

# Statisztikai adatok és elemzés a Big Data korában

Dr. Sebők Miklós

Igazgató, TK Politikatudományi Intézet

# Bevezetés

- Az előadás témái
  - A Big Data kora
  - A gazdaságstatisztikai alkalmazkodás fontossága
  - Társadalomtudományi alkalmazások
  - Összegzés

# A Big Data statisztikai kihívásai

—

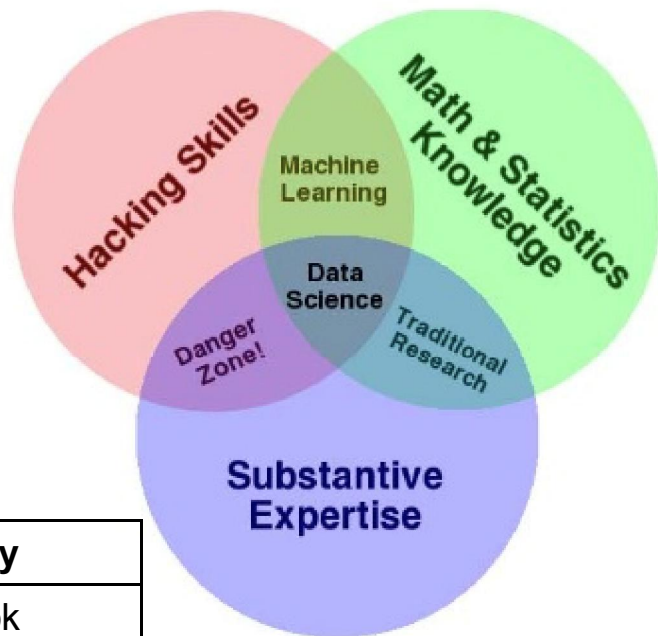
# A Big Data kihívása

- 4V - Volume, velocity, variety, veracity
- Technológiai változás – társadalmi változás – új etikai kérdések
- Új társadalmi jelenségek: manipuláció, személyes adatok, álhírek, a munka jövője
- BD: Computer architecture, networking technology, algorithms, data management, AI, ML – datafication, connectedness, networking, computer authoring, blockchain
- Minden megváltozik: adatahozzáférés, adatkezelés, adatelemzés
- Adatbányászat -» szöveg mint adat -» szövegbányászat

# Statisztika vs. adattudomány

- A nagy kérdések változatlanok:
  - Okság vs. korreláció
  - Kísérleti vs. megfigyelési adatok
  - Magyarázat vs. előrejelzés

## Drew Conway's Venn Diagram



Statisztika/ökonometria	Adattudomány
DV / IV, EV, predictors	Osztály/tulajdonságok
Minta/becslés	Tanulóhalmaz/ tanulás
Logisztikus regresszió/ klaszterezés, faktoranalízis	Felügyelt/felügyelet nélküli tanulás

# **Gazdaságstatisztikai kihívások: a fogyasztói árak esete**

—

## Decemberben 4,0%-kal, 2019-ben átlagosan 3,4%-kal nőttek az árak

2019. decemberben a fogyasztói árak átlagosan 4,0%-kal magasabbak voltak az egy évvel korábbinál. Az elmúlt 12 hónapban a szeszes italok, dohányárak, a járműüzemanyagok, valamint az élelmiszerek ára nőtt jelentős mértékben. 2019-ben a fogyasztói árak átlagosan 3,4%-kal emelkedtek az előző évhez képest.

### 12 hónap alatt, 2018. decemberhez viszonyítva:

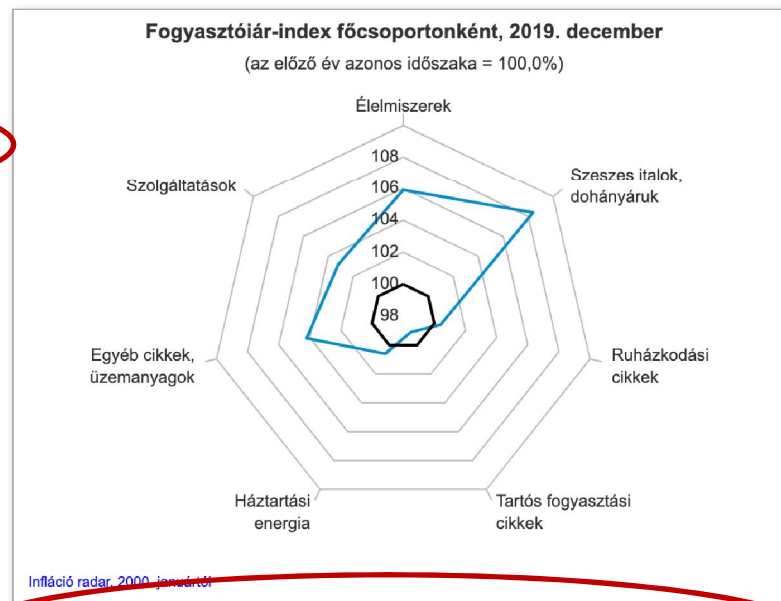
- Az élelmiszerek ára 5,9, ezen belül a sertéshúsé 23,7, a párizsi, kolbászé 13,2, a cukoré 12,2, az idényáras élelmiszereké (burgonya, friss zöldség, gyümölcs) 8,3%-kal nőtt. **A szeszes italok, dohányárak ára átlagosan 8,4, ezen belül a dohányáraké 12,3%-kal emelkedett. A szolgáltatásokért 9,2%-kal kellett többet fizetni, ezen belül a lakbér 10,5%-kal nőtt. A háztartási energia 0,6, ezen belül a tűzifa 5,9%-kal drágult. A járműüzemanyagok ára 7,7%-kal volt magasabb.**

### 1 hónap alatt, 2019. novemberhez viszonyítva:

- A fogyasztói árak átlagosan 0,3%-kal emelkedtek. Az élelmiszerek ára átlagosan 0,7%-kal nőtt, ezen belül a sertéshús 7,5, a liszt 2,6, a párizsi, kolbász 2,4, a tojás 2,3, a kenyér 2,0%-kal drágult. Az alkoholmentes üdítőitalok ára 1,1, a sajté 0,9%-kal csökkent. A szeszes italok, dohányárak ára átlagosan 0,6, ezen belül a dohányáraké 1,2%-kal emelkedett. A járműüzemanyagok ára 0,3%-kal volt magasabb.

### 2019-ben az előző évhez képest:

- A fogyasztói árak átlagosan 3,4%-kal nőttek, ezen belül az élelmiszerek ára 5,4%-kal. A legnagyobb mértékben, 8,1%-kal a szeszes italok, dohányárak drágultak. A szolgáltatások ára átlagosan 3,0, a háztartási energiáé és a ruházati cikkeké egyaránt 1,0, az egyéb cikkeké 1,3%-kal emelkedett. A tartós fogyasztási cikkekért 0,3%-kal kevesebbet kellett fizetni.



# Miért fontos a számbavétel alkalmazkodása?

- A gazdaságpolitika egyik legfontosabb siker-mércéje
- Számos gazdasági döntés épül rá, pl. nominális béremelés mértéke a várt reálbér-emelkedés fényében
- Az infláció adó is, és különösen sújtja a kisebb keresetűeket/nyugdíjasokat
- Bizonyos iparágakban teljesen domináns már az adatgazdaság (hirdetési piac, társkeresés..)
- És egyre újabbak kerülnek a célkeresztbe: fintech, agritech..





Szentkirályi Balázs  
2017. november 27. 06:52 · KÖZELET

## Nem is akkora az infláció, amikorát a KSH kimutat

GAZDASÁG



**Matolcsyék szerint meg kellene reformálni a statisztikákat, az infláció mérését**

BEFEKTETÉS BANK DEVIZA GAZDASÁG ÜZLET UNIÓS FORRÁSOK INGATLAN ZÖLD VILÁG IRODA TŐZSDE KÖTVÉNYPIAC PRIVÁT BANKOK



GAZDASÁG | 2019. november 08. 14:14

**Ellentmondások az inflációban - mit mondanak az elemzők?**

Mindennapos tapasztalat, hogy alig találni redőnyöst, hidegburkolót vagy mondjuk bébiszittert, aki ne harmadával többért dolgozna, mint egy-két évvel ezelőtt. Ehhez képest a KSH szerint a szolgáltatások az elmúlt egy évben csupán 1,6 százalékkal drágultak, ami érzésre képtelenségnek tűnik. Pedig a statisztika a szolgáltatásokhoz sorolja a lakbért is, amiről végképp tudjuk, hogy teljesen elszállt az ország legtöbb városában. Ehhez képest a hivatal kimutatása szerint csak 5 százalékkal drágult a lakásbérlet egy év alatt, ami ugye megint csak erősen ellentmond a hétköznapi tapasztalatainknak.

Vannak azonban kifutó, megszűnő termékek, gondoljunk csak az állandóan cserélődő mobiltípusokra. A KSH célja viszont az, hogy kiszűrje a minőségjavulásból származó árváltozásokat. Ha egy telefon azért drágább, mert többet tud, az nem árváltozás, hanem fejlődés. A 2010-ben vett 150 ezres ledtévénél ma sem olcsóbb az új modell, ahogy a 2000-ben vett plazmatévéhez képest sem történt meg a 73 százalékos áresés, ahogy azt a KSH az inflációs felmérésében kimutatja.

# Hogy jön mindez az adatgazdasághoz?



Mint írták, a digitális fejlődés nemcsak a gazdaság keresleti-kínálati viszonyait alakítja át, hanem az infláció- és általában a gazdasági teljesítmény mérését is új alapokra helyezheti. Erre azért is komoly szükség van, mert az ötödik ipari forradalom minden korábbinál nagyobb hatással bír, ám a következmények jelenleg még alig számszerűsíthetők.

## Korrekcíóra szorul a statisztika

A tanulmány szerint a gazdasági tranzakcióink és társadalmi interakcióink egyre nagyobb szelete kerül át a virtuális térbe, miközben a fizikai tárgyak birtoklása helyett az élmények vagy épp a legkülönbözőbb platformokhoz való hozzáférés válik értékké. Már pénz helyett is egyre gyakrabban információval fizetünk. Egy ilyen világról a 20. század közepén kialakított statisztikai rendszerek egyre kevésbé lesznek képesek megfelelő helyzetképet mutatni. Egyre több elemzés jelzi, hogy a hivatalos statisztikák a fogyasztói árak alakulását akár jelentősen is felülbecsülhetik. És ez a torzítás a jelenlegi digitalizációs, technológiai forradalomban egyre jelentősebbé válik.



# Nem csak úri huncutság: email, képek, offline



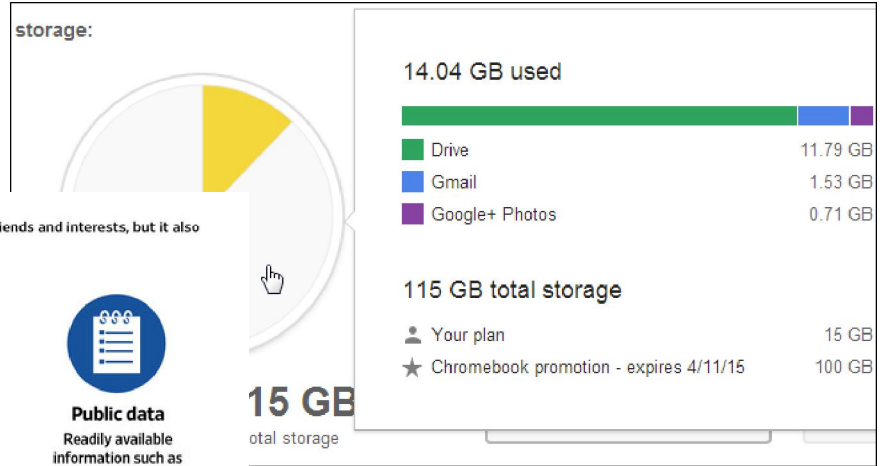
## What Facebook Knows

The social network gathers all sorts of data from its users, such as age, friends and interests, but it also collects information outside its website.



Source: the company

The Wall Street Journal



# **Társadalomtudományi kihívások: a szöveg mint adat**

---

# The Top Industries using Web Scraping

01 Retail and Manufacturing

02 Financial Research

03 Data Science

04 Marketing and Sales

05 Academic

06 Journalism

07 Real Estate





# A paradigma: a “szöveg mint adat”

- Text as Data
- Tipikus feladat-megoldás:
  - a. Nyers szöveg vs. elemezhető szöveg (unstructured-» structured)
  - b. Mintázatok keresése a feldolgozott szövegben (mátrixban)
  - c. Eredmények értékelése/értelmezése
- A tipikus átalakítási folyamat:
  - Előkészítés: szavak (features) -> számok
  - Normalizálás
  - “Sparsity” -» Dimenzió-redukció

N / D	1	.	.	.	.	d
1						
.						
.				( i ; j )		
.						
.						
n						

	ajtó	akar	akár	akarat	akárm	akart	akasztóf	akkor	ala	alá	alacsony
character(0)	0	0	0	0		0 0	0		1 0	0	0
character(0)	0	0	0	0		1 0	0		0 0	0	0
character(0)	0	0	0	0		0 0	0		0 0	0	0

N / D	1	.	.	.	.	d
1						
.						
.				( l ; j )		
.						
.						
n						

# DTM - TDM



# Törvények módosítási hálózata

Processed text

Connections

Original text

**Full title:** 2009. évi XXXI. törvény a Polgári Törvénykönyv és a Polgári Törvénykönyv hatálybalépéséről és végrehajtásáról szóló 1960. évi 11. törvényerejű rendelet, valamint ezzel összefüggésben más törvények egyes rendelkezéseinek módosításáról [LÁBJEGYZET\_1]

**Name:** 2009. évi XXXI. törvény

**Type:** law

**Other law(s) in title:**

**Likely modified law based on the title:** -

1. § (1) A Polgári Törvénykönyvről szóló 1959. évi IV. törvény (a továbbiakban: Ptk.) 81. § (2) bekezdésében az "a jogosult/jogszerű pénzügyi, gazdasági vagy piaci érdekeit sértené vagy veszélyeztetné" szövegrész helyébe az "a jogosult - ide nem értve a magyar államot - jogszerű pénzügyi, gazdasági vagy piaci érdekeit sértené vagy veszélyeztetné" szöveg lép.

(2) A Ptk. 209. §-a (4) bekezdése helyébe a következő rendelkezés lép, egyidejűleg a következő (5) bekezdéssel egészül ki, és a jelenlegi (5) bekezdésének számozása (6) bekezdésre változik:

"(4) Az általános szerződési feltételei és a fogyasztói szerződésben egyedileg meg nem tárgyalt feltételei tisztességtelenségét önmagában az is megalapozza, ha a feltétel nem világos vagy nem érthető."

"(5) A tisztességtelen szerződési feltételekre vonatkozó rendelkezések nem alkalmazhatók a főszolgáltatást megállapító, illetve a szolgáltatás és az ellenszolgáltatás arányát meghatározó szerződési kikötésekre, ha azok egyébként világosak és érthetőek."

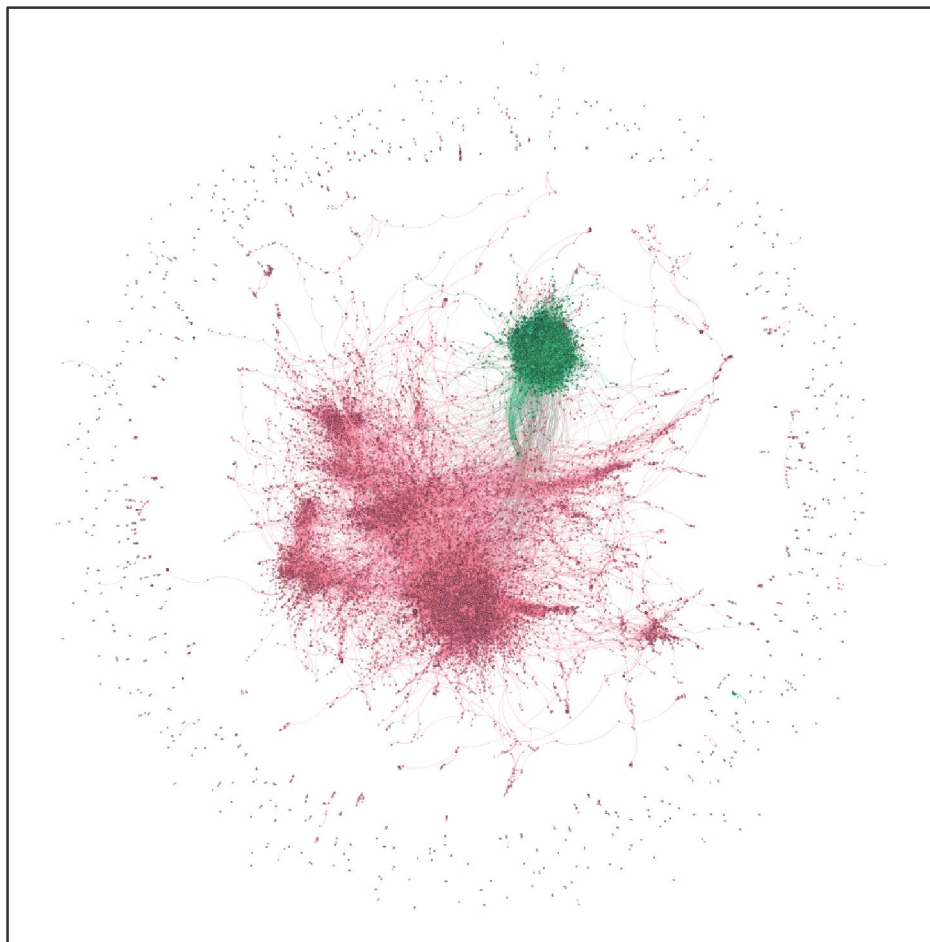
(3) A Ptk. 232. §-a a következő (2) bekezdéssel egészül ki, egyidejűleg a § (2) és (3) bekezdésének számozása (3) és (4) bekezdésre változik:

"(2) Az államháztartás airendszerében harmadik személyekkel szemben vállalt vagy az airendszereket egyébként harmadik személyekkel szemben terhelő fizetési kötelezettség esetén, ide értve a támogatási szerződésekből eredő fizetési kötelezettséget is, a kamatfizetés mellőzésére vagy a törvényesnél alacsonyabb mértékű kamat fizetésére irányuló szerződési kikötés - ha jogszabály eltérően nem rendelkezik - semmis. Ezt a rendelkezést kell alkalmazni az államháztartás airendszeréből nyújtott támogatások felhasználására kötött szerződésekben, valamint a jogszabályban meghatározott közteladatok ellátásával megbízott és e körben olyan szervezetek által vállalt fizetési kötelezettség esetében is, amilyen államháztartási szervezet alapított, vagy amelyben ilyen szervezet többségi befolyással rendelkezik."

(4) A Ptk. 301. §-a a következő (2) bekezdéssel egészül ki, egyidejűleg a § (2)-(4) bekezdésének számozása (3)-(5) bekezdésre változik:

"(2) Az államháztartás airendszerében harmadik személyekkel szemben vállalt vagy az airendszereket egyébként harmadik személyekkel szemben terhelő fizetési kötelezettség késedelmes teljesítése esetén a késedelmi kamat tekintetében is alkalmazni kell a kamatfizetés mellőzésére vagy a törvényesnél alacsonyabb mértékű kamatfizetésre irányuló szerződési kikötés semmisségére vonatkozó rendelkezést."

2. § (1) A Polgári Törvénykönyv hatálybalépéséről és végrehajtásáról szóló 1960. évi 11. törvényerejű rendelet (a továbbiakban: Ptk./Ptké.) 47. §-ának (1) és (2) bekezdése helyébe a következő rendelkezés lép:





# Záró gondolatok

—

# Az alkalmazkodás feladatai: Oktatás

- Korábban: a programozás teljes hiánya a társtud. képzéseken
- Sokat változott a világ, fejlődött a magyar oktatás
- De: társadalomtudományi képzésben még mindig nem integrált az R vagy python (legfeljebb SPSS)
- Nem kell folyamatosan újragondolni a képzések tartalmát – csak ha korszakhatáron vagyunk..... És most korszakhatáron vagyunk
- Hallgatónak is jó: több karrierpálya áll előtte



# Az alkalmazkodás feladatai: Statisztikusok

- Szemléletmód – A változatlanság lemaradást és relevancia-vesztést eredményez
- Szemléletmód-váltás területei:
  - új(ra értelmezett) fogalmak: pl. „társadalmi haladás” – humán tőke, klímaváltozás
  - új adatok: jelentős részben online platformokon zajlik a társadalom élete
  - új adatelemzési technikák: mesterséges intelligencia
- A statisztikus társadalom indulhat a legjobb helyzetből az adat korában
- De a szemléletmód alkalmazkodása és jelentős forrásbevonás is szükséges, hogy megfelelhessen a kihívásnak - együttműködések!

**Köszönöm a figyelmet!**

Dr. Sebők Miklós

[sebok.miklos@tk.mta.hu](mailto:sebok.miklos@tk.mta.hu)

[poltext.tk.mta.hu](http://poltext.tk.mta.hu)