



Területi Statisztika

Közzététel: 2018. szeptember 28.

A tanulmány címe:

A Big Data-adattudós és a területi statisztika

Szerzők:

Kilin Józsefné, Geopolitikai Tanács Közhasznú Alapítvány E-mail: statresearch@cgeopol.hu

<https://doi.org/10.15196/TS580505>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Területi Statisztika c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány, vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

- 1) A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Sztj.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
- 2) A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
- 3) A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
- 4) A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Sztj. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
- 5) A tanulmány átdolgozása, újra publikálása tilos.
- 6) A 3. a)–c.) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:

„Forrás: Területi Statisztika c. folyóirat 58. évfolyam 5. számában megjelent, Kilin Józsefné által írt A Big Data-adattudós és a területi statisztika c. tanulmány (link csatolása)”

- 7) A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem esnek szükségképpen egybe a KSH, vagy a szerzők által képviselt intézmények hivatalos álláspontjával.



A Big Data-adattudós és a területi statisztika*

Big Data Scientists and Regional Statistics

Kilin Józsefné

Geopolitikai Tanács Közhasznú

Alapítvány

E-mail:

statresearch@cgeopol.hu

A napjainkban zajló negyedik ipari forradalom, az ipar 4.0 idején a technológiák és az alkalmazások fejlesztése rohamléptekben halad, ami jelentős társadalmi és technológiai hatásokkal jár. A legtöbb társadalmat felkészületlenül érte ez a gyors ütemű fejlődés. Az adatrobbanás adta lehetőségek kiaknázásához fejlesztenünk kell tudásunkat, kreatitásunkat. A vállalati fejlődés számára elérkezett a Big Data-adatelemzési technológia, és a dolgok internete (Internet of Things – IoT) következtében hatalmas adatállományok keletkeznek. Az új eredmények átformálják a vállalkozások munkaerőigényét, a munkakörök tartalmát, miközben új szakmák születnek, mások eltűnnek. A Big Data-módszer alkalmazásával adattudósok készítenek előrejelzéseket.

Kulcsszavak:

adattudós,
Big Data-adatállomány,
kreativitás,
új gondolkodásmód

* A "Big Data a területi kutatásokban" címmel rendezett konferencián, 2018. február 27-én tartott előadás szerkesztett változata.

During Industry 4.0 – the fourth industrial revolution happening in our world today – technologies and applications are developing at an enormous pace which has significant effects on both societies and technology. Many countries were unprepared for this fast development and must enhance their knowledge and creativity in order to effectively utilize new opportunities created by data explosion. Companies have to face the challenge of analyzing Big Data, while through the Internet of Things huge files are generated. As a result, companies' labour-force needs and job profiles are changing; new professions are born, others disappear. Using Big Data methods forecasts can be predicted by data scientists.

Keywords:

data scientist,
data files in the Big Data,
creativity,
new way of thinking

Bevezetés

A Big Data-val foglalkozó munkájukban Mayer-Schönberger–Cukier (2014) megállapították, hogy az alkalmazás szempontjából még gyermekcipőben járó forradalmi módszernek a főbb jellemzői a következők:

- A Big Data alapjában véve azt jelenti, hogy eljött az idő, amikor... az adatok kerülnek a középpontba... Ez újfajta gondolkodásmódot igényel...”A Big Data-adattudósok szokásokat adatosítanak, szerkezetében változatos, olykor kusza és nem mindig pontos adatokat elemeznek, hatalmas adathalmazzal dolgoznak. Az adatállomány nagyságrendje – a bennük lévő hibás adatokkal együtt – olyan összefüggések feltárását biztosítja, amelyek korábban csak látens értéként voltak jelen.
- Fontos, hogy mi történik. Az elemzéshez hatalmas adatmennyiség áll rendelkezésre. Nem ok-okozati összefüggéseket kell keresni, hanem a szokások megfigyelésével területi statisztikai, illetve Big Data-módszerekkel kell következtetéseket levonni. Nem technológiai kérdésként (a hardverre és a szoftverre összpontosítva) tekintünk tehát a Big Data-ra” (210. old.).

A digitalizáció fejlődése, társadalmi hatása

Elérkeztünk a negyedik ipari forradalom (más néven 4.0) korába, amelyben az IoT, vagyis az okos eszközök, robotok használata, a mesterséges intelligencia tömeges terjedése mellett, adatrobbanást és adatvezérelt munkafolyamatokat prognosztizálnak a szakemberek. A létrejött adatokat a keletkezésük helyéhez kötik, és a hatalmas adatmennyiség csak Big Data-ként – matematikai és területi statisztikai módszerekkel – elemezhető digitalizált állomány.

Az IoT terjedése (Computerworld 2018b) a következő években minden területen folytatódik: az ipari, az energia- és közlekedési szektorban, az egészségügyben, valamint a kiskereskedelemben egyaránt. Naponta 2 és félmillió terabyte adat születik a világon, s az elérhető adatok 90%-a 2016–2017-ben keletkezett. Az elérhető adatoknak ugyanakkor mindössze 5%-a digitalizált. A vállalatok jelenleg a beérkező adataiknak csak a felét tárolják, és ennek a mennyiségnek is csak a felén végeznek elemzéseket.

A Tufts Egyetem kutatóinak egy csoportja, Chakravorti–Tunnard–Chaturvedi (2015) azt a kérdést fogalmazta meg, hogy hol fejlődik a digitális gazdaság a leggyorsabban. Az általuk kialakított új indexszámítási módszer alapján dolgozták ki több ország digitális fejlődési indexét. 50 ország (köztük Magyarország) 2008 és 2013 közötti digitális gazdasági mutatói alapján meghatározták az egyes országok digitális fejlődési indexét. Százaskálán jelölték az egyes országok indexét és az országok ehhez viszonyított digitális fejlődésének változását. Az eredmények alapján 4 kategóriát különítettek el: (1) kiemelkedő (Stand Out), (2) helyben járó (Stall Out), (3) kitörésre alkalmas (Break Out), (4) bizonytalan fejlődésű (Watch Out) országok.

Az említett szerzők – többek között – arra a következtetésre jutottak, hogy a legtöbb nyugat- és kelet-európai ország, továbbá Ausztrália és Japán digitális fejlődésüket tekintve – eltérő színvonalon ugyan, de – helyben járnak. Megállapították, hogy ezek az országok csak úgy tudják ismét elérni fejlődésük korábbi ütemét, ha ugrás-szerűen megduplázzák a szellemi és technológiai innovációjukat, valamint külső piacokat keresnek. Az állampolgárokról és a gazdaságról az állam hatalmas mennyiségű adattal rendelkezik, ami fontos szerepet játszik az innovációban, a digitális kapacitás fejlesztésében.

A negyedik ipari forradalom részeként gyors ütemben terjed a digitalizáció. A Világbank tanulmánya (HVG 2017b) szerint a fejlődő országokban a 4.0 miatt 40–50% lesz a megszűnő állások aránya. A magyarországi 4,1 millió munkavállalóból ez több mint 2 millió álláshelyet jelenthet.

A mesterséges intelligencia és a robotika terjedése drámaian csökkenti majd az elkövetkező évtizedben a középvezetők számát (Computerworld 2018a). 2000 és 2010 között az Egyesült Államokban 5,6 millió állás szűnt meg a gyártás területén, 85%-uk technológiai változások és automatizálás miatt. A digitalizáció következtében 2015 és 2025 között 350–400 ezer munkahely szűnhet meg Magyarországon is (Varga 2016).

Az eltűnő munkahelyek helyett újak jönnek létre, egyelőre azonban a technológiai fejlődés több állást szüntet meg, mint amennyit létrehoz. Két évtizeden belül legalább 98%-os valószínűséggel megszűnő hét foglalkozás (HVG 2018): a telemarketinges, a könyvtáros, a játékvezető és -bíró, a divatszakmai modell, a bankpénztáros, a hitelügyintéző és a könyvelő.

Az előrejelzések szerint az automatizált gyárak (vissza)települnek a fejlett országokba, és adózni is ott fognak (HVG 2017a). A felzárkózó és a fejlődő országokban – így hazánkban is – meg kell majd oldani a (az állások tömeges megszűnése miatt)

keletkező problémákat, át kell képezni, illetve új szakmákra kell képezni az állásukat elvesztett munkavállalókat.

Az előrejelzések úgy érintik a népesedést, a gazdaság és a szolgáltatás területi adatait, hogy közben megváltozik az adatok felhasználásával kapcsolatos szemlélet is. Ma már – a Big Data-nak köszönhetően – a hagyományos ok-okozati kapcsolatok feltárása mellett, az adatok másodlagos vagy újrafeldolgozásával arra keressük a választ, hogy mi történik, és a történések milyen hatással lesznek a jövőnkre. Információt gyűjtünk mindenről, ami csak létezik, majd adatosítjuk azokat. Ez lehetővé teszi, hogy új módon használjuk fel az adatokat, felfedezzük a bennük rejlő látens értékeket, összefüggéseket.

A Big Data-adattudós és -adathalmaz jellemzői

A hatalmas adatmennyiség elemeinek felhasználásához új algoritmusokat, matematikai és statisztikai módszereket kell kidolgozni. Ez a feladat a területi statisztikai elemzések módszertanát jól ismerő Big Data-adattudósokra vár.

Az adattudósokkal szembeni általános elvárások a következők:

- ismernie kell a gépi tanulási, statisztikai, adatelemzési módszereket,
- széles körű adatkezelési tudással kell rendelkeznie, mivel hatalmas mennyiségű adatokkal kell dolgoznia, amelyeknél a hagyományos adatbázis-kezelők már nem alkalmazhatók,
- értenie kell ahhoz az alkalmazási területhez, ahol a tudományát használja,
- pontosan kell megfogalmaznia az igényeket, a fejlesztésben is részt kell vennie, majd értelmeznie és prezentálnia kell a kapott eredményeket.

Ezek az elvárások egyben azt is mutatják, hogy az adattudós iránti igények eltérnek a diplomásokkal szemben támasztott általános követelményektől. A Big Data-munkakörbe könnyen tovább képezhető, megfelelő előképzettséggel rendelkező fiatalokat kell találni, akik képesek problémamegoldásra használni ezt az újszerű megközelítést, továbbá kellően kreatívak ahhoz, hogy releváns korrelációkat keresenek az előrejelzésekhez. Tökéletes előrejelzéseket lehetetlen készíteni, ezért a végző döntést mindig a menedzsmentnek kell meghoznia, és nem az adattudósnak.

Az adattudósok – kreatív ötleteik alapján, továbbá attól függően, hogy az egyes problémák megoldására milyen (Big Data-s) módszert dolgoztak ki, vagy tudásukat, módszertani ismereteiket átadják-e, oktatják-e, tanulással segítve a második generációs adattudós munkavállalókat – lehetnek:

- első generációs vagy
- második generációs adattudósok.

Képességeik, módszereik, szemléletük és tevékenységük alapján lehetnek (Schonherzbazis):

- stratégiai vagy
- operatív adattudósok.

Az e csoportokba sorolt adattudósok természetesen több szempontból is hasonlíthatnak egymásra.

A stratégiai adattudós átfogó tudással rendelkezik az üzleti teljesítmény, a stratégiai gondolkodás és a kommunikáció terén, viszont kevésbé jártas az adatbázisok létrehozásának, a megfelelő algoritmusok meghatározásának, kiválasztásának gyakorlati kérdéseiben. Jó kérdéseket tesz fel, releváns korrelációkat keres, valószínűséget határoz meg, azonban minimális informatikai-programozói ismeretekkel rendelkezik.

Az operatív adattudós nagy valószínűséggel statisztikus, matematikus vagy programozó, és a képességeit arra használja, hogy a megoldás gyorsan megszülessen, bizonyos rendszerek alkalmazásával megfelelő adatokat nyerjen és értelmezzen, majd a kérdés megválaszolásához a stratégiai adattudóssal közösen kiválassza a legfontosabb eredményeket.

Mindkét csoport képviselői tudósok, akik elméleteiket kísérletekkel tesztelik, majd rögzítik azokat.

Egy-egy Big Data-probléma megoldásához csapatmunkára van szükség, a csapat tagjainak ötlete egy következő jó ötletet eredményezhet. A Big Data-megoldások lényege az új algoritmusok, matematikai és statisztikai modellek kidolgozása és alkalmazása, majd azok informatikai rendszerekbe történő beépítése.

Kiből válhat adattudós? Egyesült államokbeli kutatások szerint a jó Big Data-elemzőket (Bitport 2015) nem a legjobb szaktudású munkatársak között kell keresni. A szaktudás – a pontossághoz hasonlóan – helyénvaló a Small Data világában, ahol nincsenek elegendő vagy helyes információk, ezért az ösztönökre és a tapasztalatokra kell támaszkodni. Ha azonban sok adat áll rendelkezésre, tapasztalatok nélkül is lehet velük boldogulni – sőt, talán még könnyebben is. A Big Data elemzésére képes szakemberek messzebbre látnak a szakmai konvencióknál, mert sok adattal rendelkeznek.

Big Data-elemzőkké tehát matematikusok, statisztikusok, illetve az Excel „zsonglőrök” válhatnak. Az Excel nagy teret enged a kreativitásnak, sőt bizonyos értelemben azt meg is követeli. Egy hatalmas adatállomány kezeléséhez azonban kevés. Az Excel megfelelő szintű alkalmazásához szükséges ismeretek tárháza sok ponton megegyezik a Big Data igényeivel, például alapvető statisztikai ismeretek, modellalkotás stb.

A Big Data-adathalmaz jellemzőinek együttesen (Computerworld 2017) három kritériumnak kell megfelelniük: (1) nagy adatmennyiség, (2) szerkezetében is változatos adatok, (3) gyors feldolgozás.

Amennyiben a felsorolt kritériumok közül bármelyiknek nem felel meg az adatállomány (például homogén szerkezetű), akkor Small Data-val van dolgunk. A gyors feldolgozás azért fontos, mert mire a statisztikai adatokat közzéteszik (legyenek akár előzetesek is), azok már többhetesek lesznek. A Big Data-elemzésnél követelmény az azonnali eredmény, de csak arra keressük a választ, hogy mi történik, és nem arra, hogy miért.

Adatvédelem, az adatokhoz való hozzáférés

A Big Data szempontjából kevésbé fontosak az adatvédelmi szabályok. A demokrácia bölcsőjének tartott Egyesült Államokban a Big Data-feldolgozások, -elemzések területén engedékenyek, viszont a Big Data lehetővé teszi számukra az ismételt azonosíthatóságot.

2018. május 25-én lépett hatályba az az EU 2016/679 adatvédelmi rendelete (EUR-Lex 2016), amely a személyes adat felhasználásának a felelősségét az adatkezelőre helyezi át, aki a személyes adat védelmének sérülése esetén nagy összegű bírságra számíthat.

Az adatgazdasággal foglalkozó, 2018. január végén rendezett konferencián megállapították (IVSZ 2018), hogy Európa hatalmas méretű ipari, gyártási kapacitással és az ahhoz kapcsolódó adatvagyonnal rendelkezik, és ezt ki kellene aknáznia. A vállalkozásoknak nem kellene félniük attól, hogy adataikat elérhetővé teszik a külvilág számára, hiszen – ezt számos nagyvállalat is példázza – csak a nyitottság lehet az alapja a hosszú távú, fenntartható növekedésnek. Amennyiben az adataikat külső felhasználók is lekérdezhették, azok elektronikus lábnyomaival még több adatot és elemezhető információt nyernének.

Összegzés

Az adatrobbanást előidéző technológiai fejlődés és az azzal járó társadalmi változások tükrében rámutattam arra, hogy az adatok újrahazsnosításánál mennyire fontos az új, már kialakulóban lévő szaktudás, a Big Data-adattudós szerepe. Az adattudósok két csoportba sorolhatók, közülük a stratégiai teszi fel a jó kérdéseket és az operatív a releváns adatok megkeresésével, elemzésével keresi a feltett kérdésekre a választ. Az adattudósok fő jellemzője az új gondolkodásmód, a hagyományokkal történő szakítás, a kreativitás és a (területi) statisztikai ismeretek, melyek birtokában leginkább előrejelzéseket készítenek. A Big Data-elemzésekhez felhasznált adatállományok halmaza hatalmas, szerkezetük változatos és gyors feldolgozást igényelnek.

IRODALOM

MAYER-SCHÖNBERGER, V.–CUKIER, K. (2014): BIG DATA – Forradalmi módszer, amely megváltoztatja munkánkat, gondolkodásunkat és egész életünket, HVG Kiadó Zrt., Budapest, 260 old.

INTERNETES HIVATKOZÁSOK

BITPORT (2015): Honnan lesz 4,4 millió big data-szakértő? <http://bitport.hu/honnan-lesz-4-4-millio-big-data-szakerto> (Letöltve: 2018. január)

- CHAKRAVORTI, B.–TUNNARD, C.–CHATURVEDI, R. S. (2015): Where the Digital Economy is Moving the Fastest <https://hbr.org/2015/02/where-the-digital-economy-is-moving-the-fastest#b03g06t20w15> (Letöltve: 2016. március)
- COMPUTERWORLD (2017): Amit a big data elemzésekről tudni érdemes <https://computerworld.hu/biztonsag/amit-a-big-data-elemzesekrol-tudni-erdemes226889.html> (Letöltve: 2018. február)
- COMPUTERWORLD (2018a): Középvezetők állását veszélyezteti a mesterséges intelligencia <https://computerworld.hu/cio/kozepvezetok-allasat-veszelyezteti-a-mesterseges-intelligencia-243095.html?hirlevel> (Letöltve: 2018. január)
- COMPUTERWORLD (2018b): IoT: a tömeges terjedés éve <https://computerworld.hu/tech/a-tomeges-terjedes-eve-242711.html> (Letöltve: 2018. január)
- EUR-LEX (2016): Az Európai Parlament és a Tanács (EU) 2016/679 Rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet) <http://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:32016R0679&from=HU> (Letöltve: 2018. február)
- HVG (2017a): Nyerhetnénk is a negyedik ipari forradalommal, de így nem fogunk http://hvg.hu/gazdasag/20170804_negyedik_ipari_forradalom_robotok_munka_ero_automatizacio_magyar_gazdasag (Letöltve: 2018. január)
- HVG (2017b): Itt a negyedik ipari forradalom, milliónyi magyar álláshely kerülhet veszélybe http://hvg.hu/gazdasag/20170802_Veszelyben_225_millio_magyar_allashely (Letöltve: 2018. január)
- HVG (2018): Hét foglalkozás, amelyről nem hitte, hogy eltűnhet http://hvg.hu/gazdasag/20180222_7_foglalkozas_amelyrol_nem_hitte_hogy_eltunhet (Letöltve: 2018. február)
- IVSZ (2018): 2025-re megduplázódik az európai adatpiac értéke <http://ivsz.hu/hirek/2025-re-megduplazodik-az-europai-adatpiac-erteke/> (Letöltve: 2018. február)
- SCHONHERZBAZIS: A Data Scientist két „alfaja” <https://schonherzbazis.hu/hirek/reszletek/A-Data-Scientist-ket-„alfaja”> (Letöltve: 2018. február)
- VARGA, M. (2016): 350–400 ezer munkahely szűnhet meg <https://mno.hu/gazdasag/varga-350-400-ezer-munkahely-szunhet-meg-1367342> (Letöltve: 2018. február)