Komplex egyetemi feladat Hallgató részére Copilot támogatással több tantárgy kapcsán

(Complex university assignment for a student with Copilot support across multiple subjects)

avagy

Az EU országok IT-biztonsági szintjének MI-alapú elemzése szoftverfrissítésre vonatkozó Google Trends-adatok alapján

(AI-based analysis of the IT security level of EU countries based on Google Trends data on software updates)

Pitlik László (MY-X team), Szilágyi Ödön (KJE)

# Bevezetés

Magánegyetemeken nagy (de bármilyen oktatási intézményben nem nulla) gyakorisággal előfordul, hogy egy-egy Hallgató több tantárgy kapcsán is lemaradásba kerül és ezen tantárgyak egymással akár összefüggésbe is hozhatók lennének, vagyis egyetlen egy nagy, komplex feladattal lehetne jegyet szerezni minden tárgyból egyszerre. A pedagógiai alapvetés egyszerű: cél az oktatókkal való közös alkotás képességének támogatása (szemben a Hallgatói önkényes tévelygéssel), a gyakorlatiasság fókuszba állítása (a KNUTH-elv mentén: vö. [https://miau.my-x.hu/miau2009/index\_tki.php3?\_filterText0=\*knuth](https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth)), ill. íróasztalfióknak készülő tesztek, elszálló pillanatokat jelentő szóbeli vizsgák helyett maradandó = publikálható teljesítmények létrehozása (vö. <https://miau.my-x.hu/miau2009/index.php3?x=e080> – IKSAD konferenciák Hallgatói részvétellel – gyakorlásként a szakdolgozatvédésre – TDK helyett/mellett).

# Esettanulmány

Prompt:

Javasolj egy konkrét problémát és egy kapcsolódó publikációra vonatkozó címet, ahol egyszerre igaz, hogy a téma érinti az operációs rendszerek kérdéskörét, az információs rendszerek kérdéskörét és az integrált vállalati információs rendszerek kérdéskörét ÉS van publikusan kellően sok adat egy minőségi problémamegoldáshoz, elemzéshez. Lehetőség szerint az adatok forrása legyen a Google Trends. A probléma megoldása legyen hasznos egy potenciálisan létezhető vállalat számára operatív és/vagy stratégiai szinten.

Copilot:

Egy érdekes és releváns probléma, amely érinti az operációs rendszerek, az információs rendszerek és az integrált vállalati információs rendszerek kérdéskörét, a következő lehet:

**Probléma:**

**Az operációs rendszerek frissítési gyakoriságának hatása a vállalati információs rendszerek teljesítményére és biztonságára: Egy Google Trends alapú elemzés**

**Kapcsolódó publikáció címe:**

**"Az operációs rendszerek frissítési gyakoriságának hatása a vállalati információs rendszerek teljesítményére és biztonságára: Integrált megközelítés Google Trends adatok alapján"**

**Probléma leírása:**

Az operációs rendszerek frissítései kritikus szerepet játszanak a vállalati információs rendszerek teljesítményének és biztonságának fenntartásában. Azonban a frissítések gyakorisága és időzítése jelentős hatással lehet a rendszerek stabilitására és a vállalati működésre. Az információs rendszerek és az integrált vállalati információs rendszerek szempontjából fontos megérteni, hogy a frissítések hogyan befolyásolják a rendszer teljesítményét és biztonságát.

**Adatok forrása:**

A Google Trends adatai felhasználhatók az operációs rendszerek frissítési gyakoriságának és a kapcsolódó keresési trendek vizsgálatára. Az adatok elemzése segíthet azonosítani a frissítések időbeli eloszlását és azok hatását a vállalati információs rendszerek teljesítményére és biztonságára.

**Hasznosság:**

A probléma megoldása operatív szinten segíthet a vállalatoknak optimalizálni a frissítési folyamatokat, csökkenteni a leállási időket és növelni a rendszerek biztonságát. Stratégiai szinten hozzájárulhat a hosszú távú tervezéshez és a kockázatkezeléshez, valamint javíthatja a vállalati információs rendszerek hatékonyságát és megbízhatóságát.

Remélem, ez a javaslat hasznos lesz számodra! Van esetleg más téma, amiben segíthetek?

Értékelés: Látszólag racionális felvetés, de az egész Copilot-javaslat mégis csak egy nagy és üres buborék mindaddig, amíg nincsenek valós adatok. S ráadásul félő ezen a ponton, hogy a frissítési statisztikák (vállalati szinten) olyan jelenségkört képviselnek, melyek nem a Google Trends objektumainak attribútumaiként értelmezhetők direkt módon. Lássuk, mikor jön rá erre a Copilot:

Prompt:

Kérlek, add meg pontosan milyen lekérdezéseket kell végrehajtanom hétköznapi felhasználóként a Google Trends-ben ahhoz, hogy az általad javasolt elemzést végre lehessen hajtani itt és most általam azonnal?

Copilot:

Természetesen! Az alábbiakban részletezem, milyen lekérdezéseket kell végrehajtanod a Google Trends-ben az operációs rendszerek frissítési gyakoriságának hatásának elemzéséhez:

**1. Operációs rendszerek frissítési gyakorisága**

* **Lekérdezés:** "Windows update frequency" vagy "operating system update frequency"
* **Időtartam:** Az elmúlt 5 év
* **Földrajzi hely:** Világszerte vagy specifikus országok (pl. USA, Magyarország)

**2. Információs rendszerek teljesítménye**

* **Lekérdezés:** "system performance after update" vagy "operating system performance impact"
* **Időtartam:** Az elmúlt 5 év
* **Földrajzi hely:** Világszerte vagy specifikus országok

**3. Információs rendszerek biztonsága**

* **Lekérdezés:** "system security after update" vagy "operating system security impact"
* **Időtartam:** Az elmúlt 5 év
* **Földrajzi hely:** Világszerte vagy specifikus országok

**4. Vállalati információs rendszerek**

* **Lekérdezés:** "enterprise information systems update impact" vagy "business information systems performance"
* **Időtartam:** Az elmúlt 5 év
* **Földrajzi hely:** Világszerte vagy specifikus országok

**Lépések a Google Trends használatához:**

1. Nyisd meg a Google Trends weboldalt.
2. Írd be a fenti lekérdezéseket egyenként a keresősávba.
3. Állítsd be az időtartamot és a földrajzi helyet a kívánt paraméterek szerint.
4. Vizsgáld meg a keresési trendeket és az adatokat, majd hasonlítsd össze az eredményeket.

Ezekkel a lekérdezésekkel releváns adatokat gyűjthetsz az operációs rendszerek frissítési gyakoriságának hatásáról a vállalati információs rendszerek teljesítményére és biztonságára vonatkozóan. Ha bármilyen további kérdésed van, szívesen segítek!

Értékelés: A Google Trends kapcsán a válasz világosan rámutat az objektumokra (vö. világszerte vagy országonként), sőt az objektum/attribútum-képzés kapcsán az elmúlt 5 év, vagyis az időhatás is helyesen kerül értelmezésre. Az, hogy az idő objektum- vagy attribútum-képző, egyelőre ég nem derült ki, de nem is volt fókuszált kérdés a robot felé. A probléma masszív gyanúja a javasolt keresési kifejezésekkel kapcsolatban áll fenn: világszerte objektum esetén BÁRMI lehet keresési kifejezés, míg országok, mint objektumok esetén olyan keresési kifejezés kell, mely a Google Trends számára előkódolt állapotban létezik, vagyis nincsenek nyelvi problémák a felhasználó számára (más-más ország, más-más keresési kulcsszó a mindenkori hivatalos nyelv(ek?!) függvényében. Emellett a keresési kifejezések TÚL hosszúak önmagukban is (vö. világszerte = objektum), vagyis félő, hogy még világszerte sem lesz találatunk. Lássuk, igaz-e a vélelem, hogy nincs semmilyen adat a Google Trends-ben világszerte sem a fenti 4 hosszú keresési kifejezésre (angolul, 5 évre – ahol a 4 tétel nem lépi túl az egyszerre lekereshető maximum 5 rétegnyi Google Trends korlátot):

Google Trends eredmények:

URL1 = <https://trends.google.com/trends/explore?date=now%201-d&q=operating%20system%20update%20frequency,system%20performance%20after%20update,system%20security%20after%20update,enterprise%20information%20systems%20update%20impact,%2Fm%2F03czj1n&hl=hu>

URL2 = https://trends.google.com/trends/explore?date=now%201-d&q=operating%20system%20update%20frequency,system%20performance%20after%20update,system%20security%20after%20update,enterprise%20information%20systems%20update%20impact,%2Fm%2F05khh&hl=hu

Megjegyzés: ahogy látható, az 5. keresési kifejezés kivételével a másik négy a javasolt karaktersorok másolata, de már a rel. hosszú, de legalább Google-Trends-kóddal rendelkező 5. réteg (=%2Fm%2F03czj1n) is adatszolgáltatási zavarokkal küzd (1. ábra – 1 napra), míg akkor, ha csak az operációs rendszerek Google Trends kódját (%2Fm%2F05khh) használjuk, akkor van adat (2. ábra – 1 napra). Az 5 éves időtávon látszólag van adat (3. ábra), de csak látszólag (vö. 2. ábra – ideális adatvagyon vizuális megjelenése – lila görbe):

A képen szöveg, képernyőkép, szoftver, sor látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

* 1. Ábra: Quasi nincs adat szcenárió 1-napra (forrás: Google Trends)

A képen szöveg, képernyőkép, szoftver, Weblap látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

* 1. Ábra: Amikor van adat szcenárió - 1 napra (Forrás: Google Trends)

A képen szöveg, képernyőkép, diagram, szoftver látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

3. ábra: Jelképesadatvagyon 5 évre (Forrás: Google Trends)

<https://trends.google.com/trends/explore?date=today%205-y&q=operating%20system%20update%20frequency,system%20performance%20after%20update,system%20security%20after%20update,enterprise%20information%20systems%20update%20impact&hl=hu>

Prompt:

Az általad javasolt hosszú keresési kifejezésekre a Google Trends semmilyen adatot nem képes szolgáltatni még világszerte sem és a Google Trends ilyen hosszú keresési kifejezést nem képes általa előrekódolt témakörként sem értelmezni, ami lehetetlenné teszi a több országot érintő lekérdezést a fordítási kihívások miatt. Kérlek tehát, adj meg olyan keresési kifejezéseket, melyeknek létezik/létezhet Google Trends témakódja és természetesen kapcsolódik az operációs rendszerek, információs rendszerek és integrált vállalati információs rendszerek területekhez egyszerre!

Copilot:

Értem, köszönöm a visszajelzést! Az alábbiakban olyan keresési kifejezéseket adok meg, amelyek valószínűleg léteznek a Google Trends témakódjai között, és kapcsolódnak az operációs rendszerek, információs rendszerek és integrált vállalati információs rendszerek területeihez:

**1. Operációs rendszerek**

* **Keresési kifejezés:** "Windows Update"
* **Keresési kifejezés:** "Linux Update"
* **Keresési kifejezés:** "macOS Update"

**2. Információs rendszerek**

* **Keresési kifejezés:** "ERP system"
* **Keresési kifejezés:** "CRM software"
* **Keresési kifejezés:** "Business Intelligence"

**3. Integrált vállalati információs rendszerek**

* **Keresési kifejezés:** "SAP ERP"
* **Keresési kifejezés:** "Oracle ERP"
* **Keresési kifejezés:** "Microsoft Dynamics"

**Lépések a Google Trends használatához:**

1. Nyisd meg a Google Trends weboldalt.
2. Írd be a fenti keresési kifejezéseket egyenként a keresősávba.
3. Állítsd be az időtartamot és a földrajzi helyet a kívánt paraméterek szerint.
4. Vizsgáld meg a keresési trendeket és az adatokat, majd hasonlítsd össze az eredményeket.

Ezekkel a keresési kifejezésekkel releváns adatokat gyűjthetsz az operációs rendszerek frissítési gyakoriságának hatásáról a vállalati információs rendszerek teljesítményére és biztonságára vonatkozóan. Ha bármilyen további kérdésed van, szívesen segítek!

Értékelés: A fenti kifejezések ellenőrzés nélkül is zömmel belső kóddal ellátottan létezőnek minősítendők első ránézésre is. A probléma innentől már csak az, hogy ezek a kulcsszavak nem támogatják a FRISSÍTÉS kulcsszó köré szőtt problémadefiníciót/címet.

Prompt:

Ezek a kulcsszavak már Google-Trends-kompatibilisek, DE ezek hogyan kapcsolódnak az eredetileg javasolt operációs rendszer-FRISSÍTÉS problémakörhöz? Kérlek, fogalmazz meg a rövid kulcsszavak adatai alapján megfogalmazható problémát, publikációs címet!

Copilot:

Értem, köszönöm a pontosítást! Az alábbiakban összekapcsolom a rövid kulcsszavakat az eredetileg javasolt operációs rendszer-frissítés problémakörével, és megfogalmazok egy releváns problémát és publikációs címet:

**Probléma:**

**Az operációs rendszerek frissítéseinek hatása az ERP rendszerek teljesítményére és biztonságára: Egy Google Trends alapú elemzés**

**Kapcsolódó publikáció címe:**

**"Az operációs rendszerek frissítéseinek hatása az ERP rendszerek teljesítményére és biztonságára: Google Trends adatok elemzése"**

**Probléma leírása:**

Az operációs rendszerek frissítései, mint például a Windows, Linux vagy macOS frissítések, jelentős hatással lehetnek a vállalati információs rendszerek, különösen az ERP rendszerek (pl. SAP ERP, Oracle ERP, Microsoft Dynamics) teljesítményére és biztonságára. Az ERP rendszerek kritikus szerepet játszanak a vállalatok működésében, és a frissítések okozta zavarok komoly következményekkel járhatnak.

**Adatok forrása:**

A Google Trends adatai felhasználhatók az operációs rendszerek frissítéseinek és az ERP rendszerek teljesítményének és biztonságának vizsgálatára. Az alábbi keresési kifejezések releváns adatokat szolgáltathatnak:

* **Operációs rendszerek frissítései:** "Windows Update", "Linux Update", "macOS Update"
* **ERP rendszerek:** "SAP ERP", "Oracle ERP", "Microsoft Dynamics"

**Hasznosság:**

A probléma megoldása operatív szinten segíthet a vállalatoknak optimalizálni a frissítési folyamatokat, csökkenteni a leállási időket és növelni a rendszerek biztonságát. Stratégiai szinten hozzájárulhat a hosszú távú tervezéshez és a kockázatkezeléshez, valamint javíthatja a vállalati információs rendszerek hatékonyságát és megbízhatóságát.

Remélem, ez a javaslat hasznos lesz számodra! Ha további kérdéseid vannak, szívesen segítek!

Értékelés: Az update jellegű kulcsszavak csak egyrésze (Windows <> Linux) rendelkezik belső kóddal a Google Trends esetén. A Copilot formálisan sugallja a kiindulási adatvagyont, de módszertanilag semmit nem mond arról, miként lehet/kell az így előálló OAM-t feldolgozni – igaz, erre fókuszált kérdés nem is hangzott még el a prompt-okban.

A képen szöveg, szoftver, Betűtípus, Weblap látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

4. ábra: Google Trends belső kóddal rendelkező kulcsszavak (forrás: Google Trends)

https://trends.google.com/trends/explore?date=today%205-y&q=%2Fm%2F025sjs0,%2Fg%2F120p9\_sj&hl=hu

Prompt:

Ha tehát pl. az EU 27 országára sikerül pl. 10 update-jellegű (X = független változó), belső Google-Trends-kóddal rendelkező keresési kifejezés mentén adatokat lekérdezni és így előáll egy 27\*10-es objektum-attribútum mátrix (OAM), valamint 11. attribútumként lekérdezzük a Google Trends-ből az IT-biztonság belső kódja szerinti adatokat (Y = függő változó) országonként, akkor kérlek írd le részletesen miként lehet ezen adatokat feldolgozni annak érdekében, hogy az egyes EU-tagországok kapcsán le lehessen vezetni az update-kókuszú X-attribútumok alapján a biztonság-fókuszú Y-változó értékeit minél pontosabban országonként?

Copilot:

Az alábbiakban részletesen leírom, hogyan lehet feldolgozni az adatokat annak érdekében, hogy az egyes EU-tagországok kapcsán le lehessen vezetni az update-központú X-attribútumok alapján a biztonság-fókuszú Y-változó értékeit minél pontosabban országonként:

**1. Adatok előkészítése**

1. **Adatok lekérdezése:** Használd a Google Trends-et, hogy lekérdezd a 10 update-jellegű keresési kifejezés (X) és az IT-biztonság (Y) adatait az EU 27 országára.
2. **Adatok exportálása:** Exportáld az adatokat CSV formátumban, hogy könnyen feldolgozhatók legyenek.

**2. Adatok tisztítása és előfeldolgozása**

1. **Hiányzó adatok kezelése:** Ellenőrizd az adatokat hiányzó értékek szempontjából, és szükség esetén pótold vagy távolítsd el azokat.
2. **Adatok normalizálása:** Normalizáld az adatokat, hogy az egyes változók összehasonlíthatók legyenek. Például használhatsz min-max normalizálást.

**3. Adatok elemzése**

1. **Korrelációs elemzés:** Vizsgáld meg a korrelációt az X-attribútumok és a Y-változó között. Használhatsz Pearson-korrelációs együtthatót vagy Spearman-rangkorrelációt.
2. **Regressziós elemzés:** Alkalmazz lineáris regressziót vagy többváltozós regressziót, hogy meghatározd az X-attribútumok hatását a Y-változóra. A regressziós modell segítségével előrejelezheted az IT-biztonság értékeit az egyes országokban.

**4. Modell validálása**

1. **Train-test split:** Oszd fel az adatokat tréning és teszt adathalmazokra (pl. 80-20 arányban).
2. **Modell illesztése:** Illeszd a regressziós modellt a tréning adatokra.
3. **Modell értékelése:** Értékeld a modell teljesítményét a teszt adatokon. Használhatsz R-négyzet értéket, RMSE-t (Root Mean Squared Error) vagy MAE-t (Mean Absolute Error).

**5. Eredmények interpretálása**

1. **Eredmények vizualizálása:** Készíts grafikonokat és ábrákat, hogy vizualizáld az X-attribútumok és a Y-változó közötti kapcsolatot.
2. **Eredmények értelmezése:** Értelmezd az eredményeket, és azonosítsd azokat az X-attribútumokat, amelyek a legnagyobb hatással vannak a Y-változóra.

**6. Jelentés készítése**

1. **Jelentés írása:** Írj egy részletes jelentést az elemzés eredményeiről, beleértve a módszertant, az eredményeket és a következtetéseket.
2. **Ajánlások megfogalmazása:** Fogalmazz meg ajánlásokat a vállalatok számára az operációs rendszerek frissítéseinek optimalizálására és az IT-biztonság javítására.

Ezekkel a lépésekkel feldolgozhatod az adatokat, és levezetheted az update-központú X-attribútumok alapján a biztonság-fókuszú Y-változó értékeit az egyes EU-tagországokban. Ha további kérdéseid vannak, szívesen segítek!

Értékelés: A Google Trends CSV-outputja keretében nincs klasszikus értelemben vett hiányzó adat, mert a NULLA valóban nullát jelent (a megfelelő felbontás mellett), míg a <1 jelzést numerikus lehet 0.5-ként értelmezni minden különösebb kockázat nélkül. A korreláció-számításra nincs semmi szükség egy regressziós modellhez. A regressziós modell a szó eredeti értelmében NEM előrejelző modell, mert az Xi és az Y között nincs időbeli eltérés, vagyis a változók mindegyik azonos-idejűséget jelent. Előrejelzés akkor lenne, ha az Xi idősávja pl. 2018-2023 lenne, míg az Y 2024-es adatokat jelentene. A tanulás:teszt bontásra (a 80:20% arányra) miért van szükség, ha a cél az Xi és az Y közötti kapcsolatok feltárása és nem az előrejelzés? Miért kellene lemondani tényadatokról 20 %-ban a megértés érdekében? Mit nyerünk a megértés kapcsán, ha tudjuk, hogy a modell teszteredményei eltérnek a tényadatoknak megtartott 80%-os minta bármilyen becslési pontossági mutatószámaitól?

Prompt:

Miért van szükség korrelációs számításokra egy többváltozós regressziós modell készítése kapcsán? Biztos, hogy kellenek a korrelációs számítások? Hiszen a regressziós modellek maguk döntenek arról, mely Xi esetén egyenes arányosság és melyik Xi esetén fordított arányosság szerint alakul majd a regressziós egyenlet koefficiensének pozitív vagy negatív előjele.

Copilot:

Igazad van, a többváltozós regressziós modellek önmagukban is képesek meghatározni az egyes független változók (Xi) hatását a függő változóra (Y), beleértve az előjelek meghatározását is. A korrelációs számítások nem feltétlenül szükségesek a regressziós modell készítéséhez, de hasznosak lehetnek az alábbiak miatt:

1. **Előzetes betekintés:** A korrelációs elemzés segíthet előzetesen megérteni az egyes változók közötti kapcsolatokat, ami hasznos lehet a modellépítés előtt.
2. **Multikollinearitás azonosítása:** A korrelációs mátrix segíthet azonosítani a multikollinearitást, amikor két vagy több független változó erősen korrelál egymással. Ez problémát okozhat a regressziós modellben, és érdemes lehet kezelni.
3. **Változók kiválasztása:** A korrelációs elemzés segíthet kiválasztani azokat a változókat, amelyek erősebb kapcsolatban állnak a függő változóval, így javítva a modell hatékonyságát.

Azonban, ha már biztos vagy benne, hogy a változók megfelelőek és nincs multikollinearitás, akkor közvetlenül is alkalmazhatod a többváltozós regressziós modellt.

**Lépések a többváltozós regressziós modell készítéséhez:**

1. **Adatok előkészítése:** Tisztítsd meg és normalizáld az adatokat.
2. **Adatok felosztása:** Oszd fel az adatokat tréning és teszt adathalmazokra.
3. **Modell illesztése:** Illeszd a többváltozós regressziós modellt a tréning adatokra.
4. **Modell értékelése:** Értékeld a modell teljesítményét a teszt adatokon.
5. **Eredmények interpretálása:** Értelmezd az eredményeket és azonosítsd a legfontosabb változókat.

Ha további kérdéseid vannak a regressziós modellezéssel kapcsolatban, szívesen segítek!

Értékelés: Per definitionem nem igaz, hogy a magas korrelációs értékeket mutató változók önmagukban mindenkor jobb becslére vezetnek, mint az alacsonyabb korrelációs értékeket mutató változókból alkotott modell becslési pontossága, s főleg nem, mint a vegyes modellek becslési pontossága.

Prompt:

A tanulás:teszt bontásra (a 80:20% arányra) miért van szükség, ha a cél az Xi és az Y közötti kapcsolatok feltárása és nem az előrejelzés? Miért kellene lemondani tényadatokról 20 %-ban a megértés érdekében? Mit nyerünk a megértés kapcsán, ha tudjuk, hogy a modell teszteredményei eltérnek a tényadatoknak megtartott 80%-os minta bármilyen becslési pontossági mutatószámaitól?

Copilot:

Nagyon jó kérdés! Valóban, ha a cél az Xi és az Y közötti kapcsolatok feltárása és nem az előrejelzés, akkor a tanulás:teszt bontás nem feltétlenül szükséges. Az alábbiakban részletezem, miért használják gyakran a tanulás:teszt bontást, és miért lehet esetleg elhagyni ebben az esetben:

**Miért használják a tanulás:teszt bontást?**

1. **Modell validálása:** A tanulás:teszt bontás segít validálni a modellt, azaz ellenőrizni, hogy a modell jól általánosítható-e új adatokra. Ez különösen fontos előrejelzési feladatoknál.
2. **Túlillesztés elkerülése:** A modell túlillesztésének (overfitting) elkerülése érdekében a teszt adatokkal ellenőrizhetjük, hogy a modell nem tanulta-e meg túl jól a tréning adatok zaját és specifikus mintázatait.

**Miért lehet elhagyni a tanulás:teszt bontást?**

1. **Kapcsolatok feltárása:** Ha a cél az Xi és az Y közötti kapcsolatok feltárása, és nem az előrejelzés, akkor az összes rendelkezésre álló adatot felhasználhatjuk a modell illesztésére. Ezáltal pontosabb képet kaphatunk a változók közötti kapcsolatokról.
2. **Adatok maximalizálása:** Az összes adat felhasználása növeli a minta méretét, ami javíthatja a statisztikai elemzések pontosságát és megbízhatóságát.

**Mit nyerünk a megértés kapcsán?**

Ha a cél a kapcsolatok feltárása és nem az előrejelzés, akkor az összes adat felhasználása segíthet:

1. **Pontosabb eredmények:** Az összes adat felhasználása pontosabb és megbízhatóbb eredményeket adhat a változók közötti kapcsolatok feltárásában.
2. **Részletesebb elemzés:** Nagyobb adatmintával részletesebb elemzést végezhetünk, és jobban megérthetjük a változók közötti összefüggéseket.

**Alternatív megközelítés**

Ha mégis szeretnél valamilyen validálást végezni, de nem szeretnél lemondani az adatok egy részéről, használhatsz keresztvalidálást (cross-validation). Ez a módszer az adatokat több részre osztja, és többször ismétli a modell illesztését és tesztelését, így minden adatot felhasználhatsz az elemzéshez.

Összefoglalva, ha a cél a kapcsolatok feltárása és nem az előrejelzés, akkor az összes adat felhasználása indokolt lehet, és nem szükséges a tanulás:teszt bontás. Ha további kérdéseid vannak, szívesen segítek!

Értékelés: Immár másodszor kapunk a Copilot-tól bölcs-rabbizmust tükröző mellébeszélést válaszul (vö. <https://vicclap.hu/vicc/11224/Rabbi_es_a_libak.html>). Vagyis a mennyi-jó-ötletem-lett-volna-még magatartásforma szakmai-etikai bűncselekmény!

# Regresszió vs. hasonlóságelemzés

A Copilot javaslata kb. 50 éves módszertani hátrányt jelent. A hasonlóságelemzés megértést támogató rétegei a XXI. század elemzésmódszertani logikáját testesítik meg ingyenes online keretrendszerre támaszkodva (vö. <https://miau.my-x.hu/myx-free/>). A hasonlóságelemzés keretében:

* Lehetőség van szakmai és/vagy korreláció-alapú döntésre az egyenes, vagy fordított arányosság nem automatizált kezelését illetően (COCO-STD szimpla attribútum-készlettel – függvényszimmetria-alapú validással).
* Lehetőség van dupla attribútumkészlettel (COCO STD) az egyenes és fordított arányosság egymással folytatott versenyének attribútumonkénti letapogatására.
* Lehetőség van az exploratív modellezés keretei között (COCO\_MCM) szimpla/dupla-attribútumkészlettel a polinomizálódás/szabályalapúság mértékének feltárására attribútumonként.

# Potenciális feladat saját ötlettel nem rendelkezők számára:

# Esettanulmány a Copilot segítségével kialakított munkaterv végrehajtása statisztikai és mesterséges intelligencia-alapú megközelítések összevetésével

…

## Az adatvagyonról

## Modellek

### Szimpla attribútumkészlet – COCO STD

### Dupla attribútumkészlet – COCO STD

### COCO-MCM

## Eredmények

## Vita

## Konklúziók

# Mellékletek

## Rövidítésjegyzék

## Ábrajegyzék

## Irodalomjegyzék

<https://miau.my-x.hu/miau/274/real_values_of_attributes.docx>

<https://miau.my-x.hu/miau/320/election2026/>

<https://miau.my-x.hu/myx-free/bevezetes.html>

[https://miau.my-x.hu/miau2009/index\_tki.php3?\_filterText0=\*knuth](https://miau.my-x.hu/miau2009/index_tki.php3?_filterText0=*knuth))

<https://miau.my-x.hu/miau2009/index.php3?x=e080>

<https://miau.my-x.hu/miau/318/taylor_swift/>

<https://miau.my-x.hu/miau/319/itsec_III_20/>

…