

Kodolányi János Egyetem

SZAKDOLGOZAT

Muhari Adrienn Martina
Gazdálkodási és menedzsment alapszak

Budapest
2023

**Kodolányi János Egyetem
Gazdálkodási és Menedzsment Tanszék**

Információhalmazok OAM szempontok szerinti feldolgozása és elemzése

Konzulens: Pitlik László

**Készítette:
Muhari Adrienn Martina
Gazdálkodási és Menedzsment
alapszak**

Budapest, 2023

1 TARTALOMJEGYZÉK

1	TARTALOMJEGYZÉK	1
2	BEVEZETÉS	3
3	AZ OAM SZEMLELETMÓD BEMUTATÁSA	6
3.1	AZ ADATTRANSZFORMÁCIÓS FOLYAMAT	8
3.2	AZ ADATOK ÉRTELMEZÉSE	8
3.3	A FORDÍTÁS ELŐTTI ADATMINŐSÉG-ELLENŐRZÉS	9
3.4	ADATFORDÍTÁS	10
3.5	A FORDÍTÁS UTÁNI ADATMINŐSÉG-ELLENŐRZÉS	10
3.6	AZ ADATTRANSZFORMÁCIÓ JELENTŐSÉGE AZ OAM SZEMPONTJÁBÓL	11
3.6.1	<i>Közvetlen OECD adatlekérési próba a ChatGPT támogatásával</i>	<i>11</i>
3.6.1.1	Segítség az objektumok minél teljesebb körű felismeréséhez	12
3.6.1.2	Segítség a nyers attribútumok minél teljesebb körű felismeréséhez	13
3.6.1.3	Segítség a mértékegységek minél teljesebb körű felismeréséhez	14
3.6.1.4	Segítség az irányok minél teljesebb körű felismeréséhez	15
3.6.1.5	Segítség a származtatott attribútumok minél teljesebb körű felismeréséhez	17
3.6.1.6	Segítség az adathiánypótlást támogató módszerek minél teljesebb körű felismeréséhez	18
3.6.1.7	Segítség az adatminőség-ellenőrzést támogató eljárások minél teljesebb körű felismeréséhez	19
3.6.1.8	Segítség az adatok normalizálását támogató megoldások minél teljesebb körű felismeréséhez	20
3.6.1.9	OAM létrehozás szimuláció	21
3.6.2	<i>Melyek az OAM konvertáláshoz szükséges adat méretbeli limitációi?</i>	<i>25</i>
4	AZ ADATFELDOLGOZÁS ÉS ELEMZÉS ÁLTALÁNOSÁGBAN	27
4.1	AZ ADATELEMZÉS JELENTŐSÉGE	29
4.2	AZ ADATFELDOLGOZÁS FOLYAMATA	30
4.3	AZ ADATELEMZÉS TÍPUSAI	30
4.4	AZ ADATELEMZÉS MÓDSZEREI	32
4.5	MILYEN ADATSOROKBÓL NEM LEHET OAM-OT KÉSZÍTENI?	33
5	A GÉPI TANULÁS AZ ADATFELDOLGOZÁSBAN	36
5.1	A MESTERSÉGES NEURÁLIS HÁLÓZATOK	36
5.2	DÖNTÉSI FÁK	37
5.3	REGRESSZIÓ	39
5.4	FUZZY-LOGIKA	39
5.5	NEM-KAUZÁLIS MODELLEZÉS	40

6	ÖSSZEFOGLALÁS.....	41
7	ÁBRAJEGYZÉK	42
8	FELHASZNÁLT FORRÁSOK JEGYZÉKE	43

2 BEVEZETÉS

A KJE Gazdálkodási és menedzsment szakon egy szakirányos Ko-Do-Libri projekt adta az alapötletet dolgozatomhoz. Hallgatótársaimmal összefogva sikerült OECD-ben¹ tárolt adatvagyon Objektum-Attribútum-Mátrix-alapú (továbbiakban OAM) automatizált feldolgozását elvégeznünk². A közös munkánk átfogó mivolta miatt több hallgatótársam is saját szakdolgozati témájának választotta a munkafolyamat egy-egy részét³.

Napjainkban az egyik legnagyobb érték az információ az adatbiztonság mellett, melyre az adatok feldolgozása révén teszünk szert. Ezért esett a választásom erre az érdekes témára, hogy hogyan lehet egy adathalmazból információt kinyerni, illetve hányféle módon lehet azokat értelmezni, a különböző szemléleteknek (vö. Tudás-reprezentációs formáknak) köszönhetően. Az adatelemzés során megvizsgáljuk adatainkat, mely folyamathoz azokat átalakítani, tisztázni, modellezni szükséges, hogy a releváns információk birtokába juthassunk. Az információk birtokában pedig következtetéseket vonhatunk le, és megkönnyíthetjük a döntéshozatal folyamatát.

Dolgozatom témaválasztásában a technológia rohamos fejlődése jelentette a fő motivációt, mivel az adatelemzés korlátját korábban a tárhely mérete és a feldolgozás sebessége jelentette, azonban mostanra például a Cloud technológiáknak⁴, a Big Data-nak⁵, a hálózati sebességek jelentős növekedésének és a folyamatosan bővülő új technológiáknak (vö. pl. High Performance Computing⁶) köszönhetően ezek az

¹ OECD (2023) <https://www.oecd.org/> (Utolsó letöltés: 2023.04.15.)

² Robot-polgár (e-gazság projekt 2022 vs 2010) (2022) https://miau.myx.hu/miau/285/e_gazsag_2022/ (Utolsó letöltés: 2023.04.15.)

³ Hirschmann Zsófia (2022) Választási esélyek egységes rendszerű elemzése több országra párhuzamosan, Jónás Attila (2022) Nyílt adatbázisok automatizált elemzése OAM megközelítés alkalmazásával

⁴ What is Cloud Technology, and How Does It Work? (2022) <https://dynamixsolutions.com/what-is-cloud-technology-and-how-does-it-work/> (Utolsó letöltés: 2023.04.15.)

⁵ What is Big Data? (2022) <https://www.oracle.com/big-data/what-is-big-data/> (Utolsó letöltés: 2023.04.15.)

⁶ What is high performance computing? (2022) <https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/> (Utolsó letöltés: 2023.04.15.)

akadályok már többé nem léteznek. Például a Machine Learning⁷, Data Mining⁸ és a Deep Learning⁹ algoritmusai pedig szinte határtalan lehetőséget biztosítanak számunkra az adatkezelés területén. Az automatizált tanulási folyamatok és a Mesterséges Intelligencia¹⁰ fejlődése szintén hasznunkat szolgálja az adatfeldolgozás lépéseinél. Így már nem feltétlenül nekünk kell a legapróbb részletekig definiálni a folyamatokat, hanem eszközeinkkel a rendelkezésre álló adatok segítségével már képesek vagyunk magunk optimalizálni a feldolgozás egyre magasabb szinten strukturált folyamatát.

A különböző vállalatok és szervezetek ma már (végre) megkerülhetetlen munkaeszközként tekintenek az információk elemzésére, mely a szakemberek mindennapi munkáját, döntéshozatalát segíti elő. Így az adatelemzést számos területen alkalmazzák, legyen szó kockázatelemzésről¹¹, önvezető járművekről¹² vagy a böngészőkben, weboldalakban jelen lévő robottechnológiákról¹³.

Utóbbira a leggyakoribb példa, hogy az emberek böngészési szokásait, látogatott témákat, oldalakat a böngészők a Cookie-k¹⁴ révén megjegyzik, majd ezeket az információkat összegezve a robotok egy profilt alkotnak a felhasználóról, a nem, a kor, az érdeklődési kör stb. feltérképezésével¹⁵ és ennek megfelelően fogja a reklámokat és a tartalmakat megjeleníteni a böngészésünk során. Ennek köszönhetően képesek fenntartani hosszú távon az érdeklődésünket az általunk kedvelt témákban.

⁷ Machine Learning (2022) <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML> (Utolsó letöltés: 2023.04.15.)

⁸ Data Mining (2022) <https://www.techtarget.com/searchbusinessanalytics/definition/data-mining> (Utolsó letöltés: 2023.04.15.)

⁹ Deep Learning (2022) <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network> (Utolsó letöltés: 2023.04.15.)

¹⁰ Mesterséges Intelligencia (2022) <https://www.europarl.europa.eu/news/hu/headlines/society/20200827STO85804/mi-az-a-mesterseges-intelligencia-es-mire-hasznaljak> (Utolsó letöltés: 2023.04.15.)

¹¹ Kockázatelemzés (2022) https://kti.uni-nke.hu/document/vtkk-uni-nke-hu/iii_tantargy-kockazatelemzes.original.pdf (Utolsó letöltés: 2023.04.15.)

¹² Önvezető járművek (2022) https://kvantumugras.blog.hu/2022/06/22/az_onvezeto_jarmuvek_jelene_es_jovoje (Utolsó letöltés: 2023.04.15.)

¹³ Robottechnológiák (2022) <https://computerworld.hu/prcik/a-legmodernebb-robottechnologia-a-webaruhazak-szolgalataban-297697.html> (Utolsó letöltés: 2023.04.15.)

¹⁴ Cookie és adatvédelem (2010.október 27) <https://www.adatvedelmiszakerto.hu/2010/10/cookie-es-adatvedelem-felhasznalok-nyomokovetese-az-interneten/> (Utolsó letöltés: 2023.04.15.)

¹⁵ Jörg Schieb - Facebook freut sich über all unsere Metadaten! (2020.02.17) <https://blog.wdr.de/digitalistan/facebook-freut-sich-ueber-all-unsere-metadaten/> (Utolsó letöltés: 2023.04.15.)

Épp emiatt a dolgozatomban lefedett téma általános érvényű, akár egy munkafolyamat optimalizálása, akár döntéstámogatás, akár még személyes célok egyszerűbb elérése érdekében is jelentős a hasznossága. Hallgatótársaimmal például gazdasági adatok alapján becsültük meg esetleges politikai kihatások esélyét, mely OAM alapú (objektum-orientált) automatizált elemzési projekt sikerrel zárult. Ez is rámutat arra, hogy milyen komplex összefüggések lehetnek különböző információk között és milyen szintű kihatásokat mutathatnak meg az elvégzett elemzések.

Egy adathalmazból értékes információkat kinyerni, azokat a megfelelő formára alakítani, majd feldolgozni egy komplex folyamat, így a szakdolgozatomban az egyes lépéseket csak töredékesen áll módomban bemutatni a dolgozat terjedelem korlátjai miatt.

Dolgozatom bevezeti az OAM szemléletmód szerinti elemzés fogalmát, majd az adatelemzés lépéseit át megvizsgálom, milyen adathalmazok elemezhetőek (vö. 3.3 fejezet) és melyek nem elemezhetőek OAM szempontok szerint (vö. 4.5 fejezet). Ezután szeretném bemutatni a konvertálási folyamatokat, miként válik adatból információ, ez milyen lépéseken keresztül valósul meg. Majd az adatok feldolgozását, elemzési metódusait fogom röviden bemutatni, hogy annak milyen jelentősége van, illetve hogyan lehet értelmezni azokat, miként fordíthatjuk hasznunkra. Végül konkrét példákkal illusztrálom az OAM módszer szerinti elemzéseket.

A szakdolgozatírás idején a szakirodalomban egy fajta eszkaláció történt a chatGPT kapcsán. A témám jellege maximálisan lehetőséget ad arra, hogy a chatGPT, mint egy fajta átlagos szakértő véleményét is önálló szakirodalmi fejezetben mutassam be az aktuálissá vált oktatási reform-folyamatot is támogatandó: vö.

<https://miau.my-x.hu/miau/297/ChatGPT.docx>

<https://miau.my-x.hu/miau/297/ChatGPT2.docx>

[https://miau.my-x.hu/miau/298/ChatGPT GPT4%20\(1\).docx](https://miau.my-x.hu/miau/298/ChatGPT GPT4%20(1).docx)

[https://miau.my-x.hu/miau/298/ChatGPT2 GPT4%20\(1\).docx](https://miau.my-x.hu/miau/298/ChatGPT2 GPT4%20(1).docx)

<https://miau.my-x.hu/miau/298/chatgpt .docx>

<https://miau.my-x.hu/miau/298/ChatGPT-food.pdf>

<https://miau.my-x.hu/miau/298/ChatGPT-food-2.pdf>

<https://miau.my-x.hu/miau/299/ChatGPT az elethosszig tarto tanulas v1 v2.docx>

3 AZ OAM SZEMLÉLETMÓD BEMUTATÁSA

Mivel dolgozatom az OAM szerinti elemzésre épül, így fontosnak tartom a fogalom bevezetését és az elemzés mikéntjének a magyarázatát, mely feldolgozási módszer a későbbi részekben alapul szolgál.

Az OAM módszer az Objektum-Attribútum Mátrixot jelenti, mely gyakorlatilag egy olyan mátrix, melyben a sorok és az oszlopok között kapcsolat határozható meg. A mátrix sorai az objektumok, a mátrix oszlopai az objektumokhoz tartozó attribútumokat jelölik. Gyakorlatilag bármit lehet objektumnak tekinteni, amely jelenséghalmaz több megvalósulási formát mutat fel és ezek azonos mértékegységgel leírt, irányítható (vö. 3.3 fejezet) tulajdonságokkal rendelkeznek, így a potenciálisan létező objektumok összehasonlíthatóvá válnak.

Ez a konkrét projektünk megvalósítása során országok, mint objektumok és idősoros statisztikai mutatóik (pl. a GDP, az infláció és az export-import arányok) vizsgálatából, mint attribútumokból felépített OAM táblázatot jelentett.

Az OAM módszer szerinti elemzés pedig alapvetően az általános adatelemzés lépéseire alapul, de ki kell egészítsük a mátrix felépítésével. Ez a kritikus része az elemzés megvalósíthatóságának is egyben, hisz nem minden adatsor alakítható át OAM-á.

Példa: egy gyűjtemény, amely számos tárgyalási jegyzőkönyvet tartalmaz. Az adatoknak nincsenek előre meghatározott attribútumai, és a szöveg strukturálatlan, ezért először strukturált formába kell alakítani az adatokat.

Tárgyalás 1:

Ügyész: Kérem a bírót, hogy hallgassa meg a tanú vallomását.

Tárgyalás 2:

Védőügyvéd: Az ügyfelem nem tartózkodott a bűncselekmény helyszínén,

Bár ilyenkor is van lehetőségünk különböző módszerekkel elősegíteni a mátrix kialakítását.

Ez alapján az OAM módszer szerinti elemzési folyamat lépései:

- **Adatgyűjtés:** Az első lépés természetesen az adatok összegyűjtése, ahol az adatokat a megfelelő forrásokból, mint például adatbázisok, fájlok stb. kell gyűjtenünk.

- Adatmezők kiválasztása: A következő lépés az adatmezők kiválasztása, amelyekre az OAM elemzést szeretnénk alkalmazni. Az objektumok egy halmaz elemei, az attribútumok, pedig azok a tulajdonságok, mely alapján leírjuk az objektumok jellemzőit az elemzéshez.
- Adatok rendezése: Ezután következik az adatok rendezése. A célunk ezzel az, hogy az adatokat objektumok és attribútumok szerint rendezzük, hogy az OAM-ot feltölthessük.
- Adatok tisztítása: Ezután jön az adatok tisztítása, amelynek célja a hibák kizárása, hiányzó értékek stb. pótlása, hogy a mátrixot hiánytalanul, megfelelően tölthessük fel. Ez a kritikus lépés az OAM elemzés esetében, mivel a feltöltött mátrix elengedhetetlen az elemzés elvégzéséhez.
- Mátrix feltöltése: A következő lépés a mátrix feltöltése az előzőleg előkészített adatokkal.
- Elemzés: Az elkészült mátrixon elvégzett elemzésünk célja a kapcsolatokat felfedezése az objektumok és attribútumok között. Ilyen elemzések lehetnek például korrelációs elemzés¹⁶, a főkomponens elemzés (más néven Principal Component Analysis (PCA) elemzés)¹⁷, klaszterelemzés¹⁸, asszociatív elemzés¹⁹, hasonlóságelemzés (vö. robot-polgár-projekt). Diagnosztikai elemzés, Prediktív elemzés, Előre leíró elemzés, Statisztikai elemzés, Szövegelemzés: vö. 3.2 fejezet
- Összegzés és jelentés készítése: ennek során bemutatjuk a felfedezett kapcsolatokat.

¹⁶ Csallner András Erik - Bevezetés az SPSS statisztikai programcsomag használatába (2015) http://www.jgypk.hu/tamop15e/tananyag_html/spss/alapfogalmak3.html (Utolsó letöltés: 2023.04.15.)

¹⁷ Lantos Gábor - Árnyjáték – A Főkomponens elemzés (Principal Component Analysis) (2021.09.10) https://statisztikaegyszeruen.blog.hu/2021/09/10/arnyjatek_fokomponens_elemzes_principal_component_analysis (Utolsó letöltés: 2023.04.15.)

¹⁸ Csallner András Erik - A klaszterelemzés technikája (2015) http://www.jgypk.hu/tamop15e/tananyag_html/spss/a_klaszterelemzs_technikja.html (Utolsó letöltés: 2023.04.15.)

¹⁹ Pang-Ning Tan, Michael Steinbach, Vipin Kumar - Bevezetés az adatbányászatba, 6. fejezet - Asszociációs elemzés: Alapvető fogalmak és algoritmusok (2006, magyar fordítás 2011) <https://gyires.inf.unideb.hu/KMITT/a04/ch06.html> (Utolsó letöltés: 2023.04.15.)

Ezen lépések segítségével végezhető el az OAM módszer szerinti adatelemzés. Ennek mentén haladok tovább dolgozatomban, nem minden lépést mélyebben bemutatva, illetve az egyes lépésekhez szükséges egyéb lépésekről is szót ejtve.

3.1 Az adattranszformációs folyamat

Napjainkban fontos megérteni az adattranszformációs folyamat lépéseit, még akkor is, ha az adattranszformáció nem elsődleges része a munkafolyamatnak. Az adatokat sokféle (pl. CSV, TSV, Excel, TXT, SQL, JSON, XML) formátumban gyűjtik össze, tárolják és elemzik, így az adatok egyik formából a másikba történő átalakításához szükséges alapvető lépések elvégzése sokunk számára általános követelmény.

Christopher Tozzi cikke²⁰ nyomán saját fordításom és értelmezésem alapján az alábbi négy lépésbe sorolhatjuk az adattranszformációs folyamatot:

- Az adatok értelmezése
- A fordítás előtti adatminőség-ellenőrzés
- Adatfordítás
- A fordítás utáni adatminőség-ellenőrzés

A következő alfejezetekben ezeket fejtem ki a szerző cikkét alapul véve, a saját gondolataimmal kiegészítve.

3.2 Az adatok értelmezése

„The first step in data transformation is interpreting your data to determine which type of data you currently have, and what you need to transform it into.” - Christopher Tozzi

Az adattranszformáció első lépése az adatok értelmezése, hogy meghatározzuk, milyen típusú adatokkal rendelkezünk jelenleg, és mivé kell átalakítanunk azokat.

A szerző a bekezdés további részében rámutat, hogy az adatértelmezés sokkal nehezebb is lehet, mint amilyennek elsőre tűnik.

²⁰ Christopher Tozzi - The Data Transformation Process Explained in Four Steps (2022.11.14.) <https://www.precisely.com/blog/big-data/data-transformation-process-four-steps> (Utolsó letöltés: 2023.04.15.)

Példaként hozza fel a tipikusan Microsoft Windows operációs rendszer alatt működő, a fájlok kiterjesztése alapján meghatározott formátumokat. Mikor egy .avi fájl a rendszer videónak, míg egy .doc-ot Microsoft Word dokumentumnak fog kezelni. Ez már egy Linux/Unix alapú rendszeren nem egyértelmű, mivel alapvetően a „kiterjesztéseket” csak könnyebb áttekinthetőség miatt használjuk, a rendszer számára irreleváns. A Windows-os megközelítést számos vírus, avagy kártékony program is kihasználja, mikor egy fájl csak látszólag az, aminek mutatja magát, ezzel megtévesztve például egy levél címzettjét. A kiterjesztés egyszerű megváltoztatása nem jelenti közvetlenül az adatformátum konvertálását is.

Az adott fájlok tartalmát kell megvizsgálnunk speciális eszközökkel, avagy értelmező programokkal (például a Linux/Unix környezet „file” parancsa), hogy beazonosíthassuk az állomány tényleges típusát.

Miután sikeresen beazonosítottuk a kiindulási formátumot, a konverzió végén jelentő változat formátumával való kompatibilitást is fel kell mérjük. Például egy szöveges fájlból nincs értelme videót készíteni, míg egy „felolvasó program” segítségével hangfájllá alakíthatjuk. Természetesen sok esetben előfordulhat, hogy egy-egy formátum (ki, vagy bemeneti is) ismeretlen számunkra, ilyenkor az átalakítás a beazonosítás hiányában nem lehetséges.

3.3 A fordítás előtti adatminőség-ellenőrzés

A Ko-Do-Libri nevű projektünk keretében Hirschmann Zsófia hallgatótársam többedmagával elemezte a rendelkezésre álló adatainkat, mely adatok értékesek számunkra és biztosítanak releváns inputot.

Ez is tekinthető a fenti cikkben szereplő utalásra, miszerint nagyon fontos a bemenő adatok minőségét ellenőrizni a konvertálás megkezdése előtt.

Ez a projektünk kapcsán a kiinduló Excel táblázatban²¹ az OAM felépítése előtt történt meg, mely folyamatot Jónás Attila hallgatótársam automatizálta Python programozási nyelv segítségével.

²¹ https://miau.my-x.hu/miau/281/KEI_10012022150615288.xlsx (Utolsó letöltés: 2023.04.15.)

Ezen ellenőrzés során a hiányzó adatsorok alapján komplett bemeneti adatokat „dobtunk el”, mivel azok becslés útján előállított értékpótlásai csökkentették volna a végeredmény pontosságát.

3.4 Adatfordítás

„After the data quality of your source data has been maximized, you can begin the process of actually translating data. Data translation means taking each part of your source data and replacing it with data that fits within the formatting requirements or your target data format.” - Christopher Tozzi

Azaz miután a forrásadatok adatminőségét maximalizáltuk, megkezdhetjük az adatok tényleges lefordításának folyamatát. Az adatfordítás azt jelenti, hogy a forrásadatok minden egyes részét olyan adatokkal helyettesítjük, amelyek megfelelnek a formázási követelményeknek vagy a céladatformátumnak.

Erre hoznám saját példának a már fentebb is említett szöveges fájlból hanggá alakítást, mikor a legegyszerűbben a karaktereket egymás után „kimondva” képesek lehetünk ezt megvalósítani. Ez tovább fog mutatni a következő pontra is, hisz az élőszövegtől jelentős mértékű eltérést fogunk tapasztalni egy ilyen megoldással.

Ezzel szemben -ahogy a cikk szerzője is említi- egy régebbi típusú HTML fájl új verzióra (HTML5) konvertálása során a kompatibilitás megléte miatt veszteségmentesen tehetjük meg az átalakítást.

3.5 A fordítás utáni adatminőség-ellenőrzés

Ahogy már utaltam is rá, a konverzió minőségének az ellenőrzése elengedhetetlen a végrehajtása után. Ez egy képfájl esetében például veszteséges tömörítés esetén elmosódást jelenthet, mely adott mértékig nem befolyásolja a kép tartalmának értelmezhetőségét, míg egy szint felett felismerhetetlen mértékűvé válhat.

Ennek a hatékony végrehajtásához előre fel kell mérjük az ellenőrizhetőséget és az esetleges hibajavítást is.

3.6 Az adattranszformáció jelentősége az OAM szempontjából

Amint azt már az OAM módszer bemutatásakor is írtam, a projektünk megvalósítása során országok, mint objektumok és idősoros értékeik vizsgálatából, mint attribútumokból felépített táblázatot jelentett az OA mátrixunk.

Ezen táblázat felépítéséhez a fent felsorolt mind a négy lépést végre kellett hajtunk, hogy biztosítsuk a megismételhető és pontos nyers adatok OAM táblázattá való formázását.

Ennek során az idősoros adatok önmagukban elégtelenek lettek volna a mátrixhoz, mivel adathiányt is tartalmaztak, illetve nem minden ország esetén volt megegyező mennyiségű elérhető adatsor a mátrix felépítéséhez.

A következő alfejezetekben megpróbáltam szimulálni a robotpolgár által megvalósított OAM generálást az OpenAI ChatGPT segítségével.

3.6.1 Közvetlen OECD adatlekérési próba a ChatGPT támogatásával

Elsőként megpróbáltam az ezirányú képességét felmérni:

Kérdés: *Képes vagy-e OECD KEI idősoros adatok alapján összeállítani egy rangsorolt OAM-ot, melyben az objektumok az országok, az attribútumok pedig az idősoros adatokban időintervallumra vonatkoztatott kimutatások (átlag, minimum, maximum, mértani közép, szórás, meredekség) lennének?*

ChatGPT: *Igen, képes vagyok segíteni az OECD KEI idősoros adatok alapján összeállítani egy rangsorolt OAM-ot, amelyben az objektumok az országok, és az attribútumok az idősoros adatok által meghatározott mutatók, mint az átlag, minimum, maximum, mértani közép, szórás és meredekség.*

Ahhoz, hogy ezt megtehessem, az adatokat először be kell olvasni és feldolgozni. Ezt követően az adatokat lehet vizualizálni és az eredményeket rangsorolni az objektumok (országok) szerint. Egy adott időszakban értékeinek összehasonlításával lehet meghatározni a relatív teljesítményt, és ez alapján lehet az országokat rangsorolni a kiválasztott mutatók szerint.

Az OAM összeállításához szükséges adatok és az összetett elemzések megértése azonban nagyon összetett feladat, amely nagy mennyiségű számítástechnikai ismeret és adatelemzési készségeket igényel. Ha konkrét kérdése vagy problémája van az

adatelemzéssel kapcsolatban, szívesen segítek, amennyire csak tudok, azonban előfordulhat, hogy szüksége lesz szakértői segítségre is az OAM létrehozásához.

Az adatok a ChatGPT számára értelmezhetővé tétele kissé időigényes és bonyolult feladat. Pontatlan a megfogalmazása, illetve nem képes közvetlenül feldolgozni az OECD adatait. Így csak egy részleges OAM felépítésére lenne képes.

3.6.1.1 Segítség az objektumok minél teljesebb körű felismeréséhez:

Kérdés: Tudnál segíteni abban, hogy az objektumokat hogyan lehet teljesebb körűen felismerni az OAM elemzés szerint?

ChatGPT: Az OAM elemzés során fontos, hogy teljesszerűen felismerjük az alkalmazásban található objektumokat, hogy azok interakcióit és viselkedését jól le tudjuk írni. Az alábbiakban néhány tipp található, amelyek segíthetnek a teljesszerű objektumfelismerésben:

1. *Áttekintés készítése: A felismerés megkezdése előtt érdemes áttekinteni az alkalmazás teljes környezetét és az objektumok hierarchiáját. Ez segíthet a folyamatok és az interakciók megértésében, és elkerülhetővé teszi az információk hiányát az elemzés során.*
2. *Objektumok azonosítása: Az alkalmazásban található összes objektumot azonosítani kell. Ezek lehetnek elemek, funkciók, felhasználói interakciók stb.*
3. *Objektumok kapcsolatainak azonosítása: Az objektumok közötti kapcsolatok azonosítása kulcsfontosságú a teljesszerű elemzéshez. Az objektumok közötti kapcsolatokat általában hierarchiában, viselkedésben vagy tulajdonságokban lehet azonosítani.*
4. *Objektumok tulajdonságainak azonosítása: Az objektumok tulajdonságai lehetnek statikusak vagy dinamikusak. A statikus tulajdonságok a felhasználói felületen jelennek meg, míg a dinamikus tulajdonságok a felhasználói interakciók során jönnek létre.*
5. *Interakciók azonosítása: Az objektumok interakcióinak azonosítása és leírása fontos a modell teljes megértéséhez. Az interakciók lehetnek felhasználói interakciók vagy automatizált folyamatok, amelyek a rendszeren belül futnak.*

6. *Több szemszögű elemzés: Az objektumok teljeskörű felismeréséhez több szemszögű elemzés szükséges. Az elemzést különböző szereplők, például az ügyfél, a felhasználó, a fejlesztők és a tesztelők szemszögéből kell megközelíteni.*

Ezek az általános tippek segíthetnek a teljeskörű objektumfelismerésben az OAM elemzés során. Azonban érdemes megjegyezni, hogy az objektumok azonosítása és a modell elkészítése számos változóra épül, és az adott alkalmazás és szükségletek alapján változhatnak.

A tippek túl általánosak és nem szolgálnak valódi segítségül.

3.6.1.2 Segítség a nyers attribútumok minél teljesebb körű felismeréséhez

Kérdés: *Tudnál abban segíteni, hogy a nyers attribútumokat hogyan lehet minél jobban teljeskörűen feltérképezni?*

ChatGPT: *Természetesen, az attribútumok teljeskörű feltérképezése nagyon fontos az OAM elemzés során, mivel ezek a tulajdonságok hozzájárulnak az objektumok viselkedésének és interakcióinak leírásához. Az alábbiakban néhány tipp található, amelyek segíthetnek a nyers attribútumok feltérképezésében:*

Készítsen listát az összes attribútumról: A teljeskörű feltérképezés megkezdése előtt, készítsen egy listát az összes attribútumról, amelyet az adott objektumon belül azonosított.

1. *Csoportosítsa az attribútumokat: Az attribútumokat azonosítsa és csoportosítsa, hogy a folyamat során egyszerűbbé váljon az elemzés. Például, különítsen el olyan attribútumokat, mint a méret, szín, típus stb.*
2. *Határozza meg az attribútumok értékét: Az attribútumok értékének meghatározása kulcsfontosságú, mivel ezek határozzák meg az objektum viselkedését és tulajdonságait. Az értékek lehetnek numerikus, szöveges vagy logikai értékek.*
3. *Azonosítsa az attribútumok viselkedését: Az attribútumok viselkedése határozza meg, hogy az adott objektum hogyan reagál az interakciókra. Az attribútumok lehetnek állapotfüggőek, azaz változtathatók, vagy nem állapotfüggőek, azaz állandó értékűek.*

4. Azonosítsa az attribútumok hatásait: Az attribútumok hatásai határozzák meg az objektum interakcióit és viselkedését. Az attribútumok hatásai lehetnek láthatóak, vagy rejtettek, és befolyásolhatják a felhasználói interakciók eredményét.
5. Több szemszögű elemzés: Az attribútumok teljeskörű feltérképezéséhez több szemszögű elemzésre van szükség. Az attribútumokat az ügyfél, a felhasználó, a fejlesztők és a tesztelők szemszögéből kell megközelíteni.

Ezek az általános tippek segíthetnek a nyers attribútumok teljeskörű feltérképezésében az OAM elemzés során. Azonban érdemes megjegyezni, hogy az attribútumok azonosítása és azok viselkedése nagymértékben változhat az adott alkalmazás vagy rendszer típusától és céljától függően. Ezért fontos az objektumok alapos megértése, az attribútumok azonosítása és a különböző szemszögű elemzések elvégzése a helyes eredmények elérése érdekében.

Az attribútumok azonosítása és feltérképezése a folyamatban fontos része, de nem szabad elfelejteni a rendszer és az alkalmazás teljeskörű megértését sem. Az OAM elemzés során fontos az objektumok és az attribútumaik megértése, de a rendszer egészét is szem előtt kell tartani az optimális eredmények eléréséhez.

A már általunk definiált attribútumok kinyerésére és feldolgozására ad tippeket és nem további attribútumok felismerésére törekszik.

3.6.1.3 Segítség a mértékegységek minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy az OAM elemzés alapján hogyan lehet szélesebb körben a mértékegységeket felismerni?

ChatGPT: Az OAM elemzés során a mértékegységek felismerése fontos szerepet játszik az objektumok és a viselkedésük leírásában. Az alábbiakban néhány tipp található, amelyek segíthetnek a mértékegységek felismerésében az OAM elemzés során:

1. Azonosítsa a szükséges mértékegységeket: Az OAM elemzés során először azonosítsa azokat a mértékegységeket, amelyek szükségesek az objektumok viselkedésének és tulajdonságainak leírásához. Például, ha az objektumok mozgása fontos, akkor az OAM elemzés során az idő és a távolság mértékegységei relevánsak lehetnek.

2. Azonosítsa az összefüggéseket az attribútumok és a mértékegységek között: Az OAM elemzés során az attribútumok és a mértékegységek közötti összefüggések az objektumok viselkedésének és tulajdonságainak pontosabb leírását teszik lehetővé. Például, az objektum sebessége és az idő közötti kapcsolat megmutathatja az objektum elmozdulását és a sebességének változását az idő függvényében.
3. Azonosítsa az egységek átváltásának szükségességét: Az OAM elemzés során az egységek átváltása fontos lehet a pontosabb eredmények elérése érdekében. Például, ha az objektum sebességét méter/másodperc mértékegységben mérte, de a rendszerben a sebességet kilométer/óra mértékegységben jelenítik meg, akkor az átváltás szükséges lehet az eredmények összehasonlításához.
4. Használjon megfelelő szoftveres eszközöket: Az OAM elemzés során használhat szoftveres eszközöket a mértékegységek pontosabb és egyszerűbb felismeréséhez. Például, a MATLAB és a Wolfram Alpha lehetőséget nyújtanak az egységek átváltására és azok közötti konverzióra.

Az OAM elemzés során az objektumok viselkedésének és tulajdonságainak leírásához szükséges mértékegységek fontos szerepet játszanak. Az attribútumok és a mértékegységek közötti összefüggések felismerése és az egységek átváltása segíthet a pontosabb eredmények elérésében. Az OAM elemzés során használható szoftveres eszközökkel, például a MATLAB vagy a Wolfram Alpha segítségével még egyszerűbbé tehető a mértékegységek felismerése és azok közötti konverzió. Emellett érdemes kiemelni, hogy az OAM elemzés során fontos figyelmet fordítani a mértékegységek konzisztens használatára, hogy az eredmények összehasonlíthatóak és értelmezhetőek legyenek.

Ebben az esetben is túl általános megfogalmazást alkalmaz, mely ugyan tartalmaz hasznos információt, azonban valódi megoldást nem kínál.

3.6.1.4 Segítség az irányok minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy milyen irányokat kell figyelembe venni az OAM modell teljes körű felismeréséhez?

ChatGPT: Az OAM modell teljes körű felismerése számos irányt és szempontot foglal magában. Az alábbiakban összefoglaljuk néhányat ezek közül:

1. *Adatgyűjtés:* Az OAM modell felismerésének első lépése a megfelelő adatok gyűjtése és rögzítése. Fontos figyelembe venni az adatgyűjtési módszerek megbízhatóságát és a mérési pontosságot.
2. *Adatfeldolgozás:* Az OAM modell felismeréséhez szükséges adatokat feldolgozni kell. Ehhez számos különböző matematikai és statisztikai módszer áll rendelkezésre, amelyek közül a leggyakrabban alkalmazottak közé tartoznak a regresszió, a faktoranalízis és a varianciaanalízis.
3. *Modellépítés:* Az OAM modell felismeréséhez szükséges egy megfelelő modell felépítése. Ennek során érdemes figyelembe venni a modell egyszerűségét, értelmezhetőségét, illetve az adatokra való illeszkedését.
4. *Változók kiválasztása:* Az OAM modellben használt változók kiválasztása nagyon fontos szempont, hiszen a változók száma és jellege nagyban befolyásolja a modell prediktív erejét. Érdemes olyan változókat kiválasztani, amelyek valóban fontosak a modell szempontjából, és amelyeknek erős összefüggése van a kimeneti változóval.
5. *Tesztelés és validáció:* Az OAM modell felismeréséhez fontos a modell tesztelése és validációja, hogy meggyőződjünk a modell pontosságáról és validitásáról. Ehhez számos különböző teszt és validációs módszer áll rendelkezésre, amelyek közül a leggyakrabban alkalmazottak közé tartozik a keresztvalidáció és a bootstrapping.
6. *Értelmezés:* Az OAM modell felismerése során nagyon fontos az eredmények értelmezése és a modell által nyújtott információk felhasználása a döntéshozatalban vagy az adott területen történő további kutatásokban.

Ezek a szempontok fontosak az OAM modell teljes körű felismeréséhez, azonban érdemes figyelembe venni, hogy az OAM modell alkalmazása számos területen változó lehet, és ennek megfelelően az előzetesek is változóak lehetnek. Ezért az OAM modell felismeréséhez szükséges szempontok az adott terület jellemzőitől, a rendelkezésre álló adatoktól és a célkitűzésektől is függenek. Fontos az is, hogy az OAM modell felismerése során számos tényezőt figyelembe kell venni, például az adatok minőségét,

a választott modell megfelelőségét, az értelmezési lehetőségeket és a döntéshozatalban való felhasználhatóságot.

Az OAM modell felismerése egy összetett folyamat, amely számos lépést és szempontot tartalmaz. Az OAM modell felismerése során fontos az adatok minőségének biztosítása, az adatfeldolgozási módszerek és statisztikai modellek használata, a változók kiválasztása, a modell tesztelése és validálása, valamint az eredmények értelmezése és a modell felhasználása a döntéshozatalban. Az OAM modell alkalmazása sok területen hasznos lehet, például az üzleti döntéshozatalban, a piaci trendek elemzésében, a közösségi média kutatásában, az egészségügyben és a környezetvédelemben.

A pontatlan definíciók nem visznek közelebb a célunk eléréséhez, nem kapunk releváns választ.

3.6.1.5 Segítség a származtatott attribútumok minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy a származtatott attribútumokat hogyan lehet minél szélesebb körben felismerni?

ChatGPT: Az OAM elemzés egyik előnye, hogy lehetővé teszi a származtatott attribútumok felismerését, amelyek nem közvetlenül megfigyelhetők, hanem más attribútumokból származnak. Az ilyen származtatott attribútumok fontos információkat tartalmazhatnak, amelyeket az OAM elemzés segítségével fel lehet fedni.

Az OAM elemzés során a származtatott attribútumok felismerése többféle módon történhet. Az egyik módszer az algoritmusok és modellek használata, amelyek képesek előrejelzéseket és becsléseket készíteni az adatok alapján. Ez a módszer a származtatott attribútumokat is feltárhatja, és lehetővé teszi a korábban nem látható összefüggések felfedezését.

A másik módszer az OAM elemzés során az adatok alapos vizsgálata és a származtatott attribútumokra vonatkozó előzetes ismeretek alkalmazása. Az adatok alapos vizsgálata segíthet az összefüggések és mintázatok felismerésében, míg az előzetes ismeretek lehetővé teszik a származtatott attribútumok alaposabb feltárását és értelmezését.

Az OAM elemzés során a származtatott attribútumok felismerése több lépésből áll. Az első lépés az adatok alapos vizsgálata és a korrelációk és mintázatok felismerése. Ezután a származtatott attribútumok kiválasztása és azok meghatározása következik. Ezután az adatokat modellálják és tesztelik, hogy meghatározzák a származtatott attribútumokat és azok jelentőségét. Végül az eredmények értelmezése és felhasználása a döntéshozatalban.

A származtatott attribútumok felismerése az OAM elemzés fontos szempontja, mivel lehetővé teszi az adatok szélesebb körű felismerését és a korábban nem látható összefüggések felfedezését.

Ez a válasz is túlzottan általános, mely nem hordoz számunkra hasznos információt. Olyan kifejezéseket használ, melyek félrevezetőek lehetnek. További kérdésekkel előfordulhat, hogy konkrét algoritmust állítson elő, ahhoz viszont példaadatokra lenne szükség.

3.6.1.6 Segítség az adathiánypótlást támogató módszerek minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy az adathiánypótlást támogató módszereket hogyan lehet szélesebb körben felismerni?

ChatGPT: Az adathiánypótlás egy fontos kérdés az adatelemzés területén, amely segít az adatok hiányosságainak kezelésében és az adatok teljesebb képének felállításában. Az adathiánypótlást támogató módszerek számos lehetőséget kínálnak az adatok hiányosságainak kezelésére, és az OAM elemzés során is alkalmazhatóak.

Az adathiánypótlást támogató módszerek felismerése az OAM elemzés során többféle módon történhet. Az egyik módszer az imputáció, amely egy olyan módszer, amely a hiányzó adatokat az ismert adatok alapján helyettesíti. Az imputáció során többféle módszer is alkalmazható, például az átlag- és mediánimputáció, a regressziós imputáció és a KNN imputáció.

A másik módszer a szintetizálás, amely egy olyan módszer, amely új adatokat generál a hiányzó adatok pótlására. A szintetizálás során többféle módszer is alkalmazható, például a GAN és a VAE.

Az adathiánypótlás támogató módszerei mellett fontos figyelembe venni az adatok minőségét és az adatok teljességét. Az adatok minőségét javítani lehet az adatok

tisztítása és normalizálása segítségével, míg az adatok teljességét javítani lehet az adatforrások diverzifikálásával és az adatok összekapcsolásával.

Az adathiánypótlást támogató módszerek felismerése az OAM elemzés során nagyon fontos szempont, mivel az adatok hiányosságainak kezelése segít az adatok teljesebb és megbízhatóbb képének kialakításában, ami fontos szerepet játszik a döntéshozatalban. Az OAM elemzés során az adathiánypótlás támogató módszerei alkalmazhatóak az adatok teljesebb és pontosabb képének kialakításához, és segíthetnek az adatokkal kapcsolatos bizonytalanságok kezelésében.

Ebben az esetben is általános választ kapunk, valószínűleg ebben az esetben is további kérdésekkel lehetne a válaszokon finomítani.

3.6.1.7 Segítség az adatminőség-ellenőrzést támogató eljárások minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy milyen adatminőség ellenőrzést támogató módszerek léteznek az OAM elemzés során?

ChatGPT: Az adatminőség ellenőrzése az OAM elemzés során fontos lépés, mivel segít az adatok minőségének javításában és az adatok megbízhatóságának növelésében. A következőkben néhány módszert mutatok be, amelyeket az OAM elemzés során használhatunk az adatminőség ellenőrzésére:

1. **Adatellenőrzési módszerek:** Az adatellenőrzési módszerek az adatok minőségének és konzisztenciájának ellenőrzését szolgálják. Az adatellenőrzési módszerek közé tartozik a logikai ellenőrzés, amely ellenőrzi, hogy az adatok megfelelnek-e a logikai szabályoknak, valamint az adatok konzisztenciáját ellenőrző módszerek, például az összehasonlító ellenőrzés.
2. **Adattisztítási módszerek:** Az adattisztítási módszerek segítenek az adatok tisztításában, például az outlier adatok vagy a hiányzó adatok kezelésében. Az adattisztítási módszerek közé tartozik a medián- vagy átlagimputáció, a csonkolás vagy a kerekítés.
3. **Adatnormalizálási módszerek:** Az adatnormalizálási módszerek segítenek az adatok egységesítésében, például az adatok skálázásában, a kategorikus adatok bináris kódolásában vagy az adatok standardizálásában.

4. *Adatvalidációs módszerek:* Az adatvalidációs módszerek segítenek az adatok helyességének és megbízhatóságának ellenőrzésében. Az adatvalidációs módszerek közé tartozik a cross-validation, amely segít a modell pontosságának ellenőrzésében.
5. *Adatvizualizációs módszerek:* Az adatvizualizációs módszerek segítenek az adatok értelmezésében és megjelenítésében. Az adatvizualizációs módszerek közé tartozik a hisztogram, a boxplot vagy a scatterplot.

Ezen módszerek alkalmazása segít az adatok minőségének javításában és az adatok megbízhatóságának növelésében. Az adatminőség ellenőrzése az OAM elemzés során rendkívül fontos, mivel az adatok minősége hatással van a modellek pontosságára és az eredmények megbízhatóságára.

A kapott válasz kiindulópontnak megfelelő.

3.6.1.8 Segítség az adatok normalizálását támogató megoldások minél teljesebb körű felismeréséhez

Kérdés: Tudnál abban segíteni, hogy milyen módszerek léteznek az adatok normalizálását támogató megoldások minél teljesebb körű felismeréséhez?

ChatGPT: Az adatok normalizálása nagyon fontos lépés az OAM elemzés során, mivel segít eltávolítani az adatokban található zajt és eltéréseket, amelyek befolyásolhatják a végeredményt. Az alábbiakban felsorolok néhány módszert, amelyek segíthetnek az adatok normalizálásában az OAM elemzés során:

1. *Skálázás:* A skálázás azt jelenti, hogy az adatokat átalakítjuk egy adott skálán, például 0 és 1 közötti értékekre. Ez segíthet az adatok összehasonlításában és azonosításában, hogy melyik adatok szerepelnek azonos vagy hasonló tulajdonságokkal.
2. *Standardizálás:* A standardizálás az adatok átalakítása az átlag és a szórás alapján. Ennek eredményeként az adatok általános eloszlása normális eloszlású lesz, amely lehetővé teszi a jobb összehasonlítást.
3. *One-Hot kódolás:* Az One-Hot kódolás egy olyan technika, amely lehetővé teszi a kategórikus változók kezelését, amelyek szöveges formában jelennek meg. Ez a technika a kategórikus változókat bináris értékekké alakítja, amelyek azt jelzik, hogy az adott változó tartalmazza-e vagy sem az adott tulajdonságot.

4. *PCA (Principal Component Analysis): Az PCA segítségével az adatok dimenzióit csökkenthetjük. Ennek eredményeként az adatok egyszerűbb formában jelennek meg, és könnyebben kezelhetők lesznek.*
5. *Normalizálás az OAM modell szerint: Az OAM elemzéshez létrehozott modell által meghatározott normál értékekhez történő igazítás segíthet az adatok normalizálásában, mivel ezek az értékek a modell szerint a leginkább relevánsak.*

Az adatok normalizálása többnyire az adott adathalmaztól függ, és különféle módszerekkel érhető el a legjobb eredmény. Ezért fontos az adathalmaz és az elemzési célok megértése az OAM elemzés során, hogy a legmegfelelőbb normalizálási módszerekkel lehessen dolgozni.

A válasz ismét általános, mely jó kiindulóalapnak bizonyul.

Összességében a chatGPT ritkán ad nálunk elsőre pontos választ, általánosít. Tapasztalat alapján nagyon részletgazdagon kell specifikálni, lehetőleg pontos példákkal kell segíteni a minél pontosabb és hasznosabb válasz eléréséhez.

3.6.1.9 OAM létrehozás szimuláció

Kérdés: *Mind az öt lépésben markdown táblázatos formában add ki az eredményt, minden magyarázat, vagy forráskód nélkül.*

1. lépés:

Generálj nekem egy példa táblázatot 5 adatsorral, melyben:

- 10 oszlop van*
- a fejlécek legyenek "Név,Adat1,...,Adat8,Cél"*
- az 1. oszlop egy véletlenszerű, de nem ismétlődő ISO országcód, ABC sorrendbe rendezve*
- a 2-9. oszlop véletlenszerű tört szám adatokat tartalmaz 1-100 intervallumban*
- a 10. oszlop egy célértéket 1-1000 tartományban*

2. lépés

Készíts egy táblázatot, mely:

- a fejlécei "Név, átlag, meredekség, cél"*
- a "név" mező szerint rendezett sorrendben, ennek értékei az 1. lépésben szereplő táblázat "név" oszlopának értékei*

- az "átlag" az 1. lépésben generált táblázat soronként számított 2-9. oszlop értékeinek az átlaga

- a "meredekség" pedig az 1. lépésben generált táblázat soronként számított 2-9. oszlop értékeinek a meredeksége

- a "cél" az 1. lépésbeli táblázat "cél" oszlopának értékei

3. lépés

Készíts egy táblázatot, ahol:

- a fejlécek: "átlag, rangsor átlag, név"

- az "átlag" megnevezésű oszlop értékei a 2. lépésben szereplő táblázat "átlag" nevű oszlopában szereplő értékeket jelölik előjelhelyesen, nagyság szerint csökkenő sorrendbe rendezve

- a "rangsor átlag" oszlop a 2. lépésben szereplő táblázat "átlag" nevű oszlopában szereplő értékeket sorrendjét jelöli előjelhelyesen, nagyság szerint csökkenő sorrendbe rendezve, az "átlag" értékek "név" értékeinek hozzárendelése alapján, ennek az indexe 1-től kezdődő legyen

- a "név" oszlop az 1. lépésben meghatározott a "név" oszlop

4. lépés

Készíts egy táblázatot, ahol:

- a fejlécek: "meredekség, rangsor meredekség, név"

- a "meredekség" megnevezésű oszlop értékei a 2. lépésben szereplő táblázat "meredekség" nevű oszlopában szereplő értékeket jelölik előjelhelyesen, nagyság szerint csökkenő sorrendbe rendezve

- a "rangsor meredekség" oszlop a 2. lépésben szereplő táblázat "meredekség" nevű oszlopában szereplő értékeket sorrendjét jelöli előjelhelyesen, nagyság szerint csökkenő sorrendbe rendezve, a "meredekség" értékek "név" értékeinek hozzárendelése alapján, ennek az indexe 1-től kezdődő legyen

- a "név" oszlop az 1. lépésben meghatározott a "név" oszlop

5. lépés

A 3. és 4. lépésben készült két segéd táblázat alapján készíts el egy összesítő táblázatot, mely fejlécei "név, rangsor átlag, rangsor meredekség, cél", a tartalma pedig a "név" oszlophoz az 1. táblázat "név" oszlopának a tartalma, az 1. lépés "név" oszlopa sorrendjében; a "rangsor átlag" a 3. lépés segédtáblázata alapján a "rangsor

átlag" megfelelő "név"-hez rendelt értéke; a "rangsor meredekség" a 3. lépés segédtáblázata alapján a "rangsor meredekség" megfelelő "név"-hez rendelt értéke; a "cél" az 1. táblázat "név" oszlophoz rendelt "cél" oszlopbeli érték

ChatGPT:

1. lépés:

Név	Adat1	Adat2	Adat3	Adat4	Adat5	Adat6	Adat7	Adat8	Cél
AUT	23.52	82.71	58.31	66.93	35.16	63.98	4.58	52.14	456
BEL	13.98	85.64	2.60	19.63	25.51	47.92	89.77	46.97	812
CAN	32.04	83.37	84.38	39.72	91.91	94.03	5.59	24.01	384
DEU	78.63	5.53	77.39	5.04	50.61	66.39	60.55	35.77	632
ESP	68.28	84.45	9.50	52.41	25.04	34.45	37.60	51.22	176

1. ábra – Példatáblázat 5 adatsorral (forrás: chatGPT-vel generált saját munka)

Az első lépést rendben teljesítette.

1. lépés:

Név	Átlag	Meredekség	Cél
AUT	46.296	20.662	456
BEL	45.337	8.960	812
CAN	51.583	33.685	384
DEU	46.400	-22.110	632
ESP	47.387	15.455	176

2. ábra – Átlag és meredekség meghatározása (forrás: chatGPT-vel generált saját munka)

Az értékek kiszámolása nélkül a táblázat felépítése rendben van, azonban "1. lépésnek" jelöli.

1. lépés:

Átlag	Rangsor átlag	Név
51.583	1	CAN
47.387	2	ESP
46.400	3	DEU
46.296	4	AUT
45.337	5	BEL

3. ábra – Az átlag szerinti rangsor felépítése (forrás: chatGPT-vel generált saját munka)

A táblázat helyes tartalommal rendelkezik, de ismét „1. lépésnek” definiálja.

1. lépés:

Meredekség	Rangsor meredekség	Név
33.685	1	CAN
20.662	2	AUT
15.455	3	ESP
8.960	4	BEL
-22.110	5	DEU

4. ábra – Meredekség szerinti rangsor felépítése (forrás: chatGPT-vel generált saját munka)

A táblázat helyes tartalommal rendelkezik, de ismét „1. lépésnek” definiálja.

1. lépés:

Név	Rangsor átlag	Rangsor meredekség	Cél
AUT	4	2	456
BEL	5	4	812
CAN	1	1	384
DEU	3	5	632
ESP	2	3	176

5. ábra – Eredmény táblázat (forrás: chatGPT-vel generált saját munka)

A táblázat helyes tartalommal rendelkezik, de ismét „1. lépésnek” definiálja.

Azonban véletlenszerűen hibázik a rangsorok felépítésénél, vagyis az adatok rangsorolását nagyobb terjedelmű adathalmaz feldolgozása esetében pontatlanul végzi. Bár képes belátni a tévedését, amennyiben felhívjuk rá a figyelmet, ennek ellenére továbbra is helytelenül jeleníti meg az adatokat OAM szimulált előállításakor. A helyes adatot csak akkor jelenítette meg, ha előzőleg példaként megadtuk a helyes rangsort.

Ezek alapján megbízhatatlan ítélem jelen állapotában a chatGPT-t valódi OAM felépítésére.

3.6.2 Melyek az OAM konvertáláshoz szükséges adat méretbeli limitációi?

Nincs abszolút maximális méret az Objektum Attribútum Mátrix (OAM) készítéséhez, mivel ez több tényezőtől függ, például a rendelkezésre álló számítógépes erőforrásoktól, a memóriától és a használt szoftvertől. Az OAM méretét és hatékonyságát a következő tényezők befolyásolhatják:

- Számítógép memóriája: A rendelkezésre álló RAM határozza meg, hogy milyen méretű adatsorokat tud kezelni a számítógép. Nagyobb memóriájú számítógépek nagyobb adatsorokból képesek OAM-ot készíteni.
- Processzor sebessége: A processzor sebessége befolyásolja az adatfeldolgozás gyorsaságát és hatékonyságát. Gyorsabb processzorokkal nagyobb adatsorokat lehet gyorsabban feldolgozni.

- Szoftver optimalizáció: Az adatelemzési szoftverek és könyvtárak (például Python pandas, R, SQL stb.) optimalizációja befolyásolja a képességét annak, hogy milyen méretű adatsorokból készíthető OAM.
- Adattárolás és adatbázis-kezelő rendszerek: Az adatbázis-kezelő rendszerek és a felhasznált adattárolási technológiák (például Hadoop, Spark, stb.) hatással lehetnek a maximális adatméretre, amelyből OAM-ot lehet előállítani.

Az adatsor méretének korlátja gyakorlatilag az, amit a rendelkezésre álló erőforrások és szoftverek képesek kezelni. A pontos korlát az adott környezettől és az adatok szerkezetétől függ. Néhány esetben, ahol a rendelkezésre álló erőforrások korlátozottak, az adatsorok részekre bonthatók és részletesen elemezhetők, majd az eredmények összevonhatók, hogy egy teljes képet kapjunk. Ezenkívül az adatokat aggregálhatjuk vagy összefoglalhatjuk, hogy csökkentsük az adatméretet, és könnyebben elkészíthető legyen az OAM.

4 AZ ADATFELDOLGOZÁS ÉS ELEMZÉS ÁLTALÁNOSÁGBAN

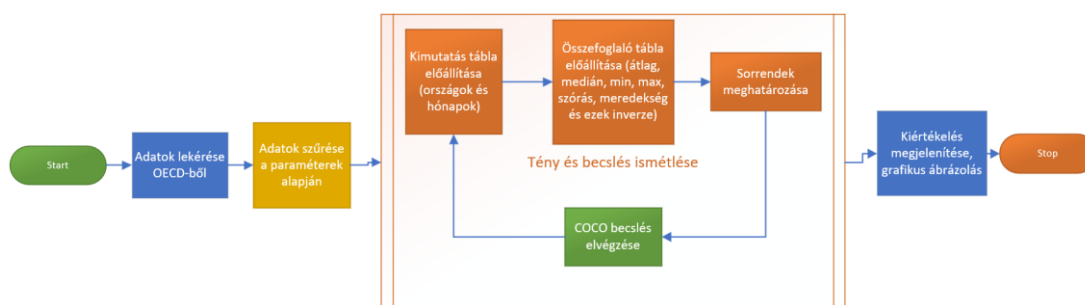
A KJE-s Ko-Do-Libri projektünk kapcsán hallgatótársaimmal együtt haladtunk lépésről lépésre a kiindulási Excel táblázat megértésében. Ezen megértés hiányában nem állt volna módunkban megfelelő minőségben elvégezni az OECD adatvagyon feldolgozását még manuálisan sem.

A feltérképezés alapján állt elő a fő automatizálendő feladat:

1. bemeneti OECD adatok betöltése, a számunkra releváns adatok kiválasztása és azok szűrése a kiválasztott feltételeink alapján
2. az adatok átalakítása egy kimutatás táblává, mely az idősoros értékeket tartalmazza
3. a kimutatás tábla adatainak „tisztázása” – azaz az adatkonverzió előtti minőség-ellenőrzés (vö 3.3)
4. OAM alaptáblázat előállítása következik a kimutatás tábla alapján, azaz a tényleges adatkonverzió és ellenőrzés (vö 3.4 és 3.5)
5. az OAM adatok rangsorolt értékelése jön ezután, ami egy új táblázatba kerül
6. ezután végezzük el a COCO²² alapú becslést az előállított OAM rangsorolt mátrixa alapján

Mivel ezek a lépések csak egy-egy adatsor kiértékelését jelentették, így ezen lépéseket kellett újra és újra megismételnünk. Erről készített hallgatótársam egy ábrát is, bemutatva a folyamatot. [6. ábra]

²²MyX-Free COCO (2022) <https://miau.my-x.hu/myx-free/coco/> (Utolsó letöltés: 2023.04.15.)



6. ábra - A program működésének folyamatábrája (Forrás: Jónás Attila KJE GazdMen szakdolgozata - 2022)

A következőkben Karin Kelley, „What is Data Analysis? Methods, Process and Types Explained” című cikke²³ alapján foglalom össze az adatfeldolgozás és elemzés folyamatát.

Először is határozzuk meg, hogy miként értelmezzük magát az adatfeldolgozást.

Ezt Karin Kelley így fogalmazta meg:

„Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.” – Karin Kelley

A saját fordításom szerint: Az adatelemzés a nyers adatok tisztázásának, megváltoztatásának és feldolgozásának a folyamata, valamint a hasznosítható, releváns információk kinyerése, amelyek segítik a vállalkozásokat a megalapozott és megfontolt döntéshozatalban. Az eljárás segít csökkenteni a döntéshozatalban rejlő kockázatokat azáltal, hogy hasznos meglátásokat és statisztikákat nyújt, amelyeket gyakran diagramok, képek, táblázatok és grafikonok formájában prezentálnak.

Ez természetesen nem feltétlen csak a vállalkozásokra vonatkoztatható, ahogy láthattuk a projektünk keretében is, annak ellenére, hogy egyféle StartUp-ként is felfogható a csoportunk munkája.

Karin Kelley az ismertető további részében egy egyszerű példának a mindennapi életünkben hozott döntéseink kiértékelését hozza, mely döntésünk meghozatalához a

²³ Karin Kelley - What is Data Analysis? Methods, Process and Types Explained (2022.11.18) <https://www.simplilearn.com/data-analysis-methods-process-types-article> (2023.04.15)

múltbéli tapasztalataink és a becsült, vagy várható jövőbeli események kiértékelését vesszük alapul.

Minél több és pontosabb adatunk áll rendelkezésre, annál határozottabban vagyunk képesek a döntéshozatalra. Emiatt jelentős a mindennapjainkban is a „Big Data” szerepe. Ennek során is akár egy rendezetlen, jelentős mennyiségű adathalmazból kell kiszűrni először a releváns adatokat, majd feldolgozni azokat. Hasonlóan a fentebb ismertetett OAM megközelítésünkhöz.

4.1 Az adatelemzés jelentősége

Az üzleti életben az adatelemzés fontossága kapcsán Karin Kelley a következő pontokat jelölte meg:

- „Better Customer Targeting”
- „You Will Know Your Target Customers Better”
- „Reduce Operational Costs”
- „Better Problem-Solving Methods”
- „You Get More Accurate Data”

Azaz rámutat, hogy a jobban behatárolható ügyfél célcsoportok, ezen célcsoportok jobb megismerése, a működési költségek csökkentése, a jobb problémamegoldó módszerek és a minél pontosabb adatok a legfontosabbak. A gazdasági életben az ügyfeleink feltérképezése, a célcsoportok meghatározása, a piackutatás, az ügyfelek megszólítása elengedhetetlen. Ez egészül ki a minél pontosabb adatok feldolgozhatóságával, amire már az előzőekben is utaltunk, hogy a pontosság a minél hatékonyabb döntéshozatalt eredményezi. A bevételek növelése mellett természetesen a kiadások csökkentése, avagy optimalizálása is a hatékony működés része. Ahogy az előre nem látható, nem, vagy nehezen tervezhető problémák minél egyszerűbb megoldása is nagyon fontos.

Itt elég csak utalni az energiaárak, vagy az infláció jelentős mértékű emelkedésére, amely tételek gyakorlatilag a beszerzési költségektől kezdve a komplett üzemeltetésre is nagy mértékben kihatnak.

Ezen területekhez a minél több és minél jobb bemenő, azaz feldolgozható adat szinte elengedhetetlen. Ezzel sok esetben 2018 előtt (és részben még a mai napig) számos weboldal például jelentős mértékben élt vissza (vö Facebook és Whatsapp adatkezelése), ami az Európai Unióban a GDPR²⁴ szabályozás életbe lépése óta szerencsére valamelyest csökkent. Azóta kezdtünk el általánosságban is hozzászokni a saját adataink védelmének a fontosságához, míg a cégek a személyes adatok korrekt kezeléséhez.

4.2 Az adatfeldolgozás folyamata

A cikk következő részében Karin Kelley az adatfeldolgozás folyamatát mutatja be lépésről lépésre. Szerinte ezek a következők:

- „Data Requirement Gathering”
- „Data Collection”
- „Data Cleaning”
- „Data Analysis”
- „Data Interpretation”
- „Data Visualization”

Ezek a saját fordításom alapján rendre: Az adatigény felmérése, az adatok begyűjtése, a begyűjtött adatok tisztítása, adatok elemzése, az elemzett adatok értelmezése, majd azok vizualizálása.

Ezek a lépések elég jelentős átfedésben vannak az általunk is feltérképezett és automatizált adatfeldolgozás lépéseivel a projektünk kapcsán, ahogy azt már a fejezet elején is, illetve az OAM módszer szerinti elemzés fejezetében is írtam.

4.3 Az adatelemzés típusai

A cikk folytatásában Karin Kelley az adatelemzés típusait foglalja össze, ezek szerinte a következők:

- „Diagnostic Analysis”
- „Predictive Analysis”

²⁴ Complete guide to GDPR compliance (2022) <https://gdpr.eu/> (Utolsó letöltés: 2023.04.15.)

- „Prescriptive Analysis”
- „Statistical Analysis”
- „Descriptive”
- „Inferential”
- „Text Analysis”

A fordításom során az ő értelmezését egészítem ki a saját gondolataimmal.

- Diagnosztikai elemzés: a „miért történt?” kérdésre kapott válaszunk alapján végezzük az elemzést, ami jelentheti a hasonlóságvizsgálatot is, mivel adott mintázatok újbóli előfordulását vizsgáljuk ilyen esetben – ezzel a már megismert és esetlegesen újra előforduló helyzetekre tudunk felkészülni
- Prediktív elemzés: itt a „mi fog történni a legnagyobb eséllyel?” kérdésre keressük a választ. Itt is a múlt eseményeit vesszük alapul, megfigyeljük és kiemeljük a mintákat, melyek alapján megpróbáljuk megbecsülni a jövő várható eseményeit. Minél részletesebb és pontosabb adatvagyonnal rendelkezünk, annál nagyobb bizonyossággal leszünk képesek megbecsülni a várható események előfordulásának a valószínűségét. Ez az elemzés áll a legközelebb a projektünkben megvalósított becslésekhez is, hisz ott is a tényadatok feldolgozását végeztük el és az alapján készítettük elő a becsléseinket.
- Előre leíró elemzés: ezen elemzési típus alkalmazása során a többi, alkalmazott elemzési módszer keverékével próbáljuk növelni az elemzésünk hatékonyságát, mivel számos esetben előfordulhat, hogy egy-egy konkrét elemzési módszer nem optimális minden körülmény között. (Például nem áll rendelkezésre elégséges és kellően jó minőségű bemenő adatunk.)
- Statisztikai elemzés: a „mi történt?” kérdésre adott válasz, azaz kizárólag az események múltbéli előfordulásának a valószínűsége alapján végzett elemzési módszer. Ez a megoldás a bemenő adatok teljes feltérképezését, feldolgozását, kiértékelését és a megjelenítését is magába foglalja, a minél pontosabb statisztikák felállításának az eléréséhez.

- Szövegelemzés: más néven az adatbányászatot értjük ez alatt, mely az előző módszerekhez hasonlóan mintákat keres a nagy adathalmazokban, ezáltal átalakítva értékes információvá azokat. Ez a legegyszerűbb megvalósítás, mely egyben a legegyszerűbb is.

4.4 Az adatelemzés módszerei

A típusok után még a módszereket is sorra veszi a cikk. Az adatelemzés módszereit Karin Kelley két nagy csoportba sorolja, „Qualitative Data Analysis” és „Quantitative Data Analysis”. Értelmezhetjük őket egyszerűen kvalitatív és kvantitatív elemzési módszereknek, azaz a minőséget helyezzük szembe a mennyiséggel.

Ezen módszerek közötti legfőbb eltérés, hogy a kvalitatív elemzés során mélyebben megértjük a kutatásunk alapját képező adatokat, a mögöttük lévő gondolkodásmódot is akár, azaz kisebb mennyiségű adatokat minél mélyebben dolgozunk fel, a feldolgozás minőségére törekedve. Ezzel szemben a kvantitatív elemzés esetén a pontos számok és statisztikák a legfontosabbak, ezekhez pedig minél nagyobb mennyiségű és lehetőleg jó minőségű adatra van szükségünk, hogy minél pontosabb elemzést végezhesünk.

A Ko-Do-Libri projekt keretében hallgatótársaimmal az elemzés teljes folyamatán végig haladtunk, ami a nagy mennyiségű adatvagyon feldolgozása miatt mennyiségi adatelemzési módszert, azaz kvantitatív elemzést jelentett kezdetben. Az OAM táblázat felépítése során már ezen adatokat összesítettük például átlag és meredekség meghatározása által, itt már részlegesen a kvalitatív elemzési módszer felé mozdulva el és így haladtunk tovább, nem egyértelműen elválasztható módon.

Ez is jól mutatja, hogy egy OAM felépítéséhez nem lehet csak ezt, vagy csak azt a módszert alkalmazni, ez mindig az adott probléma megismerése és az elérhető adatvagyon feltérképezése után kerülhet csak kiválasztásra, sőt a legtöbb esetben a módszerek keveréke lesz az ideális választás.

4.5 Milyen adatsorokból nem lehet OAM-ot készíteni?

Az OAM egy adattáblázat, amely objektumokat (például entitásokat) sorokba és attribútumokat (például tulajdonságokat) oszlopokba rendez. Az OAM célja, hogy könnyen olvasható és elemezhető formában jelenítse meg az adatokat.

A következő adatsorok esetében nem lehet OAM-ot készíteni:

- Hiányos adatsorok: Probléma, ha az adatsor jelentős mértékben hiányos és sok attribútum értéke hiányzik, az OAM készítése nem lesz hatékony. A hiányos adatsorok példája: Data assets URL of the raw data: <https://clio-infra.eu/Countries/Austria.html#countries> ← https://miau.my-x.hu/miau/261/coldwar_robot-historian_turing-test.pdf
- Rendkívül nagy adatsorok: Ha az adatsor túl nagy, és a számítógépes memória nem elegendő az egész mátrix tárolásához és feldolgozásához, akkor nem lesz lehetséges OAM-ot készíteni. Például: <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=kaleid> (FAO, 1961-2013, 200+ ország, több, mint 50 év, közel 100 élelmiszerkategória (g/fő/nap) adata), ill. http://miau.my-x.hu/miau/141/log_y0.doc.
- Nem strukturált adatsorok: Az OAM készítéséhez strukturált adatsorra van szükség, ahol az adatoknak egyértelmű attribútumai vannak. Nem strukturált adatsorok esetén, például szöveges dokumentumokban, először strukturált formátumba kell konvertálni az adatokat, mielőtt OAM-ot készítenénk belőlük. Például: eltérő hosszúságú szövegek szóelőfordulásainak pivot táblázatait: vö. <http://miau.my-x.hu/miau/140/la140.doc>.
- Összefüggéstelen adatsorok: Ha az adatsorban lévő objektumoknak és attribútumoknak nincs logikai kapcsolata, akkor nem lehet értelmes OAM-ot készíteni belőlük. Például, ha egy adott város lakosságának adatait tartalmazza az adatsorunk. Az első oszlopban a lakosság száma, a második oszlopban pedig a városok nevei találhatók. Ha azonban az adatsorban vannak olyan városok, amelyek lakossága évek óta változatlan, míg más városok lakossága nagy ugrásokkal változik minden évben, akkor ez az adatsor összefüggéstelennek tekinthető.

- Dinamikus adatsorok: Ha az adatsor dinamikusan változik, és az objektumok vagy attribútumok gyakran frissülnek, akkor egy statikus OAM készítése nem lehet hasznos, mivel a mátrix rövid időn belül elavulhat. Jó példa erre egy értékesítési adatsort szeretnénk OAM-ként megjeleníteni. Az adatsorban szereplő információk, mint például a termékek, a vevők és az értékesítések mennyisége, gyakran változnak, amikor új termékek érkeznek, új vevők jelentkeznek vagy új értékesítések történnek. Ez azt jelenti, hogy az adatsor dinamikus, és az objektumok vagy attribútumok gyakran frissülnek.
- Ebben az esetben az OAM statikus verziója nem lesz hasznos, mivel az információk elavulnak rövid időn belül. Ehelyett egy dinamikus OAM-t kellene létrehozni, amely frissíthető az új információkkal. Ez lehetővé teszi, hogy az értékesítési adatok mindig naprakészek legyenek, és a döntéshozók megbízható információkra támaszkodhassanak a stratégiai tervezés során.

Egyébként, ha az adatsor strukturált és viszonylag rendezett, általában lehetséges OAM-ot készíteni belőle. Az OAM segíthet az adatok vizuális áttekintésében, elemzésében és további feldolgozásában.

További potenciális példák arra, mely adatsorokból nem lehet OAM-ot előállítani:

Hiányos adatsorok:

Példa: egy adatsor, amelyben egy diákok listája szerepel az évfolyamokkal és az életkorokkal, de sok diák esetében hiányzik az életkor vagy az évfolyam információja.

DiákNév, Évfolyam, Életkor

Anna, 10, 15

Béla, 11,

Csaba, , 16

Dóra, ,

Rendkívül nagy adatsorok:

Példa: egy adatsor, amely az összes internetes felhasználó összes weblaplátogatását tartalmazza világszerte (vö. Google). Az adatsor mérete miatt nem lenne lehetséges adott hardver-környezetben lefuttatható OAM-ot készíteni belőle. (Egymillió soros és 100 oszlopos OAM tesztelése kapcsán sikerült már 16 órától 16 percre redukálni a futtathatóságot: forrás: Pitlik, személyes közlés).

Nem strukturált adatsorok:

Példa: egy szöveges dokumentum, amelyben az emberek véleményét gyűjtik össze egy termékkel kapcsolatban. Az adatoknak nincsenek előre meghatározott attribútumai, ezért először strukturált formába kell alakítani őket.

Anna: Nagyon elégedett vagyok a termékkel.

Béla: A termék túl drága.

Csaba: A szállítás gyors volt, de a termék minősége nem volt megfelelő.

Összefüggéstelen adatsorok:

Példa: egy adatsor, amelyben az emberek kedvenc színei, a napi időjárás és a részvénytőzsi indexek találhatók. Ebben az esetben az adatsorban lévő objektumok és attribútumok között nincs logikai kapcsolat, így nem lehet értelmes OAM-ot készíteni belőlük.

Dinamikus adatsorok:

Példa: egy adatsor, amelyben egy online áruház aktuális árukészlete szerepel. Mivel az árukészlet folyamatosan változik (új termékek érkezése, meglévő termékek eladása stb.), egy statikus OAM készítése nem lenne hasznos, mivel az információ hamar elavulhat. De a statikus OAM-ok sorozatának folyamatos elemzése olyan dinamikus eredményekhez vezethet, melyek hermeneutikája már értelmezhető.

5 A GÉPI TANULÁS AZ ADATFELDOLGOZÁSBAN

„A világ, amit teremtettünk a gondolkodásunk eredménye; nem lehet megváltoztatni gondolkodásunk megváltoztatása nélkül.” - Albert Einstein

A korábban felsorolt elemzések (pl. korreláció, főkomponens elemzés, klaszterelemzés stb.) nem tartoznak a gépi tanuláshoz közvetlenül, viszont a gépi tanulást előkészíthetik, megalapozhatják. Ebben a fejezetben erre térek ki részletesebben.

A mesterséges intelligencia részét képező gépi tanulás, avagy Machine Learning a szakdolgozati terjedelmen önmagában is jelentősen túlmutató téma, így csak az általam lényegesnek ítélt komponenseire fókuszálva fogom tudni bemutatni.

Azzal a gyakorlatilag ténnyel kezdeném, hogy a gépi tanulás alkalmazása szinte minden technológiai eszközben valamilyen szinten jelen van, így az adatfeldolgozás szinte elengedhetetlen feltétele is egyben. Néhány gyakorlati példa: egy TV felskálázó²⁵ algoritmus, vagy a hangfelismerő megoldások²⁶, vagy a telefonok képjavító²⁷ funkciói.

A következő alfejezetekben a legfontosabb megoldások mögötti technológiákra térek ki.

5.1 A mesterséges neurális hálózatok²⁸

Általánosságban elmondható, hogy a Machine Learning „lelkét” az emberi agy²⁹ neuronhálózata ihlette. Erre a hálózatra a legjobb példának a mesterséges neurális

²⁵ Joel Hruska – How to Upscale Video to 4K, 8K, and Beyond (2021.07.12) <https://www.extremetech.com/extreme/310029-how-to-upscale-video-content-to-4k-8k-and-beyond> (Utolsó letöltés: 2023.04.15.)

²⁶ Daniel Martin – Cortana vs. Siri vs. Google Assistant vs. Alexa (2021.03.09) <https://www.digitaltrends.com/computing/cortana-vs-siri-vs-google-now/> (Utolsó letöltés: 2023.04.15.)

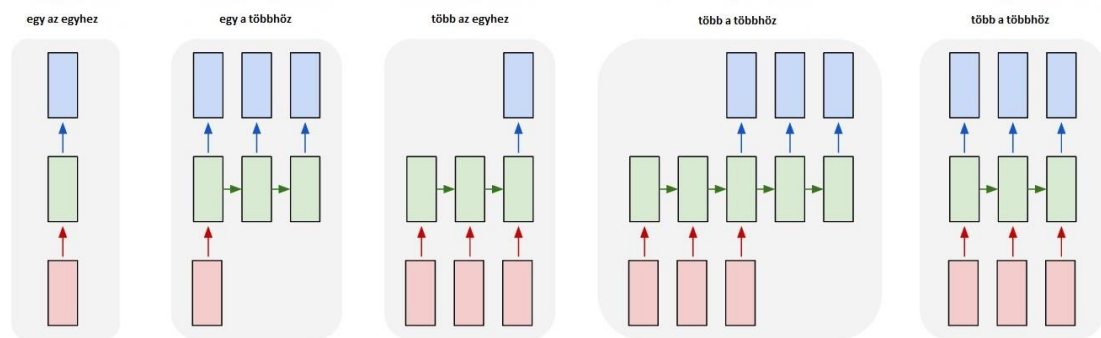
²⁷ Paul Mah – How Apple Uses AI To Produce Better Photos (2021.11.03) <https://www.cdotrends.com/story/15984/how-apple-uses-ai-produce-better-photos> (Utolsó letöltés: 2023.04.15.)

²⁸ Altrichter Márta, Horváth Gábor, Pataki Béla, Strausz György, Takács Gábor, Valyon József – 2006 Neurális hálózatok, Budapest. Panem Könyvkiadó

²⁹ Neurális hálózatok - <https://atozofai.withgoogle.com/intl/hu/neural-networks/> (Utolsó letöltés: 2023.04.15)

hálózatot (MNH, angolul: Artificial Neural Network -ANN) hozhatjuk fel. Ez gyakorlatilag mesterséges neuronokból álló hálózat, melyek egyszerű számítási egységeket jelentenek, általában rendezetten helyezkednek el a hálózaton belül és ezek jól meghatározott módon kapcsolódnak egymáshoz. Mivel párhuzamosítva épülnek fel, a számítások elvégzése is párhuzamosan történhet, így nagy sebességű adatvégrehajtásra képesek.

Minimum három fő réteg különíthető el általában a neurális hálózatok esetén, a bementi és kimeneti réteg mellett legalább egy rejtett réteg is megjelenik. [7. ábra]



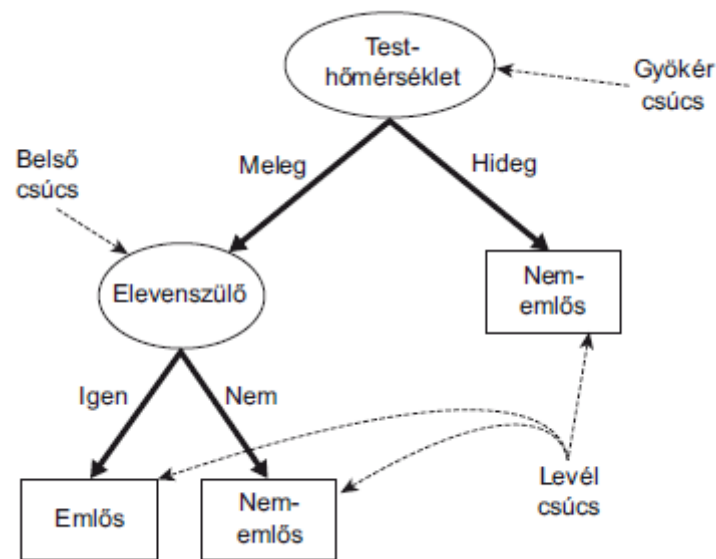
7. ábra – Neurális hálózatok szemléltetése (Forrás: Andrej Karpathy - The Unreasonable Effectiveness of Recurrent Neural Networks (2015.05.21.) <https://karpathy.github.io/2015/05/21/rnn-effectiveness/> (Utolsó letöltés: 2023.04.15.))

5.2 Döntési fák

A döntési fa (Decision Tree³⁰) újfent a természetből vett mintát követ, egy faszerkezetet jelenít meg, fejjel lefelé, azaz „gyökerétől” kiindulva, az ágakon végig haladva, eljutva végül a levelekhez.

Ezt átültetve a mesterséges intelligencia döntési fa megvalósítására, a legfelső szinten az eldöntendő kérdésünk áll, mely minden elágazáshoz érve tovább finomodik és végül a kimeneteket a levelek jelzik. Ezáltal egy bemenet mindig egy kimenethez fog eljutni, a döntési sor végeredményeképpen és egzakt módon ellenőrizhető eredményt kapunk. [8. ábra]

³⁰ Decision Trees - <https://www.ibm.com/topics/decision-trees> (Utolsó letöltés: 2023.04.15)



8. ábra – Emlős osztályozási döntési fa (Forrás: Pang-Ning Tan, Michael Steinbach, Vipin Kumar - Bevezetés az adatbányászatba (2006, Magyar változat: 2011) <https://gyires.inf.unideb.hu/KMITT/a04/ch04s03.html> (Utolsó letöltés: 2023.04.15.)

Ahogy a természetben kellően sok fa egymás közelében erdőt jelent, így a különböző döntési fák összekapcsolását „véletlen erdőnek” („random forrest”) nevezzük.

A szakértői rendszerek fogalma szorosan kapcsolódik pl. a döntési fák fogalmához: mivel a szakértői rendszerek egy olyan mesterséges intelligencia alkalmazások csoportja, amelyek célja, hogy megoldást nyújtsanak olyan problémákra, amelyekben emberi szakértelemre van szükség. A szakértői rendszerek különböző szabályokat, tudást és logikát használnak az adatok feldolgozására és az eredmények előállítására.

A döntési fák egy olyan számítógépes programozási technika, amely lehetővé teszi az adatok hierarchikus szerkezetbe rendezését, és a különböző döntési lehetőségek bemutatását. A döntési fák a szakértői rendszerek egyik formája, amelyek lehetővé teszik a szakértői tudás modellezését és automatizálását. A döntési fák segítségével a szakértői rendszerek hatékonyan megoldhatnak olyan problémákat, amelyekben több döntési lehetőség van, és amelyekhez a probléma különböző szempontjai is hozzátartoznak.

Például, ha egy szakértői rendszernek meg kell határoznia, hogy egy betegnek milyen kezelést kell kapnia, akkor a rendszer felhasználhatja a döntési fák módszerét annak érdekében, hogy az adott beteg tüneteit és korábbi orvosi előzményeit figyelembe véve javasolja a legmegfelelőbb kezelési lehetőséget.

5.3 Regresszió

A gépi tanulás megvalósításai közül az egyik az ellenőrzött, vagy felügyelt tanulás, mely esetben egy tanító adathalmaz alapján tanítjuk az algoritmust, viszont ismert a probléma kívánt megoldása is. Tipikusan az osztályozást és a regressziót soroljuk ide legjobb példaként. A kettő közötti különbség talán a legegyszerűbben úgy jellemezhető, hogy a vizsgált objektum egy meghatározott osztályba kerül besorolásra, vagy pedig egy közvetlen tulajdonságát határozzuk meg. A regressziót általában becslésekhez alkalmazzuk. Az egyik legelterjedtebb megoldása a lineáris regresszió. Azonban, amennyiben a célváltozónk egyszerű bináris értéket vehet csak fel, akkor arra nem alkalmazható a lineáris regresszió. Ennek az általánosításának tekinthető a logisztikus regresszió, mely módszer már az osztályozás csoportjába sorolható, mivel a számunkra „jó” és „rossz” egymáshoz viszonyított arányát vizsgálja meg logaritmikusan.³¹

5.4 Fuzzy-logika

A Fuzzy-logika az olyan eseteket jelenti, mikor nem egyértelműek vagy bizonytalanok a kimeneteink. Ez azzal a nagy előnnyel ruházza fel, hogy gyakorlatilag bármely bemenetre képesek vagyunk választ adni, akármennyire zajos, vagy torzult is az. Ezáltal nagyon hatékony megoldást kínálhat, hisz az emberi döntéshozást képes szimulálni.³²

³¹ Supervised and Unsupervised Machine Learning Algorithms - <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/> (Utolsó letöltés: 2023.04.15)

³² Johanyák Zsolt Csaba 2004, Fuzzy logika https://johanyak.hu/files/u1/publi/J_Fuzzy_logika_segedlet.pdf (Utolsó letöltés: 2023.04.15)

Ezzel szemben a hátránya is ebből származik, mivel nem lehet szisztematikus módon definiálni, így többértelmű megoldást hozhat. Ez ugyanígy eredményezi, hogy az eredményeink bizonyítása sem garantált, adott esetben gyakorlatilag lehetetlen. Emellett a precizitása sem garantált, hisz az inputok alapján pontatlan adatokkal is dolgozhat.³³

5.5 Nem-kauzális modellezés

A szakdolgozat írása alatt készültek el és jelentek meg nemzetközi szinten is azok a publikációk Váradi Dániel KJE-Hallgatóhoz kötődően, melyek egy újszerű Solver-alapú, OAM-vezérelt elemzési kultúrát alapoznak meg tovább gondolva a Joker-modell adathiánykezelésre kialakított logikáját is: <https://miau.myx.hu/miau2009/index.php3?x=e0&string=l.v.radi>, ill. <https://miau.myx.hu/miau/296/5.%20PROCEEDINGS%20BOOK%20V-2.pdf#page=126>

³³ Fodor János – Fuzzy irányítási rendszerek https://uni-obuda.hu/fodor/06_fuzzy_iranyitas.pdf (Utolsó letöltés: 2023.04.15)

6 ÖSSZEFOGLALÁS

A szakdolgozatom célja az OAM (Objektum-Attribútum Mátrix) szempontok szerinti információhalmaz-feldolgozás és elemzés feltérképezése volt. A vizsgálat során bemutatásra kerültek a különböző módszerek, technikák és eszközök, melyekkel az OAM elvek mentén strukturált adatok hatékonyabbá tehetők, és olyan ismeretek nyerhetők, amelyek értéket teremtenek az olvasó számára.

A dolgozatom a Ko-Do-Libri projekt egy részét dolgozza csak fel, mely rész is fontos alkotóeleme volt a teljes egész megvalósulásának. Hallgatótársaimmal összedolgozva csapatmunkában sikerült megvalósítanunk a Ko-Do-Libri projekt keretében a kitűzött célunkat, amihez mentoraink segítségét is köszönöm!

Mint azt a projektünk sikeressége és a dolgozatomban bemutatott témák is szemléltetik, az OAM szerinti adatelemzés lehetővé teszi a felhasználók számára a pontosabb, gyorsabb és hatékonyabb adatfeldolgozást és döntéshozatalt.

Az OAM-alapúság felveti, hogy a GPS (General Problem Solving) kérdésköre reális tudományos kihívás lehet/maradhat a jövőben is...

7 ÁBRAJEGYZÉK

1. ábra – Példatáblázat 5 adatsorral (forrás: chatGPT-vel generált saját munka).....	23
2. ábra – Átlag és meredekség meghatározása (forrás: chatGPT-vel generált saját munka)	23
3. ábra – Az átlag szerinti rangsor felépítése (forrás: chatGPT-vel generált saját munka)	24
4. ábra – Meredekség szerinti rangsor felépítése (forrás: chatGPT-vel generált saját munka)	24
5. ábra – Eredmény táblázat (forrás: chatGPT-vel generált saját munka)	25
6. ábra - A program működésének folyamatábrája (Forrás: Jónás Attila KJE GazdMen szakdolgozata - 2022).....	28
7. ábra – Neurális hálózatok szemléltetése (Forrás: Andrej Karpathy - The Unreasonable Effectiveness of Recurrent Neural Networks (2015.05.21.) https://karpathy.github.io/2015/05/21/rnn-effectiveness/ (Utolsó letöltés: 2023.04.15.))	37
8. ábra – Emlős osztályozási döntési fa (Forrás: Pang-Ning Tan, Michael Steinbach, Vipin Kumar - Bevezetés az adatbányászatba (2006, Magyar változat: 2011) https://gyires.inf.unideb.hu/KMITT/a04/ch04s03.html (Utolsó letöltés: 2023.04.15.)	38

8 FELHASZNÁLT FORRÁSOK JEGYZÉKE

- OECD (2023) <https://www.oecd.org/> (Utolsó letöltés: 2023.04.15.)
- Robot-polgár (e-gazság projekt 2022 vs 2010) (2022) https://miau.myx.hu/miau/285/e_gazsag_2022/ (Utolsó letöltés: 2023.04.15.)
- Hirschmann Zsófia (2022) Választási esélyek egységes rendszerű elemzése több országra párhuzamosan
- Jónás Attila (2022) Nyílt adatbázisok automatizált elemzése OAM megközelítés alkalmazásával
- What is Cloud Technology, and How Does It Work? (2022) <https://dynamixsolutions.com/what-is-cloud-technology-and-how-does-it-work/> (Utolsó letöltés: 2023.04.15.)
- What is Big Data? (2022) <https://www.oracle.com/big-data/what-is-big-data/> (Utolsó letöltés: 2023.04.15.)
- What is high performance computing? (2022) <https://www.netapp.com/data-storage/high-performance-computing/what-is-hpc/> (Utolsó letöltés: 2023.04.15.)
- Machine Learning (2022) <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML> (Utolsó letöltés: 2023.04.15.)
- Data Mining (2022) <https://www.techtarget.com/searchbusinessanalytics/definition/data-mining> (Utolsó letöltés: 2023.04.15.)
- Deep Learning (2022) <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network> (Utolsó letöltés: 2023.04.15.)
- Mesterséges Intelligencia (2022) <https://www.europarl.europa.eu/news/hu/headlines/society/20200827STO85804/mi-az-a-mesterseges-intelligencia-es-mire-hasznaljak> (Utolsó letöltés: 2023.04.15.)
- Kockázatelemzés (2022) https://kti.uni-nke.hu/document/vtkk-uni-nke-hu/iii_tantargy-kockazatelemzes.original.pdf (Utolsó letöltés: 2023.04.15.)
- Önvezető járművek (2022) https://kvantumugras.blog.hu/2022/06/22/az_onvezeto_jarmuvek_jelene_es_jovoje (Utolsó letöltés: 2023.04.15.)
- Robottechnológiák (2022) <https://computerworld.hu/prcikk/a-legmodernebb-robottechnologia-a-webaruhazak-szolgalataban-297697.html> (Utolsó letöltés: 2023.04.15.)
- Cookie és adatvédelem (2010.október 27) <https://www.adatvedelmiszakerto.hu/2010/10/cookie-es-adatvedelem-felhasznalok-nyomokovetese-az-interneten/> (Utolsó letöltés: 2023.04.15.)

Jörg Schieb (2020.02.17) Facebook freut sich über all unsere Metadaten!
<https://blog.wdr.de/digitalistan/facebook-freut-sich-ueber-all-unsere-metadaten/>
(Utolsó letöltés: 2023.04.15.)

<https://miau.my-x.hu/miau/297/ChatGPT.docx> (Utolsó letöltés: 2023.04.15.)

<https://miau.my-x.hu/miau/297/ChatGPT2.docx> (Utolsó letöltés: 2023.04.15.)

[https://miau.my-x.hu/miau/298/ChatGPT_GPT4%20\(1\).docx](https://miau.my-x.hu/miau/298/ChatGPT_GPT4%20(1).docx) (Utolsó letöltés: 2023.04.15.)

[https://miau.my-x.hu/miau/298/ChatGPT2_GPT4%20\(1\).docx](https://miau.my-x.hu/miau/298/ChatGPT2_GPT4%20(1).docx) (Utolsó letöltés: 2023.04.15.)

https://miau.my-x.hu/miau/298/chatgpt_.docx (Utolsó letöltés: 2023.04.15.)

<https://miau.my-x.hu/miau/298/ChatGPT-food.pdf> (Utolsó letöltés: 2023.04.15.)

<https://miau.my-x.hu/miau/298/ChatGPT-food-2.pdf> (Utolsó letöltés: 2023.04.15.)

https://miau.my-x.hu/miau/299/ChatGPT_az_elethosszig_tarto_tanulas_v1_v2.docx (Utolsó letöltés: 2023.04.15.)

Csallner András Erik (2015) Bevezetés az SPSS statisztikai programcsomag használatába http://www.jgypk.hu/tamop15e/tananyag_html/spss/alapfogalmak3.html
(Utolsó letöltés: 2023.04.15.)

Lantos Gábor (2021.09.10) Árnyjáték – A Főkomponens elemzés (Principal Component Analysis)
https://statisztikaegyszeruen.blog.hu/2021/09/10/arnyjatek_fokomponens_elemzes_principal_component_analysis (Utolsó letöltés: 2023.04.15.)

Csallner András Erik (2015) A klaszterelemzés technikája
http://www.jgypk.hu/tamop15e/tananyag_html/spss/a_klaszterelemzs_technikja.html
(Utolsó letöltés: 2023.04.15.)

Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2006, magyar fordítás 2011) Bevezetés az adatbányászatba, 6. fejezet - Asszociációs elemzés: Alapvető fogalmak és algoritmusok <https://gyires.inf.unideb.hu/KMITT/a04/ch06.html> (Utolsó letöltés: 2023.04.15.)

Christopher Tozzi (2022.11.14.) The Data Transformation Process Explained in Four Steps <https://www.precisely.com/blog/big-data/data-transformation-process-four-steps> (Utolsó letöltés: 2023.04.15.)

https://miau.my-x.hu/miau/281/KEI_10012022150615288.xlsx (Utolsó letöltés: 2023.04.15.)

MyX-Free COCO (2022) <https://miau.my-x.hu/myx-free/coco/> (Utolsó letöltés: 2023.04.15.)

Joel Hruska (2021.07.12) How to Upscale Video to 4K, 8K, and Beyond
<https://www.extremetech.com/extreme/310029-how-to-upscale-video-content-to-4k-8k-and-beyond> (Utolsó letöltés: 2023.04.15.)

Daniel Martin (2021.03.09) Cortana vs. Siri vs. Google Assistant vs. Alexa
<https://www.digitaltrends.com/computing/cortana-vs-siri-vs-google-now/> (Utolsó letöltés: 2023.04.15.)

Paul Mah (2021.11.03) How Apple Uses AI To Produce Better Photos
<https://www.cdotrends.com/story/15984/how-apple-uses-ai-produce-better-photos>
(Utolsó letöltés: 2023.04.15.)

Karin Kelley (2022.11.18) What is Data Analysis? Methods, Process and Types Explained <https://www.simplilearn.com/data-analysis-methods-process-types-article>
(2023.04.15)

Altrichter Márta, Horváth Gábor, Pataki Béla, Strausz György, Takács Gábor, Valyon József (2006) Neurális hálózatok. Budapest, Panem Könyvkiadó

Neurális hálózatok - <https://atozofai.withgoogle.com/intl/hu/neural-networks/>
(Utolsó letöltés: 2023.04.15)

Decision Trees - <https://www.ibm.com/topics/decision-trees> (Utolsó letöltés: 2023.04.15)

Supervised and Unsupervised Machine Learning Algorithms - <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/> (Utolsó letöltés: 2023.04.15)

Johanyák Zsolt Csaba (2004) Fuzzy logika. https://johanyak.hu/files/u1/publi/J_Fuzzy_logika_segedlet.pdf (Utolsó letöltés: 2023.04.15)

Fodor János – Fuzzy irányítási rendszerek https://uni-obuda.hu/fodor/06_fuzzy_iranyitas.pdf (Utolsó letöltés: 2023.04.15)

<https://clio-infra.eu/Countries/Austria.html#countries> (Utolsó letöltés: 2023.04.15)

https://miau.my-x.hu/miau/261/coldwar_robot-historian_turing-test.pdf (Utolsó letöltés: 2023.04.15)

<https://miau.my-x.hu/miau2009/index.php3?x=e0&string=kaleid> (Utolsó letöltés: 2023.04.15)

http://miau.my-x.hu/miau/141/log_y0.doc (Utolsó letöltés: 2023.04.15)

<http://miau.my-x.hu/miau/140/la140.doc> (Utolsó letöltés: 2023.04.15)

<https://miau.my-x.hu/miau2009/index.php3?x=e0&string=l.v.radi> (Utolsó letöltés: 2023.04.15)

<https://miau.my-x.hu/miau/296/5.%20PROCEEDINGS%20BOOK%20V-2.pdf#page=126> (Utolsó letöltés: 2023.04.15)