

Ana Ušaj

Viri podatkov: Predstavitev primera »Big data« Humanizacija masovnih podatkov

»Big data« oziroma množični ali masovni podatki so velike količine podatkov zbrane iz različnih virov namenjeni za analizo s katero se iščejo vzorci, trendi ipd. Množični podatki so lahko zelo kompleksni in jih opredeljujejo značilnosti, kot so znatna velikost, raznolikost, hitrost in verodostojnost.

V nalogi sem obravnavala vidik humanizacije masovnih podatkov na podlagi treh različnih virov, ki so predstavljeni v nadaljevanju.

1. Barry, D., et al. (24.5.2020). An Incalculable Loss. The New York Times.

Najdeno na <https://www.nytimes.com/interactive/2020/05/24/us/us-coronavirus-deaths-100000.html>?

• Vsebinski opis (vira) podatkov s poudarkom na opisu načina zbiranja oz. mehanizma, kako podatki nastanejo

Članek se nanaša na število umrlih zaradi okužbe s covid-19. V maju 2020 so v ZDA zabeležili mejnik 100.000 umrlih. V članku so zbrane besede iz osmrtnic, ki številkam dajejo imena ljudi in s drobnim podatkom iz njihovega življenja številu umrlih dodajo globlji pomen.

• Ključna spoznanja

Članek je zelo pomenljiv. Vsak dan lahko spremljamo različne podatke vendar si pogosto niti ne predstavljamo, da vsako skupno število sestavlja posamezna enota, v našem primeru človek. S člankom je bilo na enostaven način predstavljeno, kar si lahko predstavljamo, kot humanizacijo podatkov.

2. Helzlsouer, K. et al. (2020). Humanizing Big Data: Recognizing the Human Aspect of Big Data. Frontiers in Oncology.

Najdeno na <https://doi.org/10.3389/fonc.2020.00186>

• Vsebinski opis (vira) podatkov s poudarkom na opisu načina zbiranja oz. mehanizma, kako podatki nastanejo

Članek opisuje vidik humanizacije masovnih podatkov na področju genomske medicine.

Termin »big data« na področju biomedicine se nanaša na velike količine podatkov zbranih iz različnih virov, kot so evidence iz genskega testiranja, elektronske zdravstvene evidence, evidence iz testiranj pri odkrivanju sorodstvenih vezi pa tudi evidence različnih raziskav in študij.

Masovni podatki so ključni pri iskanju lastnosti in dejavnikov, ki privedejo do rakavih obolenj. Ker so lahko skupni dejavniki oziroma povzročitelji zelo redki, je potrebno zajeti čim širši obseg podatkov, da se lahko na njihovi podlagi potegne vzporednice med posameznimi primeri. Rast interesa za pridobivanje in uporabo »big data« na področju genetike izvira iz številnih odkritij in razvoja genomike, prenosu odkritij v prakso, interesa farmacevtskih podjetij za razvoj in rabo ciljno usmerjenih zdravljenj, interesa potrošnikov pri iskanju sorodstvenih vezi, medicinske uporabe.

• Dobro in slabo pri viru oz. njegove uporabe

Potencialni razlogi za nižjo kakovost masovnih podatkov tega področja so:

- pomanjkanje sistematičnosti pri zbiranju izvornih podatkov,
- različnost uporabljenih analitičnih tehnologij,
- potencialne razlike pri klasifikaciji in interpretaciji genskih različic.

Izziv predstavlja tudi trgovanje s tovrstnimi podatki. Ker obstaja velik interes za nakup podatkov, se za prodajo odločajo tudi privatna podjetja, ki izvajajo genska testiranja izven klinik. Pri tem se pojavi vprašanje, kako so ti podatki zbrani in analizirani ter kako se pri tem spoštuje posameznika, ki s podatki izpostavi svoj najintimnejši del – genski profil – potrebna je pridobitev soglasja vsakega posameznika

in zagotoviti varovanje podatkov. Vedno je potrebno upoštevati ravnovesje med dvema vrednotama: zdravjem in zasebnostjo.

Pomembna pozitivna plat uporabe teh masovnih podatkov je možnost razvoja t.i. personalizirane medicine, ki bi mu lahko po slovensko rekli kar »k posameznemu bolniku usmerjena medicinska obravnava«, ki pri zdravljenju in preventivi upošteva specifične značilnosti posameznikov, kot so genske značilnosti in življenjski slog posameznika. Za razliko od pristopa »one size fits all«, kjer je zdravljenje osnovano na povprečnem človeku, imajo pri tem pristopu zdravniki in raziskovalci možnost bolj natančne oblike zdravljenja in strategije preventive za določene skupine ljudi.

- **Ključna spoznanja**

- Delitev in agregiranje genskih podatkov v široke baze podatkov lahko zakrije dejstvo, da je osnoven vir vsakega podatka posameznik. Posamezniki zagotovijo podatke, ki so nato zbrani in analizirani z namenom da se informacija iz njih prenese spet na posameznika.
- Personalizirana medicina mora sloneti na preciznih podatkih, kar pomeni, da je pomembno da je vedno znan vir masovnih podatkov, populacija, opis in postopek analiziranja podatkov, uporabljene analitične metode in algoritmi.
- Združevanje različnih širokih baz podatkov, uporaba tehnologije za generiranje in interpretacijo masovnih podatkov v povezavi z genskim testiranjem lahko v prihodnosti zelo izboljšajo javno zdravje.

3. Barlow, B. (n.d.). Humanizing Big Data: The Key to Actionable Customer Journey Analytics. RocketSource.

Najdeno na <https://www.rocketsource.co/blog/journey-analytics/>

- **Vsebinski opis (vira) podatkov s poudarkom na opisu načina zbiranja oz. mehanizma, kako podatki nastanejo**

Za prodajo izdelka ali storitve je pomembna celotna uporabniška izkušnja. Podjetja zato zbirajo podatke iz različnih virov, da bi ugotovili, kako se obnašajo njihovi kupci. Med podatki, ki jih imajo na voljo so tudi masovni podatki, ki lahko izhajajo npr. iz analize beleženja klikov na spletnih straneh, sledenju gibanja po spletni strani, družbenih omrežjih ali pa raziskavah, ki sledijo nakupu storitve in izdelka. Zaradi tehnologije, ki omogoča dostop do ogromne količine masovnih podatkov se podjetja pogosto utapljajo v podatkih. Čeprav se velika večina podjetij trudi svoje aktivnosti in odločitve osnovati na podlagi analize zbranih podatkov pa je le manjši delež podjetij pri tem tudi dejansko uspešen.

Podatki, s katerimi podjetja skušajo razbrati potrebe kupcev, oceniti njihovo celotno uporabniško izkušnjo in razbrati, kaj vodi potrošnika do dejanskega nakupa, se zbirajo na več različnih načinov. V grobem jih lahko razdelimo na kvalitativne in kvantitativne podatke.

Masovni kvantitativni podatki so npr. število klikov na spletni strani, delež na socialnih omrežjih ipd. Podjetja pogosto zbirajo podatke na podlagi ponakupnih vprašalnikov, kjer kupci odgovorijo na vprašanja npr. ali bi podjetje priporočili svojim prijateljem in sodelavcem ali kako so bili zadovoljni s posameznimi vidiki in izkušnjami v celotnem procesu prodaje. Podjetja lahko te podatke zbirajo na različne načine, pogosto se stranke po nakupu kontaktira z vprašalnikom, ki sem ji ga pošlje preko elektronske pošte. Kvalitativni podatki dopolnjujejo kvantitativne podatke. Podjetja jih pogosto zbirajo s pomočjo intervjujev testnih skupin.

- **Dobre in slabe plati vira oz. njegove uporabe**

- Masovni kvantitativni podatki pomagajo pri določanju širšega trenda. Pridobivanje kvantitativnih masovnih podatkov je cenejše, lažje jih je zbirati, analizirati, bolj uporabni so za statistično analizo. Slaba plat teh podatkov je njihova omejenost, iz njih težko razberemo globlji pomen in jih lahko napačno interpretiramo.

- Kvalitativni podatki so bogatejši z vidika vsebine, bolj osebni, usmerjeni in nam omogočajo bolj natančen vpogled v čustvene impulze ljudi, ki vodijo do nakupa izdelka ali storitve. Na drugi strani je pridobivanje kvalitativnih podatkov dražje, podatke je težje interpretirati in odgovore oziroma odzive kupcev je težje primerjati med sabo.
- **Ključna spoznanja**
 - Cilj pridobivanja podatkov v podjetjih ni zagotoviti čim večje količine podatkov ampak na podlagi zbranih podatkov doseči čim boljšo uporabniško izkušnjo, ki predstavlja konkurenčno prednost podjetja in vodi do višjih prihodkov.
 - Pravilna interpretacija podatkov in njihova humanizacija zmanjša razliko med tem kar podjetja na podlagi podatkov mislijo o razmišljanju kupcev in tem kar kupci dejansko mislijo.
 - Kljub popularnosti masovnih podatkov ne smemo pozabiti, da so ljudje srce dejstev, števil in parametrov. Če želimo iz prikaza kvantitavnih masovnih podatkov razbrati motivacijsko obnašanje, ki vodi do dejanskega nakupa potrebujemo tudi dodatne kvalitativne podatke. Na podlagi masovnih podatkov namreč pogosto dobimo samo odgovor na vprašanje, kaj se dogaja ne pa tudi zakaj do tega pride.
 - Podjetja se utapljajo v pridobljenih podatkih – potrebna je ustrezna vizualizacija zbranih podatkov, ki jih predstavi na dovolj enostaven in uporaben način ter odgovori tudi na ključna vprašanja zakaj prihaja do določene situacije.
 - »Big Data is just that – big.« Razpolagati s širokim naborom dejstev ne pomeni, da razpolagamo s pravimi podatki za odgovor na naše vprašanje. Pomembno je, da podatke opremimo s kontekstom, da razumemo pomen, ki se skriva za številkami oziroma v našem primeru, da razumemo obnašanje kupcev.