Outlier

Clementine



2021. tél

Clementine

Láncreakció

Clementine Data Science podcast



Bevezető



A mesterséges intelligencia felfalja a világot, elveszi az emberek munkáját, és megállíthatatlan harci gépeket irányít. Ismerősek az ehhez hasonló mondatok? Az MI-vel kapcsolatos hírek legtöbbször bombasztikusan hangzó állítások, amelyek felületes rövidhírek formájában érkeznek az átlagemberhez. A kattintásvadász címadás azután megteszi hatását, de lehetetlenné válik, hogy az érdeklődők ennél jóval árnyaltabb képet kapjanak.

Mi erre vállalkoztunk, amikor elindítottuk Outlier magazinunkat. Szakmai felelősségünk, hogy ne csak az ilyen technológiákat alkalmazók, ügyfeleink, de a témával egyre többször találkozó érdeklődők számára is a realitások talaján maradjunk.

Mint minden új, formálódó, forrongó tudományterületen, az adattudományban is vannak jelentős nézetkülönbségek. Eltérőek a vélemények az egyes új eredmények hasznáról, távlatairól. A tudományos módszer egyik alapja az újdonságokkal kapcsolatos egészséges szkepticizmus. Ugyanakkor a nagy áttöréseket is fontos gyorsan felismerni. Ma az adattudomány és még inkább az üzleti alkalmazás területén is sok a sarlatán, aki a zavarosban halászva gyűjt erkölcsi és anyagi hitelt. Az értő olvasókat azonban nehezebb félrevezetni, mert van tudásuk, hátterük a kételkedéshez.

Ennek szellemében a mostani lapszámunkban is készen állunk arra, hogy hírt adjunk az MI-vel, adatokkal, elemzésekkel kapcsolatos történésekről. A terület óriási, és minden nappal egyre csak bővül. Mi szerény – de okos! – eszközeinkkel szeretnénk útmutatót kínálni: így például olvashatnak arról, miért fontos a data literacy, vagyis az adatolvasási készség; milyen tanulságokkal jár a Facebook posztjainak elemzése; de a fajsúlyosabb írások közül bátran ajánlhatjuk a szignifikancia elvével foglalkozó szakcikket is.

Köszönjük, ha megtisztelnek figyelmükkel!

Impresszum

Kiadó:

Clementine

(Statistical Products Hungary Kft.)

1115 Budapest,
Bartók Béla út 105–113. 1/B IV. emelet
Telefon: 06 1 457 0561
E-mail: info@clementine.hu

Gold Business Partner IBM

Főszerkesztő:
Törökné Nagy Monika

Fotók:
Meraki Marketing Kft.

Nyomda:
Pannónia Nyomda Kft.

A lap elkészítésében közreműködő kutatók:

Csatlós Béla
Geszler Nikolett
Hegedüs Pál
Kelemen Sára Bernadett
Kovács Gyula
Könyves Tóth Előd
Körmendi György
Pancza Judit
Zsilinszki Anna

Szerkesztőség, tartalom, dizájn

meraki
www.meraki.hu

BEKÖSZÖNŐ 3

TECHNOLÓGIA 6
A KÉK 3 ÁRNYALATA DATA LITERACY LECKÉK
/ Fejlesszük adatelemzési készségeinket!





HÍREK 10
/ Az MI etikája a WHO szerint
/ Alexa, a főnővér



TECHNOLÓGIA 12
METASZIGNIFIKANCIA
/ Egy izgalmas statisztikai jelenség háttere

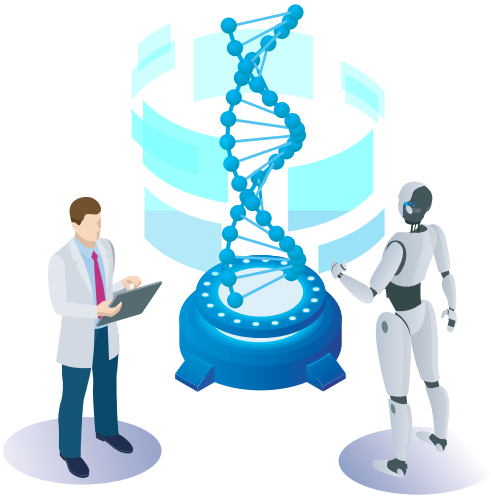


HÍREK 18
/ Jól fest az algoritmus
/ De mit adtak nekünk a méhek?



INNOVÁCIÓ 20
EGÉSZSÉGES MEGOLDÁS AZ EGÉSZSÉGMENEDZSMENT RÉSZÉRE
/ Szövegbányászat az egészségügyben

Tartalom



KITEKINTÉS 22

DR. MI RENDEL!

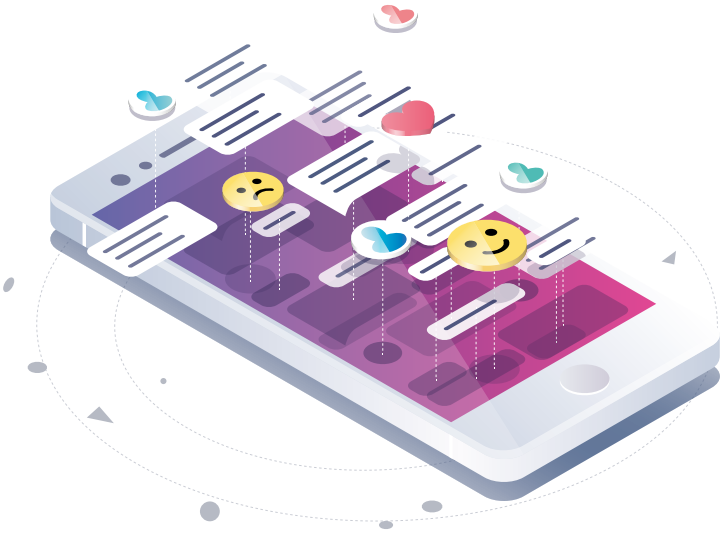
/ MI az orvostudományban és az egészségügyben



TECHNOLÓGIA 28

DATA FABRIC KONCEPCIÓ, MINT AZ ADATVEZÉRELT VÁLLALAT ALAPKÖVE

/ Bemutatkozik egy új adatmenedzsment-koncepció



KUTATÁS 24

MOSOLYGÓS, SZOMORÚ VAGY BOSSZÚS VAGY?

/ Kommunikáció hangulatjelekkel az online térben

NYEREMÉNYJÁTÉK 32

KERESZTREJTVÉNY

/ A mostani lapszám tartalmából

SZÓRAKOZÁS 34

A JÓ, A ROSSZ ÉS A SZÉP

/ MI a filmekben





DATA LITERACY

A KÉK HÁROM ÁRNYALATA
DATA LITERACY LECKÉK

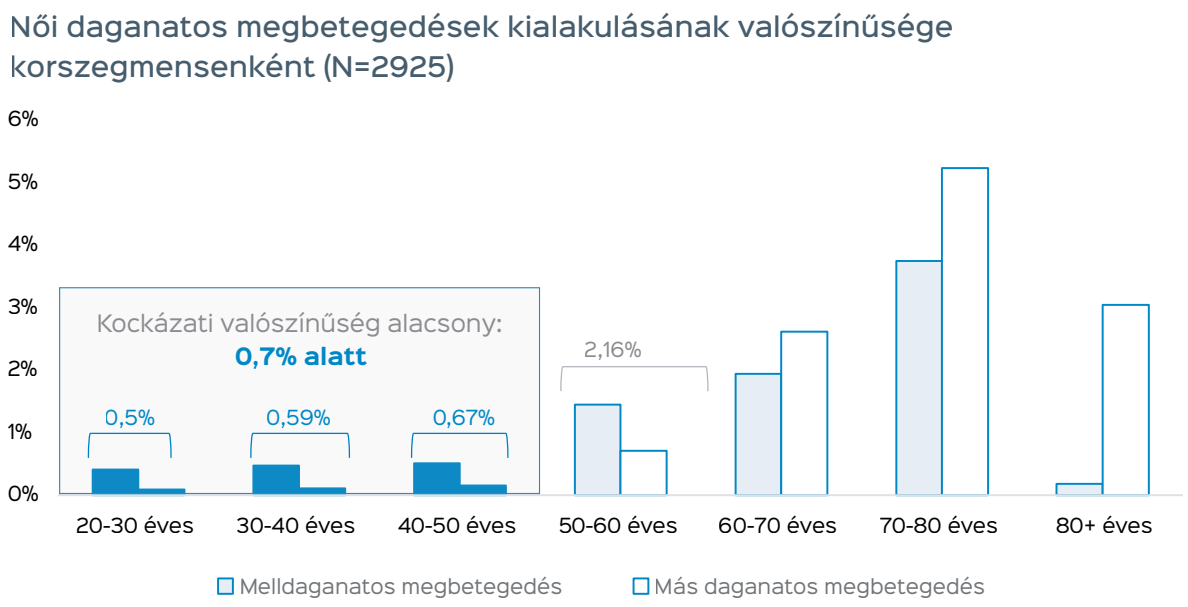
Szerző: Kovács Gyula

Egy korábbi cikkemben bővebben írtam arról, hogy mi is a Data Literacy: „adatelemzési eredmények, táblázatok, grafikonok olvasásának és értelmezésének képessége.” Többnyire azt hiszik az emberek, hogy mindenki képes táblázatokat és grafikonokat olvasni. Hogy így van, vagy nincs, nem tudom, de tegyünk egy próbát.

Egy nagyon egyszerű adattábla található lentebb. Azt mutatja, hogy egy nőnek milyen valószínűséggel lehet daganatos megbetegedése az adott korszegmensben. Akkor kezdődjön a játék! Te milyen „érdekességeket” látsz az adatokban?
Ha kész vagy, most nézzük, én mit találtam. Ehhez elkészítettem néhány diagramot, hogy jobban tudjam „olvasni” az adatokat. A diagramok ugyanis olyan vizualizációs „kódok”, amelyek a számokat agyunk számára értelmezhetőbb formában ábrázolják, és így növelik az információfeldolgozás sebességét.

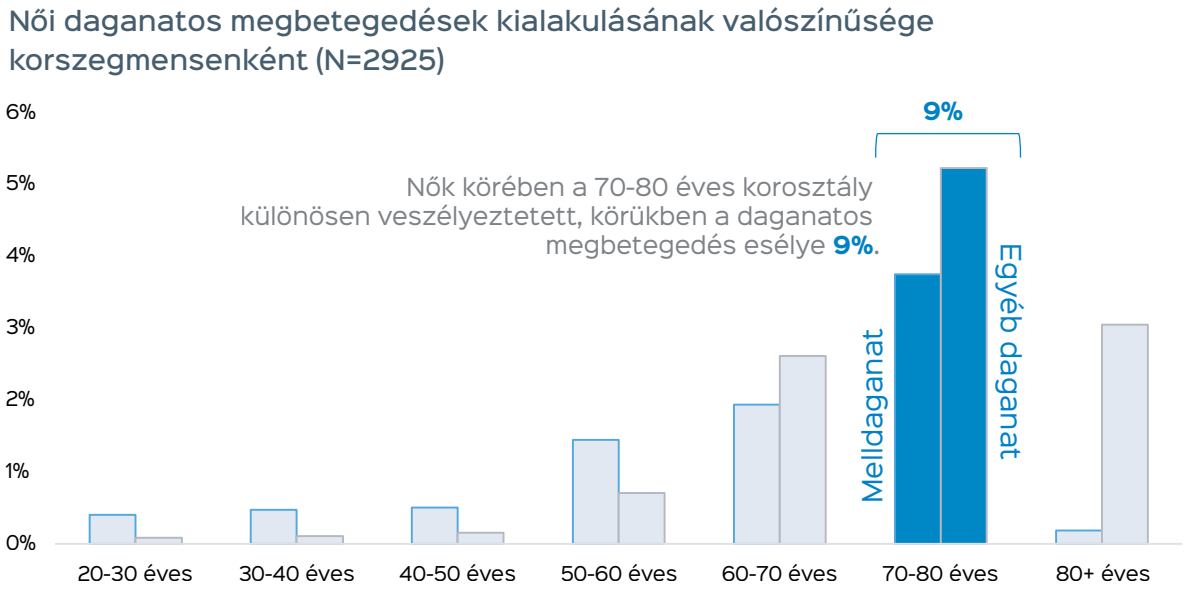
Életkor	**Adott életkorszegmensben adott megbetegedés valószínűsége**	
	Melldaganatos megbetegedés	Más daganatos megbetegedés
20–30 éves	0,41%	0,09%
30–40 éves	0,48%	0,11%
40–50 éves	0,51%	0,16%
50–60 éves	1,45%	0,71%
60–70 éves	1,94%	2,62%
70–80 éves	3,75%	5,23%
80+ éves	0,19%	3,05%

1. állítás: 50 éves kor alatt alacsony a daganatos megbetegedés valószínűsége.



A diagramon eleve jól látható, de úgynevezett preattentív attribútummal (téglalap) kiemeltem, hogy 20–50 éves kor között az oszlopok alacsonyak, azaz kicsi a megbetegedés valószínűsége. Vegyük észre, hogy a diagram úgy van kialakítva, hogy az olvasó figyelme a 20–50 éves tartományra esik, a fő üzenet szövegesen is ki van emelve („Kockázati valószínűség alacsony: 0,7% alatt"), és hogy az olvasónak legyen összehasonlítási alapja, a következő életkorszegmensnél (50–60 évesek) is ki vannak írva az értékek. Így az olvasó azonnal el tudja helyezni, hogy a 0,7% sok vagy kevés (a még mindig viszonylag alacsony kockázatú, 50–60 éves szegmensben is jóval nagyobb a daganatos megbetegedés esélye).

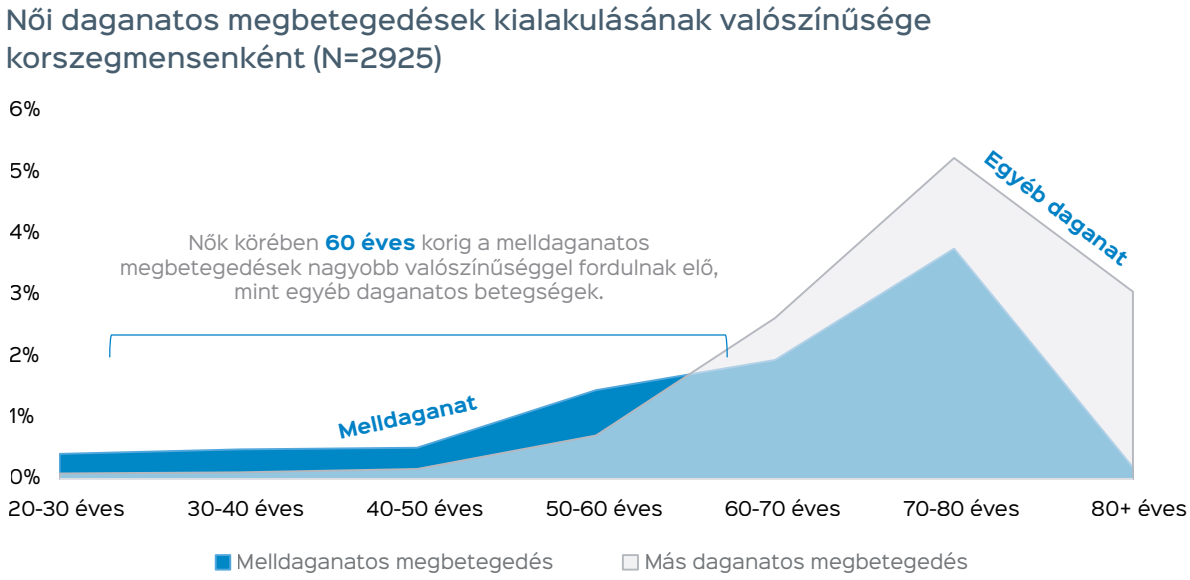
2. állítás: A 70–80 éves életkorszegmensben extrém magas a daganatos megbetegedés kockázata.





Ugyanaz az ábra, de most arra irányítjuk a figyelmet, hogy hol a legmagasabb a kockázat. Egy konkrét életkorszegmenst emeltünk ki (70–80 közöttiek), itt is szövegesen kiírjuk a megállapításunkat, és erős oszlopszínekkel irányítjuk a hallgatóság figyelmét, hogy hova nézzen. Az előző ábrához képest jóval kevesebb információ jelenik meg, így még erőteljesebb az üzenet.

3. állítás: 20–60 éves kor között az melldaganatos megbetegedések kockázata a nők körében magasabb, mint minden más daganatos megbetegedésé.



Az adattábla ugyanaz, mint eddig, de ezúttal területdiagramot alkalmaztunk, hogy kiemeljük a fenti állítást. Jól látható a diagramon, hogy 60 éves korig a melldaganatos értékek magasabbak, és utána fordul meg az arány. Itt is szövegesen emeltük ki a kulcsüzenetet. Sokszor a diagram olvashatóságát javítja, ha jelmagyarázat (legend) helyett az adatvonalra írjuk ki az adatsor jelentését („Melldaganat", „Egyéb daganat").

Mi is történt pontosan?

Van egy adattábla és három teljesen különböző diagram, ahol **mindegyik diagram egy-egy üzenetet fogalmaz meg.** Az adatolvasási készség lényege épp ez: észre tudsz-e venni az adatok között ilyen „üzeneteket"? Ha képes vagy rá, akkor ezt át tudod-e adni mások számára? Tudsz-e priorizálni az üzenetek között?

Az adatolvasási készség tehát csak részben szól rólunk. Fontos, hogy ismerjük az üzleti oldal problémáit, és így tudjunk az ő szemükkel „olvasni" adatokat. De ez csak az első lépés.

Utána kell készíteni egy olyan diagramot, amelyből a menedzserek/döntéshozók azonnal képesek leolvasni az általunk fontosnak gondolt üzenetet. (És csak egy üzenetet: egy diagram mindig egy üzenetet közvetítsen!)

Hogy ez mennyire egyszerű vagy mennyire nehéz, azt döntse el az olvasó! Jó hír azonban, hogy az adatolvasási készség fejleszthető.

Az MI etikája a WHO szerint

Az etikával, digitális technológiával foglalkozó, valamint emberi jogi szakértőkből álló nemzetközi csoport 18 hónapos munkával összeállította az egészségügyben alkalmazott mesterséges intelligenciákhoz felhasználható irányelveket. Az MI az egészségügy számos területén képes az eddigieknél hatékonyabb működtetést megvalósítani, de eközben a biztonság, hatékonyság és adatkontroll kérdését is megnyugtatóan rendezni kell.

A WHO, az Egészségügyi Világszervezet egészségügyi etikával foglalkozó szakbizottsága által gondozott kiadvány hat alapelvet határoz meg a mesterséges intelligencia használatában:

- az emberi anatómia védelmét,
- az emberi jóllét (well-being), biztonság és a közérdek előmozdítását,
- az átláthatóság, a megmagyarázhatóság és az érthetőség biztosítását,
- a felelősség és az elszámoltathatóság elősegítését,
- az elfogadás és a méltányosság biztosítását,
- valamint a reszponzív és fenntartható MI megvalósítását.



Komment

Kétségtelen, hogy hatalmas potenciál rejlik az MI orvoslásban történő felhasználásának területén, aminek határa a csillagos ég. Vagy akár az örök élet? Természetesen túlzás lenne azt állítani, hogy az MI csodát tud tenni, de az bizonyos, hogy a jelenlegi orvosi gyakorlat rendkívül hasznos kiegészítője lehet akár csak abban, hogy tehermentesíti az orvosokat a kevésbé szakértelmet igénylő feladatok alól, illetve klinikai döntéstámogató rendszerekkel segíti a pontosabb diagnózisok felállítását. De biztató eredményeket felmutató kutatások zajlanak világszerte MI-alapú EKG-elemzés, radiológiai vagy vérképalapú rákdiagnosztika, gépiasszisztens-fejlesztés, vérképalapú betegségmegelőzés támogatása, járvány-előrejelzés, gyógyszerfejlesztés és számtalan más orvosi területen egyaránt.

Egyre nagyobb a törekvés globális szinten az egészségügyi szervezetek, cégek egységes adatkezelő- és tárolórendszerben történő egyesítésének irányába is. Ha ugyanis ez minél jobban megvalósul, annál nagyobb potenciál rejlik a felhalmozódó adatok gépi feldolgozásában, hiszen annál inkább teljesülhet például a reprezentatív mintavételezés, és ezáltal az eredményeket torzító különböző hatások minimalizálhatók lennének. Talán mindenki érzi, hogy etikai szempontból nagyon ingoványos területen járunk. Sajnos nehezen vagy egyáltalán nem feloldható probléma, hogy a páciensek magánéletének és adatainak védelme akadályozza az ilyen típusú kutatásokat és fejlesztéseket, amelyek pedig az emberiség javát szolgálhatnák az egészségügyi ellátás minőségének javításával vagy akár teljesen új összefüggések feltárásával.

Éppen ezért rendkívül nagy körültekintés szükséges a fejlesztések során, de ezek ellenére azt gondolom, nagyon fontos, hogy ne ijedjünk meg ezektől a kihívásoktól, és ne hagyjuk kiaknázatlanul a rendelkezésünkre álló technológiát, amellyel életeket lehet menteni. Örömteli, hogy a WHO is kézbe vette és napirenden tartja a témát.

Zsilinszki Anna
fejlesztő

Forrás:
https://www.who.int/publications/i/item/9789240029200

Alexa, a főnővér

Nemsokára talán már a kórházakban is segíthet az Amazon cég hangfelismeréssel felruházott intelligens rendszere, az Alexa.

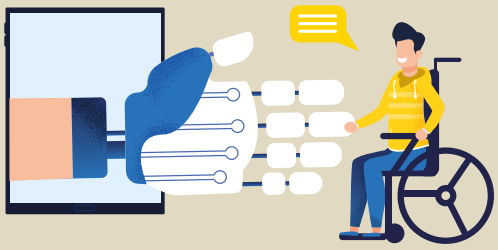
Hihetetlen logisztikai terheket ró a kormányokra a nővérhiány és az egyre idősödő, tehát egyre több kórházi ápolásra szoruló nyugati társadalom. Éppen emiatt is az amerikai Cedars-Sinai Kórházban száz beteg részvételével megszervezték egy pilotprogramot. Ennek keretében a betegek a kórtermekben elhelyezett Alexa készülékeken keresztül kommunikáltak és utasítottak: ha kellett, egy egyszerű mondattal csatornát váltottak a tévén, vagy éppen szóltak a nővérnek, hogy a mosdóba kell menniük. A világ első, ilyen jellegű kísérlete sikeresnek bizonyult. Mind a kórház, mind a betegek pozitívan nyilatkoztak az újfajta technológiai megoldásról, amelyet az Alexát működtető mesterséges intelligencia tett lehetővé. Azóta hasonló rendszereket állítottak be például a bostoni Beth Israel Deaconess Medical Centerben, a New York-i Commonwealth Care Alliance, Northwell Health intézményben és a Los Angeles-i Libertana Home Healthben is.

Forrás:
https://healthcareweekly.com/cedars-sinai-alexa-powered-platform/

Komment

A VUI (Voice User Interface), azaz a hang- és beszédalapú technológiák drámai fejlődését és térnyerését elnézve nem csoda, hogy az innováció az egészségügy területét sem hagyta érintetlenül: a hangalapú eszközök segíthetnek a betegek telefonos utánkövetésében és tájékoztatásában, az olyan adminisztratív teendőkben, mint az időpontfoglalás vagy az adatrögzítés, de akár abban is, hogy a hang elemzésével azonosítsanak bizonyos betegségeket (pl. a köhögés típusából vagy a beszédmintázatokból).
A beszéd a legtermészetesebb kommunikációs formánk, és a hangvezérlés sok szempontból kényelmesebb és gyorsabb, mint az írás: szabad marad a kezünk, párhuzamosan végezhetünk más tevékenységeket, és az olyan betegségekkel küzdőknek is elérhető, akiknek a gépelés nehézséget okoz (pl. Parkinson-kór). Ugyanakkor számos kihívás elé is állítja a tervezőket és a fejlesztőket: a beszédfelismerésnek az átlagostól eltérő helyzetekben is működnie kell (pl. beszédhibák, dialektusok kezelése). Gondolni kell arra is, hogy ki a fő felhasználói célcsoport: amennyiben az idősebb betegek, hogyan lehet az ő igényeikre szabni az eszközt és arra ösztönözni őket, hogy a személyes kontaktus mellett vagy helyett ezt is alkalmazzák? Hogyan lehet olyan személyiséggel, hanggal, beszédstílussal felruházni a gépet, ami megnyugtatja a betegeket, és bizalmat ébreszt bennük? Mi a helyzet a közterekkel és a kényes témákkal? Nem biztos, hogy a betegek egy kórteremben mások előtt hangosan szeretnék Alexával közölni, hogy mosdóba kell menniük.
Nem utolsósorban pedig ott a felelősség kérdése: egy olyan komplex és kardinális területen, mint az egészségügy, a lehető legkisebbre kell csökkenteni a hibák előfordulását, hiszen a tévedésnek súlyos következményei lehetnek az egyén egészségére és közérzetére nézve. Ugyanakkor pontos és előrelátó tervezéssel, folyamatos kutatással és ellenőrzéssel, illetve természetesen a humán kontroll megtartásával ezek a kockázatok minimalizálhatók.

Geszler Nikolett
szöveganalitikai tanácsadó





METASZIGNIFIKANCIA

Szerző: Könyves Tóth Előd

„A szignifikanciakutatás hiábavaló, hiszen ha az jön ki, ami várható, akkor felesleges, ha pedig nem, akkor biztos rosszak az adatok.”
(elkeseredett kutató nihilista kijelentése)

Megjelent az IBM Statistics 28 – benne a legfőbb újdonság a metastatisztika-menü

Az SPSS újdonságának apropóján érdemes legalább vázlatosan áttekinteni: az idők folyamán miféle ravasz fondorlatokat vezettek be a tudományos világban azért, hogy ha valami nem nyilvánvaló, de talán igaz, az korrektnek tűnő módon mégis legalább szignifikánsnak minősülhessen. A téma másik aktualitása az úgynevezett mesterséges intelligencia (MI), mivelhogy az MI hátterében sokszor matematikai statisztikai algoritmusok is rejlenek. Bár történelmileg az MI-felhajtás legalább harmadszori időszakát éljük, az MI kifejezés használata sokszor még ma is túlzásnak tűnik, hiszen kérdéses, mennyire jogos azt pusztán a számítógép, az algoritmus, a program vagy a statisztikai eljárás szinonimájaként alkalmazni.

Az ártatlanság vélelme és a nullhipotézis

Fontos a téma alapfogalmának közérthető tisztázása, de mindenekelőtt érdemes belátni, a szignifikancia fogalma nem ördögtől való. Ezzel analóg gondolkodás régóta ismeretes a bírósági gyakorlatban. Az ártatlanság vélelme a bűnösség vélelmének tilalma addig, amíg a bíróság a törvénynek megfelelően jogerősen másképp nem dönt. Az ártatlanság vélelme kötelezi a bírót a per elfogulatlan tárgyalásvezetésére. Az ártatlanság vélelme sérülhet, ha felmerül egy olyan tény bizonyossága, ami az ártatlanság esetén nagyon kis valószínűséggel volna lehetséges. A jog eme álláspontja érvényesül a szignifikancia fogalmának esetében is. Szignifikáns egy jelenség, ha egy igaznak vélt hipotézis mellett a valószínűsége nagyon kicsi, és emiatt a hipotézis elutasítható. Ez nem jelenti azt, hogy a hipotézis ellenkezője igaz, csak annyit, hogy a hipotézis elutasítása esetén kicsi a tévedés valószínűsége. Ha viszont keresve sem található olyan jelenség, amelynek szignifikáns mivolta miatt el lehetne utasítani a hipotézist, az nem jelenti a hipotézis bizonyosságát, csak annyit, hogy nincs elegendő okunk elvetni azt. A szignifikancia szintje az a kis valószínűségi érték, ami miatt már elutasítható a hipotézis.

A statisztikában szokás a kiinduló hipotézist nullhipotézisnek (H_0), ellentétét alternatív hipotézisnek (H_A) nevezni. A null jelzőt indokolják a statisztikai szignifikanciatesztek, illetve a statisztikai próbák technikája, amelyek rendszerint valamely statisztika nulla mivoltának feltételezésére épülnek. Egy statisztikai eredmény többnyire még akkor sem nulla, ha a nullhipotézis szerint nullának kellene lennie. Szignifikáns a nullától való eltérésnek az a mértéke, ami már nem magyarázható a véletlennel, így elegendő indoka lehet a nullhipotézis elvetésének, helyette az alternatív hipotézis elfogadásának. Annak az eltérésnek a valószínűsége, amelynél a nullhipotézis már elutasítható, a döntés szignifikanciaszintje. A szignifikanciaszint meghatározása egyszerűen csak megállapodás kérdése, a legelterjedtebb az 5%-os érték, de a döntés következményeinek súlyosságától függően kisebb értéket is használnak. A döntésnek négyféle esete áll fenn, H_0 elfogadása vagy elutasítása helyesen, illetve helytelenül. Elsőfajú tévedés H_0 helytelen elutasítása, másodfajú tévedés H_0 helytelen megtartása. Az elsőfajú tévedés valószínűsége maga a szignifikanciaszint, csökkentése tehát pusztán elhatározás kérdése. A másodfajú tévedés valószínűsége ismeretlen, a statisztikai minta növelésével csökkenthető. Fontos átlátni, hogy csakis a négy helyzet valószínűségének összege 100%. Ha például elutasítjuk H_0-t 5%-os szignifikanciaszinten, az egyáltalán nem jelenti H_A 95%-os valószínűségét, ha nem utasítjuk el, az viszont nem jelenti H_0 95%-os valószínűségű igazságát.

Akár H_0 igazolása, akár H_0 cáfolata a cél, ugyanaz a számítási és döntési folyamat zajlik. Sikernek tekinthető a szignifikáns eredmény, ha H_0 cáfolata, és sikertelen, ha H_0 igazolása volt a cél. Tehát értelmetlen a csakis statisztikailag szignifikáns eredménynek örülni. H_0 cáfolatának célja esetén H_0 munkahipotézisként szolgál, ugyanúgy, mint az ártatlanság vélelme a bíróságon. Ugyanaz a tárgyalási folyamat zajlik akár az ártatlanság bizonyítása, akár a cáfolata a cél, és nézőpont kérdése az ítéletnek örülni vagy sem.



A szignifikancia nem elég, avagy Parkinson törvénye érvényesül

„A másodfajú hiba annak az elmebajnak a valószínűsége, hogy valaki ezt a fogalmat kitalálja” – vélte az a matematikus, aki mintegy száz éve bevezette a szignifikanciatesztet. A „kritika” más matematikusok ötletére utalt, akik néhány évvel később kiegészítették az elméletet a másodfajú hiba fogalmával. Az ő véleményük szerint viszont a szignifikanciatesztek elméletének megalkotója „több száz évvel kiesett

az időből, de tévedései fantasztikusak". Ily kedvesen vitatkozgattak az egyébként valóban jelentős tudósok, akiknek munkássága alapján folynak azóta is a szignifikanciaszámításokra is épülő kutatások. Ha az elmélet megalkotói így kételkedtek egymás hozzáértésében, mire lehet számítani utánuk?

Néhány éve szignifikanciaválság van, és bár a problémák már a fogalom születésekor elkezdődtek, csak néhány éve csúcsosodtak azzal a megdöbbentő fejleménnyel, hogy egy tekintélyes amerikai szaklap megtagadta a szignifikanciaszámításra épülő kutatásokról szóló cikkek közlését. Hiába egyszerű a szignifikanciatesztek elve, érvényessége csak akkor garantált, ha elvégzésének kritériumai teljesülnek, és az eredményeket helyesen értelmezik. Ez pedig már nem olyan egyszerű, mint maga az elv, a jogosan alkalmazható statisztikák függenek az adatok skálázásának mérési szintjétől, az adatok eloszlásától és persze attól, hogy éppen mi a kérdés. Ezek alapján viszont akár automatikusan is meghatározható, hogy melyik statisztikai eljárás használható.

A statisztikai műveletek különösebb hozzáértés nélkül is könnyen és gyorsan elvégezhetők számítógéppel, nem meglepő hát, hogy ezen a területen is érvényesül Parkinson törvényének kiterjesztése, miszerint a munka olyan, mint a gáz: kitölti a rendelkezésre álló teret, időt, emberidőt, pénzkeretet, tárhelyet, kihasználja a teljes eszközállományt, papírkészletet, nyomtatói kapacitást stb. Sajnos gyakori az automatizmusokkal visszaélő, az alkalmazhatósági kritériumokat mellőző, esetleg az ipari mennyiségben előállított statisztikákat és azok félreértelmezéseit felelőtlenül publikáló, lényegében áltudományos tevékenység. Az eredmények félremagyarázásának leggyakoribb módja a korrelációk ok-okozatként való téves értelmezése. (Pl. A legtöbb ember ágy közelében hal meg, tehát az ágy veszélyes bútor.)

Régi görög mondás: „A gyerekek tudatlanságból dobálják a békákat, de a békák komolyan meghalnak." Ha pedig valaki szándékosan csal a statisztikai eszközökkel, az nem csak tudományos probléma. Sajnos még intézményesedett formák is léteznek, ahol a statisztikai eszközöket gazdasági vagy politikai haszonszerzés céljából szándékosan megtévesztően, csalárd módon használják. Nem a kuruzslókat kell vegzálni, az igazi probléma az, ha a képzett kutató „tudományos" tevékenysége nem különböztethető meg a kuruzslástól.

Ugyanakkor ma már létezik az úgynevezett polgáriadattudós-képzés is (Citizen Data Scientist), ahol a hagyományos elméleti alapok nélkül tanítják a matematikai statisztikai műveletek számítógéppel történő praktikus elvégzését. Ez rendben van, ha megfelelő didaktika is társul hozzá. A piaci kofáknak sem kell Newton-törvényekből vagy elektronikából vizsgázniuk, ha karos vagy elektromos mérleget használnak, de azt tudniuk kell, mit mérjenek kilóra, literre, hosszra, darabra, és hogyan olvassák le az értékeket, és az sem baj, ha a vevőnek is csak ennyi a tudása. Ilyesminek felelhet meg a statisztikában például az, hogy miféle adattípus esetén miféle középértéket szabad használni (módusz, medián, számtani átlag, négyzetes átlag stb.), és azokat hogyan kell értelmezni. A kutatások néha visszásnak tűnő világára az internet népe a „brit tudósok kimutatták" típusú mémekkel reagál. De itt akár Karinthy-féle kutatási ötletek is idézhetnénk, pl.: „A nők vagy a férfiak csókolóznak többet?", vagy „Ha létezhet barátság férfi és nő közt, akkor miért nem?"









Élni vagy visszaélni – a szignifikancia érvényessége itt a kérdés

Bármilyen adatmintából jogos leíró statisztikákat számolni, de a mintánál tágabb érvényességgel általánosítani, illetve következtetni csak a minta

által reprezentált populációra jogos, ha egyáltalán van ilyen. A reprezentatív minta megtervezése és felépítése nem hasraütésszerű véletlenen alapszik, hanem tudományos szigorúsággal megtervezett, esetleg kvótaszempontokat is figyelembe vevő eljárás. A mintából való következtetések alapjai az úgynevezett következtető statisztikák, illetve a statisztikai szignifikanciaszámítás.

Erő- és hatásméret-elemzéssel meghatározható, hogy adott elemszámot és hatásméretet feltételezve egy kiválasztott statisztikai próba mekkora valószínűséggel fog szignifikanciát jelezni. És fordítva: az is vizsgálható, hogy mekkora elemszám szükséges, hogy egy adott statisztikai próba meghatározott valószínűséggel szignifikánsnak minősítsen adott hatásmértékű statisztikát. Ezek az elemzések egyrészt meg kellene hogy előzzék a mintatervezést, másrészt a kutatási eredmények tálalásához is illene hozzátartozniuk. Komolyabban vehetők az olyan eredmények, amelyek valódi hipotéziseket erősítenek meg.

Korrektül elvégzett statisztikai műveletek esetén is keletkezhet paradox helyzet. Előfordulhat például, hogy csoportonként egyformán érvényes állításnak épp az ellenkezője igaz a csoportok összességére (Simpson-paradoxon). Paradox helyzet az is, ha egyszerre több, 5%-os szintű szignifikanciapróba közül csak 5% vagy ennél kevesebb szignifikáns. Ez ütközés a szignifikancia elvével, hiszen az 5%-os szignifikanciaszint éppen azt jelenti, hogy az esetek 5%-ában véletlenszerűen is adódhat szignifikáns eredmény.

A szignifikanciatesztek halmozásának problémája kezelhető szigorúbb szignifikanciaszint megállapításával, a post-hoc próbával. Sokféle post-hoc próba létezik, nem egyszerű közülük a választás, és előfordul, mint más statisztikai próbák esetén is, hogy a szaklap, amelybe a tanulmány készül, előírja, milyen próbákat kell használni. Post-hoc próbákkal legtöbben a varianciaelemzésnél találkoznak, ahol néha sok csoport középértékét kell egymással összehasonlítani. Ekkor kombinatorikusan felfogható módon a csoportok számának növekedésével hatványozottan nő az összehasonlítandó csoportpárok száma.

Például Bonferroni-elv esetén „a szignifikancia szintje per a próbák száma” tekinthető alkalmas döntési szintnek. Ez elfogadhatatlanul szigorú, ha a próbák száma nagyon sok. Másfajta lehetőség az LSD (Least Significant Difference), ez kábíthatja a kutatót egy olyan, minden különbséghez alkalmazható intervallummal, aminél nagyobb különbség szignifikánsnak minősül. Ez az elv viszont problémás akkor, ha lényegesen különböző elemszámok és szórások tartoznak a különböző statisztikákhoz, ekkor egy egységes intervallumhosszal mérni a különbséget értelmetlen.

Minden statisztikához számítható az azt szimmetrikusan övező megbízhatósági, konfidencia-intervallum (CI), ami azt jelenti, hogy „1 szignifikanciaszint” valószínűséggel ebbe az intervallumba tartozik a statisztikával becsült érték a populációban. CI mérete függ az aktuális elemszámtól és a statisztika szórásától is. Szemléletesen például két középérték eltérésének szignifikanciavizsgálata arról szól, menynyire „harapnak egymásba” ezek az intervallumok. Szélsőséges esetekben, ha CI-k átfedik egymás középpontjait, nem szignifikáns a különbség, ha pedig az intervallumok diszjunktak, az erős szignifikanciát jelent. Az elemszám csökkenésével tágulnak a megbízhatósági intervallumok, és megnő a másodfajú tévedés valószínűsége. Az elemszám növekedésével szűkülnek a CI-k, emiatt a nagyon kis különbségek is matematikailag szignifikánsak lehetnek. Vigyázat, a statisztikailag szignifikáns különbség nem ugyanaz, mint ami az egyébként szakmailag is figyelembe vehető lényeges különbség. Az egyik fő oka a szignifikanciatesztekre épülő kutatásokkal kapcsolatos válságnak, hogy sokszor összekeverik a matematikai és a szubsztanciális különbség fogalmát.

Fontos az interakciós hatások figyelembevétele, azaz meg kell győződni arról, hogy a minta jól definiálható csoportjaiban hasonló konstellációk alakulnak-e. (Pl. professzorok és callgirlök fizetései az életkor függvényében.) Az interakciós hatások lehetnek többdimenziósak is. Parciális jelenségek szintén befolyásolhatják az eredményeket, ekkor ezek hatásának kiküszöbölésével, azaz parciális statisztikákkal kell játszani. (Pl. lehet együttállás a strandolók és a fagyizók száma közt, de a hőmérséklet hatását kiküszöbölve kiderülhet, hogy még sincs korreláció.)

Szignifikanciák és magyarázó erők bizonyosságának fokozása

Szignifikáns statisztikai modell esetén fontos lehet megállapítani, hogy a modell hány százalékban magyarázza meg azt a jelenséget, amire éppen vonatkozik. (Szemléletesség miatt érdemes például a többváltozós regresszióra gondolni.) Sokszor eleve a modell magyarázó erejének nagysága a fő kérdés, és ha azt több változó kombinációja okozza, akkor az is kérdés, hogy melyik hány százalékban járul hozzá a magyarázó erőhöz.

A modell magyarázó erejéhez hozzájáruló változók fontosságának mértékét matematikailag többféle módon lehet korrektül megállapítani. A különféle elvek által meghatározott részesedési értékek nem egyformák, interferencia pedig az, ha a fontossági sorrend is megváltozik. Hasonló célú, de különböző statisztikai modellekhez különböző magyarázó erő tartozhat, mint ahogy az ugyanazt a hipotézist tesztelő szignifikanciatesztek is különféle érzékenységűek lehetnek.

Az érzékenységet és a hatékonyságot vizsgáló eljárások közül a legismertebb az eredetileg a radaros riasztórendszerekhez kifejlesztett módszer. Az eszköz érzékenységi szintjeihez úgynevezett ROC (Receiver Operating Characteristic)-görbéket szerkesztenek, amelyek a téves és helyes riasztások relatív gyakoriságainak mint koordinátaérték-párokkal meghatározott pontoknak a rajzolata. A görbén való elmozdulás a döntés szintjének változását jelenti, a magasabban futó görbe magasabb érzékenységre utal.

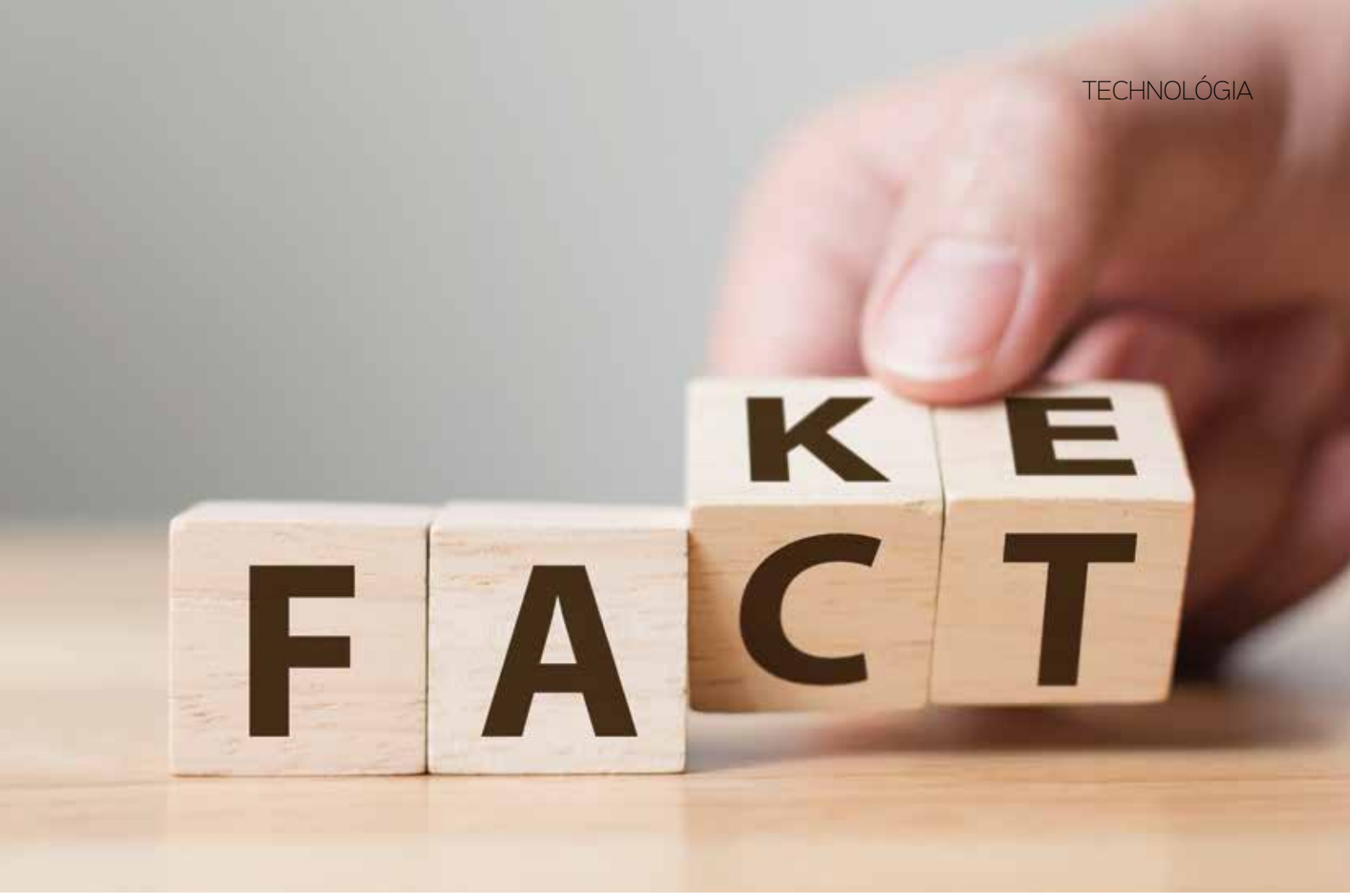
Érdemes azt is megvizsgálni, hogyan alakulnak a statisztikák és a szignifikanciaértékek, ha az adatok csak egy kicsit változnak vagy csak a minta véletlenszerű részmintáin számítódnak, vagy akár véletlenszerűen lennének szimulálva. A legegyszerűbb játék a minta véletlenszerű kettéosztása és a két részen számolt eredmények összehasonlítása. Amennyiben a kérdéses statisztikai eljárás prediktív jellegű, meg lehet vizsgálni, hogy az egyik mintarészen alkotott modell hogyan teljesít a másikon, ekkor a két véletlen részminta tanuló- és tesztmintaként szerepel.

A Jacknife, azaz zsebkéses módszer n-szer újraszámolja a statisztikákat, a k-adik lépésben épp a k-adik mintaelemet kihagyva. Ekkor az n darab „n-1” elemű mintákból számolt statisztikákból alkalmas becslési eljárással pontosabb érték adódik a kérdéses statisztikákra. Ez a „leave-one-out’” néven is ismert módszer érzékeny a durván kiugró viselkedésű (outlier) mintaelemek létezésére.

Rafináltabb játék a csizmapántoló (bootstrapping) mintavételi módszer; neve szándékosan Münchhausen báró egy kalandjára utal, aki egyszer saját csizmapántjánál fogva húzta ki magát a mocsárból. Az eljárás az eredeti mintából visszatevéses véletlen mintavétellel képez sok új mintát. Ennek hátránya, hogy az új minták csakis az eredeti mintából származó elemeket tartalmaznak, így attól lehet félni, hogy a minták nem tartalmazzák a változók kombinációinak a valóságban többféleképpen is előforduló eseteit. Viszont a sok új mintából származtatva a kérdéses statisztikákhoz átlag-, szórás-, szignifikancia- és konfidencia-intervallum is keletkezik.

A Monte Carlo-szimuláció az előbbi elv hátrányát kikerülve véletlenszerű adatokat generál úgy, hogy az azokból felépülő minta képviselje egy adott minta változóinak eloszlását és interkorrelációit. Az így keletkezett minta statisztikai elemzése, illetve kontrollmintaként felfogva azon korábbi historikus adatokhoz alkotott statisztikai modellek teljesítményének megállapítása alapvető a kockázatelemzési helyzetekben.

Megemlítendők még a Bayes-statisztikák a klasszikus statisztikáknak jóval kevesebb kritériummal alkalmazható alternatívái, matematikai hátterükben a világ leghatékonyabb algoritmusainak kikiáltott iteratív MCMC-mintaszimulációkkal (Markov Chain Monte Carlo). Szignifikanciaérték helyett Bayes-faktor számítódik, aminek jelentése, hogy a nullhipotézis valószínűsége hányszorosa az alternatív hipotézis valószínűségének. A konfidencia-intervallum helyett pedig hitelességi intervallumok lesznek. Minden klasszikus statisztikának létezik Bayes-elvű rokona, számítási igényük jóval nagyobb, ezért csak a gyorsabb számítógépek elterjedésével kezdenek elterjedtebbé válni.



Metaelemzés, metastatisztika, evidencia, metaszignifikancia

A metastatisztikai elemzés kutatásokon átívelő, több kutatás eredményeit összefoglaló vagy azokat éppen megkülönböztető jellegű statisztikai vizsgálódás. Lényeges, hogy figyelembe veszi, a különböző kutatások többnyire más mintaelemszámmal, másféle környezetben, más feltételekkel, esetleg más módszerekkel zajlottak. Egyszerű volna, ha minden kutatás eredeti mintaadatai rendelkezésre állnának, ekkor a hagyományos statisztikai eszközök elegendőek lennének.

A legegyszerűbben felfogható metastatisztikai művelet a számtani középérték kutatásonkénti mintaelemszámokkal súlyozott átlaga, feltéve, ha előbb sikerült meggyőződni arról, hogy a kérdéses változókat kutatásonként egyforma skálával mérték. Jóval összetettebb számításokat igényel viszont például a különböző kutatásokból származó regressziós együtthatók összevetése metaregressziós eljárással. A metastatisztikai elemzést meg kell előznie a szigorúan szabályozott szisztematikus szakirodalom-kutatásnak, ami biztosíthatja az adekvát kutatások elegendő mennyiségű és kiegyensúlyozott figyelembevételét. Fennáll például a publikációs torzítás veszélye, azaz ha bizonyos kutatásokat nem publikálnak, jellemzően azokat, amelyeket a kutatók sikertelennek tekintenek. A publikációs torzítás magállapításához is létezik metastatisztikai eljárás, amely képes hiányzó adatbecsléssel pszeudokutatási eredményeket imputálni, fokozva ezzel a további metastatisztikai elemzések hitelességét.

A publikációs anomáliákat kiküszöbölendő, bizonyos szaklapok csak olyan kutatásokról szóló tanulmányt fogadnak el, amit még a kutatás megkezdése előtt bejelentettek. Egy másfajta támogatási forma a metaelemzésekhez speciális kutatási adatbázisok létrehozása.

A metastatisztikai elemzés része az úgynevezett EBP (Evidence Based Practice)-evidencián alapuló gyakorlatnak. Orvosi területen nemzetközi szervezet alakult az EBP-elemzések összehangolására (Cochrane társaság, Cochrane-adatbázis). Mindezek után meglepő, de a statisztikai szakirodalom nem használja a cikk főcímeként kreált metaszignifikancia kifejezést.

Jól fest az algoritmus

Gustav Klimt a 19–20. századi bécsi szecesszió egyik legfontosabb festőjeként tevékenykedett. Sajnos képeinek körülbelül 20 százaléka elveszett, ezek közé tartozik három monumentális festménye: A filozófia, Az orvostudomány és A jogtudomány. Most azonban az MI és az emberek okos együttműködése életre kelti e három műalkotást.

Az osztrák művészt még 1894-ben kérte fel az Osztrák Oktatásügyi Minisztérium, hogy az egyetem több karát szimbolikusan megjelenítő alkotást hozzon létre a Bécsi Egyetem ünnepi termének hatalmas plafonján. A nem szokványos ábrázolás miatt egyszerre támadták és csodálták az elkészült három hatalmas festményt, amelyek végül a botrányok hatására sohasem kerültek az eredeti helyükre. A magángyűjtők által megvásárolt alkotások a második világháború vérzivataros éveiben a tűz martalékává lettek. A Google Arts & Culture Lab részlege viszont felkarolta az ügyet, és a mesterséges intelligencia, a gépi tanulás és művészettörténeti szakértők segítségével újraalkotta mindhárom festményt.

Először is kikérték Dr. Franz Smola, a bécsi Belvedere Múzeum egyik kurátorának szakvéleményét a színekről, Klimt színhasználatáról – a kutatóknak ugyanis csak a képekről akkoriban készült fekete-fehér fotók állnak rendelkezésükre. Ezután a Google-nél megírtak egy olyan algoritmust, amely Klimt stílusát imitálta. A munka csak ezután kezdődött: az algoritmussal „megetettek" 91 749 festményt, amellyel a program betekintést kapott a festészetbe, a stílusokba. A fotók alapján, emberi beavatkozással és MI gépi tanulással végül elkészült mindhárom kép. A filozófia, Az orvostudomány és A jogtudomány így teljes szépségében megcsodálható.



Forrás:
https://artsandculture.google.com/story/nwWxIZNv_3ZAIA

Komment

Ismét az ember és a gép(i algoritmusok) együttműködésére látunk kiváló példát. Tanulságos az alkalmazás létrehozásának teljes folyamata. Először kipróbáltak több, már létező algoritmust. Az egyiket a rendelkezésre álló 80 Klimt festményen tanították – a 80 kép kevésnek bizonyult a tanuláshoz, a színezés nem lett elég koherens. Aztán egy olyan színező algoritmust alkalmaztak (DeOldify), amely milliós halmazon tanult, ezek a képek viszont valós objektumokból (ember, állat, épület stb.) álltak – a színezés koherens lett, de nem eléggé Klimt. Persze ez várható volt, a festményeken nem feltétlenül valós objektumokat látunk a valós színükkel. A legjobb megoldást végül úgy kapták, hogy egy, a DeOldifyban alkalmazotthoz hasonló Deep Learning algoritmust közel 100 ezer festményen tanítottak, majd utolsó lépésben kifejezetten Klimt festményeire specializáltak. Ez azonban még mindig kevés lett volna szakértői segítség nélkül. Egy fekete-fehér képből egy színes létrehozása még akkor is rengeteg lehetséges végeredményt adhat, ha valamennyire ismerjük Klimt színezési stílusát. Ezen a ponton lépett be a folyamatba a Klimt-szakértő Dr. Smola, aki korabeli újságcikkekből, katalógusokból minden olyan említést összegyűjtött, ami a képek színezésére vonatkozott. Bizonyos szimbólumokat Klimt használt más képeken is körülbelül ugyanebben az időszakban (pl. aranykígyó a hajban). Ilyen esetben azt feltételezték, hogy az arany ugyanazon árnyalata jelent meg a kiszínezendő festményeken is. (Megjegyzem, ez sem gyenge feltételezés: mi van, ha aznap Klimt olyan hangulatban volt, hogy más árnyalatot használt, vagy másfajta festéket kapott a boltban?) Az ehhez hasonló pluszinformációk jelentősen csökkentették a lehetséges színezések számát. Végül az így létrehozott színezésekből azokat tartották meg, amelyekre Dr. Smola is rábólintott. Az eredményt érdemes megnézni, nekem tetszik. Kérdésként felmerülhet: ki lehet-e váltani Dr. Smola szerepét valamilyen gépi megoldással a folyamatban? Részben biztos: információ kinyerése újságcikkekből (szöveganalitika), szimbólumok megfeleltetése egymásnak (képfelismerő algoritmusok); de még mindig ott van az a bizonyos utolsó bólintás.

Hegedüs Pál vezető elemző, Clementine

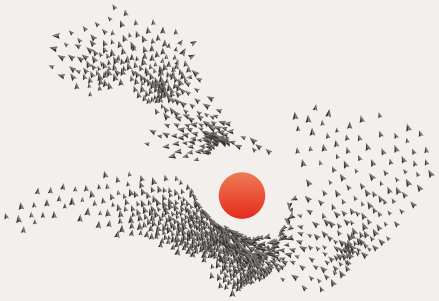
De mit adtak nekünk a méhek?

Állati csapatgondolkodási leckék az MI-hez

Karinthy Frigyes híres, 1917-es elbeszélésében az író a humán gondolkodás egyik negatívumát mutatta be. A Barabbás című novellája szerint ugyanis a Krisztust megfeszítő tömeg egyenként a názáreti Jézust akarja felmenteni, de együtt mégis „Barabbást!" kiáltanak a római helytartó „Kit bocsássak szabadon?" kérdésére.

A modern példabeszéd jól kifejezi, hogy még ha kardinális kérdésekről is van szó, a homo sapiens képtelen együtt, nagy mennyiségű egyedszámban optimális, a faj egészét érintő pozitív döntéseket meghozni. Ha azonban az állatvilágra nézünk, pozitív példák sokaságával találkozhatunk. Ezek közül is kiemelkedik – és mindenki számára érthető analógia – a méhek kaptártudata, a túlélés miatti tökéletes együttműködés. Nem véletlen, hogy pont a méhek inspirálták a néhány évvel ezelőtt létrejött Unanimous AI nevű amerikai cég megalakulását is. A startup célja nem kisebb, mint hogy az emberek együttműködésébe beillesszen olyan mesterséges intelligenciát, amely egyfajta rajt képez a méhekhez hasonlóan a döntési hálózatba bevont emberekből, és irányítja a mechanizmusokat. A technológia az Artificial Swarm Intelligence elnevezést kapta.

Forrás:
https://bigthink.com/the-future/swarm-intelligence-ai-honeybees/



Komment

Méhek világa, érdekes hely. Sir David Frederick Attenborough évtizedek óta mutatja be jobbnál jobb természettudományos ismeretterjesztő filmek garmadáján keresztül, hogy mi történik az állatok világában. Neil deGrasse Tyson az univerzum megértésében segít, és egymással nem egyeztetve a méhek mindennapjairól mindketten a legmélyebb tisztelettel és csodálattal beszélnek.

Odáig is el lehet menni, hogy bizony van a földön olyan faj, amelynek a kommunikációját érdemes lenne megtanulni, mert nem klasszikus nyelvi eszközökkel elképesztően bonyolult információk átadására képes: olyanokénak, amelyeket emberi gondolkodásmóddal folyékonyan nem tudunk megérteni. Komoly barátságra törekedve a méhekkel, pillanatok alatt eljutunk a rajintelligencia fogalmához. Témánkat tekintve azt gondolnák, hogy jó helyen járunk, de egyet még kell ugrani: a homo sapiens nem rajfaj, ezért kell a mesterséges raj is, nem csak az intelligencia.

Az állatvilágnak olyan 250 millió éve volt, hogy a túlélésnek azt az útját járja, amely a speciális kommunikációs képességek, döntési folyamatok kifejlesztésével járt. Azt mondjuk, hogy a méheket vesszük alapul: igen, de az legfeljebb olyan, mint amikor egy film egy könyvet vesz alapul, és ha még a szándék meg is lenne, nem birtokoljuk a méhek könyvét. Mi az esszenciális eltérés? Az Unanimous AI anonimmá teszi az egyén döntését, és rajintelligencia-módszerek segítségével, a humán reakciókat figyelve több döntési alternatíva közül egy döntés felé terel. Kiválóan működik, gyors és hatékony. A klasszikus kiscsoportos csapatmunka döntési hatékonysága fölé emelheti a teljesítményt.

Mi marad ki? A személyes előadás. A meggyőzőképesség. A méhek egyik legnehezebb döntése a költözés. A felderítőik saját tapasztalataik alapján érvelnek a legjobb új otthon helye mellett. A sok felderítő fáradhatatlanul ontja az információkat, amiket a többiek képesek megérteni, és ezek alapján döntenek; együtt. Értik, hogy mi történt. Magyarázható, mint egy döntési fa. Az Unanimous AI szépen vizualizálja, hogy hogyan alakult a döntési folyamat dinamikája, a miérteket még csak nem is az algoritmus homálya fedi, hanem az egyének döntései, amelyek a fejekben maradnak. Addig, amíg ezen a szinten vagyunk, a technológia megmarad kuriózumnak, nem pedig egy óvatosan settenkedő diszruptív erő – igen, rád nézek, kvantum.

Csatlós Béla IT-támogató



EGÉSZSÉGES MEGOLDÁS AZ EGÉSZSÉGMENEDZSMENT RÉSZÉRE

SZÖVEGBÁNYÁSZAT AZ EGÉSZSÉGÜGYBEN

Szerző: Pancza Judit, vezető elemző

Ha egy vállalat kiemelt figyelmet akar fordítani munkavállalói egészségének megőrzésére, elengedhetetlen a múltbeli adatok elemzése, trendek feltárása és értelmezése, ehhez pedig egy olyan adatbázis létrehozására van szükség, amely megbízhatóan és egységesen tartalmazza a korábbi évek egészségügyi adatait.

Az egészségügyi szöveges adatok elemzése talán a legnagyobb kihívás az összes szöveganalitikai feladat közül. Más területeken is jellemzőek a rövidítések, elírások – például a social media vagy a fórumoldalakon folytatott kommunikációban –, de a latin és magyar megnevezések keveredése, vegyes rövidítése és fonetikus leírási formái rendkívüli módon megnehezítik bármilyen algoritmus működését.
A továbbiakban bemutatott projektünk célja egy vállalat 15 év alatt összegyűlt, ömlesztett szöveges formátumban tárolt egészségügyi adatainak konzisztens, elemezhető adatbázisba rendezése. Ez lehetővé teszi eddig rejtett információk, trendek feltárását, ami hatékonyan támogathatja az egészségmenedzsment területét.
Az adatbázis létrehozása itt az esetleírások alapján történő BNO-besorolást (betegségek nemzetközi osztályozása) jelentette, azaz hogy a vizsgálat során készített leírás szövege alapján megállapítottuk, milyen betegségről van szó. A feladatot **szöveganalitikai technológiákra** alapozva végeztük el, elsősorban az SPSS Modeler segítségével.
A szöveges leírások alapján felépítettünk egy egészségügyi szótárat, amely betegségenként tartalmazza azok lehetséges előfordulásait, leírási (elírási) formáit és szinonimáit, például:

- I10H0 – Magasvérnyomás-betegség ← hipertónia, hypertonia, magas vérnyomás stb.
- J0290 – Felső légúti fertőzések ← pharingitis, pharygitis, laringitis, larynghitis, felső légúti fertőzés stb.
- K4090 – Lágyéksérv ← lágyéksérv, hernia ing, hernia inguinalis, inginális hernia stb.

Az elkészült egészségügyi szótár 252 betegséget, 6055 szót és kifejezést, valamit ezekhez több mint 11 ezer szinonimát tartalmaz.

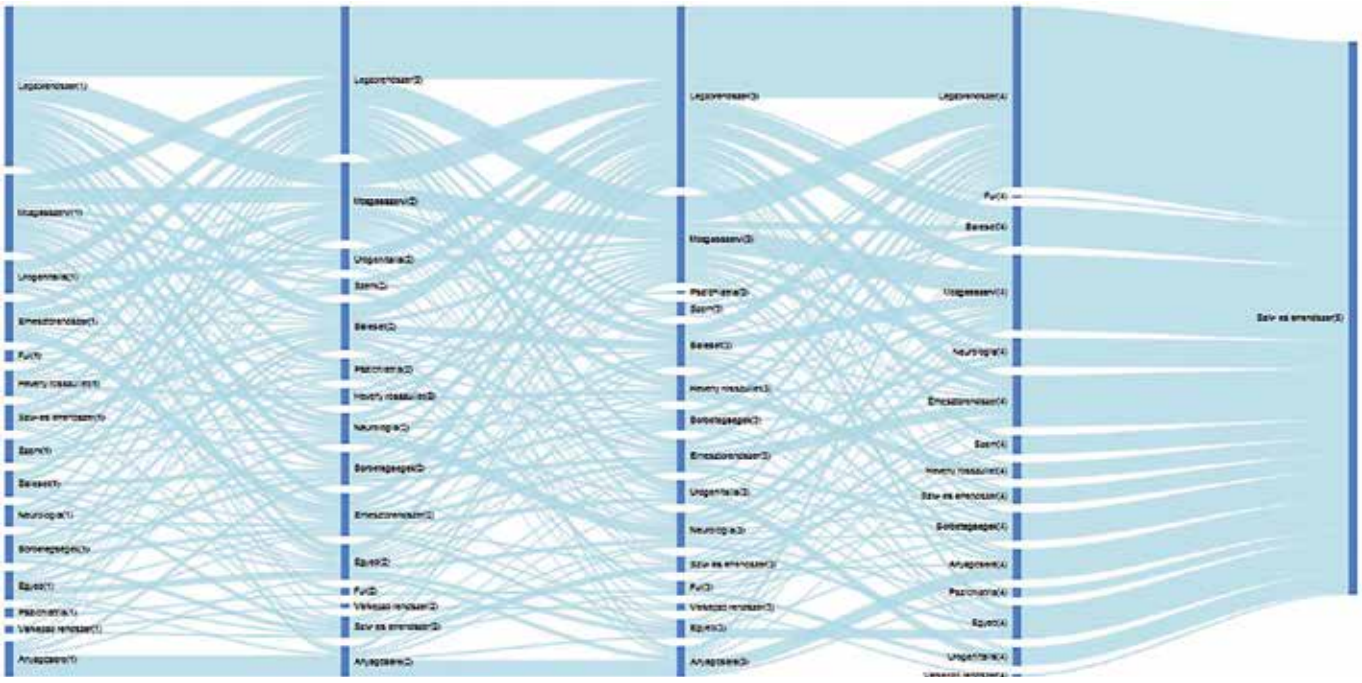
A szótár elkészültével a szöveganalitikai technológia révén lehetővé vált, hogy a szöveges esetleírásokhoz megállapítsuk a megfelelő BNO-kódot. A gépi besorolás egyik hatalmas előnye, hogy megbízhatóan és konzisztensen rendeli azokhoz a szöveg alapján a megfelelő betegséget, a BNO-t, így **ugyanahhoz a diagnózishoz minden esetben ugyanazt a betegséget rendeli.** A másik pedig, hogy az elkészült folyamat bármikor újra felhasználható, de **teljes mértékben automatizálható** is: a szótár alapján akár valós időben képes az újonnan keletkező szövegen elvégezni a megfelelő betegség vagy betegségcsoport hozzárendelését.

A BNO-besorolás helyességét és a szótár tartalmát is két független orvos segítségével ellenőriztük. A validáláshoz leválogattunk egy ezer vizsgálati leírásból álló, véletlenszerűen kiválasztott mintát, amelyet az orvosok manuálisan besoroltak; ezt vetettük össze a gépi eredménnyel. A gépi besorolás az első orvos esetében 96,6%-ban volt helyes, a második orvos esetében pedig 96,7%-ban. Ezek az eredmények alátámasztották módszertanunk és szótárunk megbízhatóságát.

A feldolgozás eredményeképpen létrejött adatbázis már bármilyen statisztikai vagy adatvizualizációs eszközzel tovább elemezhető. Jól látszott például, hogy a leggyakoribb betegségcsoport a szív- és érrendszeri megbetegedések, azon belül is a magasvérnyomás-betegség, amelynek előfordulási aránya főleg a 30-as életkor végével jelentősen nőtt.

Mivel ez volt a leggyakoribb **betegségcsoport, felrajzoltuk az érintett „beteg életútját" is:** azonosítottuk, mikor diagnosztizálták először magasvérnyomás-betegséggel (I10H0), azaz mikor járt ezzel a problémával először vizsgálaton, majd megnéztük, hogy az előző öt vizsgálatán milyen betegséggel kezelték. A megelőző betegség a legtöbb esetben légzőrendszeri volt, ami azzal magyarázható, hogy ez a második leggyakrabban előforduló betegségcsoport. Ami érdekességként kiemelhető, az a mozgásszervi és anyagcsere-betegségek megjelenése a megelőzők között.

A betegségutak leírása a prevencióban jelent hatékony segítséget: célzott programok indíthatók a szív- és érrendszeri megbetegedések korai megelőzésére. A további statisztikai elemzések segítségével pedig azonosíthatók a veszélyeztetett személyek, csoportok vagy az egyes munkakörök rizikófaktorai is. De nemcsak munkavállalói oldalon hasznosíthatók az eredmények: az ellátási trendek elemzésével megalapozottan tervezhető az egészségmenedzsment területe is.



1. ábra: Magasvérnyomás-betegek megelőző betegségei

DR. MI RENDEL

MI AZ ORVOSTUDOMÁNYBAN ÉS AZ EGÉSZSÉGÜGYBEN

Szerző: Kánai András, Jövőkutató



Évről évre rengeteg, valódi áttöréssel kecsegtető felfedezést jelentenek be az orvostudományban. Finomodnak a módszerek és az eljárások is. A heurékapillanatok nagy része mögött emberi leleményességet, kísérletezőkedvet, bátorságot és jól működő technológiai apparátust találunk. Ma már nem számít ritkának a mesterséges intelligencia használata a gyógyításban, a megelőzésben és a kutatásban sem. Sőt, épp az ellenkezője felé tartunk: az MI térhódítása az olcsóbbá és gyorsabbá váló informatikai háttér, a számítási kapacitás növekedése miatt feltartóztathatatlannak látszik. Felsejlik előttünk egy sokkal hatékonyabb alapokon működő egészségügy körvonala is. Ebben a cikkünkben néhány példát mutatunk a lelkesedésre okot adó trendekre.

A bioforradalom küszöbén

Minden, MI-vel és orvoslással kapcsolatos hír közül az egyik legfigyelemreméltóbb a világ legnagyobb keresőóriásához kötődik. 2020 novemberében ugyanis hatalmas eredményt jelentettek be a Google anyavállalatához tartozó Deepmind szakemberei. Az általuk kifejlesztett AlphaFold MI-algoritmus képes lett arra, hogy megjósolja az élet építőkockáinak tartott proteinek háromdimenziós alakba rendeződését. Mivel egy protein alakja meghatározza a funkcióit is, ennek ismeretében a kutatók már pontosabb képet kaphatnak arról, hogy mire is jó az adott protein. (Az inzulinnak a cukorszint-szabályozásban vagy az antitesteknek a koronavírusok elleni védekezésében betöltött szerepe is nagyban függ a protein formájától.) Ez a mai számítástechnikai kapacitások és a jól megírt MI nélkül eddig lehetetlen vállalkozásnak tűnt.

A szakértők bioinformatikai forradalomról, a személyre szabott gyógyszergyártás hajnaláról beszélnek. Eközben a befektetők is letették a voksukat: 2020-ban az MI-vel támogatott molekuláris és gyógyszerkutatásokba minden eddiginél több invesztíció, 13,8 milliárd dollárnyi tőke érkezett.

Kérdezze meg (mesterséges) orvosát, gyógyszerészét!

A nyugati világ elöregedése annyi pácienst eredményez, hogy a közeljövőben a diagnosztikát is át kell gondolni. Az egyik megoldás, hogy ki se mozdulunk otthonról. A szingapúri Veredus Laboratories eszköze, a VerePLEX™ Biosystem például többféle fertőzés házi tesztelését teszi lehetővé. Ebben annak speciálisan erre a célra kifejlesztett chipje és kifinomult algoritmusa, képfeldolgozó képessége a kulcs. Az orvos- és nővérhiány leküzdésére a másik hatásos eszköz a gépi tanulásra épített képfeldolgozás. Már több hír is szólt például arról, hogy egyes daganatok korai felismerésében az erre megtanított algoritmus kevesebbet hibázik, mint az ember – nem fárad el a szeme és az agya. Ráadásul minél több adatot kap felcímkézve (például egy CT „tumor” vagy „nem tumor”), annál jobb döntést hoz.

Gyógyszerfejlesztés felgyorsítva

Az emberiség nem szakmabeli része örömmel, megdöbbenéssel vagy kétkedéssel fogadta a koronavírus elleni vakcinák viszonylag gyors előállítását. Pedig nincs sem csoda, sem összeesküvés a háttérben: már az 1990-es évek végén óriási molekula-adatbázisokkal rendelkeztünk. Az MI gépi tanulás módszerével a már többszázmilliósra nőtt adatbázis modellezése felgyorsítja az eddig évekig tartó, dollármilliárdokba és gyakori kudarcokba kerülő folyamatot. A mesterséges intelligencián és a gépi tanuláson alapuló megközelítések alapvető szerepet játszhatnak a továbbiakban, hiszen a költség-, idő- és eszközigényes laborkísérletek helyett a sokkal gyorsabb számítógépes modelleket használhatjuk.



Személyre szabott géntechnológia

Az emberi gén szerkesztése, manipulálása hihetetlen távlatokat nyit. A veszély abban rejlik, hogy az ebben élvonalbeli CRISPR-technológiában tisztában legyünk a nem várt mellékhatásokkal, és a megfelelő RNS-t használjuk. A legkevésbé veszélyes mellékhatásokkal rendelkező RNS gondos kiválasztása a CRISPR-rendszer alkalmazásának egyik fő szűk keresztmetszete. A gépi tanulási modellek bizonyítottan a legjobb eredményeket adják, hogy ezt a problémát minimálisra csökkenthessük.

MOSOLYGÓS, SZOMORÚ VAGY BOSSZÚS VAGY?

KOMMUNIKÁCIÓ HANGULATJELEKKEL AZ ONLINE TÉRBEN (RÉSZLET EGY KUTATÁSBÓL)

Szerző: Kelemen Sára Bernadett, Junior elemző



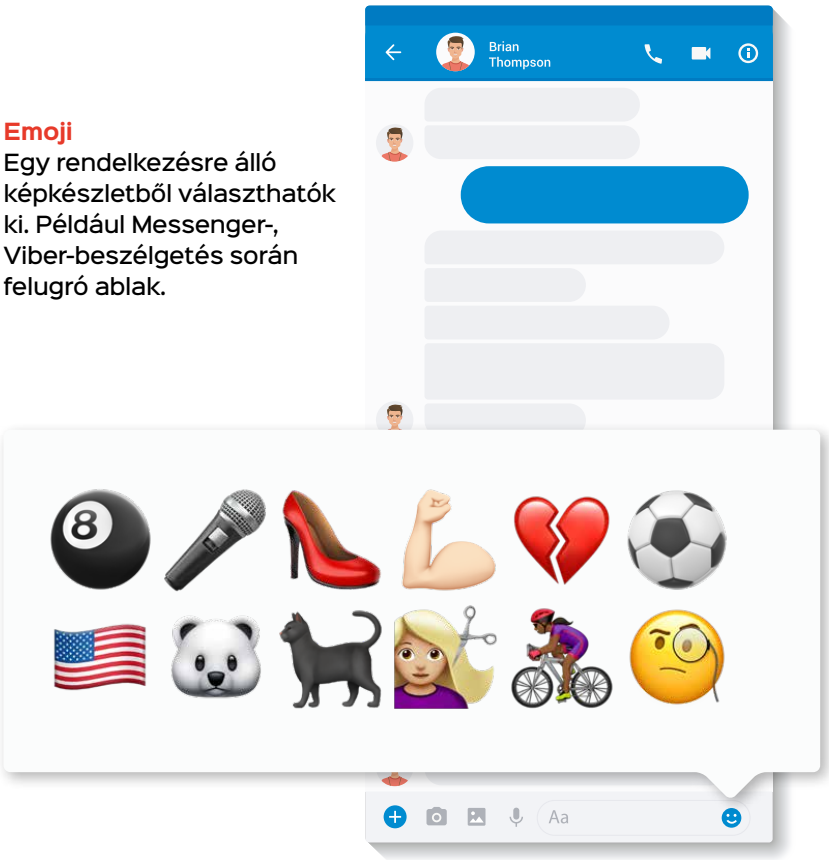
Egyre népszerűbbé válik a közösségi oldalak felhasználói által hagyott digitális lábnyomok elemzése, hiszen egyszerűen elérhető, hatalmas adatmennyiségről van szó. Napjainkban a Twitter- vagy akár Facebook-felhasználók posztjaiból kiinduló, az ő nyelvhasználati szokásaikat kutató publikációk száma is megnövekedett. Cikkünk is egy, az online térben történő nyelvhasználatra vonatkozó kutatás, amelynek fókusza, hogy milyen nemi különbségek tapasztalhatók a hangulatjelek használata terén a vizsgált szövegekben.

Chateléskor, blogoláskor, kommentek vagy posztok írásakor olyan dolgokat is karakterek által kell kifejeznünk, amit szemtől szembeni kommunikáció esetén nem tennénk. A gesztus és a mimika hiánya ösztönözte a ma már mindennapjaink szerves részévé vált emotikonok és emojik megjelenését.

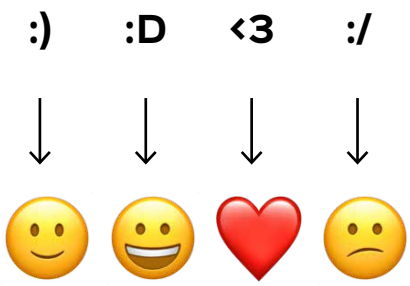
A szövegbe ágyazott hangulatjelek kétféle módon jöhetnek létre: egy részük megfeleltethető billentyűkombinációknak, míg vannak, amelyeket a rendelkezésre álló képkészletből választhat ki a felhasználó. Emotikonoknak nevezzük azokat, amelyeket a billentyűzet segítésével, konkrét karakterek bevitelével rögzítünk, majd ezeket a használt eszköz vagy alkalmazás képpé alakítja. A legelterjedtebb emotikon például a kettőspont bezáró zárójellel, amit ma már szinte minden felület mosolygóvá alakít. Az emojik pedig azok a szövegbe ágyazott apró képek, amelyeket egy rendelkezésünkre álló képkészletből, általában egy felugró ablakból választunk ki. Ezek készlete általában igen sokszínű, túlmutat a konkrét, arcot ábrázoló hangulatjeleken. Ide tartoznak a különféle ételek, tevékenységek, zászlók, sportok képi megjelenítései.

Az online nyelvhasználat kutatása szinte egyidős az online szöveges fórumok létrejöttével. A témában fellelhető legkorábbi publikációk közel húsz évvel ezelőtt születtek. Huffaker és Calvert a *Journal of Computer-Mediated Communication* nevű folyóiratban 2005-ben közölt kutatásukban[1] tinédzserek webblogjain keresztül vizsgálták a serdülők nyelvhasználatát, érzelmi kódokkal történő önkifejezésüket. Tanulmányukban elemezték az emotikonok használatát is. Megállapításaik közül az egyik legáltalánosabb, miszerint az érzelmi kódok, emotikonok alkalmazása jobb benyomást tesz a szerzőről a társakra, az olvasókra. Tehát a vizsgált tinédzserek esetén pozitívabb megítélést kapott az a személy, aki az online térben közzétett szövegét érzelmi, hangulati jelekkel egészítette ki. A nők és a férfiak kommunikációs szokásait elemezve a szerzők megfogalmazták, hogy a tinédzser fiúk

Emoji
Egy rendelkezésre álló képkészletből választhatók ki. Például Messenger-, Viber-beszélgetés során felugró ablak.



Emotikon
Billentyűkombinációknak feleltethetők meg.
A felhasznált eszköz/ alkalmazás képpé alakítja őket.



[1] *Huffaker, D. & Calvert S. (2005). Gender, Identity, and Language Use in Teenage Blogs. Journal of Computer-Mediated Communication, 10/2.*

általában direktebb és erőteljesebb stílust, míg a lányok indirektebb, intimebb interakciós kifejezésmódot használnak. Ez megfeleltethető az Eckert & McConnell-Ginet 2003-as írásában[2] olvasható megfigyeléssel, miszerint a nők online bejegyzéseik esetében gyakran olyan nyelvi stílussal élnek, amely azt sugallja, hogy félénkek, kevesebb önbizalommal rendelkeznek.

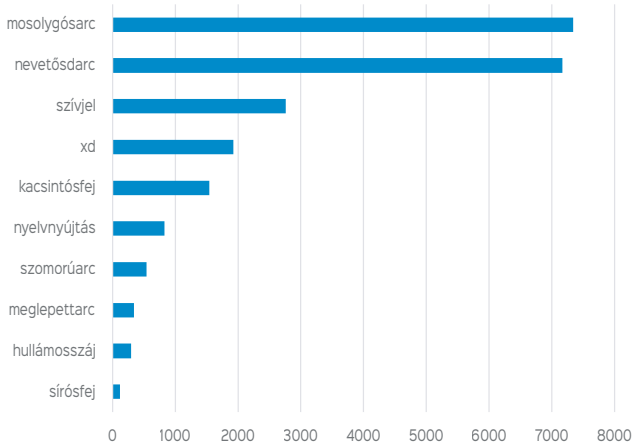
Lee 2003-ban publikált kutatása[3] az online szövegekben használt grafikus elemekre vonatkozik. Eredményei alapján a nők több grafikus megjelenítést, emotikont használtak érzelmeik kifejezésére, mint a férfiak. Ez egybevág azzal a hétköznapi feltételezéssel, megfigyeléssel, miszerint a nők általában a mindennapi életben is szívesebben kommunikálják az érzelmeiket, mint a férfiak.

Ennél érdekesebb az a kutatási vonal, amelyen Lee tovább vizsgálta az érzelmek online térben zajló kommunikációját. Nem csupán azt nézte meg, hogy ki a szöveg szerzője, hanem a fogadó fél nemét is figyelembe vette az elemzésben. Így vizsgálta a férfi–férfi, a férfi–nő, valamint a nő–férfi, nő–nő kommunikációs irányokat. A nők esetében az emotikonok használatának gyakoriságában nincs kimutatható különbség az alapján, hogy nő vagy férfi a fogadó fél. A nők általában használják a szövegközi érzelmi kódokat, akár férfihez, akár nőhöz szólnak. Ez nem így van a férfiak esetében. Eredményei szerint a férfi–férfi kommunikáció során kifejezetten ritka az emotikonok használata, ám a férfi–nő kommunikáció alkalmával már lényegesen gyakoribb. Érdekes kérdés, hogy vajon a férfiak a nőkkel való kommunikáció során automatikusan alkalmazkodnak a nők emotikonos stílusához, vagy ez tudatos döntés részükről, azaz megtanulták, hogy a nőkkel való interakció sikeresebb, ha érzelmeiket direktebb módon is kifejezik.

A szakirodalmi kitekintés után néhány saját kutatási eredményt mutatnék be a témához kapcsolódóan. Facebook-posztokat elemezve arra is választ szerettem volna kapni, hogy melyek a leggyakrabban használt emotikonok és emojik. Milyen arányban használják ezeket a férfiak, és milyen arányban a nők?

A kutatáshoz 147 magyar Facebook-felhasználó posztjait használtam fel, amelyeket egy Magyarországon zajló 2019-es kutatás során gyűjtöttek össze. Az elemzéshez felhasznált 149 471 digitális lábnyom a résztvevők Facebook-használatának teljes időtartamát lefedik, azaz a regisztrációjuktól kezdve a letöltés pillanatáig a felületen végrehajtott összes aktivitásukat, beleértve azokat is, amiket később eltávolítottak. Néhány esetben ez több mint tízévnyi adatot jelent. A mintát 110 nő és 37 férfi tette ki. Az átlagos életkor 30 év, a legfiatalabb egyén 18 éves, a legidősebb pedig 71 éves.

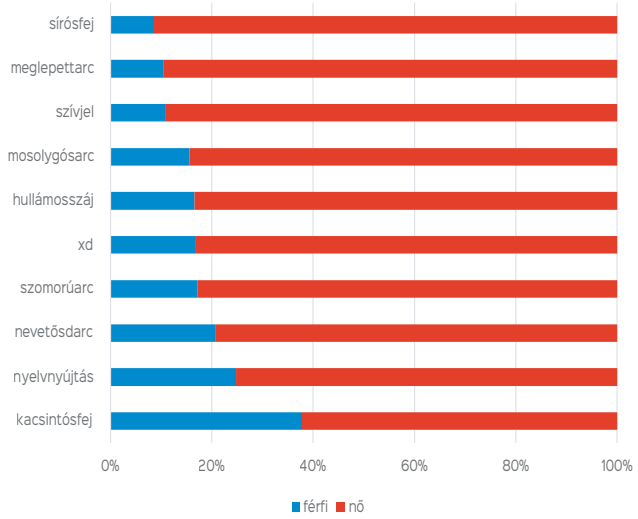
A szöveganalitikai elemzések elvégzésére igen jól használható a Python nevű, nyílt forráskodú programozási nyelv, de az alapvetően angol nyelvű szövegekre fókuszál. Ezzel szemben a Clementine-nál használt, az SPSS Modeler szoftverbe ágyazott Clemtext kifejezetten a magyar nyelvű szövegek elemzésére készült. A szövegközi hangulatjelek elemzésére tökéletesen alkalmasak a szöveganalitikai módszerek, mint például a Modelerben is elérhető TLA-módszer, csupán szükséges a vizs-



Emotikonok megjelenésének gyakorisága a vizsgált posztokban

[2] *Eckert, P., & McConnell-Ginet, S. (2003). Linking the linguistic to the social. In P. Eckert, & S. McConnell-Ginnet, Language and Gender. Cambridge: Cambridge University Press.*
[3] *Lee, C. (2003). How Does Instant Messaging Affect Interaction Between the Genders? Stanford, CA: The Mercury Project for Instant Messaging Studies at Stanford University.*

gálandó hangulatjelek szövegszerű előkészítése. Az általam vizsgált posztok esetében emotikonok terén közel 30 különböző előfordulásról lehetett beszélni. A felhasználók sok esetben egy „eredeti” emotikonnak rengeteg más, kissé eltérő alakját alkalmazták. („Eredeti”: ismétlődések, felesleges szóközök és plusz karakterek nélküli megfelelője egy-egy hangulatjelnek). Emojikból rengetegfélét használtak a mintában szereplő felhasználók.



Emotikonok megjelenésének gyakorisága nemek szerint

Az emotikon- és emojihasználati tendenciák alapján az adatbázisban szereplő felhasználókra az emotikonok nagyobb számú használata, ellenben az emojik változatosabb alkalmazása volt jellemző. A leggyakoribb emotikonok a derűs hangulat kifejezésére szolgáltak. Ezek közül is a legnépszerűbbek a „mosolygós arc”-ként 🙂 és „nevetős arc”-ként 😀 kódolt érzelmet kifejező ábrák voltak, amelyek együttesen az emotikonokat tartalmazó posztok 30 százalékában fordultak elő. E kettő mellett gyakori volt még az „xD”, és a 😉 „kacsintós fej” alkalmazása. Emojik esetében a halmozás gyakrabban megfigyelhető esemény volt, azaz amikor egy bizonyos poszton belül egymás után több különböző emojit alkalmazott az adott felhasználó. A leggyakrabban használt emoji aránya meg sem közelítette a leggyakrabban használt emotikonét. De gyakran alkalmazták a „nevetős arc”, 😀 „szívecskés szem” 😍 és „kacsintós fej” 😉 hangulatjeleket. Az egy bejegyzésen belüli emojicsoportosítás inkább a nők esetében volt megfigyelhető. Leginkább az egyes ruhadarabokat 👗, körmöst 💅, fodrászt 💇 szimbolizáló kis képek követték egymást, de a jókedvet kifejező arcok, a szeretetet kifejező szívjelek halmozása is sok esetben megfigyelhető volt. A férfiak esetében pedig inkább az olyan, rájuk talán jobban jellemző emojik voltak megfigyelhetők, mint a söröskorsó 🍺 vagy az autó 🚗.

Bármi is áll a fókuszban, akár az érzelmi tartalom, akár más tényező, fontos szem előtt tartani a kutatások lehetséges korlátait. Elképzelhető, hogy az aktuálisan vizsgált emberek más nemi szerepekkel, más kommunikációs stílussal rendelkeznek, mint az előttük vizsgált generációk. Az internetes nyelvhasználat folyamatosan változik. Napjainkra már 3633 különböző emojit különböztetünk meg, a kutatás kezdetekor ez a szám jóval alacsonyabb volt, a 2010-es évek elején még az ezret sem érte el. Számukkal együtt a használatuk gyakorisága is folyamatosan nő. Kutatásom idején szinte elképzelhetetlennek tűnt, hogy a privát megnyilvánulásokon kívül mást vizsgáljak, például munkahelyi e-maileket. Mostanra már egy ilyen kutatást sem tartok annyira távolinak, talán pont errefelé tart a világ.



DATA FABRIC KONCEPCIÓ MINT AZ ADATVEZÉRELT VÁLLALAT ALAPKÖVE



Szerző: Huszti Dániel, Data & AI Technical Sales IBM

Az adatvezérelt vállalat olyan, mint egy szuperautó, amelyet mindenki szeretne birtokolni, viszont csak keveseknek adatik meg. Amikor ilyen vállalatokra gondolunk, rögtön olyan sikertörténetek jutnak eszünkbe, mint a Netflix, az Amazon, a Facebook és az Uber. Amennyiben kicsit jobban a siker mélyére tekintünk, láthatjuk: ezek a cégek hihetetlen mennyiségű pénzt költöttek a legjobb technikai szakértők alkalmazására, és hosszú éveken át mindent megtettek az ügyféladatok gyűjtéséért, tisztították azokat, és az alapoktól kiindulva fejlesztették szolgáltatásaikat.

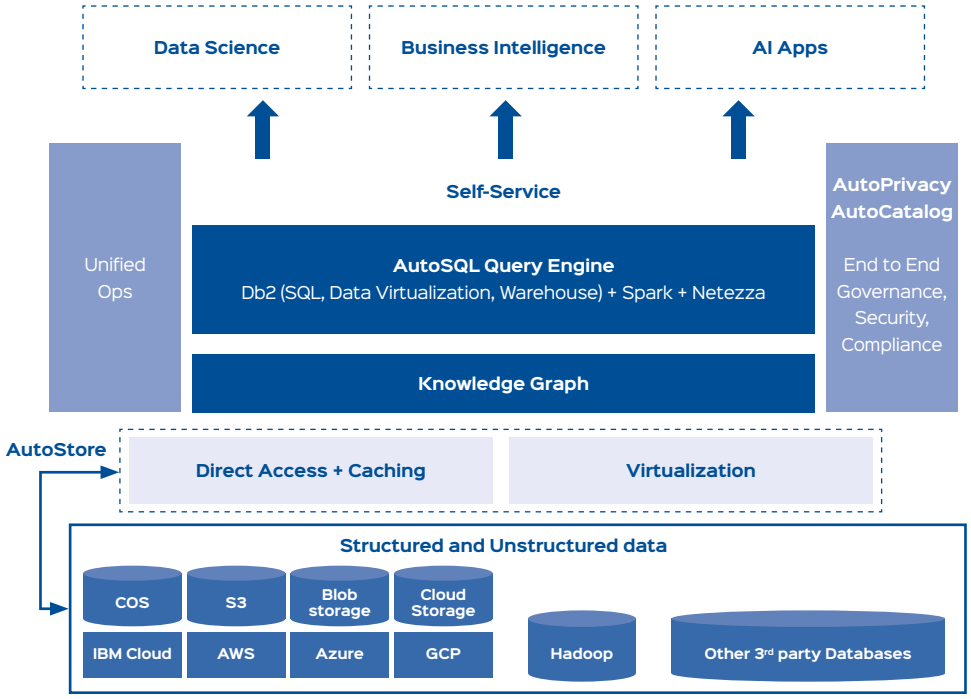
A klasszikus modellben működő vállalatok számára ez a megközelítés nehezen megvalósítható, mivel a stabil alapok és az üzleti működési modellek már rendelkezésre állnak. Egy bank, biztosító rendelkezik termékekkel és ehhez kapcsolódóan nagy mennyiségű tranzakciós, kártérítési, csalási és egyéb, üzletileg releváns adatokkal. Jól kialakított hűségkártyaprogramja révén egy kiskereskedelmi vállalat szintén nagy mennyiségű ügyféladat birtokában áll. Utóbbiaknak fontos lenne arra törekedniük, hogy a meglévő kis létszámú adatelemzőn felül minél több, kiváló üzleti tudással rendelkező szakember nyisson az adatelemzési eszközök felhasználásának irányába, ezáltal csapatsporttá varázsolva az adatelemzést, megágyazva az adatvezérelt kultúrának.

Amennyiben az üzleti felhasználók a teljes adatelemzési folyamatot az adatok előkészítésétől, tisztításától kezdve az elemzések elvégzéséig egy egységes platformon tudják megvalósítani, akkor teljes körű governance, azaz felügyelet biztosítható az adatok és folyamatok felett. Az IBM Data Fabric koncepciója ebben tud segítséget nyújtani.

Új Data Fabric-koncepció

A Data Fabric egy adatmenedzsmentarchitektúra-koncepció, amely heterogén adatforrásokból (akár hibrid vagy multi-cloud környezetben) álló elemzések hatékony és felügyelt elvégzését támogatja. Segítségével az önkiszolgáló és tradicionális üzleti elemzésekhez, mesterségesintelligencia-modell építéséhez szükséges adatok egy helyen elérhetők, automatizáltan elemzésre előkészíthetők, és különféle adatelemzési szolgáltatások (adatbányászat, BI vagy mesterséges intelligencia) számára továbbíthatók.

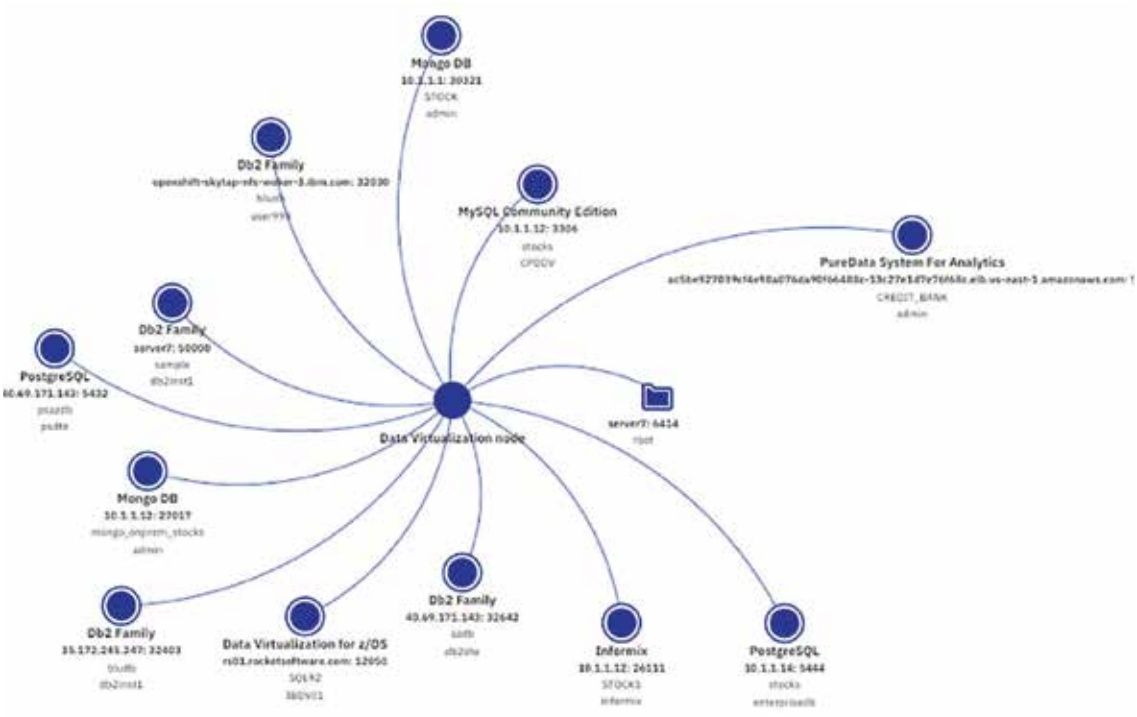
Az IBM Data Fabric koncepciója az IBM Cloud Pak for Data integrált adatelemzési platformjára épül, amely széles körű végfelhasználói réteg számára képes egységes, szolgáltatásorientált platformot biztosítani. A platform ezeket a folyamatokat MI alapú automatizmusokkal képes megtámogatni, hogy az adatok üzleti fogalmak alapján is kereshetővé váljanak, több heterogén adatforrást egységesen lehessen kezelni, elemzések készítésére alkalmas legyen, és mindemellett az adatok védelmére is nagy hangsúlyt fektet. A megoldás lokálisan, privát vagy akár publikus felhőben szolgáltatásként is elérhető.



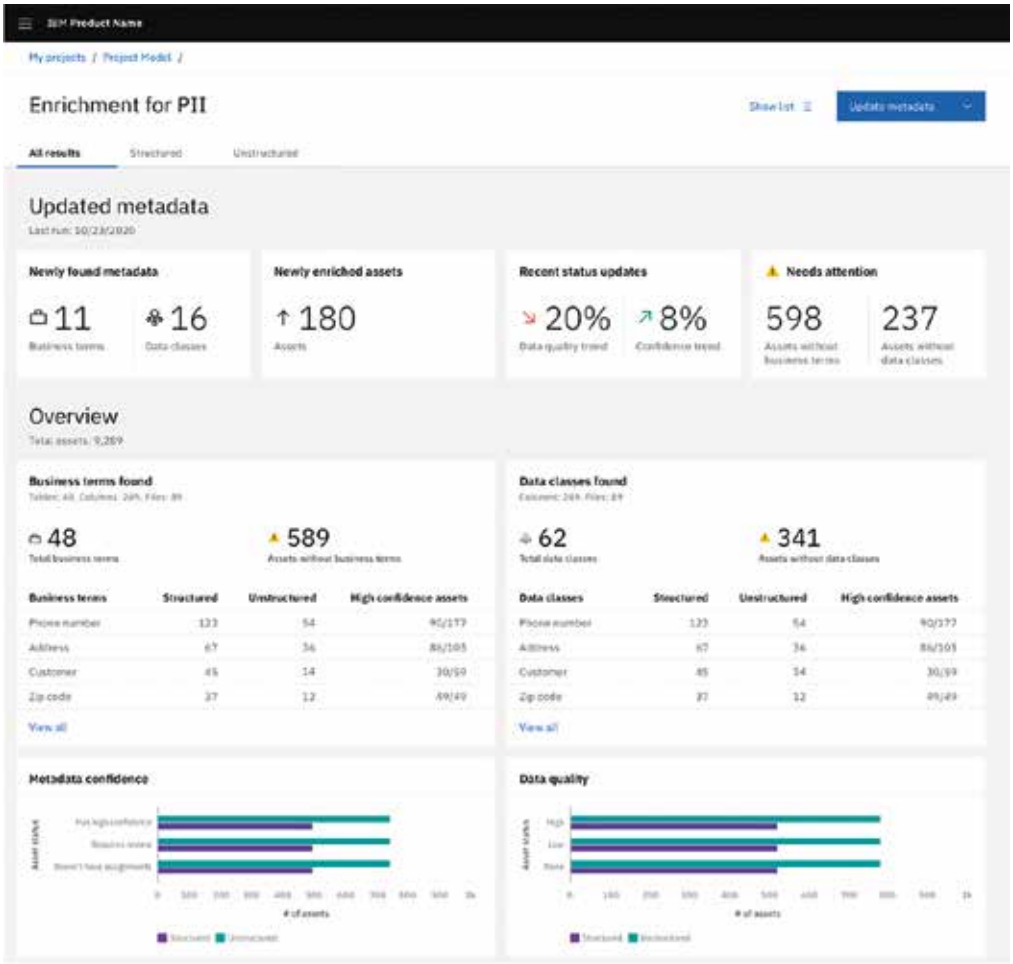
Data Fabric-koncepció az IBM Cloud Pak for Data platformon

Az IBM Cloud Pak for Data platform mesterséges-intelligencia-támogatással segít az üzleti adatok automatizált előkészítésében, így az üzleti felhasználóknak csak az elemzéssel szükséges foglalkozniuk:

- **Adathozzáférhetőség és felhasználhatóság:** A Data Virtualization (AutoSQL) szolgáltatása egy olyan univerzális lekérdező motor, amely több heterogén adatforrás logikailag egységes kezelését biztosítja adatmozgatás vagy replikáció nélkül. Automatizálja az adatok elérésének, frissítésének és egyesítésének módját bármilyen adatforráson (adatbázis-kezelők, felhős tárolók, adattárházak, data lake-ek, Excel- és egyéb, nem strukturált adatok), fejlett cache menedzsment képessége révén minimalizálja az adatmozgatást.



Logikailag összekapcsolt heterogén adatforrások



Üzleti adatok klasszifikálása és az adatminőség-vizsgálat áttekintése

- **Üzleti adatok katalogizálása:** A Watson Knowledge Catalog (AutoCatalog) szolgáltatása automatizált adatklasszifikáció révén az adatokhoz üzleti fogalmakat képes rendelni, így felépítve és katalógusba rendszerezve a vállalati adatvagyont. Az üzleti felhasználók ezáltal az üzleti fogalomra való keresés révén egyszerűen megtalálhatják az elemzéshez szükséges adatokat, így nagymértékben csökkenthető az adathoz elérés ideje, növelhető az elemzési kedv, a hatékonyság, és egyaránt biztosítható a törvényi megfelelőség (pl. GDPR). Az adatvagyon-katalógus ezenfelül képes támogatni az adatok automatizált minőség-ellenőrzését, jelezve az adatminőségi problémákat, ezáltal garantálva az üzleti felhasználók hatékony munkavégzését.
- **Adatbetöltés és -tisztítás:** A DataStage adatbetöltő és Match360 MI-algoritmusokkal támogatott törzsadatkezelési szolgáltatás segít az üzletileg kritikus adatok előkészítésében, tisztításában, legyen szó akár önkiszolgáló vagy tradicionális elemzésről.
- **Adatvédelem és biztonság:** A Watson Knowledge Catalog (AutoPrivacy) szolgáltatása mesterségesintelligencia-támogatás révén felismeri az érzékeny, személyes adatokat, és az illetéktelenek számára automatikusan elmaszkolja azokat, így garantálva megfelelő védelmét.

Data Fabric kihasználása az elemzés során

Az IBM Cloud Pak for Data az imént bemutatott képességeken felül több olyan üzleti elemzés készítésre alkalmas integrált szolgáltatással rendelkezik, mint:

- **Watson Studio:** Teljes körű MI-életciklus támogatása, amely az open source adatelemzési eszközökön (Python, R, Spark) felül grafikus (SPSS Modeler) és automatizált modellépítési (AutoAI) képességei révén széles körű elemzői felhasználást tesz lehetővé.



- **Cognos Analytics** és **Planning Analytics:** Fejlett vizualizációs és tervezési képességek.
- **Watson Assistant** és **Discovery:** Virtuális-asszisztens építése és fejlett szövegelemzési, keresési képességek.

Az előkészített adatok bármilyen külső adatelemzőn keresztül is tovább analizálhatók. Ilyenkor a több adatforrást átölelő, logikailag egységes adatok egy JDBC/ODBC-interfészen keresztül integrálthatók a meglévő informatikai megoldásokhoz.

Esettanulmányok

Az IBM Cloud Pak for Data platform előnye a teljesen integrált szolgáltatásokban rejlik, ezért több különféle üzleti problémára nyújt megoldást:

- Adatvagyon-gazdálkodás bevezetése révén folyamatok felügyelete és törvényi szabályozási megfelelőség támogatása
- Önkiszolgáló és tradicionális adat-előkészítés, adattisztítás
- Önkiszolgáló és tradicionális elemzések támogatása (pl. ad-hoc elemzés vagy BI-riport), akár több heterogén forrásrendszer együttes használatával és automatikus szenzitív adatmaszkolással
- Statisztikai elemzések és MI-modellek építése
- Üzleti felhasználók számára a MI alapú döntések elmagyarázása, ezáltal növelve a technológiába vetett hitet
- SAS-modellek kiváltása, költségcsökkentés
- Szöveganalitikai és virtuális asszisztens által megvalósított külső (ügyfélszolgálati asszisztens) és belső folyamatautomatizációja (felhasználóknak MI által támogatott tudástár)

Az integrált platform széles körű eszköztára mellett teljes körű governance-t biztosít a teljes adatelemzési ciklusra, nagyban elősegíti az adatvezérelt vállalattá válási folyamatot.

További információk

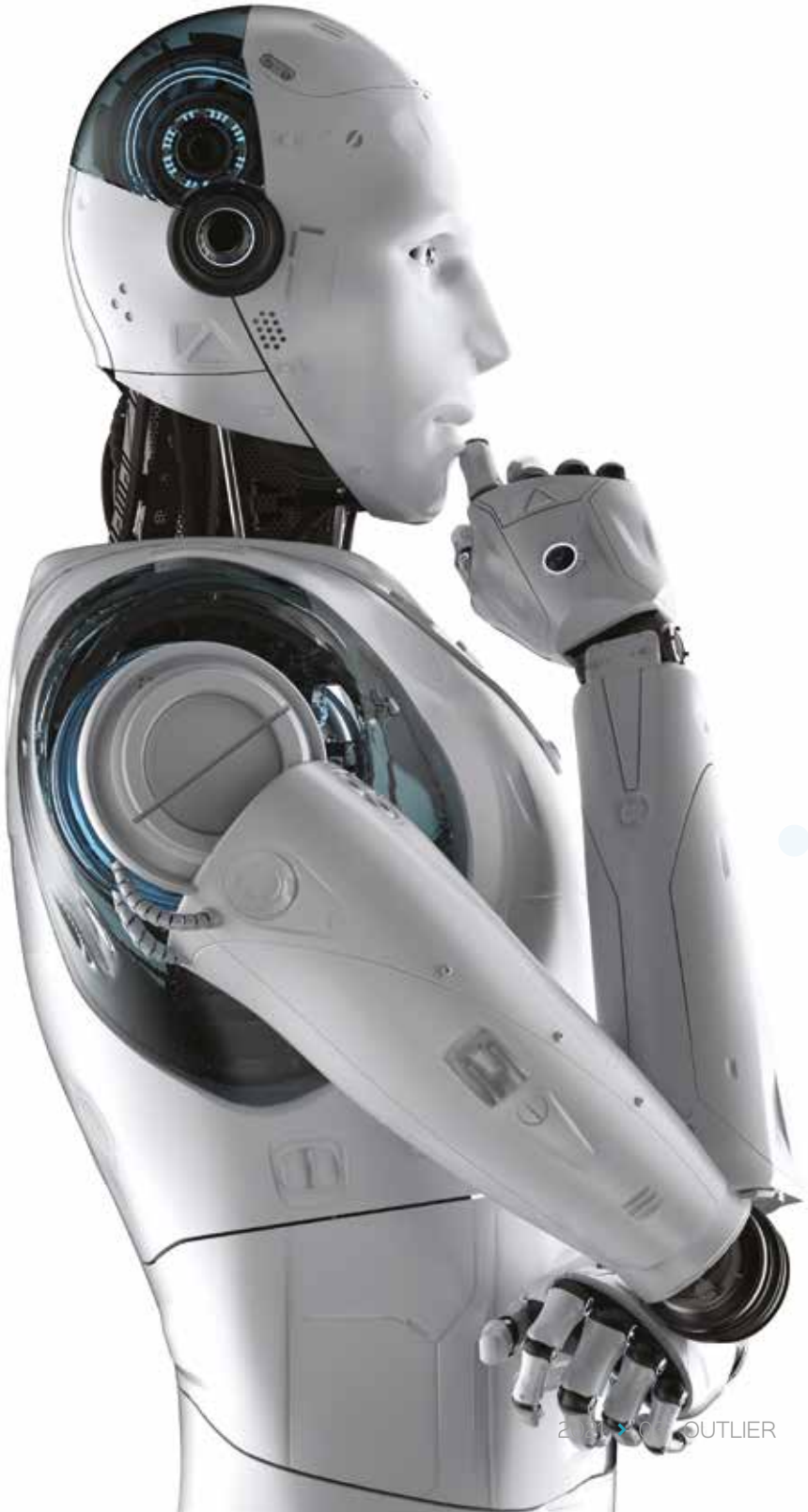
Az imént röviden bemutatott IBM-eszközkészlet hatékonyan képes támogatni az adatvezérelt vállalatok alapjainak letételét. A szolgáltatások IBM Cloud Pak for Data platformon lokálisan, privát felhőből vagy felhős szolgáltatásként egyaránt elérhetők.
További információkért keresse fel az IBM-et, vagy forduljon a Clementine-hoz!



Forrás:
https://www.ibm.com/analytics/data-fabric

MILYEN ÉRDEKESSÉGEKET OLVASOTT A MAGAZINBAN?

Tegye próbára emlékezetét, vagy lapozzon vissza a nemrég olvasott cikkekhez, és keresse meg a válaszokat az Outlier magazin mostani számában! Kérjük, ékezetek nélkül oldja meg a feladványt, a rejtvény megfejtését, ami a Clementine egyik megoldása és amit kék háttérszínnel jelöltünk, küldje el az info@clementine.hu címre. A helyes megfejtést beküldő első öt olvasónknak ajándékcsomaggal kedveskedünk. Kellemes időtöltést kívánunk!



Vízszintes

3 Hangulatjelek, amelyeket a szövegben billentyűzet segítségével, adott karakterkombinációval rögzítünk.

6 Az Egészségügyi Világszervezet angol nyelvű rövidítése.

8 Ezen változó szerinti szegmensekre olvashatunk elemzést a nők daganatos megbetegedéseire vonatkozóan.

9 Szövegbe ágyazott, hangulatot, szándékot, tevékenységet kifejező képek, amelyeket egy rendelkezésre álló képkészletből kiválasztva szúrunk be a szövegbe.

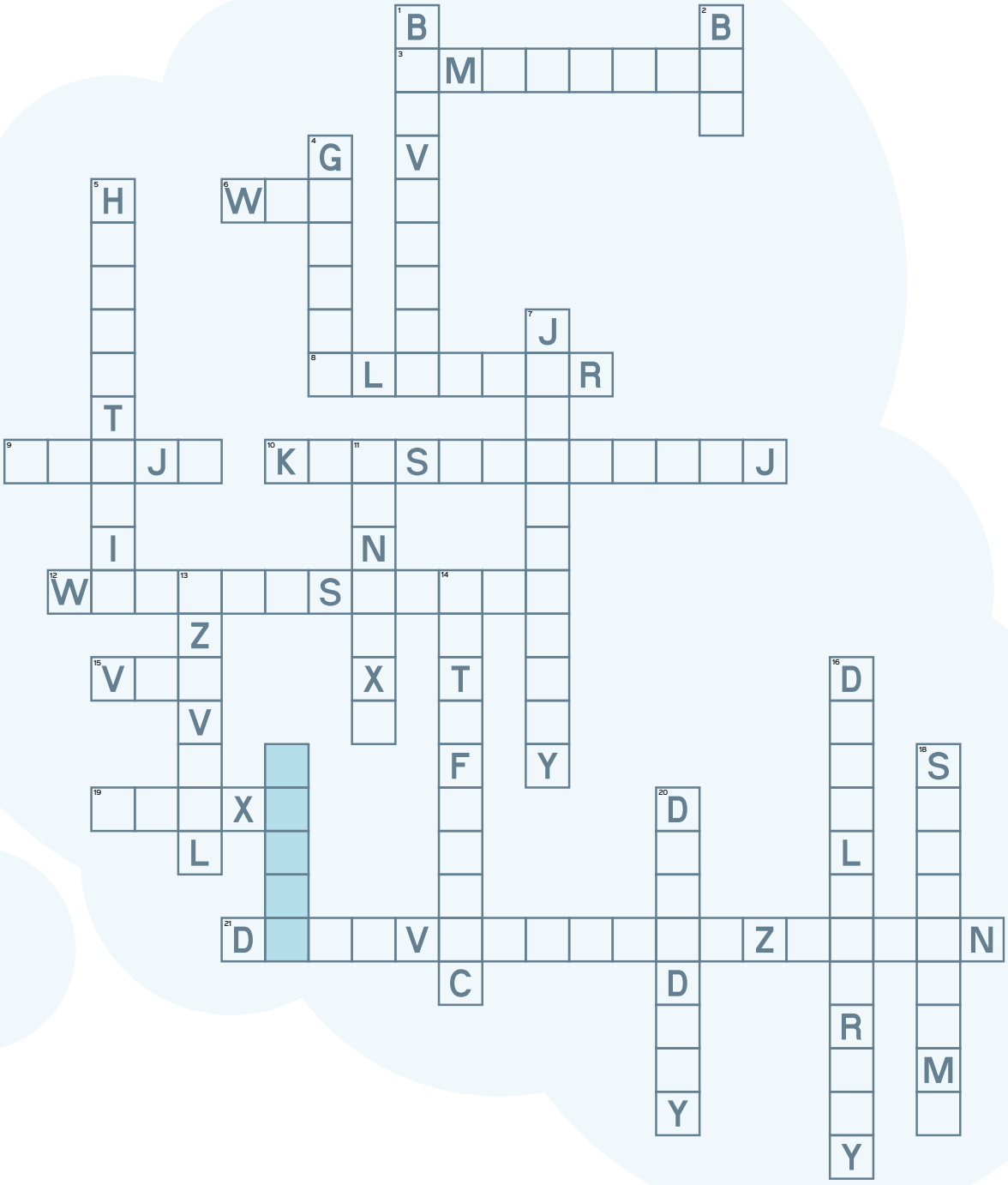
10 A bemutatott kutatás szerint mely hangulatjelet használják legnagyobb arányban a férfiak?

12 Az IBM Cloud Pak for Data része, amely szolgáltatás széles körű adatelemzői felhasználást tesz lehetővé.

15 Rövidítés, Voice User Interface.

19 Az Amazon cég hangfelismeréssel felruházott intelligens asszisztense.

21 Az IBM szolgáltatása, amely az adatforrások egységes kezelését adja adatmozgatás és replikáció nélkül.



Függőleges

1 Dr. Franz Smola professzor ennek a múzeumnak a kurátora.
2 Rövidítés a betegségek nemzetközi osztályozására.
4 Melyik nagy cég állt a Gustav Klimt festmények újraalkotása mögé?
5 A magasvérnyomás-betegség latin megnevezése.
7 Az egykori tervek szerint Gustav Klimt három festménye került volna a Bécsi Egyetem ünnepi termének plafonjára. Ezek címe: A filozófia, Az orvostudomány és A ...
11 A Clementine őszi konferenciája.
13 A bemutatott kutatás szerint melyik emotikon a leggyakrabban használtak sorában a harmadik?
14 Az IBM adatmenedzsment-architektúrájának koncepciója.
16 Az adatelemzési eredmények, táblázatok, grafikonok olvasásának és értelmezésének képessége.
18 Rokon értelmű kifejezés, azaz amikor egy fogalmat több különféle szóval is kifejezhetünk.
20 A színező algoritmus neve, amit a bécsi egyetemnek szánt Gustav Klimt festmények újraalkotásakor használtak.

A jó, a rossz és a szép
MI A FILMEKBEN

A mesterséges intelligencia annyira az életünk része lett, hogy már fel sem tűnik. Nem így a science fiction filmekben, ahol egyelőre még elképzelhetetlen személyiségjegyekkel ruházzák fel alkotóik. Most három típust mutatunk be, amelyekkel a filmvásznon vagy a képernyőn találkozhatunk.

A jó: J.A.R.V.I.S.

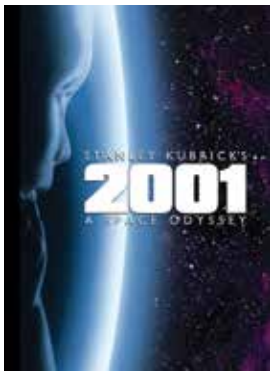
Vasember, pontosabban Tony Stark teremtménye a Marvel Univerzumból minden ember álma. A (férfi) gépi hangon megszólaló mesterséges intelligencia Iron Man kreálmánya, és mint ilyen, hűen segíti alkotóját. (Érdekesség, hogy a mozaikszó kibontva a következő: Just A Rather Very Intelligent System.) Stark néhai inasa, Edward Butler után adta a nevét a frissen létrehozott nyelvi interfésznek, amely nagy segítséget nyújtott a fogságban megalkotott páncélja létrehozásában. Ezek után J.A.R.V.I.S.-t feltöltötte a védőfelszerelésébe, és már a különböző harcok



során is igénybe vehette szolgálatát. Az engedelmes és jól nevelt MI azután többször megmentette a néha nem túl óvatos mágnás-feltaláló életét, sőt megalkotott egy külön kémiai elemet is, elemezte az ellenfelek gyengeségeit, és sok filmben kulcsszerepe volt Vasember győzelmeinek. Jarvisnak neveztek el egyébként a való világban egy hirdetéseket, blogposztokat író MI-t (jarvis.ai).

A rossz: HAL 9000

Annyira adta volna magát, hogy a legendás Skynet kerüljön ide! Nem ő kezdte azonban az emberek szisztematikus gyilkolását: e kétes dicsőség az 1968-as *2001: Űrodüsszeia* című film MI-jét illeti. A filmben ugyanis David Brownman és Frank Pool űrhajósok úgy döntenek, hogy lekapcsolják a HAL 9000 névre hallgató mesterséges intelligenciát, annak meghibásodását feltételezve. Az MI azonban tud



szájról olvasni, és meghiúsítja az emberek szándékát: megöli Poolt, és nem engedi be az űrhajóba a világűrből visszatérő Browmant. Érve így szól: „Ez a küldetés túl fontos ahhoz, hogy hagyjam, hogy veszélyeztesd."

A szép: Samantha

A közeljövőben játszódó, 2013-as, *A nő* című filmben a hivatásos levélíró Theodore beleszeret Samanthába, az operációs rendszerbe. Persze könnyebben érthető mindez, ha főhősünk magányos és zárkózott személyiségére pillantunk, amelyet tűzbe hoz egy okos, gyorsan tanuló, barátságos MI jelenléte. Fun fact: az eredetiben Samantha hangját Scarlett Johansson kölcsönözte az oprendszer számára.



Clementine



Hanga gyorsan kapcsol

Hanga a Clementine hang- és szöveganalitikai fejlesztéseken alapuló virtuális ügyfélszolgálati megoldása. Segítségével ügyfelei gyorsan célba érnek.



Clementine

Clementine | 1115 Budapest, Bartók Béla út 105–113. 1/B | info@clementine.hu | www.clementine.hu