsük és hogy az esetszámra releváns becslést tudjunk adni. Ehhez egyfajta regressziós eljárásra lesz szükségünk.

Mindemellett egy-egy betegség kialakulását, beleértve az infarktusokat is, nemcsak a pillanatnyi globális paraméterek befolyásolják, hanem az előző időszakot jellemzők is. Emiatt a paramétereinket időben visszamenőleg is meg kell vizsgáljuk és a regresszió részévé kell tegyük. Meteorológusokkal is egyeztetve azt a tanácsot kaptuk, hogy az eseményt megelőző néhány napot érdemes figyelembe venni. Nagyon hosszú intervallum esetén ugyanis annyiféle egyéb hatás is bejön, hogy nem lesz megállapítható, valóban az éppen használt bemeneti változó van-e kapcsolatban a vizsgált kimeneti változóval

Jelen modellünkben az eseményt megelőző hét időjárásának hatását vizsgáljuk paraméterenként. Annak megállapításához, hogy az adott nap időjárási paraméterei milyen súllyal szerepeljenek, Poisson-regressziót használunk, melyet több egészségügyi cikkben is

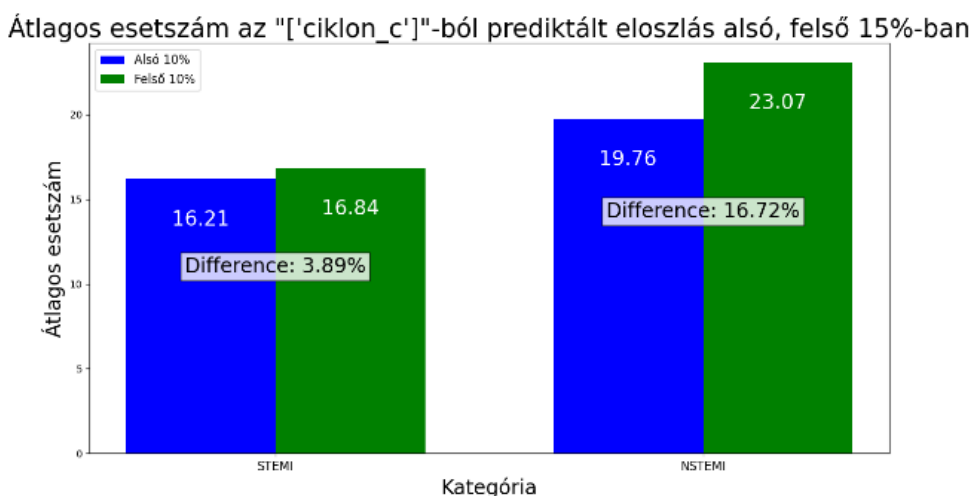
megjelenik [15]. (Lineáris regressziót is megpróbáltunk alkalmazni, de nem vagy kevésbé használható eredményeket nyújtott, mint a Poisson-regresszió.)

A Poisson-regresszió feltételezi, hogy a kimeneti változó Poisson-eloszlású. A várható esetszám (λ) a regresszióban a következőképpen fejezhető ki az x_i bemeneti változókból egy hetet tekintve:

$$\ln(\lambda) = C + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_7 x_7,$$

ahol, λ az esetszám várható értéke, x_i az $(i - 1)$ nappal korábban mért értéke az időjárást jellemző paraméternek, β_i az adott naphoz tartozó súly a regresszióban, C pedig egy jellemző konstans.

A regresszió elvégzése után azt tapasztaljuk, hogy önmagában a valódi esetszám így is kis korrelációt mutat a prediktált esetszámmal, de azt megvizsgálhatjuk már, hogy a Poisson-regresszió által kis és nagy esetszámúnak prediktált napok és azok valódi esetszáma közt mi a különbség. Ezt a STEMI és NSTEMI esetek számára külön-külön is el fogjuk végezni. Az előző fejezethez hasonlóan a regresszió által becsült esetszámok szerint sorba rendezzük napjainkat és vesszük ezen napok alsó, felső 15%-át. Azért nem a felső 10%-ot vizsgáljuk csak, mert egyszerűen azt tapasztaltuk, hogy ebben a nagyobb intervallumban is kimutathatóak jelentősebb különbségek. Emellett a 10%-nyi kiugró értékeket egy rendkívül fokozatként akartuk kezelni. Ekkor a következő ábrához hasonlóan tudjuk kiértékelni a paramétereinket [3.1-es ábra]:



3.1-es ábra

Poisson regresszió szerinti kis és nagy esetszámmal rendelkező napok valódi átlagos esetszáma és különbsége STEMI/NSTEMI lebontásban

Az alsó/felső differenciák a releváns paramétereknél STEMI/NSTEMI lebontásban [3.1-es táblázatban]:

Paraméter	STEMI differencia	NSTEMI differencia	Poisson regresszióban résztvevő napok súlyai $[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7]$ STEMI; NSTEMI
ciklon_c	3.90%	16.70%	['-0.017', '0.005', '0.023', '0.011', '-0.011', '-0.021', '-0.024']; ['0.025', '0.038', '0.043', '0.039', '0.061', '0.008', '0.003']
AQI	13.86%	17.14%	['0.007', '-0.011', '0.000', '0.000', '-0.002', '0.002', '-0.002']; ['0.010', '-0.010', '-0.001', '-0.002', '0.001', '0.003', '0.002']
T avg [C°]	15.3%	21.45%	['0.007', '-0.008', '0.002', '-0.006', '0.004', '-0.001', '-0.007']; ['0.008', '-0.014', '0.005', '-0.006', '0.007', '-0.005', '-0.007']
WS [m/s]	7.11%	8.49%	['-0.018', '0.016', '0.006', '-0.005', '0.014', '-0.004', '0.004']; ['0.014', '0.011', '0.028', '0.001', '0.022', '-0.005', '0.014']
ΔT napok közti [C°]	7.35%	8.45%	['0.003', '0.009', '-0.003', '-0.001', '0.001', '-0.001', '-0.001']; ['-0.000', '0.006', '0.004', '0.007', '0.001', '0.001', '0.003']
ΔT nap folyamán [C°]	8.24%	5.73%	['0.000', '-0.003', '0.002', '-0.000', '-0.003', '0.004', '0.002']; ['0.001', '0.004', '-0.003', '-0.004', '0.005', '0.000', '0.001']
Nyomás [hPa]	4.90%	4.31%	['-0.002', '0.004', '-0.005', '0.004', '-0.003', '0.002', '0.001']; ['-0.006', '0.007', '-0.006', '0.001', '-0.001', '0.001', '0.004']

3.1-es táblázat

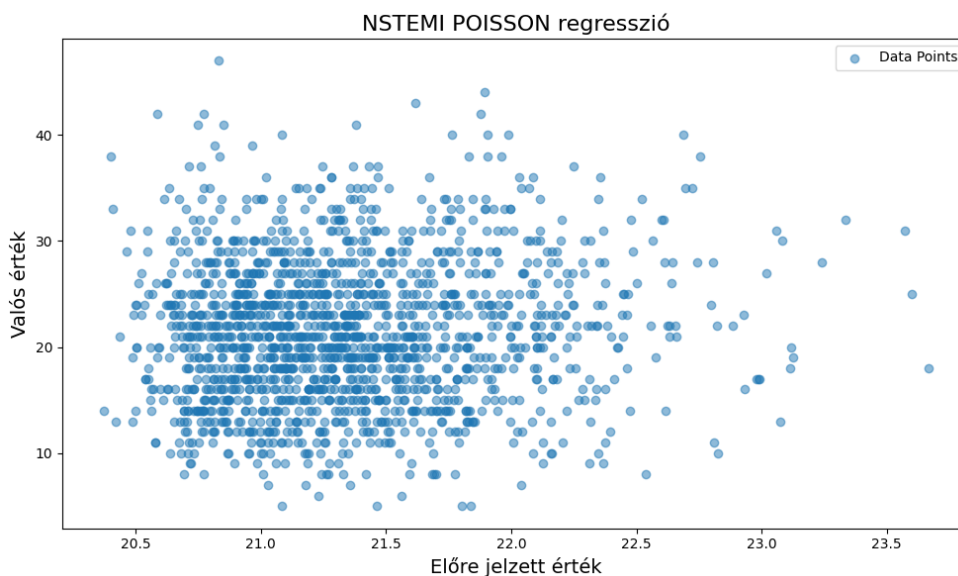
A táblázatunkban látható, hogy valamilyen szinten minden paraméter változása kapcsolatban van az esetszámmal, továbbá a legtöbb paraméternél azt tapasztaljuk, hogy az NSTEMI esetszám jóval érzékenyebb a változásokra, mint a STEMI. Ez a fajta érzékenységbeli különbség dolgozatunk további vizsgálatainál is megfigyelhető lesz.

Ezen eljárásoknál még sokszor érdemes megnézni a regresszió együtthatóit, mert sokat elárulhatnak a hatás jellegéről. Az átlaghőmérsékletnél a nagyobb abszolútértékű együtthatók rendre negatívak, ezzel utalva arra, hogy egy fordított arányosság áll fenn a hőmérséklet és az esetszám közt [9], viszont egy előjelváltakozás is megfigyelhető. Ez arra is utalhat, hogy a hőmérséklet váltakozásának is nagy szerepe van. Ezt láthatjuk a ΔT napok közti [C°] változónál, mely az előző nap és az adott nap átlaghőmérsékletei különbségének veszi az abszolút értékét. Itt minden súly pozitív utalva arra, hogy a változások folyamatos terhet jelenthetnek a szervezeteknek. Ennek ellenére a ΔT nap folyamán [C°] változó mely a napközbeni hőingadozást kvantifikálja furcsa súlyokkal rendelkezik, hiába várnánk azt, hogy a napközbeni

hőingás is egyértelműen hozzájáruljon az esetszám növekedéséhez. Ennek persze számos oka lehet, elképzelhető, hogy szimplán az index számításának a módszere nem megfelelő.

3.2 Robusztusabb regressziós modell, XGBoost

Az előző fejezet Poisson-regressziója képes volt trendeket kimutatni adatainkban és rámutatni, hogy sok légkörfizikai és -kémiai tényező bírhat információértékkel. Ettől függetlenül esetszámhelyes becslésére látszólag nem volt alkalmas [3.2-es ábra]. Ennek egyszerű oka, hogy az esetszám közel sem egyszerű függvények szerint írható le a globális paraméterek által.

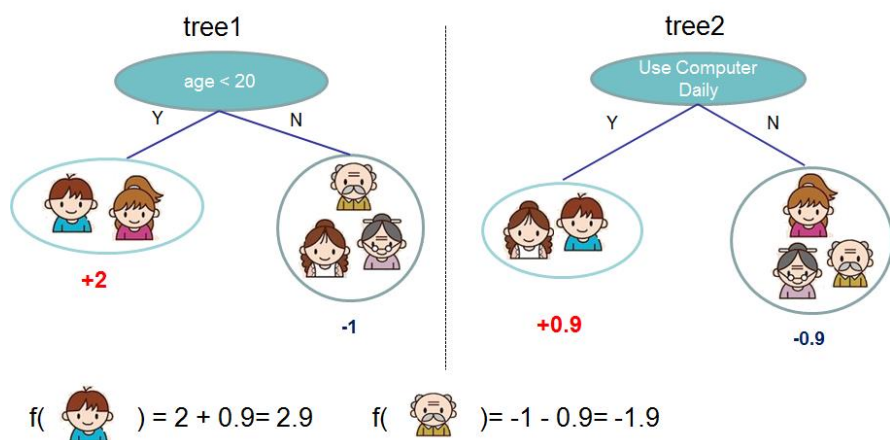


3.2-es ábra

Ahogy a [3.2-es ábrán] is láthatjuk a regresszió szimplán levágta az értékeket és egy nagyon kis tartományon belül próbálta jóslni az esetszámot. Ez nem alkalmas az említett becslés elkészítésére.

Ezért robusztusabb, gépi tanulás alapú módszerekhez fogunk folyamodni. Itt már az összes időjárást jellemző paramétert a modell bemeneteként szeretnénk kezelni és megnézni, hogy ezen tényezők közül melyek befolyásolják jobban a kimenetet. Ehhez először egy fully connected neurálishálózattal próbálkoztunk, de mivel minden napra csak egy adat jutott nekünk, így összesen kb. 3000 tanító adatvektorunk volt. Ez sajnos nem elég ahhoz, hogy a neurális hálónk rá tudjon tanulni a komplex rendszert leíró függvényekre.

Ennek okán egy másik metódushoz, az úgynevezett XGBoost-hoz [1-3] folyamodtunk. Ez egy gépi tanulási algoritmus, mely a tanulása alatt döntési fákat hoz létre és így jut el egy

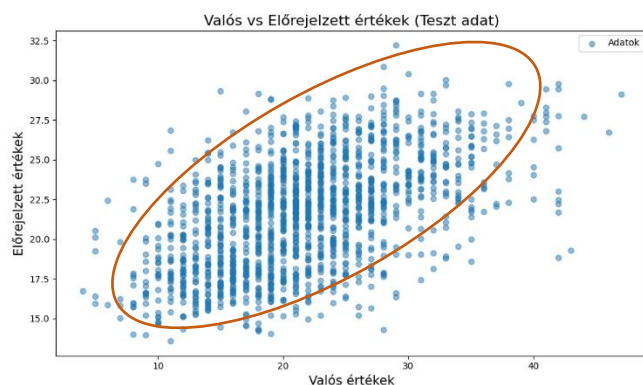
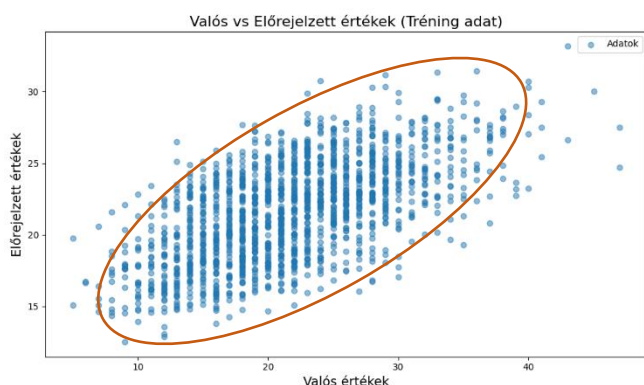


3.3-as ábra [3]

kimenthez. A döntési fák a [3.3-as ábra] -n látható elven működnek. Miért is jó ez nekünk? Emlékezzünk vissza a kiugró napoknál vizsgált eloszlásfüggvényekre. Néhány gyakran huplis, fura formát vesz fel, de a döntési fák segítségével ki tudjuk vágni azokat a régiókat az eloszlásokból amikor az esetszám

kisebb vagy éppen magasabb. Ezeket összetéve egy sokkal pontosabb regressziót kaphatunk. Emellett a modellnek egy plusz bemeneti paraméterként az előző napok esetszámát is odaadjuk. Ennek köszönhetően olyan tendenciákra is rá tud tanulni, melyek nem mutathatók ki az időjárásból.

A regressziós eljárás hyper paramétereinek, scikit-learn programcsomag [4]-beli grid-search algoritmussal történő illesztését és a tanítást követően a következő identitás grafikont kapjuk az NSTEMI esetszámra [3.4-es ábra]:

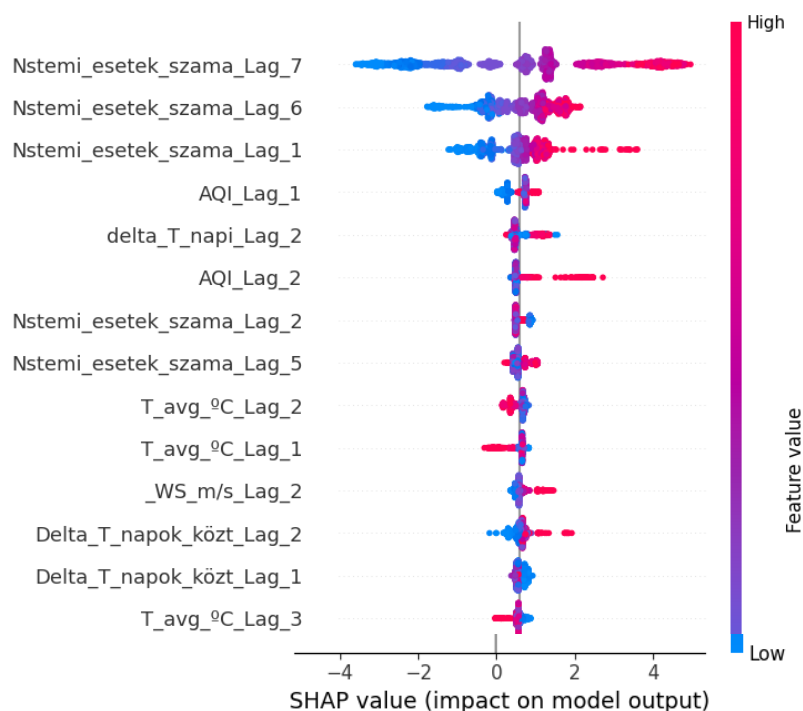


3.4-es ábra

A regresszió abszolút és NSTEMI esetszámra jól működik. A tesztadatokon a prediktált és a valódi értékek közt az R^2 érték 0.25 és 0.3 környékén mozog. Ezt ilyen nehezen leírható rendszereknél kimondattan jó eredményként kezelhető. A STEMI adatokra viszont nem lehet

ezzel az eljárással sem jól illeszteni. Úgy tűnik, hogy ezen események bekövetkezése sokkal kevésbé vagy sokkal bonyolultabban függ az általunk vizsgált globális paraméterektől.

Most nézzük meg, mely paraméterek fontosak a modellünk szempontjából. Ehhez az SHAP [5-6] értékeket fogjuk felhasználni, melyek számítása egy Pythonba integrálható könyvtárként [5-6] is lehetséges. Ez egy olyan eljárás, mely során egyes paraméterek értékének változtatásával vagy teljes elhagyásával megvizsgáljuk, mely paraméterekre érzékeny leginkább a modellünk.



3.5-ös ábra

Ezen az ábrán [3.5-ös ábra] láthatjuk a modell kimentet leginkább befolyásoló különböző paramétereket SHAP értékek szerint. A LAG a paraméter neve mellett csak annyit jelent, hogy hány nappal korábbi értéket veszünk figyelembe (1-es időjárás Lag az aznapi időjárást, de az 1-es esetszám Lag az egy nappal korábbi esetszámot jelöli). A

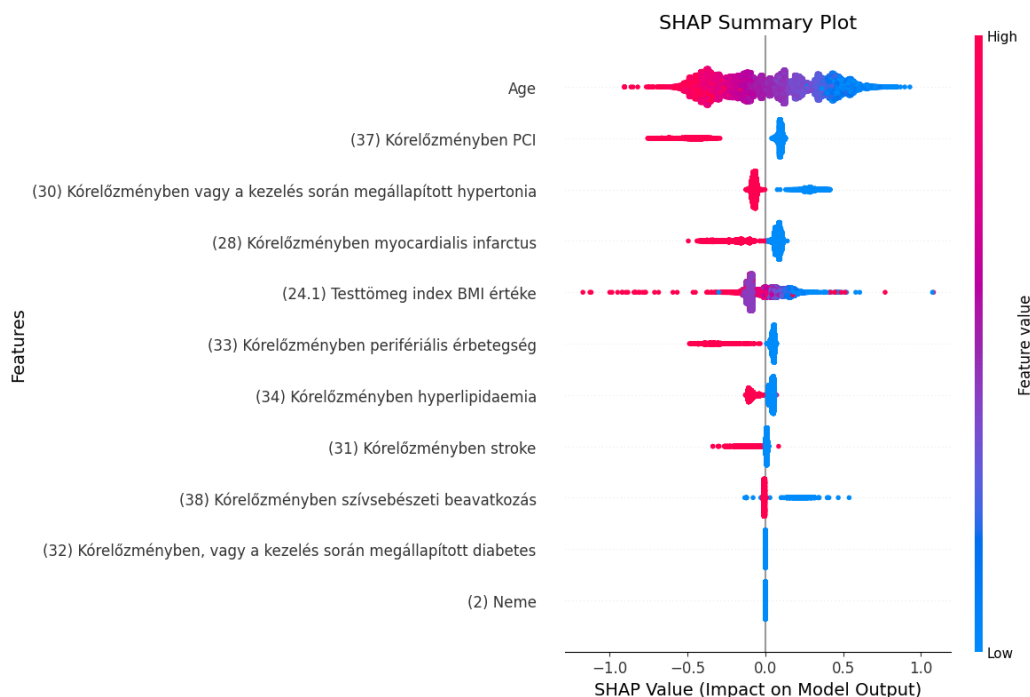
pontok színe adja meg, hogy az adott paraméterünk értéke értelmezési tartományán belül mekkora volt, míg a grafikonon való vízszintes elhelyezkedés azt mondja meg nekünk, hogy az esetszámot növelte vagy csökkentette. A [3.5-ös ábrán] láthatjuk például, hogy amennyiben az előző napok esetszáma (*Nstemi_esetek_szama_Lag_i*) nagyobb volt a vizsgált nap esetszáma is nagyobb lesz a modell szerint. Amennyiben megvizsgáljuk a *T_avg* értékeket, melyek a napi átlaghőmérsékletet reprezentálják egy fordított arányosságot figyelhetünk meg, azt látjuk, hogy a nagyobb piros értékek a grafikon bal oldalán helyezkednek el, azaz esetszám csökkenést eredményeznek.

4. STEMI / NSTEMI-t meghatározó paraméterek

Ebben a szekcióban egy olyan módszert dolgozunk ki, melynek célja, hogy az ismert egyéni paraméterek alapján – tehát az EKG eredmények ismerete nélkül - meghatározza, hogy amennyiben egy betegben kialakul az infarktus az STEMI vagy NSTEMI típusú lesz-e.

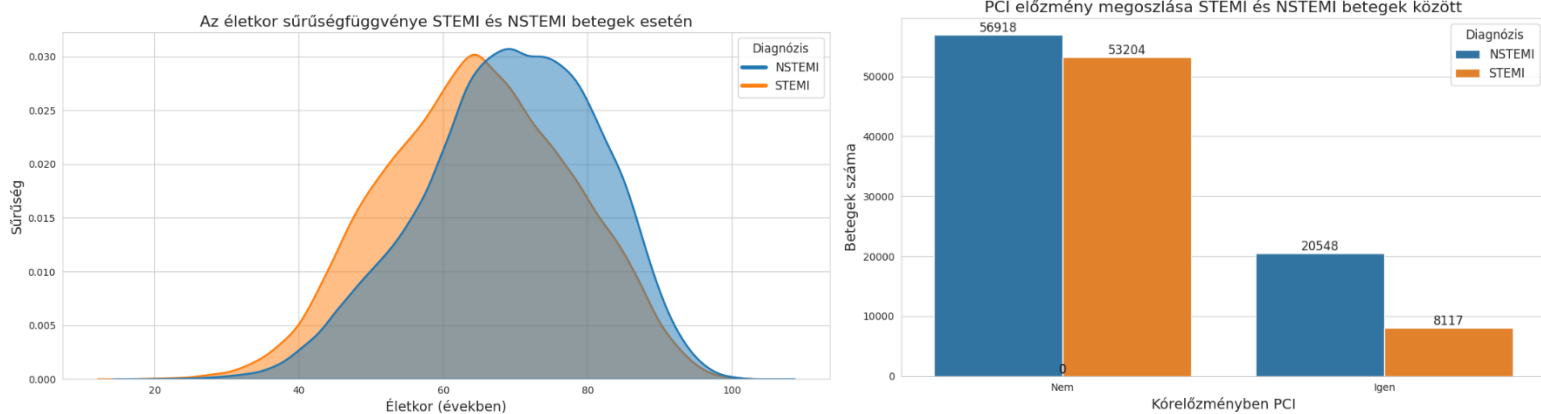
4.1 Klasszifikáló algoritmus

A klasszifikáció bemeneti paraméterei az ismert kórelőzmények [4.1-es ábra] lesznek, az életkor és a BMI értéke. Itt is először a korábban már leírt XGBoost nevű eljárással alkottunk modellt. Ekkor azt kaptuk, hogy az NSTEMI eseteket 72%-os míg a STEMI eseteket 49%-os pontossággal tudjuk eltalálni. Az egyértelmű, hogy modellünk nem tudja még jól megmondani, hogy mi az, „ami STEMI-vé tesz” egy esetet. Ezt követően megpróbálkoztunk még egy hybrid modellel is mely az XGBoost kimeneti hibáját beteszi egy neurális hálóba, mint bemenet az eredeti bemenetek mellett, hogy picit simítson rajtuk. Ez a módszer sem javított a klasszifikáción. Az SHAP értékeket az alábbi grafikonon láthatjuk [4.1-es ábra].



4.1-es ábra

Az ábrán a negatív értékek tartoznak az NSTEMI-hez, míg a pozitívak a STEMI-hez. A modellünk szerint az életkornak lehet a legnagyobb jelentősége és még a korábbi PCI-k rendelkeztek nagy információ értékkel. Ennek ellenére azt még nem értjük, hogy miért nem tudjuk jól meghatározni, hogy az eset SETMI-e. Ahhoz, hogy erre jobb rálátással rendelkezünk, ábrázoljuk a STEMI/NSTEMI eloszlást a két legrelevánsabb paraméterünk függvényében [4.2-es ábra].



4.2-es ábra

Azt láthatjuk, hogy az életkornak tényleg van szerepe a STEMI/NSTEMI eldöntésében. Emmelett természetesen a PCI meglétének is, viszont, amennyiben már volt PCI az NSTEMI esélye sokkal magasabb, hiányában viszont a STEMI/NSTEMI arány az kb. 1:1-hez. Ez azt jelenti nekünk, hogy amennyiben a kórelőzmény jelen volt, a modellünk magabiztosan rá tudta mondani, hogy az eset NSTEMI, ezzel elérve a 72%-os pontosságot, viszont hiányában nehezen tudott dönten a típusok közt. Ha ennek tudatában újra egy pillantást vetünk SHAP értékeket tartalmazó [4.1-es ábrára], ténylegesen az látszik a többi kórelőzménynél is, hogy meglétük erősen az NSTEMI esetfajtra utalhat, de hiányukban nehéz egy döntést meghozni.

5. Ok-okozatiság

5.1 Ok-okozati vizsgálatok

Az ok-okozati összefüggések feltárásai napjainkban egyre nagyobb figyelmet kapnak a tudományos kutatás és az alkalmazott statisztikai elemzések területén, különösen azokban az esetekben, amikor a változók közötti tényleges okozati kapcsolat megismerése kritikus a helyes döntéshozatal szempontjából. Az egyszerű korreláción alapuló elemzések, mint például a Poisson-regresszió vagy a prediktív gépi tanulási algoritmusok – például az XGBoost – gyakran nem alkalmasak arra, hogy feltárják az adatok mögötti oksági kapcsolatokat, mivel csupán a

változók közötti összefüggéseket mutatják, anélkül, hogy bármilyen irányított okozati összefüggésre utalnának. Ezen igények kielégítésére fejlesztették ki például a DoWhy csomagot [16, 17], amely célzottan az ok-okozati elemzések támogatására alkalmas eszközkészletet kínál.

5.1.1 Az ok-okozati elemzés alapjai és módszertani kérdései

Az ok-okozati elemzés alapvetően arra a kérdésre keresi a választ, hogy egy adott változás hatására milyen következmények várhatók egy másik változóra nézve, illetve hogy milyen mértékben befolyásolja az egyik változó a másikat. Az ok-okozati kapcsolat nem pusztán statisztikai összefüggés: a cél, hogy tényleges oksági összefüggéseket azonosítsunk, azaz kimutassuk, hogy az egyik változó módosítása közvetlenül okozza a másik változó módosulását. Ehhez számos különféle módszertani megközelítést alkalmaznak, amelyek között szerepelnek a kísérleti és kvázi-kísérleti módszerek, instrumentális változók, valamint különféle súlyozási és ok-okozati gráf alapú technikák.

5.1.2 Kísérleti és kvázi-kísérleti megközelítések

A kísérleti megközelítések, mint például a randomizált kontrollált kísérletek (RCT-k), az ok-okozati kapcsolat felismerésének egyik legmegbízhatóbb módját jelentik. Ebben a megközelítésben a kontrollcsoport és a beavatkozási csoport randomizált módon kerül kiválasztásra, így az elemzés során a különböző csoportok közötti eltérések ténylegesen a vizsgált beavatkozás hatásának tudhatók be. Mivel a mi vizsgálataink retrospektívek voltak, illetve az időjárást nem állt módunkban befolyásolni, így ez a megközelítés nem volt alkalmazható.

Amennyiben az RCT-k alkalmazása nem lehetséges, a kvázi-kísérleti megközelítések – mint például a propensity score matching, azaz hajlandósági pontozásos illesztés, vagy a difference-in-differences módszer – nyújtanak megoldást. Ezek a technikák úgy próbálják kezelni az elfogultságot, hogy a hasonló jellemzőkkel rendelkező megfigyeléseket párosítják vagy különbségeik vizsgálatán keresztül becslik a tényleges okozati hatást.

5.1.3 Instrumentális változók és fordított valószínűség súlyozás

A kvázi-kísérleti módszerek mellett fontos szerepet kap az instrumentális változók módszere is, amely lehetőséget ad arra, hogy olyan változók segítségével vizsgáljuk az okozati hatást, amelyek külső hatásként befolyásolják a vizsgált változókat. Az instrumentális változók módszerét tipikusan akkor alkalmazzák, amikor egy változó közvetlenül nem manipulálható vagy mérhető, illetve amikor a változó és a célváltozó között fennáll egy rejtett zavaró tényező. Emellett a fordított valószínűség súlyozás (Inverse Probability Weighting, IPW)[18] is gyakran használt eljárás, amely a mintavételi torzításokat és a kiválasztási elfogultságot próbálja mérsékelni. Az IPW során minden megfigyelés egy súlyt kap, amely arányos az adott megfigyelés valószínűségével, hogy egy bizonyos kezelési csoportba kerüljön, így pontosabban becsülhető a kezelési hatás.

A DoWhy csomag az ok-okozati elemzés új megközelítését nyújtja azáltal, hogy lehetőséget ad az oksági gráfok felépítésére és azok elemzésére. Az oksági gráf (Directed Acyclic Graph, DAG) alapú elemzés során a változók közötti hipotetikus ok-okozati kapcsolatok explicit módon kerülnek megadásra, majd ezeket a kapcsolatrendszereket különféle elemzési és cáfoló tesztekkel lehet ellenőrizni. A DoWhy használata során a kutatók először meghatározzuk az oksági gráfot, amely reprezentálja az egyes változók közötti feltételezett kapcsolatokat. Ezt követően a DoWhy különböző technikákat kínál az oksági összefüggések vizsgálatára, beleértve az "identification" és "estimation" lépéseket, valamint a "refutation" teszteket, amelyek célja, hogy megkérdőjelezzék az oksági gráfok helyességét. Ezek a tesztek, például az "placebo treatment test" vagy a "subset test," segítenek megerősíteni vagy megcáfolni a modell feltételezéseit, növelve ezzel a következtetések hitelességét.

5.1.4 Következmény-elemzés (Counterfactual Analysis)

Az egyik leghasznosabb módszer, amelyet a DoWhy és más ok-okozati elemzési technikák is támogatnak, a következményelemzés, amely arra keresi a választ, hogy mi történne, ha egy adott körülmény eltérő lett volna. Ez az elemzési forma az "ellentétes tények" vizsgálatán alapul: például, ha egy személy nem kapott egy bizonyos kezelést, de vizsgálni szeretnénk, hogy mi történt volna, ha megkapta volna azt. A következményelemzés során a kutatók lehetőséget kapnak arra, hogy megbecsüljék az alternatív valóságokat, vagyis olyan szituációkat, amelyek a való életben nem történtek meg, de amelyek befolyásolhatják a vizsgált

változó alakulását. Ezáltal a következményelemzés mélyebb megértést nyújt a változók közötti okozati kapcsolatokról, hiszen így nem csupán a megfigyelt adatokra korlátozódik az elemzés. [19]

5.1.5 A DoWhy előnyei a hagyományos modellekkel szemben

A DoWhy és más hasonló ok-okozati elemzési keretrendszerek előnyei az egyszerű statisztikai vagy gépi tanulási modellekhez képest abban rejlenek, hogy képesek tényleges okozati összefüggéseket feltárni. A Poisson-regresszió és az XGBoost például korrelatív vagy prediktív modellezési eszközök, amelyek jól működnek a prediktív feladatoknál, de nem alkalmasak oksági következtetések levonására. Ezzel szemben a DoWhy oksági gráf alapú megközelítése lehetőséget nyújt arra, hogy az adatok mögötti rejtett összefüggések feltárásával mélyebb megértést kapjunk az egyes változók közötti tényleges okozati hatásokról. A következő szempontok különösen előnyösnek bizonyulnak:

- Az ok-okozati irány feltárása: Míg a statisztikai modellek általában nem képesek az okozati irány meghatározására, addig a DoWhy explicit módon az ok-okozati összefüggések azonosítására épül, így alkalmasabb a tényleges hatások feltárására.
- Zavaró változók kezelése: Az oksági modellek figyelembe veszik a zavaró tényezők hatásait, így csökkentik a hibás következtetések lehetőségét.
- Beavatkozások hatásának értékelése: A DoWhy képes hipotetikus forgatókönyveket is elemezni, ezzel lehetőséget nyújtva arra, hogy a beavatkozások tényleges hatását becsüljük, szemben a statisztikai modellek prediktív megközelítésével.
- Cáfoló tesztek: A DoWhy által biztosított cáfoló tesztek (például a placebo tesztek) lehetővé teszik, hogy megvizsgáljuk az oksági modell feltételezéseinek helyességét, ezzel növelve a következtetések megbízhatóságát.

Összességében elmondható, hogy az ok-okozati elemzések – különösen a DoWhy által nyújtott eszközök és módszerek – jelentős előnyt kínálnak a hagyományos statisztikai és gépi tanulási modellekkel szemben, amikor az oksági kapcsolatok felismerése a cél. Az ok-okozati elemzési technikák lehetővé teszik, hogy ne csupán a változók közötti összefüggéseket térképezzük fel,

hanem pontosabb képet kapjunk arról, hogy egy változó milyen tényleges hatást gyakorol a másokra. Az ilyen típusú elemzések különösen fontosak azokban az esetekben, amikor a vizsgált változók közötti kapcsolat befolyásolhatja a döntéshozatalt, vagy amikor a beavatkozások következményeit kell pontosan felmérni.

5.2 Oksági vizsgálat eredménye adatbázisainkon

Az infarktusokkal kapcsolatos vizsgálatainkban is szeretnénk volna megérteni, hogy a megelőző számításokból kapott összefüggések mögött ok-okozati kapcsolat áll-e. Ezért a regressziós számításoknál kapott jelentős paraméterek esetén ok-okozati elemzést végeztünk.

Ezzel az alábbi megállapításokat igazoltuk:

1. A levegőminőségi index (Air Quality Index, AQI) és az esetszám között szignifikáns ok-okozati kapcsolat van. A levegő minőségének romlása növeli az infarktus bekövetkezésének esélyét. Itt érdemes megjegyeznünk, hogy hazánkban - szerencsére – ritkán van olyan erős légszennyezettségi körülmény, ami érdemben emelné az esetszámot.
2. A Wind Effect Index (WEI) és az esetszám között szignifikáns ok-okozati kapcsolat van. A WEI egy az általunk is vizsgált légkörfizikai paraméterekből számított dimenziótlan index, mely a hőérzettel van kapcsolatban.
3. Láttuk korábban, hogy az eseményt megelőző időszakban elszenvedett hőmérsékletváltozások abszolútértékei hatással vannak az esetszámmra. Ennek ok-okozati vizsgálatával azt kaptuk, hogy a hőmérsékletváltozások abszolútértékének 3 napos szummázással hatása van az esetszámmra. Az oksági kapcsolat a nyári esetszámváltozásra szignifikáns, míg 7 napos összegzéssel az egész évre kiterjesztve azt kaptuk, hogy van oksági hatás, de nem szignifikáns.

Fontos még megjegyezni, hogy a nem ok-okozati összefüggésnek is lehet információ értéke.

6. Összegzés

Az új adatelemzési módszerek lehetővé tették az egészségügyi adatok korábbinál pontosabb elemzését, a kapott eredmények szignifikáns voltának vizsgálatát.

Dolgozatunk során a rendelkezésre álló adatbázisok feltérképezését követően egy mozgó átlagokon alapuló algoritmussal kiszűrtük a kiugró esetszámmal rendelkező napokat és megvizsgáltuk egyes paraméterek függvényében a betegszám eloszlását az említett napokon. Ez utalást adott arra, hogy mely paraméterek lehetnek számunkra különösen fontosak. Ezt követően Poisson-regresszió segítségével megmutattuk, hogy lehetséges, akár egészen jelentős esetszámbeli különbségeket kimutatni egyes időjárást jellemző tényezők hatásáról. A regresszió javítása érdekében XGBoost segítségével is összeállítottunk egy esetszám prediktív modellt, mely amellet, hogy relevánsan volt képes az esetszám becslésére, segített megismerni a leginkább fontos paramétereket. Eredményül kaptuk, hogy a betegszám alakulása az előző napokban kimondottan nagy információ értékkel rendelkezik az adott nap esetszámáról. Elemzéseinket STEMI/NSTEMI bontásban is végeztük és azt tapasztaltuk, hogy az NSTEMI esetszám sokkal jelentősebben érzékeny a különböző általunk mért paraméterekre, mint a STEMI. Egy STEMI/NSTEMI kórelőzmények szerinti klasszifikációs algoritmust is elkészítettünk szintén XGBoost segítségével. Klasszifikációs algoritmusunk nem nyújtott kimagasló pontosságot, de segítségével könnyen észrevehettünk trendeket, miszerint az infarktus bekövetkezésekor a kórelőzmények léte és a magas életkor mind az NSTEMI kockázati faktorai közé tartoznak. Végezetül relevánsnak ítélt paramétereinken ok-okozati vizsgálatokat is végeztünk, ezzel mélyebb kapcsolatokat kimutatva, az időjárás jellemző tényezők hatása közt.

Összességében kidolgoztunk és bemutattunk egy olyan eljárásomagot, amely segítségével számos rizikófaktor és csoport azonosítható, az időjárási paraméterek gyenge. sok esetben nehezen kimutatható hatása számszerűsíthető. A kapott eredmények az infarktus, mint kórkép pontosabb megismerésében segíthetnek, de hasznosak lehetnek a betegellátás szervezésében is.

A bemutatott eljárások általánosak: más kórképekre is könnyen átvihetők.

Megjegyzés: A dolgozat tartalmának egy részét már előadtuk a Szegeden megrendezett SciComp24 konferencián.

7. Köszönetnyilvánítás

Ezennel megköszönném Dr. Szilágyi Brigittának és Sipos Bencének a segítségét, melyet a dolgozat elkészítése alatt nyújtottak nekem.

Emellett a kutatás megvalósulását az Egyetemi Kutatói Ösztöndíj Program támogatta, a Kulturális és Innovációs Minisztérium, valamint a Nemzeti Kutatási, Fejlesztési és Innovációs Alap segítségével.

A tanulmányt az RRF-2.3.1-21-2022-00014 Éghajlatváltozás Multidiszciplináris Nemzeti Laboratórium támogatta.

A TKP2021-EGA-02 számú projekt a Kulturális és Innovációs Minisztérium Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott támogatásával, a TKP2021-EGA pályázati program finanszírozásában valósult meg.

8. Irodalomjegyzék

- [1] <https://xgboost.readthedocs.io/en/stable/> (XGB dokumentáció)
- [2] <https://www.nvidia.com/en-eu/glossary/xgboost/> (XGB kvalitatív leírás)
- [3] <https://xgboost.readthedocs.io/en/stable/tutorials/model.html> (XGB magyarázó ábra)
- [4] https://scikit-learn.org/dev/modules/generated/sklearn.model_selection.GridSearchCV.html
(scikit-learn grid search dokumentáció)
- [5] <https://pypi.org/project/shap/#citations> (SHAP weboldal)
- [6] Lundberg, S.M., Erion, G., Chen, H. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* **2**, 56–67 (2020). <https://doi.org/10.1038/s42256-019-0138-9> (SHAP cikk)
- [7] <https://my.clevelandclinic.org/health/diseases/22233-nstemi-heart-attack> (Stemi/Nstemi ekg)
- [8] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html> (KS-teszt Python dokumentáció)
- [9] Bhaskaran, Krishnan, et al. "Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry." *Bmj* 341 (2010).
- [10] Mohammad MA, Koul S, Rylance R, Fröbert O, Alfredsson J, Sahlén A, Witt N, Jernberg T, Muller J, Erlinge D. Association of Weather With Day-to-Day Incidence of Myocardial Infarction: A SWEDEHEART Nationwide Observational Study. *JAMA Cardiol.* 2018 Nov 1;3(11):1081-1089. doi: 10.1001/jamacardio.2018.3466. PMID: 30422202; PMCID: PMC6583065.
- [11] Gebhard C, Gebhard CE, Stähli BE, Maafi F, Bertrand MJ, Wildi K, Fortier A, Galvan Onandia Z, Toma A, Zhang ZW, Smith DC, Spagnoli V, Ly HQ. Weather and risk of ST-elevation myocardial infarction revisited: Impact on young women. *PLoS One.* 2018 Apr 9;13(4):e0195602. doi: 10.1371/journal.pone.0195602. PMID: 29630673; PMCID: PMC5891074.

- [12] Rus, AA., Mornoş, C. The Impact of Meteorological Factors and Air Pollutants on Acute Coronary Syndrome. *Curr Cardiol Rep* 24, 1337–1349 (2022). <https://doi.org/10.1007/s11886-022-01759-5>
- [13] Li, N., Ma, J., Liu, F. et al. Associations of apparent temperature with acute cardiac events and subtypes of acute coronary syndromes in Beijing, China. *Sci Rep* 11, 15229 (2021). <https://doi.org/10.1038/s41598-021-94738-9>
- [14] https://www.ksh.hu/stadat_files/nep/hu/nep0010.html (központi statisztikai hivatal honlapja)
- [15] Crowther, M.J., Riley, R.D., Staessen, J.A. et al. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol* 12, 34 (2012). <https://doi.org/10.1186/1471-2288-12-34>
- [16] Sharma, Amit, and Emre Kiciman. "DoWhy: An end-to-end library for causal inference." *arXiv preprint arXiv:2011.04216* (2020).
- [17] Blöbaum, Patrick, et al. "DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models." *Journal of Machine Learning Research* 25.147 (2024): 1-7.
- [18] Mansournia, Mohammad Ali, and Douglas G. Altman. "Inverse probability weighting." *Bmj* 352 (2016).
- [19] Menzies, Peter, and Helen Beebee. "Counterfactual theories of causation." (2001).
- [20] 13. **International Journal of Cardiology**. [Temperature Effects on STEMI] (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5891074/>)