

Primjena tehnologije velikih podataka: stajališta hrvatskih građana o društvenim dobrobitima i štetnim posljedicama

Mesić, Karla

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, Faculty of economics Split / Sveučilište u Splitu, Ekonomski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:124:415833>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-02-07**

Repository / Repozitorij:

[REFST - Repository of Economics faculty in Split](#)



UNIVERSITY OF SPLIT



EKONOMSKI FAKULTET
SVEUČILIŠTE U SPLITU



**FACULTY OF ECONOMICS,
BUSINESS AND TOURISM**
UNIVERSITY OF SPLIT

DIPLOMSKI RAD

**PRIMJENA TEHNOLOGIJE VELIKIH PODATAKA:
STAJALIŠTA HRVATSKIH GRAĐANA O DRUŠTVENIM
DOBROBITIMA I ŠTETNIM POSLJEDICAMA**

Mentor: doc. dr. sc. Tea Mijač

Studentica: Karla Mesić

Split, svibanj 2024.

SADRŽAJ

1. UVOD	1
1.1. Problem istraživanja.....	1
1.2. Predmet istraživanja.....	3
1.3. Istraživačka pitanja.....	5
1.4. Ciljevi istraživanja.....	6
1.5. Metode istraživanja.....	7
1.6. Doprinos istraživanja.....	8
1.7. Struktura rada.....	9
1.8. Sadržaj rada.....	9
2. PREGLED LITERATURE	11
2.1. Veliki podaci.....	11
2.1.1. Definicija pojma „big data“.....	11
2.1.2. Karakteristike velikih podataka.....	12
2.1.3. Vrste velikih podataka.....	13
2.1.4. Prikupljanje velikih podataka.....	14
2.1.5. Obrada i analiza velikih podataka.....	17
2.1.6. Podaci o građanima (engl. Citizen data).....	19
2.1.7. Pravna regulativa velikih podataka.....	21
2.1.8. Društvena kvantifikacija kroz povijest.....	24
2.1.9. Podatkovni sistemi.....	26
2.1.10. Komparativan prikaz praksi korištenja europskog gdpr-a i kineskog scs.....	28
2.2. Prednosti tehnologije velikih podataka.....	34
2.2.1. Poboljšanje proizvoda i korisničke usluge.....	34
2.2.2. Medicinski napredak.....	35
2.2.3. Financijska zaštita.....	37
2.2.4. Predviđanje kriminala.....	38
2.2.5. Obrazovanje.....	40
2.2.6. Urbani razvoj.....	42
2.3. Nedostaci tehnologije velikih podataka.....	44
2.3.1. Gubitak privatnosti i nadzor.....	44
2.3.2. Društvena diskriminacija.....	48
2.3.3. Sigurnosni rizici.....	50
2.3.4. Utjecaj na ponašanje.....	52
2.3.5. Loš rezultat predviđanja.....	52
3. REZULTATI ISTRAŽIVANJA	55
3.1. Demografski rezultati.....	55
3.2. Stajališta ispitanika o velikim podacima.....	57
3.3. Odgovori na postavljena istraživačka pitanja.....	62
4. RASPRAVA	68
5. ZAKLJUČAK	70

LITERATURA	71
POPIS SLIKA, TABLICA I GRAFOVA.....	78
SAŽETAK	79
SUMMARY.....	80
PRILOG 1 – ANKETA	81

1. UVOD

1.1. Problem istraživanja

Svijet je usred podatkovne revolucije. Sa svakom sekundom koja prođe, količina digitalnih informacija koja teče između poduzeća, vlada i ljudi brzo se širi. Podaci sada sve više podupiru naš svakodnevni život: koristimo ih za online kupovinu, za praćenje zdravstvenih podataka, kao i osobnih bankovnih računa. Predstavljajući ključno *oružje* u borbi protiv COVID-19 pandemije podaci su postali sve važniji aspekt međunarodne trgovine i globalnih odnosa (Government Office for Science, 2020).

Danas se sve više spominju veliki podaci (engl. Big Data) koji predstavljaju svakodnevne informacije podatke, odnosno činjenice o svijetu koji nas okružuje postajući tako moćna i rastuća sila u društvu. Veliki podaci su skup podataka različitih izvora i vrsta, velikog obujma, složenosti i potencijalne vrijednosti, što se teško može obraditi i analizirati u očekivanom roku (Moorthy i suradnici, 2015). Veliki podaci kao novi strateški resurs u digitalnom dobu i važan čimbenik koji pokreće inovacije, mijenja ljudsku proizvodnju i način života. Područja primjene velikih podataka su raznolika i pokrivaju gotovo sve aspekte društvenog života, uključujući zdravstvo, genetiku, komunikaciju, meteorologiju, kreditiranje i društveni kontakt. Veliki podaci uglavnom imaju tri izvora (Kocijan, 2014): vlada, poduzeća i otvoreni izvor (engl. Open source). Među te izvore spadaju i baze s otiscima prstiju, DNA baze, zapisi aviokompanija, zapisi obrazovnih ustanova, transakcije kreditnih kartica, Facebook stranice, e-mailovi, zapisi ustanova javnog zdravstva... Znanstvena analiza temeljena na velikim podacima ne samo da može pružiti učinkovitu podršku za odlučivanje nadležnima, već i temeljito mijenja sve slojeve života (Shi, 2022).

Porast obujma (velikih) podataka doveo je do značajnih promjena u tehnološkoj industriji, a te su promjene imale i pozitivne i negativne društvene učinke na ljudsku populaciju.

S pozitivne strane, veliki podaci omogućili su tehnološkim tvrtkama da razviju nove proizvode i usluge koje živote ljudi mogu učiniti lakšim i praktičnijim (George & Paul, 2020). Na primjer, aplikacije za pozivanje prijevoza kao što su Uber i Lyft koriste velike podatke za spajanje vozača s putnicima u stvarnom vremenu, što ljudima olakšava kretanje. Platforme društvenih mreža poput Facebooka i Twittera koriste velike podatke za personalizaciju sadržaja i oglašavanja, što može pomoći ljudima da otkriju nove proizvode i usluge za koje inače možda nisu znali. U današnje vrijeme, velike količine podataka postale su sveprisutne u gotovo svim sektorima, uključujući i financijsku industriju. Ova količina podataka donosi sa sobom i mnoge izazove, no isto tako i pruža brojne prednosti, posebno u detekciji financijskih prevara (Bay i sur., 2006). Korištenje tehnologije velikih podataka u financijskoj

industriji pruža mogućnosti za brže i učinkovitije otkrivanje različitih vrsta prijevара, što može pomoći u smanjenju gubitaka i povećanju sigurnosti (Tao i sur., 2019).

Međutim s negativne strane, veliki podaci također predstavljaju brojne rizike i izazove. Jedan od najvećih rizika je mogućnost kršenja privatnosti jer tvrtke prikupljaju i analiziraju ogromne količine osobnih podataka. Prema Mayer-Schönberger i Cukier (2013) također postoji tamna strana velikih podataka koja predstavlja značajne izazove konvencionalnim pravnim instrumentima zaštite privatnosti. Tehnologija velikih podataka može predstavljati alat za nadzor te povećava kapacitet organizacija za praćenje života pojedinaca, narušavanje anonimnosti i diskriminaciju na temelju spola, rase, dobi, lokacije i drugih čimbenika. Društveno razvrstavanje stanovništva pomoću tehnologija velikih podataka možemo vidjeti u primjeru gdje na temelju uspostavljene korelacije velikih podataka radnici koji duže putuju na posao su klasificirani kao oni koji prije daju otkaz. Je li onda pravedno odbijati kandidate za posao s dužim putovanjem na posao? Također, je li prihvatljivo da političke stranke donose zaključke o tome kako biste mogli glasovati, na temelju analize Facebook prijatelja, Twitter pratitelja i drugih dokaza o društvenom grafikonu? (Bennet & Bayley, 2013). Spomenuti rizik od algoritamske pristranosti, gdje modeli strojnog učenja mogu diskriminirati određene skupine ljudi na temelju njihove rase, spola ili drugih čimbenika može dovesti do društvene nejednakosti i produžiti postojeće predrasude u društvu.

Prikupljanje i korištenje podataka o pojedincima i skupinama ljudi (podataka o građanima) posebno se ubrzava (van der Sloot i sur., 2016). Više nego ikad, ostavljamo digitalni trag kad god pretražujemo internet, putujemo i kupujemo; u gotovo svakom aspektu našeg svakodnevnog života. Primjerice, podaci o građanima (engl. citizen data) predstavljaju ogromnu potencijalnu vrijednost za potrošače, tvrtke i javna tijela (Herschel & Miori, 2017). U digitalnom dobu, niz podataka o građanima sada se može mnogo lakše koristiti za širi skup svrha i za svrhe koje nisu bile inicijalno predviđene. To također može uključivati i zlobne namjere i bez pažljivog upravljanja štetu pojedincima, društvu i nacionalnoj sigurnosti.

1.2. Predmet istraživanja

Kao što je prethodno navedeno, u današnjem digitalnom dobu, osobni podaci postaju sve vrijedniji i važniji u poslovanju, marketinškim strategijama i u političkim procesima. U ovom radu istražiti će se *svjesnost građana o dijeljenju svojih podataka te njihove percepcije prednosti kao i prijetnji koje pojedinac može ostvariti*. U Hrvatskoj, kao i u drugim zemljama, postoji pravna regulativa koja štiti osobne podatke građana, no ipak, mnogi ljudi nisu upoznati s načinima prikupljanja podataka putem algoritma i tehnologije velikih podataka te ne razumiju koje su posljedice ove prakse (Gruschka i suradnici, 2018).

Prema podacima Eurostata iz 2018. godine, postotak korištenja digitalnih tehnologija u poduzećima (>10 zaposlenika) na području EU odnosno postotak korištenja tehnologije velikih podataka u Hrvatskoj čini 10 %, dok je najveći postotak korištenja zabilježen za državu Malta (24 %), Nizozemska (22 %) te Irska (20 %) i Belgija (20 %).

Neki od primjera korištenja tehnologije velikih podataka u Hrvatskoj:

- Financijski sektor- Hrvatska poštanska banka (HPB) koristi osobne podatke klijenata odnosno tehnologiju velikih podataka i sofisticirane analitičke metode za ocjenu kreditnog rizika, kako bi bolje procijenila kreditnu sposobnost klijenata i osigurala kvalitetniju ponudu proizvoda i usluga. Osim toga, HPB bilježi podatke o transakcijama klijenata kako bi otkrila prijevare i druge nezakonite aktivnosti (HPB, 2022).
- Zdravstveni sektor- Hrvatski zdravstveni sustav prikuplja velike količine podataka o zdravstvenom stanju građana, kao što su medicinske povijesti, dijagnoze, liječenje, laboratorijski rezultati, itd. Ovi podaci se mogu analizirati kako bi se identificirale prilike za poboljšanje kvalitete skrbi, optimizacija troškova i učinkovitosti, te razvoj personaliziranih terapija i usluga (Vlada Republike Hrvatske, 2018).
- Sektor za javni red i sigurnost-provodi se obrada velikih količina strukturiranih i nestrukturiranih podataka danas, između ostalog, može se bolje predviđati i prevenirati kriminalitet kao i druga potencijalno opasna društvena ponašanja. Tehnologija je također omogućila da se u ranoj fazi detektiraju potencijalne terorističke prijetnje (Kopal & Karas, 2017).

Strategija digitalne Hrvatske za razdoblje do 2032. godine pretpostavlja primjenu naprednih tehnologija kao što su 5G/6G, umjetna inteligencija (engl. *artificial intelligence - AI*), strojno učenje (engl. *machine learning*), računarstvo u oblaku (engl. *cloud computing*), tehnologija velikih količina podataka i tehnologija ulančanih blokova (engl. *blockchain*) u javnom i privatnom sektoru, ali i ostaje otvorena za implementaciju nekih budućih disruptivnih tehnologija koje će se pojaviti u promatranom

periodu. Primjena navedenih naprednih tehnologija omogućit će kvalitetniju obradu i uporabu podataka što će, zauzvrat, doprinijeti učinkovitosti u radu javnih tijela, kreiranju javnih politika utemeljenih na podacima, personalizaciji javnih usluga, administrativnom rasterećenju, efikasnijoj komunikaciji javnih tijela i građana te boljim prilikama za suradnju javnog i privatnog sektora (Narodne novine, 2023).

Stoga, predmet ovog istraživanja je analizirati percepciju i razumijevanje hrvatskih građana o načinima korištenja osobnih podataka kao i spremnost dijeljenja takvih podataka za različite svrhe koje s jedne strane mogu olakšati život pojedinca i poboljšati društvenu dobrobit, a s druge mogu oštetiti pravo privatnosti građana.

1.3. Istraživačka pitanja

Temeljem definiranog predmeta i problema istraživanja, postavljaju se sljedeća istraživačka pitanja.

1. Utječe li razina razumijevanja korištenja osobnih podataka na **percepciju društvene dobrobiti** koju stvara upotreba tehnologije velikih podataka?

Istraživačkim pitanjem se nastoji utvrditi ima li razumijevanje građana o mogućim daljnjim upotrebama osobnih podataka tehnologijom velikih podataka utjecaj na percepciju opće dobrobiti koja može u konačnici unaprijediti život pojedinca kao i utvrditi je li samo razumijevanje takvih tehnologija djeluje na percepciju društvene dobrobiti.

2. Utječe li razina razumijevanja korištenja osobnih podataka na **percepciju štetnih posljedica** koju stvara upotreba tehnologije velikih podataka?

Istraživačkim pitanjem se nastoji utječe li razumijevanje građana o tehnologiji velikih podataka na percepciju štetnih posljedica za život pojedinca.

3. Utječe li razina razumijevanja korištenja osobnih podataka **na spremnost dijeljenja** takvih podataka?

Istraživačkim pitanjem se nastoji utvrditi je li razumijevanje građana o mogućoj daljnjoj upotrebi osobnih podataka tehnologijom velikih podataka ima utjecaj na spremnost dijeljenja osobnih podataka.

1.4. Ciljevi istraživanja

Osnovni cilj istraživanja je temeljem prikupljenih podataka razumijevanje mogućnosti iskorištavanja tehnologije velikih podataka, inteligentnih algoritama te njihovog utjecaja na ostvarivanje društvene dobrobiti, posebno u pogledu poboljšanja kvalitete života te učinkovitosti različitih procesa i usluga, ali ujedno predstavljajući i izazov u vidu 'curenja' podataka građana te zlouporaba osobnih podataka gdje se narušava pravo privatnosti pojedinca. Stoga, svrha je svakako postići ravnotežu između koristi društva u cjelini i negativnih posljedica narušavanja privatnosti, uzimajući u obzir moguće etičke, pravne i socijalne implikacije.

Nadalje, anketnim istraživanjem hrvatskih građana istražiti će se **upoznatost i razumijevanje građana s konceptom tehnologije velikih podataka** i načinima na koje se prikupljaju osobni podaci putem algoritama i drugih tehnologija. Analizirat će se percepcija hrvatskih građana o korištenju takve tehnologije te njihova **spremnost** davanja osobnih podataka u svrhu javnog dobra kao i mogućnost negativnih ishoda dijeljenja takvih podataka, kako bi se bolje razumjeli **rizici i izazovi koje ova tehnologija može donijeti**.

Kroz prikupljanje primarnih podataka, analizirat će se podaci o tome koliko su građani upoznati s tehnologijom velikih podataka, koliko su zabrinuti zbog mogućnosti zlouporabe podataka te koliko su spremni prihvatiti takvu tehnologiju ukoliko im ista može pomoći u području poboljšanja poslovanja, zdravstvene skrbi te opće, financijske i druge sigurnosti.

1.5. Metode istraživanja

Diplomski rad sastoji se od teorijskog i empirijskog dijela rada, gdje se oba dijela sastoje od različitih metoda koje su potrebne kako bi se konstruirao određeni rezultat na kraju provedenog istraživanja. U svrhu definiranja problema i predmeta istraživanja, napravljen je pregled literature. Korišteni radovi pretraživali su se uz pomoć ključnih riječi poput: veliki podaci, osobni podaci, privatnost, zaštita, prednosti i nedostaci velikih podataka. Prvi dio rada temelji se na teorijskom dijelu gdje se korištenjem brojne literature, znanstvenih časopisa, znanstvenih spoznaja te internetskih podataka želi definirati puka teorija odnosno žele se ustanoviti i pojasniti određene stavke koje su potrebne za razumijevanje svrhe ovog rada te zaključka na kraju.

Dodatno, znanstvene metode koje će se koristiti za pisanje ovog diplomskog rada su sljedeće (Zelenika, 2000) :

- Metoda analize putem koje će se složeniji sudovi, pojmovi i zaključci raščlaniti na jednostavnije sastavne dijelove i elemente za sebe i u odnosu na druge dijelove.
- Metoda sinteze kojom će se provesti postupak spajanja jednostavnijih misli te zaključaka u složenije.
- Metoda komparacije kojom će se uspoređivati srodne ili slične činjenice, pojave te odnosi kako bi se utvrdile razlike i sličnosti među istima. Navedena metoda omogućava istraživačima da dođu do raznih uopćavanja te novih zaključaka.
- Deskriptivna metoda predstavlja postupak opisivanja ili ocrtavanja činjenica, predmeta ili odnosa u okolini bez znanstvenog objašnjavanja i tumačenja. Navedena metoda se koristi u početnim fazama rada dok joj vrijednost daljnjim izlaganjem raste.

Drugi dio rada temelji se na primjenama kvantitativnih metoda, odnosno korištenje anketnog istraživanja. Odnosno, uz pomoć Kvantitativnog istraživanja analizirat će se prikupljeni kvantitativni podaci (Creswell & Creswell, 2018). Primarni podaci prikupljat će se uz pomoć anketnog upitnika te će se po obradi podataka moći odgovoriti na postavljena istraživačka pitanja. Anketnim upitnikom prikupljat će se podaci o hrvatskim građanima kako bi se ispitala njihova percepcija i razumijevanje tehnologije velikih podataka.

Percepcija (stav) će se ispitati izjavama na koje će ispitanici izražavati slaganje upotrebom Likertove mjerne skale u rasponu vrijednosti od 1 do 5.

Prikupljeni podaci analizirat će se uz pomoć SPSS softvera te Excela.

1.6. Doprinos istraživanja

Digitalizacija koja u današnje vrijeme eksponencijalno raste iz godine u godinu zasigurno ima brojne pozitivne ali i negativne posljedice. Istraživanje ove teme moglo bi pomoći u razumijevanju načina na koji se digitalni sustavi za prikupljanje velikih podataka mogu koristiti. Odnosno, istražiti će se utječe li razumijevanje tehnologije velikih podataka na percepciju prednosti te na percepciju rizika. Dodatno, istraživanje na ovu temu moglo bi pomoći u **promicanju informiranosti i javne rasprave o korištenju digitalnih sustava te u podizanju svijesti o potencijalnim prednostima i rizicima povezanim s ovakvim sustavima gdje bi se na primjeru već postojećih država mogli spoznati uvidi koji bi doprinijeli budućem digitalnom razvoju Hrvatske i drugih zemalja.**

Rezultati ovog istraživanja mogu pomoći u boljem razumijevanju percepcije građana o prikupljanju osobnih podataka putem algoritama i tehnologije velikih podataka te mogu dati smjernice za poboljšanje prakse prikupljanja i korištenja osobnih podataka u Hrvatskoj.

Konačno, istraživanje će pomoći u razumijevanju percepcija dobrobiti i rizika velikih podataka s perspektive građana. Odnosno, istražiti stavove građana o tome može li primjena tehnologije velikih podataka poboljšati svakodnevni život građana i društvo u cjelini, ali i istovremeno osigurati zaštitu prava na privatnost. Ovo će biti od izuzetne važnosti u razvoju politika i zakona koji će regulirati prikupljanje i korištenje osobnih podataka.

1.7. Struktura rada

Rad je podijeljen u šest poglavlja zajedno s uvodom i zaključkom. U uvodnom dijelu rada definirani su predmet i problem istraživanja, nakon čega su postavljena istraživačka pitanja koja će biti obrađena u empirijskom dijelu rada. Nastavkom rada određeni su ciljevi i metode istraživanja koje će se koristiti u ovom diplomskom radu. Drugo poglavlje posvećeno je teoretskom pregledu literature o pojmu Veliki podaci, njegovim karakteristikama, upotrebi i prikupljanju. Dodatno, navest će se prednosti i nedostaci tehnologije velikih podataka odnosno prikaz svih dobrobiti i štetnih posljedica koje ovakva tehnologija može priuštiti pojedincu uzimajući u obzir pravo na privatnost i društvenu korist. Treće poglavlje donosi rezultate istraživanja odnosno empirijski dio istraživanja putem anketnog upitnika na temelju kojeg će odgovoriti na postavljena istraživačka pitanja te donijeti određeni zaključci. Četvrto poglavlje donosi raspravu ovog istraživanja te na kraju posljednje poglavlje predstavlja zaključak ovog diplomskog rada.

1.8. Sadržaj rada

1. UVOD

1.1. Problem istraživanja

1.2. Predmet istraživanja

1.3. Istraživačka pitanja

1.4. Ciljevi istraživanja

1.5. Metode istraživanja

1.6. Doprinos istraživanja

1.7. Struktura rada

2. PREGLED LITERATURE

2.1. Veliki podaci

2.1.1. Definicija pojma „Big Data“

2.1.2. Karakteristike Velikih podataka

2.1.3. Vrste Velikih podataka

2.1.3. Prikupljanje Velikih podataka

2.1.4. Obrada i analiza Velikih podataka

2.1.5. Podaci o građanima (engl. citizen data)

2.1.6. Pravna regulativa Velikih podataka

2.1.7. Društvena kvantifikacija kroz povijest

2.1.8. Podatkovni sistemi

2.1.9. Komparativan prikaz praksi korištenja europskog GDPR-A i kineskog SCS

2.2. Prednosti tehnologije velikih podataka

2.2.1. Poboljšanje proizvoda i korisničke usluge

2.2.2. Medicinski napredak

2.2.3. Financijska zaštita

2.2.4. Predviđanje kriminala

2.2.5. Obrazovanje

2.2.6. Urbani razvoj

2.3. Nedostaci tehnologije velikih podataka

2.3.1. Gubitak privatnosti i nadzor

2.3.2. Društvena diskriminacija

2.4.3. Sigurnosni rizici

2.5.4. Utjecaj na ponašanje

2.5.5. Loš rezultat predviđanja

3. REZULTATI ISTRAŽIVANJA

3.1. Demografski rezultati

3.2. Stajališta ispitanika o Velikim podacima

3.3. Odgovori na postavljena istraživačka pitanja

4. RASPRAVA

5. ZAKLJUČAK

LITERATURA

POPIS SLIKA, TABLICA I GRAFOVA

SAŽETAK

SUMMARY

PRILOZI

2. PREGLED LITERATURE

2.1. Veliki podaci

Neprestani napredak informacijske tehnologije kao i povećanje sveprisutnosti mreža i elektroničkih uređaja rezultira golemom količinom podataka iz raznih izvora i u različitim formatima. Količina podataka koja je danas dostupna je mjerena u zetabajtima (ZB): mjera jednaka jedan trilijun gigabajta (GB) i ekvivalentno kapacitet pohrane podataka od oko 250 milijardi DVD-a (Alharti i sur.,2017). Brzim razvojem aplikacija društvenih medija, računalstva u oblaku, internet stvari, mobilnog računalstva, e-trgovine, senzorskih podataka i drugih tehnologija dovodi do pojave tehnoloških inovacija u velikim podacima. Kako se naš digitalni otisak nastavlja eksponencijalno širiti, veliki podaci postali su središnja točka u raznim područjima, od poslovanja i zdravstva do znanosti i vlade. Sposobnost iskorištavanja, analiziranja i izvlačenja smislenih uvida iz ogromnih skupova podataka ima potencijal za otključavanje neviđenih prilika, katalizirajući inovacije i informirajući procese donošenja odluka (Alalawneh & Alkhatib, 2020).

Takvi podaci omogućavaju brojne prilike, ali i neprilike današnjici što će se detaljno obraditi nastavkom ovog diplomskog rada.

2.1.1. Definicija pojma „big data“

Neviđeni rast u obujmu, raznolikosti i brzini kojom se podaci generiraju i prenose tijekom proteklog desetljeća dovela je do rođenja koncepta veliki podaci. Hashem i dr. (2015.) definiraju velike podatke kao „skup tehnika i tehnologije koje zahtijevaju nove oblike integracije za otkrivanje velikih vrijednosti skrivenih unutar različitih, složenih i velikih skupova podataka.“ Veliki podaci kombinacija su strukturiranih, polustrukturiranih i nestrukturiranih podataka koje prikupljaju organizacije, a koji se mogu iskorištavati radi informacija i koristiti u projektima strojnog učenja, prediktivnom modeliranju i drugim naprednim analitičkim aplikacijama(Botelho & Bigelow, 2022).

Veliki podaci kakve danas poznajemo prvi put se spominju 2001.godine nakon objave Douga Laneya o „3V“. IT analitičar Doug Laney postavio je okvir za razumijevanje izazova koji su tada nastajali u upravljanju podacima za poslovanje, a koji se sada sveprisutno nazivaju "3 V's velikih podataka": volumen, brzina i raznolikost. Jednostavno rečeno, podaci su se počeli stvarati opsežno, brzo i raznolikih oblika dovoljnih da preplave relacijske baze podataka i druge konvencionalne načine upravljanja informacijama poduzeća. Mnogi drugi autori produžili su „3 V“ model i, kao rezultat toga, više značajki velikih podataka, poput vrijednosti, vjerodostojnosti, složenosti i nestrukturiranosti, dodanih na popis (Kache i sur., 2017). Prema Stantonovoj definiciji(2013); "Veliki podaci su pojam koji opisuju veliku količinu podataka – i strukturiranih i nestrukturiranih – koji svakodnevno preplavljaju tvrtku. Međutim nije količina podataka ono što je važno. Važno je što organizacija radi s podacima koji su važni odnosno kako ih iskorištava", dok autor Dumbill (2013) ističe da „podaci postaju veliki kada prelaze kapacitet tradicionalnih sustava za obradu baza

podataka te zahtijevaju primjenu alternativnih metoda obrade.“

2.1.2. Karakteristike velikih podataka

Jednostavno rečeno, veliki podaci su veći, složeniji skupovi podataka, posebice iz novih izvora podataka. Ovi su skupovi podataka toliko opsežni da tradicionalni softver za obradu podataka jednostavno ne može njima upravljati. Nadalje, ove ogromne količine podataka mogu se koristiti za rješavanje poslovnih problema što prije nije bilo moguće. S ovime, se također slažu Mayer-Schönberger i Cukier navodeći kako se veliki podaci odnose na novootkrivenu sposobnost prikupljanja i analiziranja golemih količina informacija koji su bili nemogući prije samo nekoliko godina. Ova sposobnost mijenja način na koji živimo, radimo i razmišljamo.

U zadnjih nekoliko godina velikim podacima dodani su mnogi V-ovi, kao što se navodi na sljedećem prikazu:



Slika 1. Karakteristike velikih podataka prema Mayer-Schönberger & Cukier(2013)

Karakteristike velikih podataka mogu se sažeti na sljedeće komponente (Curtis, 2023):

- **Volumen:** veliki podaci obuhvaćaju ogromne količine informacija koje se generiraju svakodnevno u različitim sektorima poduzeća što dovodi do kompleksnosti analize takvih podataka.
- **Brzina:** velika brzina generiranja, prikupljanja i obrade podataka kako bi se pravovremeno zadovoljile sve potrebe tržišta.
- **Raznolikost:** uključujući strukturirane, polustrukturirane i nestrukturirane podatke. Kao na

primjer, tekstualni podaci, slike, audio i videozapisi, podaci s društvenih mreža, senzorski podaci, i tablični podaci - sve to čini raznolikost podataka.

- Vjerodostojnost: veliki podaci često uključuju podatke s različitih izvora, pa je važno osigurati vjerodostojnost i kvalitetu tih podataka zahtijevajući procjenu analitičara u vezi usklađenosti, točnosti i istinitosti podataka.
- Vrijednost: analiza velikih podataka trebala bi pridonijeti donošenju informiranih poslovnih odluka ili omogućiti otkrivanje novih uzoraka i trendova što čini njihovu iznimnu vrijednost u poslovanju.
- Varijabilnost: promjenjivost u brzini, volumenu i vrsti podataka koji dolaze u sustav; dnevna ili sezonska fluktuacija u podacima, kao i neočekivane promjene u brzini prikupljanja podataka.
- Vizualizacija: kvaliteta prezentacije podataka ključna je za stvaranje lako razumljivih i djelotvornih prikaza podataka.

2.1.3. Vrste velikih podataka

Izvori podataka za analitiku velikih podataka višestruki su i raznoliki, u rasponu od tradicionalnih baza podataka do senzora u stvarnom vremenu, od kojih svaki služi različitoj svrsi u podatkovnom ekosustavu. Jedna od stvari koja velike podatke čini stvarno velikima je to što dolaze iz velikog broja različitih izvora nego što je to bio slučaj prije. Mnogi od izvora su web izvori, uključujući zapise, klikove i društvene medije. Naravno, korisničke organizacije godinama prikupljaju web podatke. Ali za većinu organizacija je to bila jedna vrsta gomilanja podataka kao što su npr. RFID podaci iz opskrbnog lanca. Međutim ono što se promijenilo jest da daleko više organizacija sada analizira velike podatke umjesto da ih samo gomila; podaci se danas generiraju i obrađuju na složenijoj i sofisticiranijoj razini. **Strukturiranim** podacima sada se pridružuju i **nestrukturirani podaci** (slike, glazba, ljudski jezik) i **polustrukturirani** podaci (XML, RSS izvori). Postoje i podaci koje je teško kategorizirati jer dolaze s audio ili video zapisa i drugih uređaja. Osim toga, višedimenzijski podaci mogu se izvući iz skladišta podataka kako bi se velikim podacima dodao povijesni kontekst. To je mnogo eklektičnija mješavina tipova podataka nego što je analitika ikada vidjela. Dakle, s velikim podacima, raznolikost je jednako velika kao i količina. Osim toga, raznolikost i volumen imaju tendenciju da potiču jedno drugo. Strukturirani podaci održavaju hegemoniju nad drugim vrstama podataka. Većina podataka kojima se danas rukuje putem analitičkih platformi spada u rubriku strukturirani podaci. Ovdje se prvenstveno radi o tablicama i drugim podatkovnim strukturama relacijskih baza podataka (Russom,2011).

- Strukturirani podaci: imaju određena unaprijed definirana organizacijska svojstva i prisutni su u strukturiranoj ili tabelarnoj shemi, što olakšava analizu i sortiranje. Osim toga, zahvaljujući svojoj

unaprijed definiranoj prirodi, svako polje je diskretno i može mu se pristupiti zasebno ili zajedno s podacima iz drugih polja. To čini strukturirane podatke iznimno vrijednima, što omogućuje brzo prikupljanje podataka s različitih mjesta u bazi podataka.

- Nestrukturirani podaci: uključuju informacije bez unaprijed definiranih konceptualnih definicija i standardne baze podataka te se ne mogu lako interpretirati ili analizirati. Nestrukturirani podaci čine većinu velikih podataka i sadrže informacije poput datuma, brojeva i činjenica. Primjeri nekih velikih podataka ove vrste uključuju video i audio datoteke, mobilne aktivnosti, satelitske slike i No-SQL baze podataka. Fotografije koje postavljamo na Facebook ili Instagram te videozapisi koje gledamo na YouTubeu ili bilo kojoj drugoj platformi pridonose rastućoj hrpi nestrukturiranih podataka.
- Polustrukturirani podaci: su hibrid strukturiranih i nestrukturiranih podataka. To znači da nasljeđuju nekoliko karakteristika strukturiranih podataka, ali bez obzira na to sadrži informacije koje nemaju određenu strukturu i nisu u skladu s relacijskim bazama podataka ili formalnim strukturama modela podataka. Na primjer, JSON i XML tipični su primjeri polustrukturiranih podataka (Bay Atlantic University, 2023).
- Podaci u stvarnom vremenu: streaming data: ovo su podaci koje kontinuirano generiraju tisuće izvora podataka, koji obično šalju zapise podataka istovremeno i u malim veličinama (često u kilobajtima). Ovakvi podaci uključuju informacije kao što su one koje generiraju korisnici pomoću mobilnih ili web aplikacija, kupnje putem e-trgovine i informacije s društvenih mreža ili geoprostornih usluga.
- Serija podataka, batch data: Za razliku od podataka u stvarnom vremenu, serija podataka uključuje obradu podataka koji su prikupljeni tijekom određenog vremenskog razdoblja. Akumuliraju se prije početka analize, za razliku od „streaming data“ koji se analiziraju u stvarnom vremenu. Ovakvi podaci su korisni kada je prihvatljivo imati određeno kašnjenje između prikupljanja podataka i izvlačenja uvida (Sundararajan, 2022).

2.1.4. Prikupljanje velikih podataka

Nova tehnologija sada je dio naših života na način na koji prije 20 godina nismo mogli ni sanjati. Tehnologija vezana uz podatke, obradu podataka, informacije i komunikaciju poboljšava naše društvo na mnogo načina. Nastaje novo informacijsko društvo: svijet u kojem je sve mjerljivo i u kojem su ljudi i gotovo svaki uređaj koji vam padne na pamet povezani 24/7 putem interneta (Mayer-Schönberger & Cukier, 2013). Ova mreža veza i senzora pruža fenomenalnu količinu podataka i nudi fascinantne nove mogućnosti koje se zajedno nazivaju veliki podaci.

Prikupljanje velikih podataka u suvremenom razdoblju u oštroj je suprotnosti s povijesnim metodama prikupljanja podataka. U prošlosti je prikupljanje podataka često bilo ograničeno ručnim procesima, ograničenom tehnologijom i manjim skupovima podataka. Danas je tehnološki napredak revolucionirao prikupljanje podataka, omogućujući sakupljanje ogromnih količina informacija neviđenim brzinama. Pojava Interneta stvari (IoT), sofisticiranih senzora i snažnih računalnih mogućnosti omogućila je prikupljanje podataka u stvarnom vremenu iz različitih izvora, od IoT uređaja i platformi društvenih mreža do strojno generiranih podataka. Osim toga, raširena uporaba web analitike, mobilnih aplikacija i naprednih tehnika rudarenja podataka promijenila je način na koji organizacije i istraživači pristupaju prikupljanju podataka. Za razliku od povijesnih praksi koje su se oslanjale na ručna istraživanja i ograničene evidencije transakcija, suvremene metode iskorištavaju potencijal različitih izvora, pružajući bogatiji i dinamičniji krajolik za uvide (Rahmati, 2016). Međutim, s ovom inovacijom dolazi odgovornost za rješavanje pitanja privatnosti i etičkih pitanja u sve većem području prikupljanja velikih podataka.

Za tvrtke i javne organizacije veliki podaci nadilaze lansiranje novih proizvoda ili usluga: radi se o tome kako biti relevantan u potpuno drugačijem svijetu. U makroekonomskom smislu razloga za veliki optimizam ima. Nalazimo se usred klasičnog ekonomskog fenomena: Schumpeterove *krektivne destrukcije* (Schumpeter 1934.), što znači da možemo postići inovaciju tek nakon što se stari modeli pokvare. To je ono što se događa u gotovo svim sektorima. Glazbena industrija bila je jedna od prvih koju je poremetio iTunes, a nedavno i Spotify. Posljednjih godina vidjeli smo kako druge industrije slijede taj primjer. Booking.com i AirBnB donijeli su radikalne promjene u industriju putovanja. Uber na sličan način djeluje kao katalizator promjena u taksi industriji. Svi ovi primjeri pokazuju da nove tvrtke donose promjene na tržišta usvajanjem novih pristupa podacima. Uzmimo za primjer Uber. U biti, ova tvrtka se ne bavi taksi djelatnošću. Bavi se prodajom podataka o korisnicima mreži taksista i obrnuto, optimizacijom taksi mreže uvođenjem veće inteligencije podataka. Kupac je zapravo proizvod. Teška su točna predviđanja o budućnosti u vremenima velikih transformacija, ali postoji čvrsto uvjerenje da se svaka industrija djelomično temelji na informacijama i da će prije ili kasnije biti pogođena na usporedive načine. Tehnologija velikih podataka je katalizator ovog trenda.

Tok podataka koji dolazi s bilo koje vrste uređaja poput robotskih proizvodnih strojeva, mikrofona koji osluškiju ili video kamera koje traže određeno lice u gomili potvrđuje kako prikupljanje velikih podataka u stvarnom vremenu nije novost; mnoge tvrtke godinama prikupljaju podatke o strujanju klikova s web stranica, koristeći podatke o klikovima kako bi posjetiteljima tog web mjesta dale preporuke za daljnju kupnju (van der Sloot i sur., 2016). Internet stvari (IoT) proširio je stvaranje tokova podataka u stvarnom vremenu. IoT se odnosi na mrežu fizičkih objekata - 'stvari' - ugrađenih sa sensorima, softverom i drugim tehnologijama u svrhu povezivanja i razmjene podataka s drugim uređajima i sustavima putem interneta. Ovi uređaji variraju od običnih kućanskih predmeta do sofisticiranih industrijskih alata.

Računalstvo u oblaku ima za cilj transformirati IT arhitekturu (Abadi 2009.), dok Big Data ima za cilj transformirati proces donošenja odluka. Dobit računalstva u oblaku obično se opisuje na razini tehnologije: dovođenje ogromnih resursa za računalstvo i pohranu pod koncentrirano upravljanje, što se kasnije može pružiti na razini infrastrukture, platforme ili softvera kao detaljna usluga klijentima. Ista paradigma opisuje ogroman porast mrežno povezanih senzora u našem društvu, ugrađenih u sve vrste proizvoda, uređaja i strojeva kako bi ih učinili "pametnima". Očekuje se da će do 2030. postojati jedan trilijun senzora, a podaci koje oni prikupljaju bit će dominantan dio Big Data (Chen i sur., 2013). Nadalje, tehnologija velikih podataka igra bitnu ulogu u nedavnim razvojjima koji se možda neće odmah povezati s velikim podacima. Kao na primjer sve više spominjana tehnološka singularnost (tj. trenutak kada će inteligentni sustavi premašiti ljudski intelektualni kapacitet) ili blok lanac, mehanizam koji stoji iza bitcoina koji omogućuje pouzdane transakcije bez treće strane (Desai, 2015).

Pojam *datafikacije*, koji su uveli Mayer-Schönberger i Cukier (2013), obuhvaća proces prikupljanja informacija o svemu što nas okružuje (kao što su GPS lokacije inhalatora koji bilježe podatke o okolišu neprikladnom za astmatičare, praćenje tjelesnih podrihtavanja kod neuroloških pacijenata radi praćenja njihovog zdravstvenog stanja). Nakon toga, te informacije se transformiraju u podatkovni format kako bi ih se moglo kvantificirati i dalje analizirati. Time se fokus informacijske tehnologije pomiče iznova s tehnologije na samu informaciju, naglašavajući kontinuitet ljudskog nastojanja da mjeri, bilježi i analizira svijet u kojem živimo (Kocijan, 2014).

Nekoliko nas indikacija navodi na vjerovanje da je rast fenomena velikih podataka neizbježan. Tehnološki razvoj ima tendenciju pobuditi veća očekivanja, a budući da mnogi od nas postaju ovisni o praktičnosti, udobnosti i jednostavnosti korištenja novih usluga, konkurencija će potaknuti daljnji napredak u tehnologiji i aplikacijama velikih podataka.



Slika 2. Izvori velikih podataka, dostupno na: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji190595>

2.1.5. Obrada i analiza velikih podataka

Tvrtke sve više pohranjuju, obrađuju i analiziraju podatke kako bi iz njih generirali vrijednosti značajne za uspjeh poslovanja. Veliki podaci nisu novost, ali učinkovito analitičko korištenje velikih podataka jest. Kako bi to uspješno izvršile potrebno je izabrati neku od tehnologija velikih podataka.

Postoje brojne tehnologije za obradu velikih podataka poput (Rahmati, 2016):

- Hadoop: Okvir otvorenog koda za paralelnu obradu podataka na vrlo skalabilnim klasterima poslužitelja. Hadoop je posebno prikladan za evaluacije gdje je potrebno provesti detaljne analize.
- Cloudera: Sveobuhvatan portfelj testiranih aplikacija otvorenog koda koje se mogu instalirati i kojima se može vrlo jednostavno upravljati na web sučelju putem Cloudera Cluster Managera. Organizacije koriste provjerena rješenja i imaju fleksibilnost za uključivanje novih Big Data tehnologija u postojeće procese.
- Apache Hive: Skladište podataka za Hadoop. Apache Hive premješta podatke iz relacijskih baza podataka u Hadoop putem SQL dijalekta HiveQL. Ključne funkcije uključuju sažimanje, postavljanje upita i analizu podataka.

- Cloudera Impala: Skalabilni i distribuirani alat za upite podataka za Hadoop. Ad-prednosti: Upiti u stvarnom vremenu bez potrebe za premještanjem ili transformacijom podataka.
- MongoDB: Jedna od vodećih NoSQL baza podataka na tržištu iz prostora otvorenog koda. Baza podataka opće namjene omogućuje dinamičan razvoj i visoku skalabilnost.

Danas se sve više spominje analitika u stvarnom vremenu koja se odnosi na sposobnost obrade i evaluacije podataka u trenutku kada ulaze u sustav, pružajući trenutne uvide i omogućavajući brzo djelovanje. Ova promjena paradigme s retrospektivne analize na trenutne uvide nudi duboke prednosti. Poduzeća sada mogu trenutačno reagirati na tržišne promjene, personalizirati interakcije s klijentima u hodu i optimizirati operacije s izvorima podataka uživo, mijenjajući način na koji industrije funkcioniraju. Obrada velikih podataka u stvarnom vremenu ne samo da ubrzava tempo kojim organizacije posluju, već također povećava preciznost njihovih odgovora. Omogućuje dinamično i agilno okruženje u kojem se odluke donose na temelju podataka, a prilike se iskorištavaju čim se ukažu. Ova neposredna generacija uvida nije samo konkurentska prednost; brzo postaje poslovna potreba ukoliko se želi ostati relevantnim na današnjem brzom digitalnom tržištu.

Obrada podataka u stvarnom vremenu u komparaciji s tradicionalnom odnosno serijskom obradom podataka čini znatne razlike. Ključna razlika između ove dvije vrste obrade podataka je vremenski okvir za povratnu informaciju i akciju. Obrada u stvarnom vremenu, ili obrada toka, obrađuje podatke kako stignu, pružajući kontinuirani unos, obradu i izlaz. Idealan je za scenarije u kojima je pravodobnost kritična. Dok serijska obrada, s druge strane, uključuje prikupljanje podataka tijekom određenog razdoblja, zatim njihovu obradu kao jednu seriju. Ova je metoda prikladnija za zadatke koji nisu vremenski kritični, gdje je učinkovitije obraditi podatke u velikim dijelovima (Sundararajan, 2022). Iako obrada u stvarnom vremenu omogućuje trenutnu radnju, može biti složenija i zahtijevati više resursa. Serijska obrada je jednostavnija i isplativija, ali joj nedostaje neposrednost sustava u stvarnom vremenu.

Prije same obrade podaci moraju proći kroz proces ETL-a, što je kratica za Extract, Transform, Load- kamen je temeljac skladištenja podataka i analitike desetljećima. Tradicionalno, ETL se bavio izdvajanjem podataka iz različitih izvora, njihovim pretvaranjem u strukturirani format i učitavanjem u bazu podataka ili skladište podataka za analizu. Međutim pojavom velikih podataka i sam proces ETL-a trebao je evoluirati pa je danas korak transformacije u ETL-u ključan za analitičku spremnost jer uključuje čišćenje, deduplikaciju, standardizaciju i obogaćivanje podataka. Ovaj korak osigurava da su podaci točni, dosljedni i u prikladnom obliku za analitičare i podatkovne znanstvenike kako bi došli do značajnih uvida (Sundararajan, 2022).

2.1.6. Podaci o građanima (engl. Citizen data)

Podaci o građanima odnosno osobni podaci generiraju se neprestano kroz interakcije s velikim rasponom javnih i privatnih usluga, sa i bez znanja i pristanka ljudi. Osobni podaci obuhvaćaju sve informacije koje su povezane s pojedincem čiji se identitet može utvrditi ili je već utvrđen. Čak i kad su ti podaci deidentificirani, šifrirani ili pseudonimizirani, ali ih je moguće iskoristiti za otkrivanje identiteta osobe, i dalje se smatraju osobnim podacima i podliježu Općoj uredbi o zaštiti podataka (Europska komisija, 2024).

Uspon internetske ere omogućio nam je da besplatno učimo, tražimo informacije i ostanemo povezani s ljudima diljem svijeta. Teško je zamisliti dan modernog zaposlenog stanovnika bez dnevne ponude vijesti, vremenske prognoze, društvenih medija i elektroničke razmjene poruka. Zašto bi bilo koja profitna korporacija nudila ove usluge besplatno? Zato što im možete dati nešto zauzvrat umjesto novca, a to su vaši podaci. Vaši interesi, vaša lokacija, vaši obrasci ponašanja, vrijeme vaše online aktivnosti - sve su to vrijedna sredstva.

Kako se ti podaci koriste? U dobrom slučaju, pojedinci dobivaju preporučene proizvode i usluge koje stvarno trebaju, osjećaju se zaštićeno dok koriste uslugu dijeljenja automobila i doprinose istraživanju ponašanja. U ne tako dobrom slučaju, utjecaj na pojedinca može biti kroz poticaj na donošenje određenih političkih odluka i uvjeravanje da kupuju proizvode na temelju svoje nesigurnosti. Problem također može nastati i u agresivnom marketingu; kada osjetljivije informacije poput podataka o kreditnoj kartici, privatnih adresa i povijesti bolesti procure i budu ranjive. U oba slučaja, suprotna strana, korporacija koja prodaje svoje proizvode, državna agencija ili politička stranka plaća za dobivanje podataka o građanima.

Primjeri onoga što podrazumijevamo pod podacima o građanima uključuju (Government Office for Science, 2020):

- Jedinstveni identifikatori (npr. broj putovnice)
- Zajednički identifikatori (npr. ime, datum rođenja, adresa)
- Biometrijski podaci (npr. DNK, otisci prstiju)
- Medicinska, obrazovna ili druga evidencija

Podaci generirani kroz interakciju s uslugama ili uređajima:

- Podaci generirani ili promatrani kroz interakciju s uslugama ili uređajima, kao što su o povijest pregledavanja interneta, kolačići za praćenje, IP adrese
- Video podaci (npr. CCTV snimke)

- Podaci o korištenju komunalnih usluga (npr. iz pametnih brojila)
- Podaci generirani kroz interakciju s uređajima Interneta stvari (IoT) (npr. glasovne snimke)
- Podaci o potrošačima (npr. Ponašanje na mrežnoj stranici pri online kupnji)
- Podaci s društvenih mreža
- Podaci o lokaciji (npr. putem aplikacija za praćenje fitnessa)

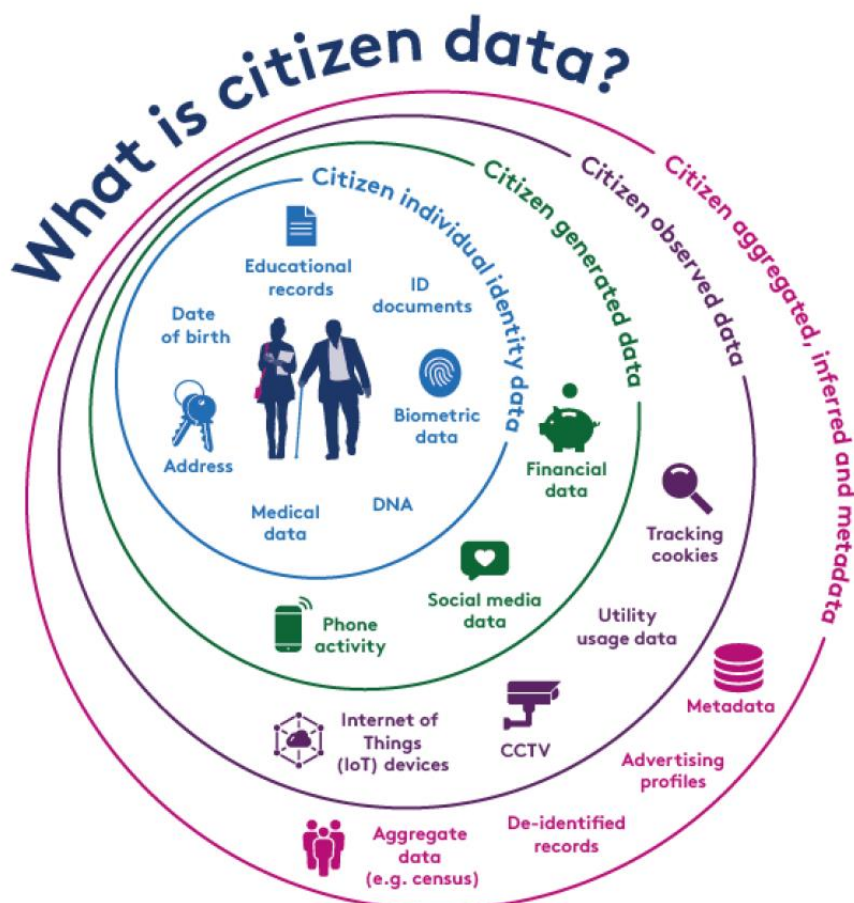
Ostali manje značajni primjeri:

- Zaključci, predviđanja i pretpostavke izvedene iz podataka o ljudima (npr. digitalni profili koji se koriste za ciljano oglašavanje)
- Meta podaci koji se odnose na podatke o ljudima (npr. kada i kako su podaci generirani)
- Deidentificirani podaci (npr. medicinska dokumentacija s uklonjenim ili promijenjenim identifikacijskim poljima, za korištenje u znanstvenom istraživanju i planiranju)
- Zbirni podaci kao što su informacije o popisu stanovništva, čak i ako su navodno anonimizirani

Prilikom pružanja usluge, npr. usluga e-pošte, tvrtka može prikupljati i obrađivati osobne podatke sadržane u e-pošti. Prikupljeni podaci također se mogu odnositi na npr. jezik kojim korisnik govori ili korisnikov mobilni telefon, stvarna lokacija (ili čak informacije specifične za uređaj, kao što su model hardvera, verzija operativnog sustava, jedinstveni identifikatori uređaja i informacije o mobilnoj mreži). Osim toga, kada korisnik pohranjuje svoje digitalne, a ponekad i osobne datoteke koristeći „Cloud Computing“ (na primjer, Dropbox, Google Drive, Sky Drive, i-Cloud) pružatelj, tj. tvrtka koja nudi uslugu u oblaku, može obrađivati podatke sadržane u korisničkim datotekama pohranjenim u „oblacima“ (Morozov, 2012). Konačno, obilje podataka koji se odnose na korisnikovo zdravlje, kretanje ili općenito životne obrasce (npr. broj otkucaja srca, krvni tlak ili čak vrijeme spavanja) mogu se prikupljati i obrađivati sve dok korisnici, u pratnji svojih pametnih uređaja i odabirom iz bezbrojnih aplikacija mjere svoje svakodnevne tjelesne aktivnosti (Bottis & Bouchagiar, 2018).

Stoga bezbrojne online aktivnosti, standardna značajka svakodnevnog života, uključuju proizvodnju i obradu neviđene količine osobnih podataka. Iako je dvojbeno predstavljaju li nečiji zabilježeni otkucaji srca osobne podatke, mnoge, ili možda većina vrsta informacija koje su gore opisane kao primjeri, zapravo su osobni podaci prema Općoj uredbi o zaštiti podataka EU-a. To je zato što u doba velikih podataka, prikupljanje ogromne količine istih omogućuje tvrtkama izvođenje brojnih zaključaka koji se odnose na jednu osobu i omogućuje identifikaciju fizičke osobe. Pod uvjetom da se informacija koju tvrtka prikupi odnosi na fizičku osobu koju je moguće identificirati, izravno ili neizravno, ta je informacija osobni

podatak. Drugim riječima, kriterij koji mora biti zadovoljen, a koji podatke "čini" osobnima nije stvarna identifikacija, već sposobnost da se, izravno ili neizravno, identificira jedna osoba. Ukratko, ako postoji mogućnost identifikacije pojedinca na kojeg se gore spomenuti „snimljeni broj otkucaja srca” odnosi, podaci su osobni, a posebno zdravstveni podaci i u potpunosti su regulirani GDPR-om (Bottis & Bouchagiar, 2018).



Slika 3. Podaci o građanima (Government Office for Science, 2020)

2.1.7. Pravna regulativa velikih podataka

GDPR (engl. *General Data Protection Regulation*) odnosno Opća uredba o zaštiti podataka je uredba Europske unije o privatnosti podataka u Europskoj uniji i Europskom gospodarskom prostoru. GDPR je važna komponenta zakona EU-a o privatnosti i ljudskim pravima, posebno članka 8. Povelje o temeljnim pravima Europske unije (E-građani, 2024). Dana 25. svibnja 2018. godine GDPR je formalno stupio na snagu, okončavajući dvogodišnje prijelazno razdoblje koje je slijedilo nakon konačnog usvajanja regulacije u travnju 2016. godine. GDPR je prozvan "jednim od najrobusnijih zakona o privatnosti podataka na svijetu" koji postavlja novi globalni standard za prikupljanje, pohranu i korištenje podataka (Pardes, 2018).

Njegov dvostruki cilj je "unaprijediti prava pojedinaca na zaštitu podataka i poboljšati poslovne mogućnosti olakšavanjem slobodnog protoka osobnih podataka na digitalnom jedinstvenom tržištu," (Službena internetska stranica Europske Unije, 2020).

Nadalje, Zakon o digitalnim uslugama (engl. The Digital Services Act) ima za cilj stvoriti sigurniji digitalni prostor u kojem su zaštićena temeljna prava korisnika i uspostaviti ravnopravne uvjete za poslovanje. Regulira internetske posrednike i platforme kao što su društvene mreže, platforme za dijeljenje sadržaja, aplikacije za trgovanje te platforme za putovanja i smještaj putem interneta. Glavni cilj je spriječiti ilegalne i štetne aktivnosti na internetu te širenje dezinformacija. Osigurava sigurnost korisnika, štiti temeljna prava i stvara pošteno i otvoreno okruženje na internetskim platformama. Za građane to znači: bolja zaštita temeljnih prava, veća kontrola i izbor, snažnija zaštita djece na internetu te manje izloženosti ilegalnom sadržaju. Od 17. veljače 2024. pravila Zakona o digitalnim uslugama primjenjuju se na sve platforme. Od kraja kolovoza 2023. ta su pravila već bila primjenjivana na određene platforme s više od 45 milijuna korisnika u EU (10% populacije EU), takozvane Vrlo velike online platforme (Europska komisija, 2024).

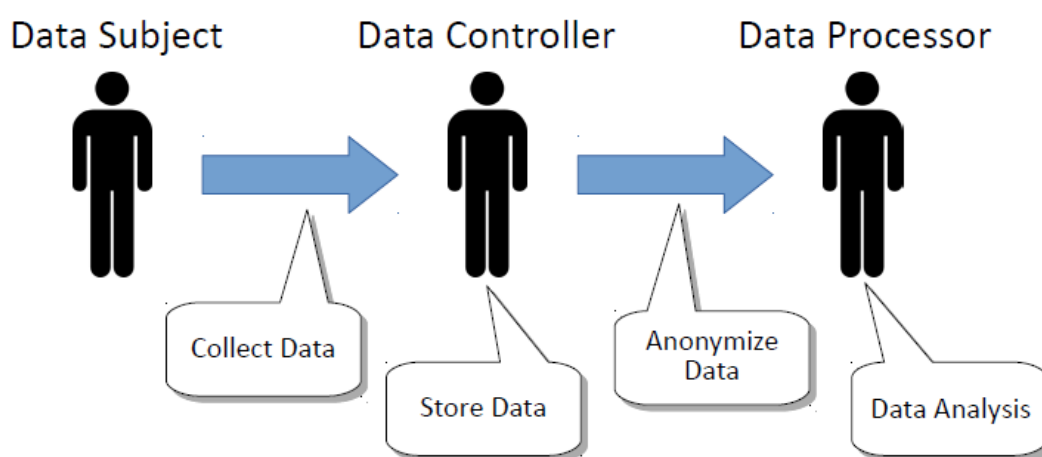
Zaštita fizičkih osoba u vezi s obradom osobnih podataka temeljno je pravo. Članak 8. stavak 1. Povelje o temeljnim pravima Europske unije („Povelja”) i članak 16. stavak 1. Ugovora o funkcioniranju Europske unije (TFEU- Treaty on the Functioning of the European Union) propisuju da *svatko ima pravo na zaštitu osobnih podataka* koji se odnose na njega ili nju (Official Journal of the European Union, 2016). Načela i pravila o zaštiti fizičkih osoba u vezi s obradom njihovih osobnih podataka trebaju, bez obzira na njihovo državljanstvo ili prebivalište, poštivati njihova temeljna prava i slobode, posebno njihovo pravo na zaštitu osobnih podataka. Ova Uredba ima za cilj pridonijeti ostvarenju područja slobode, sigurnosti i pravde ekonomske unije, gospodarskom i društvenom napretku, jačanju i konvergenciji gospodarstava unutar unutarnjeg tržišta te dobrobiti fizičkih osoba. Direktivom 95/46/EZ Europskog parlamenta i Vijeća nastoji se uskladiti zaštita temeljnih prava i sloboda fizičkih osoba u pogledu aktivnosti obrade i osigurati slobodan protok osobnih podataka između država članica (Official Journal of the European Union, 2016).

Brzi tehnološki razvoj u posljednja dva desetljeća donio je nove izazove za zaštitu osobnih podataka. Razmjeri dijeljenja i prikupljanja podataka eksponencijalno su porasli, ponekad se odvijaju na globalnoj razini, a pojedinci sve više svoje osobne podatke čine javno dostupnima. Gospodarska i društvena integracija koja proizlazi iz funkcioniranja unutarnjeg tržišta također je dovela do znatnog povećanja prekograničnog protoka podataka. Kako bi se u potpunosti uzela u obzir sva ova kretanja i promicalo digitalno gospodarstvo, potrebno je osigurati visoku razinu zaštite osobnih podataka, dok je u isto vrijeme omogućeno slobodno kretanje takvih podataka. U slučaju osobnih podataka koji se koriste u svrhe provođenja zakona, postoji sve veća potreba da tijela u državama članicama obrađuju i razmjenjuju

podatke u sklopu borbe protiv transnacionalnog kriminala i terorizma. U tom kontekstu, jasna i dosljedna pravila o zaštiti podataka na razini EU-a poput GDPR-a ključna su za poboljšanje suradnje između tih tijela (European Council, 2024).

S pravnog gledišta, Opća uredba EU o zaštiti podataka (GDPR), koja je stupila na snagu u svibnju 2018. relevantna je za sve organizacije unutar Europske unije (EU), europsko ekonomsko područje (EEA) te također organizacijama iz drugih zemalja, ako obrađuju podatke europskih građana. Dakle, GDPR ima učinak na većinu velikih tvrtki u svijetu. GDPR regulira prikupljanje, pohranjivanje i obradu osobnih podataka. Osobni podaci su svi podaci koji se mogu povezati s određenom fizičkom osobom. To uključuje ne samo izravne osobne identifikatore (npr. puno ime, matični broj), već i neizravne identifikatore poput telefonskih brojeva, IP adresa ili fotografija s osobama koje se mogu identificirati. Podaci koji ne uključuju takve identifikatore obično se smatraju anonimnima i izvan su opsega GDPR-a. GDPR bez davanja precizne ili konkretne definicije anonimnosti smatra skup podataka anonimnim kada je moguća samo ponovna identifikacija s velikim uložnim naporom ili malo vjerojatnim sredstvima (Gruschka i sur., 2018).

U EU, privatnost i zaštita podataka su temeljna prava, jednaka drugim pravima poput slobode izražavanja, i kao takva su zakonski zaštićena. Nisu apsolutna prava, već se uravnotežuju s drugim pravima, zakonima EU-a i prioritetima poput nacionalne sigurnosti - ali moraju biti predviđena od strane nacionalnih vlada i izražena su u regulativama na razini EU, uključujući GDPR. Ta prava mogu se smatrati sastavnim dijelom formiranja integriranog digitalnog jedinstvenog tržišta unutar EU-a, omogućujući slobodan protok osobnih podataka te harmoniziran pristup koji pruža stabilnost poslovnim operatorima, a istovremeno osnažuje građane i gradi povjerenje u upotrebu podataka (Official Journal of the European Union, 2016).



Slika 4. Proces obrade osobnih podataka (Gruschka i sur., 2018)

Za obradu osobnih podataka GDPR definira niz pravnih, organizacijskih i tehničkih zahtjeva te predlaže različite metode. Ovdje su opisana najrelevantnija načela (Gruschka i sur., 2018). Prije svega, u većini

slučajeva obrada osobnih podataka dopuštena je samo ako je nositelj podataka dao privolu. Iznimke se primjenjuju kada je obrada podataka izričito dopuštena zakonom ili propisom ili osigurava „vitalne interese nositelja podataka“. Dodatno, dana privola mora biti ograničena na određenu svrhu obrade podataka. Kontrolor podataka (subjekt koji je odgovoran za prikupljanje podataka) ne može definirati previše generičku svrhu obrade podataka niti naknadno samovoljno promijeniti namjenu. Drugo načelo obrade podataka je minimizacija podataka koja se odnosi na ograničavanje prikupljanja, pohranjivanja i korištenja osobnih podataka na podatke koji su relevantni, primjereni i što je još važnije nužni za provedbu svrhe za koju se podaci obrađuju. Vrijedno je napomenuti da se pseudonimizacija izričito spominje kao mjera minimiziranja podataka. U pseudonimiziranim podacima, parametri koji se mogu identificirati zamjenjuju se drugim (nasumično) generiranim identifikatorima. To obično nema nikakav negativan utjecaj na proces ruđenja podataka i po mogućnosti bi ga trebao pokrenuti voditelj obrade podataka prije prijenosa podataka procesoru podataka. Ako je potrebno povezati rezultate obrade podataka, to može postići kontrolor podataka, budući da sadrži pseudonime mapiranja (također poznate kao pseudo-pretraživačka tablica) prema parametrima koji se mogu identificirati. Osim anonimiziranih podataka, pohrana kod voditelja obrade mora biti u skladu s GDPR-om korištenjem tehnika koje štite podatke u mirovanju (npr. šifriranje i čvrsta kontrola pristupa). Nadalje, GDPR zahtijeva „odgovarajuće tehničke i organizacijske mjere kako bi se osigurala razina sigurnosti primjerena riziku“, što obično uključuje primjenu tehnika poput enkripcije podataka, kontrole pristupa, fizičke zaštite i (opet) pseudonimizacije. Proširenje načela minimizacije podataka je načelo ograničenja pohrane, koje ograničava trajanje pohrane podataka na određeno (potrebno) razdoblje. U kontekstu obrade podataka, nadalje se mora uzeti u obzir da procesi automatskog donošenja odluka s utjecajem na pojedince (članak 22.) kao i obrada iznimno osjetljivih podataka, poput biometrijskih podataka (članak 9.), zahtijevaju „izričit“ pristanak subjekta podataka. Novčane kazne koje se mogu nametnuti kao kazne za nepoštivanje temelje se na određenim člancima Uredbe koje je organizacija prekršila. GDPR također daje pojedincima pravo na naknadu za materijalnu i/ili nematerijalnu štetu nastalu kršenjem Uredbe. Konkretno, kontrolori i obrađivači podataka suočavaju se s administrativnim kaznama (Gruschka i sur., 2018):

- do 10 milijuna eura ili 2% godišnjeg globalnog prometa (koji god je iznos veći) za povrede članaka: 8 (uvjeti za pristanak djece), 11 (obrada koja ne zahtijeva identifikaciju), 25-39 (opće obveze procesora i kontrolera), 42 (certifikacija), i 43 (certifikacijska tijela).
- do 20 milijuna eura ili 4% godišnjeg globalnog prometa (koji god je iznos veći) za povrede članaka: 5 (načela obrade podataka), 6 (pravne osnove za obradu), 7 (uvjeti za privolu), 9 (obrada od posebne kategorije podataka), 12-22 (prava nositelja podataka), i 44-49 (prijenos podataka u treće zemlje).

2.1.8. Društvena kvantifikacija kroz povijest

Društvena kvantifikacija je proces mjerenja i analize socijalnih fenomena, interakcija i ponašanja pojedinaca unutar društva pomoću kvantitativnih metoda i podataka. Ova praksa uključuje upotrebu brojeva, statističkih alata i matematičkih modela kako bi se dobila kvantitativna i objektivna perspektiva o društvenim pojavama. Društvena kvantifikacija može obuhvaćati istraživanje društvenih mreža, analizu javnog mnijenja, ekonomske trendove i druge aspekte ljudskog ponašanja unutar društvenog konteksta (Wiedemann, 2013). Ovaj pristup omogućuje dublje razumijevanje društvenih dinamika kroz sistematičko prikupljanje, obradu i interpretaciju kvantitativnih podataka. Postoje najmanje tri vala društvene kvantifikacije.

Prvi val društvene kvantifikacije podrazumijeva brojeve i statistiku koju su koristili europski narodi za kategorizaciju stanovništva u devetnaestom stoljeću. Rane prakse vodili su prvenstveno vlade i javni sektori, pokušavajući prebrojati ljude i dalje generirati znanje o tom društvu. Tijekom kasnog devetnaestog stoljeća, potrošački kreditni rejting prvotno je razvijen u Sjedinjenim Državama kako bi se ocijenila kreditna sposobnost pojedinaca i tvrtki. Početkom 20. stoljeća, uspon socijalnih država dodatno je unaprijedio prakse kvantifikacije poput popisa stanovništva i javne statistike u Europi i Sjedinjenim Državama (Lazarsfeld, 2019).

Drugi val razvoja dogodio se 1960-ih kada su agencije za kreditni rejting postupno usvajale relacijske baze podataka i informacijsku infrastrukturu za kompjutorizaciju financijskog kreditnog izvješćivanja. Tome je pridonio i tzv. pokret društvenih pokazatelja čiji je cilj mjerenje društvenih vrijednosti pomoću brojeva i podataka. U usporedbi s prvim valom, privatne korporacije i tržišta igraju sve važniju ulogu u pretvaranju društvenog života u metrike i podatke. Sustave kreditnog bodovanja (npr. FICO) i stalni nadzor razvili su privatni sektori za procjenu hoće li ljudi kasniti s plaćanjem svojih dugova i za izgradnju „bodovanog društva” (Citron i Pasquale, 2014). Dok je prvi val kvantifikacije usmjeren na brojanje ljudi na razini populacije, drugi val bio je usmjeren na stvaranje profila i mjerenja na razini pojedinca.

Veliki podaci i računalne revolucije promiču razvoj trećeg vala nudeći automatizirano donošenje odluka i masivne podatke ukorijenjene u institucionalnim infrastrukturama i digitalnim platformama. Masivne baze podataka se agregiraju i istražuju radi identificiranja obrazaca i korelacija koje su prije bile nemoguće. Na primjer, kreditna izvješća agencije su proširile sustave bodovanja prikupljanjem podataka koji se odnose na nefinancijska plaćanja, digitalne otiske i psihometrijsko testiranje. Medijske industrije također usvajaju digitalne alate za poboljšanje analize publike. Štoviše, tehnologijom tvrtke sada mogu predvidjeti individualna ponašanja putem prediktivnih algoritama. Kao rezultat toga, ljudi se stalno procjenjuju u kontekstu statističke procjene i dalje transformiraju u automatizaciju i statističku analitiku (Wiedemann, 2013). Društvena kvantifikacija kao oblik socijalne kontrole stvara načine gledanja i poznavanja ljudi i dio je društvene infrastrukture koja regulira i usmjerava pojedinca i organizacijsko ponašanje (Espeland &

Sauder, 2007).

Indija je na primjer pokrenula biometrijsku identifikacijsku bazu podataka pod nazivom Aadhaar za pružanje javnih usluga i ocjenjivanje građana (Bhadani & Jothimani, 2016). Provedba zakona u SAD-u koristi analitiku velikih podataka za izračunavanje rezultata rizika i transformaciju praksi nadzora (Brayne, 2017). Nadalje, EU je predložio projekt iBorderCtrl za otkrivanje prijevare i jačanje granične kontrole pomoću umjetne inteligencije (Sánchez-Monedero&Dencik, 2019). Očito je da se trenutna **društvena kvantifikacija usredotočuje na automatizirane tehnologije i skladišta podataka za procjenu i nadzor pojedinaca** u stvarnom vremenu. Nadalje, Narodna Republika Kina (NR Kina) prakticira društvenu kvantifikaciju za procjenu i klasifikaciju ljudi (Leibold, 2020). Dang'an (osobni dosjei), na primjer, koriste se za prikupljanje osobnih podataka poput obrazovanja i zaposlenja. Drugi sustav Hukou (registracija kućanstva) može odrediti imaju li ljudi pristup stanovanju, zdravstvenoj zaštiti i obrazovanju u gradovima. Nedavno je Kina agresivno koristila digitalne tehnologije za nadzor i kategorizaciju građana. Na primjer, lokalne su vlasti pokrenule "datoteke o moralu građana" za bilježenje i procjenu dobrih djela susjeda. Godine 2020. kineske su vlade surađivale s tehnološkim divovima Alibaba i Tencent za sprječavanje širenja COVID-19 kvantificiranjem povijesti putovanja i zdravstvenog stanja ljudi (Jiang, 2020). Etničko razvrstavanje, dronovi i GPS uređaj za praćenje korišteni su za jačanje nadzora u Xinjiangu (Leibold, 2020). Razvoj ovih inicijativa ukazuje kako se uvelike ulaže u iskorištavanje digitalnih tehnologija za društvenu kvantifikaciju i društveno upravljanje.

2.1.9. Podatkovni sistemi

Pojam podatkovni sustav imaće različite asocijacije za one s različitim strukovnom pozadinom. Tehnolozi i znanstvenici mogu razmišljati o računalnim sustavima; poslovni i IT stručnjaci mogu razmišljati o arhitekturi poduzeća; pravni i politički analitičari mogu razmišljati o zakonima, propisima i politici koja oblikuje što se može, a što ne može učiniti s podacima. Drugi mogu imati posve drugačije koncepcije pojma. Takve varijacije u stajalištima i jeziku ključni su izazov u razumijevanju problema koji se odnosi na cijeli sustav. Zauzimamo široki pogled na podatkovne sustave, što uključuje sve sljedeće i više od toga: ljudi, procesi i tehnologije uključeni u prikupljanje, otkrivanje, pohranjivanje, analiziranje, povezivanje i dijeljenje podataka o građanima. To uključuje pojedince koji prikupljaju podatke (npr. u interakciji između liječnika i pacijenta), senzore i druge tehnologije za generiranje podataka, tehničke standarde, sustave koji pohranjuju i obrađuju podatke i stupanj do kojeg se skupovi podataka konsolidiraju i/ili dijele između različitih strana (Government Office for Science, 2020).

U sljedećem dijelu obradit će se jedan od najvećih svjetskih podatkovnih sustava - Kina te će se prikazati načini prikupljanja, korištenja i dijeljenja podataka o građanima.

Navest će se nekoliko primjera kako Kina prikuplja i iskorištava podatke o svojim građanima u različite

svrhe. Projekti vlade u Kini se usredotočuju na podatke u svrhu sigurnosti. Program Skynet koristi podatke s nadzornih kamera diljem zemlje i tehnologiju prepoznavanja lica u svrhu provođenja zakona i borbe protiv terorizma. Ovo se proširuje projektom Sharp Eyes, čiji je glavni cilj proširenje pokrivenosti na ruralna područja, uključujući korištenje mobilnih telefona i televizora. U nekim slučajevima gdje su ovi sustavi implementirani, prijavljeno je smanjenje kriminala. Izvještaji sugeriraju da je Kina najveće tržište za opremu za sigurnost i nadzor, s velikim državnim nabavama (Creemers, 2018).

Nadalje **sustav društvenog kreditiranja u Kini primjer je kako se digitalne tehnologije, veliki podaci, inteligentni algoritmi te računalne platforme mogu koristiti za oblikovanje ekonomskog i društvenog ponašanja**. Sustav koristi digitalne platforme i analitiku podataka za praćenje i ocjenjivanje ponašanja građana, uključujući financijske transakcije, online aktivnosti i društvene interakcije. Ovi se podaci koriste za dodjeljivanje rezultata koji mogu utjecati na sposobnost osobe da pristupi određenim uslugama, poput kredita, zajmova ili putnih dozvola. **Kina danas suočava s nestabilnošću financijskog kreditnog sektora**, ograničenjima financijskih usluga te s prijetnjom društvene nestabilnosti. Kako bi tim problemima stala na kraj godine 2014. Kinesko državno vijeće najavilo je provedbu takozvanog 'Socijalnog kreditnog Sustava' (Social Credit System), koji bi svakom kineskom građaninu i poduzeću pridružio 'Osobni kreditni rezultat' (Hvistendahl,2017). Prema (Creemersu, 2018.) ovisno o nečijem PCS-u (Personal Credit Score), dodjeljuje se ponašanje koje se smatra "poželjnim" ili "nepoželjnim", pojedinci će biti nagrađeni ili kažnjeni (npr. jeftiniji/skuplji krediti, brži/sporiji birokratski procesi, stavljanje na crvenu listu/ na crnu listu itd.) Kao rezultat, 'poželjna' ponašanja bi se poticala, a 'neželjena' bi se obeshrabrivala. Ovakav sistem koristi online i offline tehnologiju nadzora koja se često prikazuje kao "alat za masovni nadzor" koji olakšava razvoj ove mnogoljudne države. Ovakav sustav temelji se na prikupljanju maksimalne količine podataka o građanima i procjeni pouzdanosti stanovnika na temelju njihovog financijskog, društvenog i online ponašanja. Navodno je namijenjen agregiranju financijskih, policijskih, komercijalnih, društvenih medija i drugih podataka kako bi se pratila usklađenost pojedinaca i poduzeća s pravnim, moralnim i stručnim obvezama. Postoje dokazi da se podaci SCS-a dijele s poduzećima, a vlada koristi crne liste da spriječi dužnike da kupuju luksuznu robu, kao i dokazi o prikupljanju velikih količina podataka o pojedincima, poput DNA baza podataka i obavezne biometrije za stanovnike regije Xinjiang (Leibold, 2020). O sustavu socijalnog kreditiranja u europskoj literaturi mogu se pročitati samo kritički stavovi, iako su mišljenja osoba koje su pod tim sustavom – kineskih državljana – prilično pozitivna. U tom kontekstu ne treba zaboraviti na veliku razliku u mentalitetu Azijata i Europljana (Bazina,2020).

U Kini, zakoni o privatnosti i zaštiti podataka se razvijaju, ali uglavnom se koriste za kibernetičku i nacionalnu sigurnost. Glavni zakon u ovom području je Zakon o kibernetičkoj sigurnosti iz 2016. godine, koji propisuje lokalizaciju podataka, ograničava prijenos podataka preko granica, te utvrđuje mjere za osiguranje podataka od kibernetičkih napada i zloupotreba. Dolazi s standardom za zaštitu podataka koji

pruža smjernice o zahtjevima za pristanak, deidentifikaciji, pravima na brisanje i minimalizaciji podataka, između ostalog. Nadalje Zakon o elektroničkoj trgovini iz 2018. godine proširuje ovo pravo pružanjem potrošačima prava na ispravak i brisanje korisničkih informacija te na kraju PIPL (Personal Information Protection Law) koji označava Zakon o zaštiti osobnih podataka. To je zakon koji je stupio na snagu 1. rujna 2021. godine i predstavlja značajno zakonodavstvo o zaštiti privatnosti i osobnih podataka u Kini. Navedeni zakon postavlja jasne smjernice i zahtjeve za prikupljanje, pohranu, obradu i prijenos osobnih podataka te postavlja pravila koja se odnose na prava pojedinaca u vezi s njihovim osobnim podacima. Međutim i dalje nije poznato koliko se ovaj zakon provodi unutar Kineske vlade i korporacija koji i dalje imaju omogućen pristup svim podacima o građanima, pogotovo ukoliko se koriste za nacionalnu sigurnost. Industrijske inicijative, poput Plan razvoja AI iz 2017., "Made in China 2025" i "Internet Plus", imaju cilj učiniti Kinu globalnim liderom u naprednim tehnologijama temeljenim na podacima, poput umjetne inteligencije i informacijskih i komunikacijskih tehnologija (Kostka, 2019).

<i>Individual Rewards</i>	<i>Individual Punishments</i>
Lower tax rates	Travel restrictions
Discounts on utilities	Blocking purchases of train/plane tickets
Deposit-free rentals	Visa restrictions
Lower interest rates	Hotel restrictions
Faster check-in at hotels and airports	Throttled internet speeds
Faster internet speeds	Restricted access to higher education
Increased access to public services	Job restrictions
Discounts on public transportation	Public shaming and blacklisting
Faster processing of travel visas	Credit restrictions
Shorter wait time at hospitals	Higher taxes and loan interest rates
Increased visibility on dating apps	Restrictions on property ownership
<i>Firm Rewards</i>	<i>Firm Punishments</i>
Commendations and positive publicity	Warnings
Removal of red tape and reduction of state regulation	Blacklisting mechanisms
Access to markets for public services	Market withdrawal and shutdown of e-commerce accounts
Preferential bidding on public contracts	Circulation of criticism to business partners
Granting of accreditations and qualification certifications	Public shaming/censure
Policy support	Red tape and increased administrative burdens
Administrative approvals	Unfavourable loan conditions
Tax incentives	Higher taxes than compliant competitors
Access to preferential credit services	Restrictions on stock or bond investments
Access to investment opportunities	Decreased opportunity to participate in publicly funded projects
Open markets and unrestricted foreign investment opportunities	Mandatory government approval for investments, even in sectors where market access is not usually regulated
Expedited processing of permits and visas	Managers denied tickets for high-speed rail or international business flights

Tablica 1. „System of carrots and sticks“ (Hvistendahl, 2017)

2.1.10. Komparativan prikaz praksi korištenja europskog gdpr-a i kineskog scs

Tehnološki divovi, potpomognuti slabim regulatornim nadzorom, proširili su kapacitete za prikupljanje i analizu osobnih podataka. To je rezultiralo novim skupom dinamika moći i logike akumulacije koje se zajedno nazivaju nadzornim kapitalizmom. Kao odgovor, EU i Kina su usvojile ključne politike o velikim podacima s implikacijama za budući društveni i ekonomski razvoj. Opća uredba o zaštiti podataka (GDPR) u Europi predstavlja reaktivni odgovor, afirmirajući pojedinačnu privatnost i postavljajući ograničenja za korporativnu upotrebu osobnih podataka. Nasuprot tome, kineski sustav društvenog kreditiranja (SCS) predstavlja proaktivni odgovor, kombinirajući nadzorne arhitekture i tehnologije umjetne inteligencije u svrhu vođenja države (Aho & Duffield, 2020).

U posljednjih nekoliko godina, Kina i Europska unija krenule su različitim putovima u svojim digitalnim ekonomijama kroz ključne politike o velikim podacima. Kineski sustav socijalnog kreditiranja (SCS) pozicioniran je kao značajan projekt socijalnih i ekonomskih reformi s ciljem oblikovanja razvoja nacije u doba informacija. S druge strane, Europska opća uredba o zaštiti podataka (GDPR) djeluje kao sveobuhvatan okvir za zaštitu podataka i privatnosti, utječući na to kako tvrtke i države prikupljaju i koriste osobne podatke. Ovi projekti upravljanja predstavljaju radikalno različite pristupe konceptualizaciji podataka, potaknuti globaliziranim širenjem 'nadzorne kapitalizacije'. Taj fenomen, karakteriziran neuređenim prikupljanjem i analizom podataka od strane tehnoloških tvrtki, rezultirao je koncentracijom moći nad društvenim i ekonomskim ponašanjima diljem svijeta. SCS i GDPR javljaju se kao odgovori na taj kontekst, označavajući tvrdnje o državnoj kontroli nad digitalnim sektorima s dubokim implikacijama za društveni i ekonomski napredak. Usporedba SCS-a i GDPR-a otkriva šire normativne izjave o tome kako svako upravno tijelo percipira velike podatke i povezane tehnologije. Kao rezultat toga, Europa pokušava ograničiti utjecaj nadzorne kapitalizacije putem GDPR-a, dok Kina prihvaća njegove logike za povećanu državnu upotrebu, stvarajući dva različita oblika kapitalizma koji djeluju na temeljito različitim setovima logika (Aho & Duffield, 2020). Od iznimke je važnosti shvatiti kako tehnologije velikih podataka preoblikuju društva i kako društva reagiraju na preuređivanje tih tehnologija.

S obzirom na brzu digitalizaciju **kineskog gospodarstva**, računalni znanstvenik Kai Fu Lee (2018) opisuje Kinu kao 'Saudijsku Arabiju podataka', aludirajući na nevjerojatnu količinu podataka koje kineski građani proizvode svakog dana. Tvrdi da je dubina digitalne integracije u Kini neusporediva u odnosu na ostatak svijeta, budući da mobilna plaćanja zamjenjuju gotovinu u većini svakodnevnih ekonomskih transakcija. S obzirom na to da su pametni telefoni učinili svakodnevni život razumljivim putem tehnoloških arhitektura nadzorne kapitalizacije, kineska država sada želi koristiti te kapacitete kao novi izvor moći i kontrole. Prema Kineskom državnom vijeću cilj sustava je "široko oblikovanje guste atmosfere u cijelom društvu u kojoj se održavanje povjerenja i lojalnosti potiče odnosno smatra se poželjnim, dok je njihovo kršenje sramotno, osiguravajući tako da iskrenost i pouzdanost postanu svjesne norme ponašanja među svim ljudima" (Kostka, 2019). Kako bi se postiglo željeno društveno i tržišno ponašanje, SCS nastoji osigurati

dobro ponašanje između ekonomskih subjekata, kao i usklađenost s propisima i sudjelovanje u vladinim programima i politikama. Ponašanje subjekata kontrolira se poticanjem dobrog ponašanja i poticanjem dobrovoljnog usklađivanja i sudjelovanja u državnim programima i politikama - vizijom samoregulacije SCS-a. Na mnogo načina, SCS se može promatrati kao sustav upravljanja ponašanjem čija funkcionalnost proizlazi iz principa prisilne samoregulacije koji su prvi predstavljeni u Benthamovu prijedlogu panoptikona iz 1791. godine, u kojem se subjekte uvodi u percepciju stalnog nadzora snagom arhitektonskog dizajna, mijenjajući njihovo ponašanje ne silom, već oblikom psihološke manipulacije. Bez obzira jesu li subjekti kriminalci, potrošači ili tvrtke, pod budnim okom Velikog Brata, dolazi do samoregulacije ponašanja. Kao što Benthamov panoptikon povećava učinkovitost zatvora dok smanjuje broj čuvara, SCS predstavlja sredstvo za povećanje učinkovitosti državnih institucija, dok se istodobno odriče potrebe za proširivanjem institucijskog osoblja (Zuboff, 2015). S obzirom na sposobnost praćenja i prilagodbe SCS-a u stvarnom vremenu zahvaljujući digitalnom prikupljanju podataka, sustav će omogućiti kineskoj državi mogućnost brze implementacije novih politika i programa. Vlada može brzo promatrati učinke svojih politika i intervencija te se prilagođavati kako bi maksimizirala učinak, omogućujući proces brzog učenja u regulatornom procesu.

S obzirom na veliki obujam prikupljenih podataka, SCS bi s vremenom mogao duboko zadirati u živote građana i pružiti niz uvida u politike povezane s dobrobiti građana i zaposlenika. Primjerice, prakse zdravlja zaposlenika u tvrtki mogle bi se mjeriti kroz broj transakcija koje zaposlenici obavljaju u medicinskim ustanovama. Visoki troškovi radne snage na lijekove i zdravstvenu skrb mogu služiti kao pokazatelj loših praksi zdravlja na radnom mjestu, te bi socijalna kreditna ocjena nezdrave tvrtke bila smanjena sukladno tome. Da bi podigla ocjenu, tvrtka bi bila potaknuta na usvajanje politika koje poboljšavaju relevantne uvjete rada, a uspjeh tih politika mogao bi se analizirati u stvarnom vremenu nastavkom praćenja medicinskih transakcija zaposlenika. Osim toga, putem nadzora osobnih komunikacija i društvenih medija, digitalni algoritmi teoretski bi mogli mjeriti stavove i mentalno zdravlje radnika, omogućujući stvaranje složenih metrika koje bi potencijalno mogle revolucionirati odnose na radu i prakse ljudskih resursa. Potencijali za regulaciju tvrtki praktički su neograničeni unutar sustava koji dopušta neograničeni nadzor i intervenciju (Aho & Duffield, 2020).

Kako to primjećuje kineski milijarder Jack Ma Yun, osnivač Alibaba Grupe: „U proteklih 100 godina došli smo do uvjerenja da je tržišna ekonomija najbolji sustav, ali po mom mišljenju, u sljedeća tri desetljeća doći će do značajnih promjena, a planirana ekonomija postat će sve veća. Zašto? Jer, uz pristup svakoj vrsti podataka, možda ćemo moći pronaći nevidljivu ruku tržišta“, (Thornhill, 2017). Sa ogromnim količinama podataka, povratnim informacijama u stvarnom vremenu i algoritmima strojnog učenja koji mogu obrađivati i razumjeti rezultate, SCS može pomoći tvrtki da odgovori na fluktuacije tržišta gotovo trenutno. Koristeći SCS, ekonomski planeri sada imaju potencijal usmjeravati ponašanje tvrtki

mijenjanjem nagrada i kazni kako bi odgovarale promjenama globalnih gospodarskih okolnosti. U trenucima ekonomske recesije, SCS bi se možda prilagodio tako da tvrtke suočavaju s manje kazni. Na sličan način, nagrade se mogu ponuditi kao oblik ekonomskog poticaja. Ove mikrointervencije mogu se prilagoditi pojedinačnim industrijama i mogu se koordinirati širom cijele ekonomije. Svaka intervencija i podešavanje SCS-a sakuplja ogromne količine podataka i analizira ih kako bi se razumjele kako vidljive, tako i nevidljive posljedice djelovanja, poboljšavajući time razumijevanje projekta ekonomije kao sustava i sposobnost implementacije učinkovitije akcije u budućnosti. Što je duže vrijeme sustav u funkciji, to postaje učinkovitiji (Creemers, 2018).

Unificirajući prakse digitalne zaštite gledajući **europsko gospodarstvo**, opći cilj GDPR-a je povećati stupanj kontrole koji obični građani imaju nad svojim osobnim podacima, te razumijeti kako se oni prikupljaju i koriste u doba gdje su podaci postali jedan od najvažnijih inputa. Regulacije postavljaju značajna ograničenja na način kako tvrtke mogu prikupljati i koristiti osobne podatke, djelotvorno ometajući prakse nadzornog kapitalizma. Individualizam se smatra važnim sastavnim dijelom političkog djelovanja prisutnog u mnogim modernim, prosperitetnim zapadnim društvima. Unutar liberalne demokratske države, svakodnevne odluke i preferencije građana temelje se prije svega na ostvarivanju osobnih interesa, a jedna od glavnih uloga države je zaštita individualnih sloboda. Tijekom doba prosvjetiteljstva kasnog 17. i 18. stoljeća u zapadnoj Europi i šire, mislioci poput Voltairea, Rousseaua, Lockeja, Humea i Kanta razvili su koncept osobne slobode koji će imati dubok utjecaj na zapadne političke, društvene, ekonomske i kulturne prakse. Ideologija liberalizma kristalizirala se oko ideje društva utemeljenog na prirodnom pravu nasuprot predrasudama, privilegijama i tiraniji; razumu i sekularnosti nasuprot vjeri; inherentnoj dobroti čovjeka; toleranciji; te državi kao instrumentu neprestanog napretka. Filozof John Locke opisao je mjesto pojedinca unutar toga, definiranog društvenim ugovorom; suradnjom između građanina i vlade koja osigurava funkcionalno društvo i osobna prava njegovih sastavnica kroz zastupanje i vladavinu prava (Satra, 2019). Iako je pojednostavljenje propisa za tvrtke značajan motiv za razvoj GDPR-a, u središtu zakonodavstva je potvrda prava pojedinca kao glavne jedinice putem koje se izražava sloboda. Naime, GDPR je opisan kao "plemenit i nužan" zakon koji "štiti osobne podatke viđene kao atribut pojedinca i vraća nam prava i slobode". Glavni strateški stručnjak za kibernetičku sigurnost Tom Pendergast prikazuje GDPR kao najnoviji potez u "novom hladnom ratu oko zaštite podataka" gdje su osobni podaci "valuta modernog doba". Njegove riječi su: „Natječu se one društva koja vjeruju da pojedinci imaju apsolutno pravo kontrolirati svoje osobne podatke - vježbati istu vrstu vlasti nad podacima kao što to čine nad vlastitim tijelima ili vlastitim vlasništvom - i one koje vjeruju da su osobni podaci roba koja se može trgovati na otvorenom tržištu i stoga podložna istim tržišnim silama kao i drugdje...“ EU čvrsto stoji na strani interesa pojedinca (Tankard, 2016).

Međutim i dalje zabrinjava zloupotreba podataka od strane države, kao što je već spomenuto u slučaju

marginalizirane ujgurske populacije u Xinjangu i potencijalno na širim razinama. Nadzor velikih podataka poboljšava zahvaćanje i kategorizaciju razlika, stvarajući potencijal za sistemsku socijalnu polarizaciju kako se ljudski subjekti identificiraju i sortiraju prema vrijednosti i riziku stvarajući digitalne podklase (Hvistendahl,2017). GDPR se suočava s sličnim logističkim i operativnim izazovima s obzirom na njegov ambiciozan opseg. Prema priznanju Europske unije, tvrtke se čini da GDPR više tretiraju kao pravnu zagonetku, kako bi sačuvale svoj način rada; umjesto da prilagode svoj način rada kako bi bolje zaštitile interese onih koji koriste njihove usluge. Doista, prema pravilima GDPR-a, "voditelj regulatora" multinacionalnih tvrtki mora se nalaziti u zemlji gdje tvrtke imaju svoje "glavno sjedište", što je za većinu velikih tvrtki, uključujući Google, Facebook, Twitter i Microsoft, Irska. Unatoč tisućama navodnih povreda privatnosti podataka, Irska Agencija za zaštitu podataka sporije poduzima mjere provedbe, radi čega su neki izrazili zabrinutost zbog regulatornog zarobljeništva. Ova nesrazmjernost između identifikacije i djelovanja čini se uobičajenom diljem EU-a, s izvješćima koji ukazuju na više od 60 000 obavijesti o povredi podataka, ali samo 61 izrečenu kaznu, uglavnom niske vrijednosti (Kretschmer i sur.,2021).

GDPR na snazi je od svibnja 2018. godine. Kao jedan od najopsežnijih zakonodavnih akata o privatnosti, izazvala je mnogo rasprava o utjecaju koji bi mogla imati na korisnike i pružatelje usluga na internetu, posebno zbog velike količine osobnih podataka obrađenih u tom kontekstu. Osim toga, sažimamo da online usluge češće pružaju sredstva svojim korisnicima za odustajanje od obrade podataka, ali redovito otežavaju praktičan pristup takvim sredstvima putem suviše složenog i ponekad nezakonitog oblikovanja sučelja. Bitno je spomenuti da većina web usluga nije primijenila dovoljne promjene u svojim politikama osobnih podataka kako bi u potpunosti bile usklađene s GDPR-om. Iako su većina web stranica ažurirale obavijesti o svojoj politici privatnosti, povećavajući time transparentnost i primjetno je povećana upotreba obavijesti o kolačićima, rijetko se pruža potrebna kontrola nad osobnim podacima korisnicima. Ova situacija može proizaći iz neslaganja između pružatelja usluga i tijela za zaštitu potrošača u vezi s tumačenjem GDPR-a (Kretschmer i sur.,2021).

Hu i Sastry (2019) ističu da mnoge web stranice u EU koriste "tamne obrasce", manipulativne dizajne sučelja, koji utječu na korisnike da pristanu na praćenje suprotno njihovim željama. To naglašava da kvantitativni učinci opaženi možda stvarno ne predstavljaju informirane i dobrovoljne izbore korisnika. Takve manipulativne prakse također otkrivaju da 30% web stranica ne pruža mogućnost odbijanja kolačića, dok više od 40% olakšava njihovo prihvaćanje. Korisnici često imaju malo praktičnog izbora kad se suoče s obavijestima o kolačićima, jer su dizajnirane kako bi poticale prihvaćanje praćenja putem kolačića. S obzirom na ove čimbenike, promjene u praćenju trećih strana i korištenju kolačića možda se ne mogu isključivo pripisati GDPR-u. Postojanost kolačića za praćenje i prevalencija tamnih oblika sugeriraju da su potrebni dodatni napor kako bi se osiguralo da korisnici mogu donositi iskrene i informirane odluke o svojoj privatnosti (Kretschmer i sur.,2021).

Uz vaš pristanak, mi i [naš 779 partneri](#) koristimo kolačiće ili slične tehnologije za pohranu, pristup i obradu osobnih podataka kao što su Vaša posjeta ovoj web stranici, IP adrese i identifikatori kolačića. Neki partneri ne traže Vaš pristanak za obradu Vaših podataka i oslanjaju se na svoj legitimni poslovni interes. Možete povući svoj pristanak ili se usprotiviti obradi podataka na temelju legitimnog interesa u bilo kojem trenutku klikom na "Saznajte više" ili u našoj Pravilima o privatnosti na ovoj web stranici.

Mi i naši partneri obrađujemo podatke kako slijedi:

Personalizirano oglašavanje i sadržaj, mjerenje oglašavanja i sadržaja, uvidi u publiku i razvoj usluga, Pohrana i/ili pristup podacima na uređaju, Precizni geolokacijski podaci i identifikacija putem skeniranja uređaja



Slika 5. Primjer prihvaćanja „kolačića“ na internetskoj stranici tportal.hr

2.2. Prednosti tehnologije velikih podataka

Organizacije diljem svijeta sve više shvaćaju da će sposobnost analize i korištenja velikih i složenih skupova podataka biti najvažniji izvor konkurentske prednosti u 21. stoljeću. Veliki podaci imaju potencijal pružiti bolje iskustvo korisnicima, poboljšati internu učinkovitost te konačno unaprijediti profitabilnost i konkurentnost organizacija u svim industrijama. Organizacije mogu koristiti velike podatke kako bi postale pametnije i inovativnije na načine koji nisu bili mogući prije dolaska "zettabyte ere" (LaValle i sur., 2011). U ovom poglavlju prikazat će se sve prednosti koje veliki podaci donose za pojedinca od opće sigurnosti, financijskih dobiti do zdravstvenih koristi.

2.2.1. Poboljšanje proizvoda i korisničke usluge

Danas je postalo neophodno za organizacije kako bi ostale konkurentne u brzo rastućoj i promjenjivoj okolini analizirati i optimizirati korisničko iskustvo te na taj način unaprijediti kvalitetu svojih proizvoda. Da bi uspjele u tome moraju poznavati tehnologiju i mogućnosti analize velikih podataka. Postoje brojne strategije koje to omogućuju poput analize korisničkog ponašanja gdje tvrtke prikupljaju podatke o interakcijama korisnika s proizvodima ili uslugama kako bi stvorile dublje razumijevanje njihovih potreba, navika i preferencija. Analizom velikih podataka koji su uglavnom dohvaćeni online organizacije prilagođavaju svoje proizvode prema stvarnim potrebama svojih korisnika. Putem online ponašanja korisnika odnosno „klika miša“ danas tvrtke mogu predvidjeti buduće ponašanje kupca na njihovoj web stranici te im preporučiti prilagođene proizvode koji im u stvarnom vremenu trebaju. Na taj način veliki podaci omogućuju stvaranje personaliziranih iskustava za korisnike na temelju njihovih individualnih preferencija. Ovakvo pristupanje doprinosi povećanju zadovoljstva korisnika i jačanju njihove lojalnosti. Nadalje, putem tehnologije velikih podataka organizacije mogu identificirati slabosti u dizajnu korisničkog sučelja te osigurati da korisnici njihovih aplikacija i web stranica imaju intuitivno i ugodno iskustvo korištenja. Povijesni podaci o kupovini, recenzije korisnika također predstavljaju veliku prednost organizacijama putem kojih tvrtke prediktivnom analitikom mogu donositi informirane odluke o tome kako poboljšati značajke proizvoda i razviti nove proizvode koji će zadovoljiti trenutne potrebe tržišta. Automatizirano praćenje kvalitete omogućuje identifikaciju potencijalnih problema ili nedostataka, čime se smanjuje vrijeme reakcije i poboljšava ukupno zadovoljstvo korisnika.

Svake godine gotovo 100 milijuna ljudi posjeti parkove Walt Disney diljem svijeta. Godine 2013. tvrtka je predstavila ulaznicu za park u obliku narukvice opremljene senzorom radio frekvencijske identifikacije (RFID). RFID senzori slični su barkodima u svrhu, s razlikom što su RFID senzori mali bežični računalni uređaji koji automatski i bežično mogu prenositi digitalne informacije o objektu kojem su pričvršćeni, na značajnu udaljenost i bez potrebe za vidnim kontaktom. Ove narukvice opremljene RFID-om, nazvane „Magic Bands“, koje pružaju posjetiteljima pogodnosti poput preskakanja redova i unaprijed rezerviranja vožnji. Dok pružaju ove pogodnosti, Magic Band također bilježi kretanje posjetitelja i olakšava automatsko

prikupljanje velikih količina složenih i vrijednih podataka o posjetiteljima i njihovim interakcijama s Disney parkovima. Disney koristi ove skupove podataka u kombinaciji s nekoliko alata za analizu velikih podataka kako bi proučavao ponašanje kupaca, uključujući povijest kupnje, vrijeme čekanja i preferencije. Ove informacije pomažu tvrtki pružiti bolja iskustva posjetiteljima parka i poboljšati učinkovitost svog marketinga. Sve to nadalje pomaže privući više kupaca i povećati prihode po kupcu (Van Rijmenam, 2014). Ovim primjerom vidljivo je kako veliki podaci igraju ključnu ulogu u poboljšanju korisničkih usluga i proizvoda te ostvarenju konkurentske prednosti na tržištu.

2.2.2. Medicinski napredak

Zbog izuma novih tehnologija u medicini i zdravstvu tijekom posljednjih desetljeća, poput tehnika digitalne slikovne dijagnostike i dijagnostike molekularne biologije, te uspostave registara pacijenata i elektroničkih zdravstvenih kartona u mnogim zemljama u Europi i šire, došlo je do brzog rasta količine i složenosti podataka, kako u medicinskim istraživanjima, tako i u kliničkoj praksi. Klasični pristup istraživanju više nije dovoljan za suočavanje s izazovima analize podataka (Majnarić i sur., 2021). Metode i tehnike iz područja strojnog učenja i analize velikih podataka pristupa umjetnoj inteligenciji počele su se pojavljivati kako bi pružile alternativna rješenja. Ovaj pristup omogućava upravljanje podacima velike količine, visoke raznolikosti i složenosti revolucionirajući način pružanja, upravljanja i istraživanja zdravstvene skrbi. Algoritmi i alati iz ovog istraživačkog pristupa kompilirani su iz različitih analitičkih područja, uključujući matematiku, statistiku i računalne znanosti, kako bi omogućili računalima analizu skupova podataka koji su velikih volumena, zahtijevaju analizu visoke brzine i pokazuju visoku razinu raznolikosti i složenosti (visoka dimenzionalnost, kompleksni odnosi i druge komplikacije)(Pham i sur., 2017).

Sve više se potiče takozvana „precizna medicina“, označena s 4P: Personalizirano, Prediktivno, Pervencijsko i Participativno - individualizirana procjena pacijenta i tretmana, suprotno paradigmi "jedna veličina odgovara svima". Kako bi se takva usluga pružila pacijentima potrebna je analiza velikog obujma podataka u stvarnom vremenu koji sadržavaju genetske, kliničke i životne informacije na temelju kojih se prilagođava i optimizira plan liječenja svakom pojedinom pacijentu individualno. Personalizirani pristup pospješuje djelotvornost samog liječenja smanjujući nuspojave što na kraju rezultira uspješnosti poboljšanja ishoda pacijenta (Majnarić i sur., 2021). Tehnologija velikih podataka osnažila je medicinska istraživanja mogućnošću analize ogromnih skupova podataka zapanjujućom brzinom što je ubrzalo proces istraživanja i otkrića novih lijekova i terapija.

Prediktivna analitika zajedno s algoritmima strojnog učenja omogućuje prognoziranje bolesti. Analizom povijesnih podataka pacijenata, okolišnih čimbenika i genetskih predispozicija predviđaju se vjerojatnosti pojavljivanja bolesti. Ovakav proaktivan pristup omogućuje rane intervencije, modifikacije životnog stila i preventivne mjere čime se smanjuje opterećenje bolestima kako za pojedince, tako i za zdravstvene

sustave (Xiao i sur., 2018).

Klinički sustav za potporu odlučivanju (engl. Clinical decision support system- CDSS) odnosno je aplikacija koja analizira informacije kako bi pružila pomoć zdravstvenim stručnjacima u donošenju odluka i poboljšanju kvalitete skrbi pacijenata. Ovaj sustav predstavlja varijaciju sustava za potporu odlučivanju (engl. Decision support system, DSS), često korištenog u poslovnom upravljanju. U postupku korištenja kliničkih sustava za potporu odlučivanju, može se primijeniti rudarenje podataka kako bi se analizirala medicinska povijest pacijenta u kombinaciji s relevantnim kliničkim istraživanjima. Tehnologija velikih podataka omogućila je funkcioniranje ovakvog sustava za potporu u odlučivanju što u konačnici doprinosi poboljšanju opće koristi društva (Awati i sur., 2024).

U proteklih nekoliko godina primjećuje se sve veći trend u korištenju metoda dubokog učenja (DL) i podataka iz elektroničkih zdravstvenih kartona (eHRs) za zadatke poput fenotipizacije pacijenata, otkrivanja ili klasifikacije karakteristika bolesti i predviđanja kliničkih ishoda na temelju longitudinalnih nizova događaja, što se očekuje da će poboljšati rješavanje zadataka povezanih s multimorbiditetom (Xiao i sur., 2018). Na primjer, Meng i sur. (2019) koristili su informacije o terapijskim linijama za pacijente s rakom iz velikog skupa podataka osiguravajućih zahtjeva kako bi identificirali putanje liječenja. Autori su stvorili algoritam za izvođenje terapijskih linija na razini pacijenta i agregirali su ove informacije pomoću metoda klasteriranja i vizualizacije podataka kako bi izvodili temporalne fenotipe i podržavali predviđanje progresije bolesti. Nguyen i sur. (2016) razvili su modificirani model konvolucijske neuronske mreže (CNN - convolutional neural network) za predviđanje vjerojatnosti ponovnog prijema u bolnicu, temeljen na informacijama o medicinskoj povijesti korištene kao slijed koncepata.

Veliki podaci u farmaceutskom i zdravstvenom sektoru postaju sve očitiji. Komercijalne prednosti koje ovakvi podaci pružaju potiču farmaceutske i zdravstvene tvrtke da prihvate ovu tehnologiju. Procjene sugeriraju da je tržišna vrijednost tehnologije velikih podataka u zdravstvu iznosila 29,30 milijardi američkih dolara u 2020.godini i procjenjuje se da će doseći 59,10 milijardi američkih dolara do 2028. godine, rastući po godišnjoj stopi od 9,12% od 2021. do 2028. godine (Padma, 2023).

Primjerice, Genentech (važan igrač u biotehnološkoj industriji s vodećom praksom u području znanosti o podacima) ulaže u upravljanje podacima i alate za analizu podataka. Genentech je izgradio platformu za velike podatke koja može pregledavati milijarde zdravstvenih zapisa pacijenata u sekundama. Genentech ozbiljno zapošljava i razvija ljude s potrebnim stručnostima, surađujući s sveučilištima i tvrtkama poput Data Incubatora kako bi zapošljavao i educirao znanstvenike za podatke, te sada djeluje kao poduzetnički tim. Primjer njihova trenutnog rada uključuje stvaranje baze podataka o prošloj kohorti pacijenata koji su formalno dijagnosticirani s karcinomom. Grupa je proučila podatke kako bi identificirala rezultate različitih podtipova pacijenata i postupaka liječenja. To je pomoglo Genentechu da otkrije kako različite promjene

u biomarkerima i različiti postupci liječenja utječu na rezultate medicinskih eksperimenata u stvarnom svijetu. Ti podaci konačno su podržali ozbiljan napredak u medicini. Genentech je iskoristio podatke iz stvarnog svijeta i u drugim područjima liječenja, poput neuroznanosti, gdje je razvoj lijekova izuzetno izazovan, kako bi bolje prepoznao različitost uzoraka bolesti, simptoma tijekom vremena, stope napredovanja bolesti i odgovore na liječenje, te kako bi numerirao dinamiku troškova kako bolesti napreduju. Tradicionalno, farmaceutske tvrtke koriste SAS programere koji provode precizna istraživanja kliničkih ispitivanja na dosljedan i organiziran način. Financiranje za ove programere bilo je zadovoljavajuće, s obzirom da su klinička ispitivanja osmišljena kako bi odgovorila na pitanja o učinkovitosti i sigurnosti uz svježije i filtrirane skupove podataka (Copping & Li, 2016).

Spoj tehnologije velikih podataka i medicine je transformacijska snaga koja potiče napretke koji su nekada bili nezamislivi. Od ubrzanja istraživanja do omogućavanja personaliziranih tretmana i poboljšanja ishoda pacijenata, veliki podaci mijenjaju pejzaž zdravstva. Dok nastavljamo koristiti moć podataka, budućnost nosi obećanje učinkovitijeg, učinkovitijeg i pacijentom usmjerenijeg zdravstvenog sustava.

2.2.3. Financijska zaštita

Financijski sektor je vitalno područje primjene, a sigurnost velikih podataka zauzima čvrsto mjesto u brojnim financijskim industrijama. Da bi financijska organizacija uspješno upravljala velikim podacima, mora ih redovito održavati sigurnima i usklađenima s regulatornim zahtjevima (Gutierrez, 2014). Pružanje zaštite ogromnoj i rastućoj količini važnih podataka te sposobnost pronalaženja i analiziranja kako bi se identificirale moguće prijetnje danas je važnije nego ikad. Financijske organizacije moraju zaštititi pohranu, prijenos i upotrebu korporativnih i osobnih podataka putem različitih poslovnih aplikacija, uključujući e-bankarstvo i e-komunikacije osobnih informacija, datoteka i dokumenata. Iako su sve vrste poslovanja ranjive na napade kriminalaca, upravo su sigurnosni propusti u financijskim organizacijama oni koji izazivaju najviše medijske pažnje, javne kritike i zabrinutost zakonodavaca.

Primjerice, HSBC (The Hongkong and Shanghai Banking Corporation) uložio je i surađivao s SAS Fraud Managementom kako bi iskoristio velike podatke u otkrivanju prijevara u transakcijama na bankomatima i prijevornim aktivnostima. To pomaže smanjiti globalne gubitke od prijevara i prijetnji. HSBC je počeo koristiti SAS Fraud Management kao temelj za otkrivanje prijevara u stvarnom vremenu i upravljanje prijevarama diljem svoje mreže. Rješenje je zaživelo u SAD-u, Europi i Aziji te štiti 100% transakcija kreditnim karticama u stvarnom vremenu. HSBC planira pojačati mjere kako bi obuhvatio prijevare kroz različite poslovne i trgovinske mreže. Nije iznenađujuće da se borba protiv svih oblika prijevara, od plaćanja karticama, online ugovora, tržišta, transakcija pa čak do prijevora prve strane (korisnika), uzdigla na vrh korporativne agende. Prema Dereku Wyldeu, voditelju Odjela za rizik od prijevora, globalne sigurnosti i rizika od prijevora za HSBC, značajno su smanjili broj slučajeva prijevora, iznude i krađa

intelektualnog vlasništva na desecima milijuna debitnih i kreditnih kartica, što značajno premašuje postavljene ciljeve (Tao i sur., 2019).

Medicinske informacije vrijede deset puta više od broja vaše kreditne kartice na crnom tržištu. FBI je upozorio pružatelje zdravstvene skrbi da se čuvaju od cyber napada nakon što je jedan od najvećih američkih operatera bolnica, Community Health Systems Inc, izvijestio da su hakeri provalili u njegovu računalnu mrežu i ukrali osobne informacije 4,5 milijuna pacijenata. Stručnjaci za cyber sigurnost izražavaju zabrinutost zbog toga što kibernetički kriminalci sve više pretražuju američku zdravstvenu industriju, koja ima brojne tvrtke i poslovanja koja i dalje ovise o zastarjelim računalnim mrežnim sustavima koji nisu opremljeni najnovijim značajkama i alatima za kibernetičku sigurnost. Kako kibernetički napadači razvijaju inovativne pristupe za zaradu novca, zdravstvena industrija postaje privlačnija meta zbog mogućnosti prodaje ogromnih setova privatnih i osjetljivih podataka radi ostvarivanja dobiti. Bolnice imaju nisku razinu sigurnosti, pa je relativno lako hakerima doći do velike količine osobnih podataka radi medicinskih prijevара, zbog uvjerenja da zdravstvena industrija nije cilj (Tao i sur., 2019). Ovo ne znači da je sigurnost velikih podataka nevažna u drugim industrijama, već pokazuje da nijedna industrija nije izuzeta od povreda podataka sve dok se podaci prikupljaju u velikim količinama u toj određenoj industriji i da svaka industrija treba postaviti ulaganje u sigurnost podataka kao svoj prioritet broj jedan.

2.2.4. Predviđanje kriminala

Veliki podaci, odnosno "veliki skupovi podataka", postali su sve više dostupni u posljednjim godinama i primjenjuju se u velikom broju industrija. U provedbi zakona postoji mnogo načina prikupljanja i obrade takvih podataka koji mogu poboljšati, pa čak i potpuno zamijeniti postojeće metode analize i cjelokupni radni proces. Zahvaljujući tehnologiji velikih podataka, koja omogućava prikupljanje i obradu velikih količina strukturiranih i nestrukturiranih podataka, između ostalog, sada je moguće bolje predviđati i sprječavati kriminal, kao i druge oblike potencijalno opasnog društvenog ponašanja.

Kada razmatramo velike podatke, internet predstavlja najopsežniju bazu podataka na svijetu, pretvarajući se u svojevrsni "rudnik zlata" kada ga sagledamo s aspekta provedbe zakona. U suvremenom svijetu, sve više informacija pohranjuje se u oblaku (tj. "Cloud") i na mreži. Sami podaci nisu jedini izvor informacija relevantnih za kriminalistička istraživanja. Često su važne društvene mreže, otvorene baze podataka, specifični izvori podataka za određenu zemlju, poput katastra, telefonskog imenika, adresa građana, registra tvrtki i slično, kao i informacije o mapiranju, online video i foto arhive (repozitorije), podaci o mreži, razne usluge chatova ili e-mail (Kopal & Karas, 2017).

Los Angeles Police Department je jedan od desetaka gradova diljem SAD-a koji pokušava predviđati gdje će se zločin dogoditi - i tko će budući kriminalci biti - na temelju podataka o prošlim zločinima i uhićenjima.

Jedan napor, poznat kao Operacija LASER, koja je započela 2011., analizira informacije o prethodnim počiniteljima tijekom dvogodišnjeg razdoblja, koristeći tehnologiju koju je razvio tajanstveni analitički tim tvrtke Palantir, i boduje pojedince na temelju njihovih dosjea. Ako ste ikada bili u bandi, to je pet bodova. Ako ste pod uvjetnom kaznom ili uvjetnom slobodom, još pet bodova. Svaki put kad vas zaustavi policija, svaki put kad kucaju na vaša vrata, to vam može donijeti još bodova. Što su bodovi veći, to je veća vjerojatnost da završite na nečemu što se zove "Bilten za kronične prekršitelje", popisu ljudi za koje podaci kažu da su najviše u opasnosti da ponovno počine prekršaj i da ih treba pažljivo nadzirati (Brayne, 2017).

Sličan primjer dogodio se i u Lijepoj našoj gdje se provelo istraživanje kojem je svrha bila analizirati lokacije izvršenja kaznenih djela u području Grada Zagreba i Republike Hrvatske (Split, Osijek), kao i istražiti promjene u strukturi i dinamici izvršenih kaznenih djela tijekom proteklih 10 godina (2007. – 2016.). Tijekom istraživanja, prikupili su se i geokodirali podaci o mjestima izvršenja kaznenih djela u Republici Hrvatskoj. Analiza je obuhvatila i pristup dostupnih karakteristika počinitelja i žrtava kaznenih djela, te fenomenologiju samih kaznenih djela. Ovo istraživanje predstavilo je početak sustavne vremenske i prostorne analize izvršenja kaznenih djela u Hrvatskoj, pružajući uvide u konfiguraciju područja/kvartova, upotrebu i utjecaj prostora, interakciju između počinitelja i žrtava, izradu profila počinitelja i geografskih profila kaznenih djela, te konačno predviđanje i planiranje preventivnih mjera. Osim toga, doprinijelo je identifikaciji najranjivijih lokacija te objasnilo razloge zašto se kaznena djela pojavljuju u određenom, a ne u drugom području (Kopal & Karas, 2017). Pretpostavka je da će ovakva vrsta istraživanja temeljena na velikim skupovima podataka zajedno s rudarenjem istih smanjiti zabrinutost i nerealno procjenjivanje prisutnosti kriminala među javnošću, pružajući jasnu, transparentnu i pouzdanu sliku o dinamici kriminaliteta. Policiji će također pomoći u prepoznavanju područja na koje treba posebno usmjeriti svoje resurse u prevenciji kaznenih djela.

Proaktivno djelovanje policije ključno je za održavanje stabilnosti sigurne okoline, posebice u kontekstu prevencije i suzbijanja terorizma, što predstavlja ključni mehanizam protuterorističkog djelovanja. Terorizam se smatra jednom od najopasnijih društveno-političkih i sigurnosnih prijetnji, s dubokim posljedicama koje značajno utječu na društvo. Stoga su moderni protuteroristički sustavi razvijenih država usmjereni na proaktivno djelovanje kako bi se eliminirale sve okolnosti koje potiču razvoj terorizma. Proaktivni modeli mogu se operacionalizirati kroz izradu strateških i operativnih dokumenata koji precizno definiraju ulogu svakog dionika sustava u primjeni općih mjera usmjerenih na suzbijanje određenih sigurnosnih prijetnji. Takvi dokumenti, koji propisuju aktivnosti pojedinih tijela, prilagodljivi su novonastalim okolnostima te se uz minimalne izmjene mogu primijeniti na različite specifične situacije određene izvorom i oblikom prijetnje. Baze podataka, poput onih o modalitetima počinjenja kaznenih djela terorizma, kriminalnih djela povezanih s terorizmom, korištenim sredstvima izvršenja, načinu njihova pribavljanja, financiranju terorističkih aktivnosti i indikatorima društvenih promjena koje mogu potaknuti

radikalne i terorističke ideologije, igraju ključnu ulogu u definiranju proaktivnog pristupa terorizmu. Sustavi koji raspolažu dovoljnom količinom kvalitetnih operativnih saznanja organiziranih kroz analitičke baze pružaju učinkovitu analitičku podršku predviđanju budućih sigurnosnih prijetnji, pri čemu je uspostavljen sustav indikatora od izuzetne važnosti. Metodološki okvir, koji postavlja pravila za definiranje mjerljivih ulaznih i izlaznih rezultata, predstavlja analitički preduvjet za rad s zahtjevnim informacijskim resursima (Kopal & Karas, 2017).

Pravni uvjeti u vezi s obradom određenih podataka su postavljeni presedanima Europskog suda za ljudska prava, kao što je primjerice u vezi pristupa podacima o IP adresama, u vezi nadzora računala zaposlenika te u vezi prikupljanja podataka s GPS uređaja građana. Podaci iz otvorenih izvora nisu se smatrali posebnom kategorijom, ali sveobuhvatan pristup obradi tih podataka postaje predmetom pravnog interesa. Vijeće Europe je 2017. godine donijelo Smjernice o zaštiti pojedinaca u okruženju analize velikih skupova podataka, što potvrđuje važnost ovog područja za pravno uređenje. Neki teoretičari analitike velikih skupina podataka kritiziraju opsežne mogućnosti nadzora građana, ali zanemaruju činjenicu da nastanak podataka proizlazi iz tehnološkog razvoja digitalnih sustava, a ne nužno od intervencije tijela vlasti (Kopal & Karas, 2017). Neki pravni sustavi uveli su tzv. digitalna prava (poput prava zaboravljenog ili prava na zastari podataka na internetu, kako predlaže Mayer-Schönberger), no ostaje nejasno koje tijelo bi provodilo zaštitu i na temelju kojih kriterija bi se netko mogao pozivati na ta prava. Razvoj u ovom području pokazuje da pravno uređenje u brojnim sustavima još uvijek traži odgovarajuća rješenja za standardizaciju u regulaciji pravnih osnova, uvjeta za provedbu i uporabljivosti rezultata.

Policija sve više koristi ove metode kako bi se prilagodila novom digitalnom okruženju, primjerice kroz analizu videonadzora koji integrira prepoznavanje registracija vozila i drugih obilježja s raznim bazama podataka radi identifikacije neobičnih pojava. Unatoč kritikama koje ističu mogućnost zloupotrebe, analitičke metode često se koriste u reaktivnom modelu kriminalističkog istraživanja, uključujući praćenje kretanja žrtava putem uređaja za komunikaciju, podataka iz prometnih sustava, uređaja za sportske aktivnosti, bankovnih kartica, društvenih mreža i slično (Kopal & Karas, 2017). Unatoč tome, u recentnoj literaturi nedostaje empirijskih istraživanja o stvarnom doprinosu analitičkih metoda u otkrivanju ili istraživanju konkretnih kaznenih djela te koliki je njihov doprinos u prevenciji.

2.2.5. *Obrazovanje*

Edukacijsko rudarenje podataka (engl. Education Data Mining - EDM) je interdisciplinarno istraživačko područje stvoreno kao primjena rudarenja podataka u obrazovnom sektoru. Koristi različite metode i tehnike iz strojnog učenja, statistike, rudarenja podataka i analize podataka kako bi analiziralo podatke prikupljene tijekom nastave i učenja. Edukacijsko rudarenje podataka je proces pretvaranja sirovih podataka iz velikih obrazovnih baza podataka u korisne i značajne informacije koje se mogu koristiti za

bolje razumijevanje studenata i njihovih uvjeta učenja, poboljšavanje podrške nastavi, kao i za donošenje odluka u obrazovnim sustavima (Zorić, 2020).

Količina podataka prikupljenih i pohranjenih u mnogim obrazovnim institucijama postala je prevelika, te analiza obrazovnih podataka više nije mogla biti obavljena ručno. EDM je relativno novo područje koje proizlazi iz primjene tehnika rudarenja podataka na obrazovnim podacima. Prva međunarodna istraživačka konferencija o EDM-u održana je u Montrealu, Kanada, 2008. godine. Časopis *Journal of Educational Data Mining* počeo je s objavljivanjem 2009., a Međunarodno društvo za edukacijsko rudarenje podataka osnovano je 2011. godine (Zorić, 2020). Od tada, EDM nastavlja rasti iz različitih istraživačkih područja poput rudarenja podataka i strojnog učenja, prepoznavanja uzoraka, psihometrije i drugih područja statistike, umjetne inteligencije, vizualizacije informacija i računalnog modeliranja.

Konačni cilj EDM-a je poboljšati obrazovni proces i objasniti obrazovne strategije radi boljeg donošenja odluka. Postoje različite definicije EDM-a, ali sve imaju zajedničko to da je riječ o interdisciplinarnom istraživačkom području koje koristi različite metode i tehnike iz strojnog učenja, statistike, rudarenja podataka i analize podataka kako bi analiziralo velike podatke prikupljene tijekom nastave i učenja te otkrilo prethodno nepoznate informacije, odnose i uzorke u velikim skladištima podataka (Zorić, 2020).

Edukacijsko rudarenje podataka je mlado istraživačko područje koje postaje sve popularnije zbog svog potencijala. Obrazovni podaci mogu se koristiti kako bi se pomoglo profesorima, poboljšali nastavni planovi, razumjelo ponašanja studenata, unaprijedio nastavni proces, poboljšali e-learning sustavi, identificirali razlozi odustajanja, podržalo donošenje odluka, itd. Istraživanje edukacijskog rudarenja podataka može se podijeliti u dvije glavne kategorije, jednu koja se odnosi na analizu ponašanja tijekom učenja i atributa koji utječu na uspješno učenje, dok je glavni cilj drugog istraživanja pronaći prediktivni model za studentske performanse (Ali, 2013).

Kumar i Chadha (2011) predstavili su empirijsko istraživanje primjene velikih podataka u visokom obrazovanju u kojem su pokušali identificirati potencijalna područja primjene tih tehnika. Zaključili su da su potencijalne primjene: organizacija nastavnog plana, predviđanje upisa studenata u obrazovni program, predviđanje studentskog učinka, otkrivanje varanja na online ispitima te identificiranje abnormalnih ili pogrešnih vrijednosti. Autor Ali (2013) istaknuo je sljedeće koristi edukacijskog rudarenja podataka: identificiranje uzoraka ponašanja studenata, preferencija i potreba za kolegijima, odabir specijalizacije, predviđanje konačnih rezultata studenata, automatsko istraživanje podataka i izradu profila studenata.

Kovačić (2012) istraživao je socio-demografske varijable (dob, spol, etnička pripadnost, obrazovanje, radni status i invaliditet) te okolinu učenja koje mogu pomoći u identifikaciji uspješnih i neuspješnih studenata.

Ovo istraživanje zaključilo je da klasificiranje studenata na temelju informacija prije upisa može pomoći u prepoznavanju studenata koji su u riziku od odustajanja te predložiti programe savjetovanja i mentorstva kako bi im se pomoglo postići uspjeh.

Prednosti i primjene edukacijskog rudarenja podataka su brojne. Najčešća primjena edukacijskog rudarenja podataka uključuje: poboljšanje procesa učenja, unaprjeđenje završetka tečaja, podršku studentima pri odabiru usmjerenja, izradu profila studenata, otkrivanje problema koji dovode do odustajanja, usmjeravanje prema studentima, razvoj kurikuluma, predviđanje studentskih performansi i podršku donošenja odluka tijekom upisa studenata.

Razumijevanje ponašanja studenata i načina na koji uče može pomoći upravljanju obrazovanjem u poboljšanju trenutačnih studijskih programa i općenito obrazovne prakse. Analizom obrazovnih podataka, kao i analizom važnosti utjecaja pojedinih varijabli, različiti modeli rudarenja podataka mogu se koristiti kao podrška donošenju odluka u obrazovanju, pridonoseći tako uspješnijem studiranju i poboljšanju kvalitete obrazovanja. Rezultati edukacijskog rudarenja podataka mogu pomoći sveučilištima da učinkovitije rasporede resurse i djelotvorno usmjere sudionike obrazovnog sustava kojima je takva podrška potrebna.

Tehnike rudarenja podataka razvijene su kako bi automatski otkrile skriveno znanje i prepoznale obrasce iz podataka. Edukacijsko rudarenje podataka može se koristiti za klasificiranje i predviđanje uspjeha studenata, napuštanja studija, kao i učinkovitosti nastavnika. Može pomoći odgajateljima u praćenju akademskog napretka radi poboljšanja nastavnog procesa, pomoći studentima pri odabiru smjera te omogućiti učinkovitije obrazovno upravljanje. Edukacijsko rudarenje podataka može se koristiti kako bi privuklo i zadržalo studente dovodeći do profitabilnosti sveučilišta. Analiza podataka studenata ključna je za otkrivanje, prepoznavanje i razumijevanje koje su nastavne prakse učinkovite, a koje to nisu.

2.2.6. Urbani razvoj

Veliki podaci i napredne računalne simulacije mogu se koristiti kako bi se razumjelo kretanje ljudi kroz gradove. U dobu transformacijske mobilnosti, ovo može pomoći u boljem oblikovanju strategija i investicijskih planova za upravljanje postojećom urbanom infrastrukturom i predviđanje budućeg planiranja urbane infrastrukture. Veliki fokus prometnih planera, urbanista i geografa jest razumjeti kako urbani sustavi funkcioniraju i razvijaju se kroz modele pojedinačnih dnevnih urbanih aktivnosti. U gradovima je stoga osnovna uloga velikih podataka olakšati mehanizme protoka informacija za učenje i koordinaciju između heterogenih pojedinaca. S ogromnim napretkom u IKT (Informacijsko-komunikacijske tehnologije), računalnom oblaku i informacijskoj tehnologiji, veliki podaci pružaju gradovima više prilika da budu što pametniji u kratkom vremenskom periodu. Sustavi temeljeni na IoT-u (Internet of Things) koriste ogromnu količinu podataka generiranu pametnim sustavima za izgradnju

pametnih gradova i urbanog planiranja za budućnost. Ovi podaci generiraju se iz senzora pametnih kuća, mreža vozila, senzora za praćenje vremena i vode, pametnih sustava parkiranja, praćenja objekata itd. Pametni gradovi koriste informacijsku i komunikacijsku tehnologiju za prikupljanje, obradu i primjenu podataka na određenom području. ICT igra važnu ulogu u pametnim gradovima čineći podatke prikupljene komponentama informacijske tehnologije odmah dostupnim. To je moguće komuniciranjem s povezanim uređajima i razmjenom podataka bez potrebe za internetom, bežičnom vezom ili bilo kojim drugim komunikacijskim sredstvom. Pametni gradovi koriste informacijsku i komunikacijsku tehnologiju za promjene u urbanim uslugama analizom podataka i povratnim informacijama od građana. Veliki podaci utječu na različite sektore gradova, uključujući prijevoz, javnu sigurnost, gradske proračune i mnoge druge. Podaci s mobilnih uređaja i povezanih automobila mogu pomoći u otkrivanju nedostataka u javnom prijevozu i pružanju prilika za poboljšanje. Praćenjem trendova, trenutne uporabe i predviđenog rasta, planeri urbanog prijevoza pomoću tehnologije velikih podataka mogu predvidjeti gdje proširiti kapacitet i dodati nove rute. Optimizacija urbanog razvoja postaje jednostavnija i efikasnija uz prikupljene podatke i pametne aplikacije koje uvelike doprinose tom prikupljanju. Veliki podaci se koriste za optimizaciju urbanih operacija, upravljanje resursima i u konačnici trebali bi doprinijeti poboljšanju svakodnevnog života građana (Strielkowski i sur.,2021).

2.3. Nedostaci tehnologije velikih podataka

Prema Furmanu i Simcoe (2015), veliki podaci definirani su kao sposobnost prikupljanja ogromne količine podataka s različitih izvora radi istraživanja, analiziranja i razumijevanja potrošača, te prognoziranja tržišne potražnje i pristupačnih cijena za potrošače. S druge strane, način poslovanja tvrtki promijenjen je ovom novom tehnologijom. Kao rezultat toga, cijelo društvo se promijenilo, s tehnologijom velikih podataka koja donosi brojne spomenute koristi, ali i generira izazove koji će se obraditi u ovom poglavlju.

Često nam se govori da su podaci nova nafta. No, za razliku od nafte, podaci nisu tvar koja se nalazi u prirodi. Podaci se moraju prisvojiti. Obrada društvenih odnosno osobnih podataka osigurava "prirodnu" pretvorbu svakodnevnog života u tok podataka. Rezultat može dovesti do novog društvenog poretka, temeljenog na kontinuiranom praćenju, i pruža neviđene prilike za gubitak privatnosti, nadzor, društvenu diskriminaciju i utjecaj na ponašanje.

2.3.1. Gubitak privatnosti i nadzor

Prema Kamenici, Mullainathanu i Thaleru (2011), što više informacija mobilni operateri i sakupljači podataka prikupljaju i analiziraju, to bolje razumiju ponašanje potrošača. Ako im se pruži dovoljno vremena da prikupe dovoljno podataka, analitika podataka može razumjeti potrošača više nego on sam. Calo (2014) objašnjava personaliziranu reklamu kao sposobnost tvrtki da detektiraju kada je najbolje vrijeme za dostavu reklame ili kako izgraditi emocionalni odnos s potrošačima na temelju njihovog osobnog iskustva. Duhigg (2012) primjećuje da Target može razlikovati trudne potrošače na temelju njihovog kupovnog zapisa i izravno im dostaviti povezane reklame o proizvodima za bebe.

Ovi veliki podaci omogućuju tvrtkama da direktno dosegnu svoje ciljane potrošače i uštede milijune dolara koje gube na ljude koji neće kupiti njihove proizvode ili usluge, bez obzira na to koliko se trude, jer tvrtke mogu analizirati trenutnu situaciju potrošača, poput trudnoće, razvoda, selidbe, i tako dalje. Prije ere velikih podataka, marketinški istraživači osmišljavali bi, isporučivali, prikupljali i analizirali ankete kako bi razumjeli svoje potencijalne potrošače. Taj bi proces trajao mjesecima, čak i godinama, kako bi pružio grubo razumijevanje na temelju uzorka koji nikada nije bio dovoljno velik. Svi poslovni planovi i marketinški dizajni temeljili bi se na tim informacijama, sve dok veliki podaci nisu to promijenili.

Veliki podaci mijenjaju način na koji tvrtke posluju, izrađuju marketinške planove i dostavljaju reklame. S velikim podacima, tvrtke mogu direktno pronaći potencijalnog potrošača, dostaviti reklamu u odgovarajuće vrijeme i naplatiti pristupačnu cijenu za sklopljeni dogovor. Na primjer, kada potrošači pretražuju internetsku trgovinu za poslovno odijelo, reklame se ne pojavljuju nasumično, već u logičkom rasporedu ovisno o tome što su prije kupili. Ova tehnologija povezuje potrošače s proizvodima ili uslugama koje traže. Potrošači mogu imati koristi od personalizirane reklame jer im štedi vrijeme i daje mogućnost

bržeg pronalaska onoga što im je potrebno.

Međutim osim pozitivnih postoje i negativne karakteristike zadiranja u privatnost pojedinca. Tijekom pružanja usluge, primjerice, usluge e-pošte, tvrtka može prikupljati i obrađivati osobne informacije sadržane u e-pošti. Prikupljeni podaci također mogu obuhvaćati, primjerice, jezik kojim korisnik govori, mobilni telefon ili stvarnu lokaciju (čak i informacije specifične za uređaj, poput modela hardvera, verzije operativnog sustava, jedinstvenih identifikatora uređaja i informacija o mobilnoj mreži). Osim toga, kada korisnik pohranjuje svoje digitalne, a ponekad i osobne, datoteke koristeći računarstvo u oblaku (primjerice, Dropbox, Google Drive, Sky Drive, i-Cloud), pružatelj, tj. tvrtka koja nudi uslugu u oblaku, može obrađivati podatke sadržane u korisnikovim datotekama pohranjenim u "oblacima". Naposljetku, mnogobrojni podaci koji se odnose na korisnikovo zdravlje, kretanje ili samo obrasce života (npr. otkucaji srca, krvni tlak ili čak vrijeme spavanja) mogu se prikupljati i obrađivati sve dok korisnici, uz pomoć pametnih uređaja i brojnih aplikacija (Mayer-Schönberger & Cukier, 2013), mjere sebe tijekom svojih svakodnevnih fizičkih aktivnosti. Stoga, brojne internetske aktivnosti, standardna značajka svakodnevnog života, uključuju proizvodnju i obradu nezapamćenog obujma osobnih podataka. Iako je upitno jesu li zabilježeni otkucaji srca nečiji osobni podaci, mnogi, ili možda većina vrsta informacija opisanih gore kao primjeri, zapravo su osobni podaci prema Općoj uredbi o zaštiti podataka EU. To je zato što u doba velikih podataka, prikupljanje ogromne količine podataka omogućuje tvrtkama da izvuku mnoge zaključke koji se odnose na jednu osobu i omogućuje identifikaciju prirodnih osoba. Ako se podaci koje je prikupila tvrtka odnose na prirodnu osobu, koja se može identificirati, izravno ili neizravno, ti se podaci smatraju osobnim podacima. Drugim riječima, kriterij koji se mora ispuniti, a koji čini podatke osobnim, nije stvarna identifikacija, već sposobnost izravne ili neizravne identifikacije jedne osobe. Ukratko, ako postoji mogućnost identifikacije osobe, na koju se gore spomenuti "zabilježeni otkucaji srca" odnose, podaci su osobni, posebno zdravstveni podaci (George & Paul, 2020) i potpuno su regulirani GDPR-om. Nakon što su prikupili masu, ponekad, osobnih podataka koje korisnici proizvode "samo svojim postojanjem", mnoge tvrtke se ponašaju kao "vlasnici" tih informacija, razmjenjujući ili ih dalje obrađuju. U tom slučaju, neki znanstvenici čak govore o krađi humanističke imovine, ovu krađu počinili su privatni poduzetnici, dok drugi tvrde da bi fizičke osobe trebale primiti pravednu naknadu za prikupljanje, obradu, razmjenu i korištenje svojih osobnih podataka, budući da ne bi trebalo biti besplatnih obroka kada je u pitanju kršenje privatnosti (George & Paul, 2020). Osim privatnih tvrtki zabrinjavajući je podatak da vlade također generiraju osobne podatke svojih građana čuvajući ih u masovnim bazama podataka. Jedan ilustrativan primjer je američka Nacionalna sigurnosna agencija (engl. National Security Agency; NSA), koja se prema istraživanju Washington Posta iz 2010. godine navodi da sakuplja i zadržava ogroman broj elektroničke pošte, telefonskih razgovora i drugih oblika svakodnevne komunikacije, procijenjen na 1,7 bilijuna podataka. (Mayer-Schönberger & Cukier, 2013). Također ukoliko podaci nisu adekvatno zaštićeni i osigurani, postoji rizik od neovlaštenog pristupa ili zloupotrebe poput hakiranja baze podataka što može

rezultirati procurivanjem osjetljivih osobnih informacija o građanima, što može dovesti do krađe identiteta, financijskih prevara ili drugih oblika zloupotrebe podataka.

Želja za prikupljanjem informacija o kupcima nije nova za Target ili bilo kojeg drugog velikog trgovca. Desetljećima, Target je prikupljao ogromne količine podataka o svakoj osobi koja redovito ulazi u jednu od njegovih trgovina. Kad god je to moguće, Target dodjeljuje svakom kupcu jedinstveni kod - poznat interno kao ID gostiju - koji prati sve što kupuju. "Ako koristite kreditnu karticu ili kupon, ispunite anketu ili pošaljete zahtjev za povratom, nazovete liniju za pomoć korisnicima ili otvorite e-poštu koju smo vam poslali ili posjetite našu web stranicu, zabilježiti ćemo to i povezati s vašim ID-om gosta, želimo znati sve što možemo," prema riječima Andrewa Pola jednog od zaposlenika Targeta.

Također povezan s vašim ID-om gosta su demografske informacije poput vaše dobi, jeste li u braku i imate li djecu, u kojem dijelu grada živite, koliko vam treba vremena da dođete do trgovine, vaša procijenjena plaća, jeste li se nedavno preselili, koje kreditne kartice najčešće koristite i koje web stranice posjećujete. Target može kupiti podatke o vašoj etničkoj pripadnosti, radnoj povijesti, časopisima koje čitate, jeste li ikada proglasili bankrot ili se razveli, godini kada ste kupili (ili izgubili) svoju kuću, gdje ste išli na fakultet, o kojim temama razgovarate na mreži, preferirate li određene marke kave, papirnatih ručnika, žitarica ili jabučnog soka, vašim političkim uvjerenjima, navikama čitanja, dobrotvornim donacijama i broju automobila koje posjedujete (u priopćenju, Target je odbio identificirati koje demografske informacije prikuplja ili kupuje). Sve te informacije su beskorisne, bez nekoga tko će ih analizirati i razumjeti. Tu dolazi Andrew Pole i deseci drugih članova Odjela za analizu potrošača u Targetu (Duhigg, 2012).

Gotovo svaki veći maloprodajni lanac, od trgovina prehrambenih proizvoda do investicijskih banaka i američke pošte, ima odjel za prediktivnu analitiku posvećen razumijevanju ne samo kupovnih navika potrošača, već i njihovih osobnih navika, kako bi ih što učinkovitije tržili. Živimo u zlatnom dobu istraživanja ponašanja. Nevjerojatno je koliko se sada može shvatiti o tome kako ljudi razmišljaju.

Nadalje, još davne 2009. godine operaciju masovnog nadzora, kodnog naziva KARMA POLICE, pokrenuli su britanski špijuni bez ikakve javne rasprave. Bio je to samo jedan dio golemog globalnog internetskog špijunskog aparata koji je izgradila britanska agencija za elektroničko prisluškivanje, GCHQ (Government Communications Headquarters). Jedan sustav izrađuje profile koji prikazuju nečiju povijest pregledavanja weba. Drugi analizira komunikaciju putem instant poruka, e-poštu, Skype pozive, tekstualne poruke, lokacije mobitela i interakcije na društvenim mrežama. Izgrađeni su zasebni programi za praćenje "sumnjivih" Google pretraživanja i korištenja Google karata (Gallagher, 2015). Ovo je jedan od primjera koliko daleko prikupljanje i nadzor osobnih podataka može ići.

Tijekom posljednjih 30 godina, gradovi diljem Azije, Europe i Sjeverne Amerike postali su zasićeni video

nadzorom odnosno CCTV kamerama što je financirano od strane države. Zahvaljujući popularnosti CCTV-a na javnim područjima, sada je gotovo nemoguće proći ulicom u velikim gradovima poput Pekinga, Londona i New Yorka, a da vas ne nadziru kamere. Daleko od toga da se smatraju neuobičajenima, kamere za nadzor javnih područja - poput prometnih znakova i uličnih svjetiljki prije njih - brzo su postale prihvaćena značajka urbanog života (Goold, 2020). S ovim postupnim proširenjem i normalizacijom CCTV-a, također je došlo do rastuće prihvaćenosti činjenice da pojedinci imaju malo ili nimalo očekivanja privatnosti na javnim prostorima. Dok su u prošlosti ljudi možda uživali u određenoj privatnosti i anonimnosti dok su hodali ulicom ili sjedili na klupi u parku, to jednostavno više nije slučaj u mnogim gradovima. U zemljama poput Australije, Kanade, Ujedinjenog Kraljevstva i Sjedinjenih Američkih Država, zakon postavlja malo ili nikakvih ograničenja na sposobnost policije i drugih državnih agencija da postavljaju CCTV kamere i provode video nadzor javnih područja (Goold, 2020). Kao rezultat toga, širenje CCTV-a imalo je učinak pretvaranja mnogih urbanih prostora u mjesta stalnog i intenzivnog državnog nadzora. Privatnost i anonimnost u javnosti, za mnoge ljude koji žive u gradovima barem, postala je stvar prošlosti.

Na pozadini općeg, uglavnom nereguliranog nadzora javnih područja sada vidimo uvođenje nove vrste tehnologije nadzora: kamere za nošenje na tijelu policije. Obično pričvršćene na prsa policijske uniforme, kamera za nošenje na tijelu može se koristiti za snimanje interakcija između policije i članova javnosti, kao i za dokumentiranje događaja u tijeku i pojedinosti mjesta zločina (Bud, 2016). S obzirom da se velik dio policijskog rada obavlja na javnim prostorima poput ulica i parkova, kamere za nošenje na tijelu mogu se razumjeti kao proširenje postojećih mreža CCTV-a na javnim područjima. Međutim, budući da su ove kamere za nošenje na tijelu pričvršćene za policijske službenike - a ne fiksne za stacionarni stup ili zgradu - mogu se koristiti i u prostorima koji bi inače mogli biti smatrani privatnima, poput domova, hotelskih soba ili radnih mjesta. Za mnoge borce za građanske slobode i zagovornike privatnosti, rastuća popularnost kamera za nošenje na tijelu policije izaziva ozbiljne zabrinutosti. Iako su pobornici ove nove tehnologije brzo istaknuli da kamere za nošenje na tijelu mogu pomoći u činjenju policije odgovornijom, koristiti se za snimanje zločina u tijeku i pomoći u očuvanju dokaza, drugi su tvrdili da predstavljaju značajnu prijetnju pojedinačnoj privatnosti. Kao što su to učinile i CCTV kamere prije njih, kamere za nošenje na tijelu izlažu pojedince na javnim prostorima intenzivnom i često vrlo invazivnom nadzoru. Osim potencijalnog utjecaja na autonomiju i dostojanstvo - što je ključno za mnoge opise vrijednosti privatnosti - kamere za nošenje na tijelu također imaju potencijal da sputaju politički govor i odvrate pojedince od korištenja zakonitih prava na slobodu izražavanja, religije i udruživanja. Doista, jedna od najvećih opasnosti neograničenog masovnog nadzora je potencijalni efekt hlađenja političkog diskursa, i na sposobnost grupa da izraze svoja stajališta putem prosvjeda i drugih oblika mirnih građanskih akcija (Goold, 2020). Osiguravajući da postoji ograničenje na ono što država može razumno očekivati da zna o nama, privatnost ne samo da pomaže u zaštiti individualne autonomije, već osigurava i da smo slobodni

koristiti tu autonomiju u ostvarivanju drugih temeljnih prava. Otišavši dalje, iako kamere za nošenje na tijelu mogu učiniti važne aspekte policijskog rada na razini ulice vidljivijima, također imaju implikacije za privatnost na radnom mjestu policajaca koji ih nose. U mnogim aspektima može se tvrditi da kamere za nošenje na tijelu predstavljaju sljedeći, prirodni korak u evoluciji nadzora javnih područja. Za razliku od CCTV kamera, međutim, kamere za nošenje na tijelu su mobilne. Gdje je domet CCTV kamere obično ograničen njezinim fiksnim položajem, vidnim poljem i sposobnošću zumiranja, kamere za nošenje na tijelu nemaju tih ograničenja. Domet kamere za nošenje na tijelu ograničen je samo mobilnošću službenika kojem je pričvršćena (Goold, 2020). Kamere za nošenje na tijelu ne samo da dovode u pitanje postojeće granice između javnih i privatnih prostora, već i naše pretpostavke o tome za koga su i što su javni prostori. Jednostavno rečeno, dolaskom mobilnog nadzora moramo se pitati: ako se državi dopusti daljnje jačanje njenog stiska na javnim prostorima putem tehnologija, što to znači za budućnost tih prostora i narativa koji iz njih proizlaze?

2.3.2. Društvena diskriminacija

Jedan od najalarmantnijih, ali još uvijek nedovoljno istraženih aspekata tehnologija velikih podataka je rizik od potencijalne diskriminacije. Iako "ne postoji univerzalno prihvaćena definicija diskriminacije", termin se općenito odnosi na postupke, prakse ili politike koje nameću relativnu nepovoljnost osobama zbog njihovog članstva u istaknutoj društvenoj ili prepoznatoj ranjivoj grupi na temelju spola, rase, boje kože, jezika, religije, političkog mišljenja, etničke manjine, itd. (Vandenhole, 2005). Razlikujemo izravnu diskriminaciju (tj. postupaka koji diskriminiraju manjine ili ugrožene skupine na temelju osjetljivih diskriminatornih obilježja povezanih s pripadnošću grupi kao što su rasa, spol ili seksualna orijentacija) i neizravnu diskriminaciju (tj. postupaka koji mogu namjerno ili slučajno diskriminirati manjinu, iako se eksplicitno ne spominju diskriminatorna obilježja) (Favaretto i sur., 2019). Također treba istaknuti blisku povezanost između diskriminacije i nejednakosti, budući da nepovoljnost uzrokovana diskriminacijom nužno dovodi do nejednakosti između razmatranih grupa.

Istraživanje diskriminacije u tehnologijama rudarenja velikih podataka nije novo, ali je dobilo zamah u posljednje vrijeme, posebno nakon objavljivanja izvješća Bijele kuće iz 2014. godine, koje je oštro upozorilo da bi diskriminacija mogla biti nehotičan ishod tehnologija velikih podataka. Od tada, mogući diskriminatorni ishodi profiliranja i ocjenjivanja sve više su privukli pažnju šire javnosti. Na primjer u Sjedinjenim Američkim Državama sustav tehnologije koji se koristi za procjenu budućeg rizika od ponovnog počinjenja među optuženicima otkriveno je da diskriminira crnce. Isto tako, u Ujedinjenom Kraljevstvu, algoritam koji se koristi za donošenje odluka o pritvoru otkriven je da diskriminira ljude s nižim prihodima. No, i više aplikacija usmjerenih prema građanima, poput aplikacije Boston's Street Bump, koja je razvijena za otkrivanje rupa na cestama, potencijalno su diskriminatorne. Koristeći se pametnim telefonom, aplikacija riskira povećanje društvene podjele između četvrti s većim brojem starijih ili manje

imućnih građana i onih bogatijih područja s više mladih vlasnika pametnih telefona (Favaretto i sur., 2019).

Tehnologije velikih podataka također bi mogle propagirati postojeće socijalne i geografske povijesne disparitete i nejednakosti, na primjer, povećavajući isključenost etničkih manjina iz društvenog angažmana, pogoršavajući životne uvjete ekonomski ugroženih, proširujući ekonomsku razliku između siromašnih i bogatih zemalja, isključujući neke manjine iz zdravstvene skrbi ili pružajući fragmentiranu i nepotpunu sliku populacije putem tehnologija rudarenja podataka (Brayne, 2017). Prema Newellu i Marabelliju, pojedinci bi mogli neobjašnjivo i neočekivano biti isključeni iz određenih prilika, iskorišteni na temelju njihovih nedostataka, te nepravedno tretirani putem ciljane reklame i profiliranja. Izdvojit ćemo dva glavna nova oblika diskriminacije: prvo, ekonomska ili marketinška diskriminacija, tj. nejednako postupanje prema različitim potrošačima na temelju njihovih kupovnih navika ili nejednakosti u cijenama i ponudama koje se daju potrošačima na temelju profiliranja, poput osiguranja ili stambenih kredita; drugo, diskriminacija na temelju predviđanja zdravlja, tj. nejednako postupanje ili diskriminacija pojedinaca na temelju prediktivnih, a ne stvarnih, zdravstvenih podataka.

Pojavom digitalnog jaza, odnosno razlici između onih koji imaju kontinuiran i spreman pristup internetu, računalu i pametnim telefonima i onih koji to nemaju, stvara se uzrok nejednakosti, nepravde ili diskriminacije. Nedostatak resursa ili računalnih vještina, starija dob, geografska lokacija i niski prihodi identificirani su kao mogući uzroci ovog digitalnog jaza. Nadalje određeni pojedinci mogu biti izostavljeni "isključenjima velikih podataka", misleći na one pojedince "čiji se podaci ne prikupljaju niti analiziraju redovito jer se ne uključuju redovito u prakse stvaranja podataka". Na istu notu, Bakken i Reame (2016) tvrdili su da se podaci uglavnom prikupljaju od bijelih, obrazovanih ljudi, izostavljajući rasne manjine poput crnaca i latino-amerikanaca. Također treba spomenuti da prilikom zapošljavanja mnoge tvrtke analiziraju podatke o kandidatima putem algoritama kako bi odlučile koje kandidate pozvati na razgovor, a koje ne. Algoritmi pritom mogu nepravedno preferirati određene skupine ljudi na temelju različitih karakteristika poput rase, spola ili dobi.

Diferencijalno određivanje cijena je praksa naplaćivanja različitih cijena različitim kupcima, što nazivamo "diskriminacijom cijena". Svakodnevni primjeri uključuju popuste za starije osobe u kino-dvorani i skuplje karte za poslovne putnike u zadnji trenutak (Furman, 2015). Tehnološki napredak otvara mogućnost da će u budućnosti tvrtke moći koristiti analizu velikih podataka kako bi otkrile i ponudile potrošačima njihovu individualnu rezerviranu cijenu (tj. najvišu cijenu koju bi svaki potrošač bio spreman platiti, s obzirom na njihove preferencije i dostupni приход). To može generirati neke zanimljive koristi, poput boljeg stanja stvari u smislu jednakosti kako u dobrobiti tako i u resursima, kao i povećane društvene dobrobiti. Međutim, ove koristi nadmašuju razmatranja o relacijskoj jednakosti.

Ekonomisti proučavaju diferencijalno određivanje cijena već mnogo godina, i iako se čini da će veliki

podaci promijeniti praksu određivanja cijena u praksi, nisu promijenili temeljne principe. Možda iznenađujuće, ti principi sugeriraju da je diferencijalno određivanje cijena često dobro i za tvrtke i za njihove kupce. Kada cijene odražavaju kupčevu sposobnost plaćanja, prodavači često mogu poslužiti kupce koji bi inače bili isključeni s tržišta, kao što je to slučaj s financijskom pomoći za studente na temelju potreba (Furman,2015).

Prednosti diferencijalnog određivanja cijena ukazuju na to da ono može imati važnu pozitivnu ulogu u cjelokupnom gospodarstvu. Međutim, diskriminirajuće određivanje cijena može postaviti teške izbore i izazvati ozbiljne zabrinutosti za pravednost i privatnost, pogotovo kada potrošači nisu svjesni kako prodavatelji koriste informacije o njima, ili kada se cijenjenje temelji na faktorima izvan kontrole pojedinaca. Jedan način ograničavanja nepravednih ili netočnih primjena velikih podataka u ovom kontekstu jest pružiti potrošačima veću vidljivost u vrste informacija koje tvrtke prikupljaju i veću kontrolu nad načinom njezine upotrebe.

Tvrtke profiliraju potrošače grupirajući ih te tako provode cjenovnu diferencijaciju na temelju različitih grupa. Sada je moguće da preciznost tvrtki u grupiranju, prikupljanju i analiziranju podataka postane tako dobro usavršena da će tvrtke moći profilirati potrošače s potpunom preciznošću i dodijeliti ih grupama od jednog člana. Drugim riječima, moguće je da će tehnološki napredak u prikupljanju i analizi podataka omogućiti cjenovnu diferencijaciju na temelju personalizacije.

The Wall Street Journal izvijestio je 2012. godine da Staples trgovina prikazuje različite cijene ljudima na temelju njihove lokacije. Staples.com je prikazivao nižu cijenu potrošačima koji žive unutar 20 milja udaljenosti od konkurentne trgovine (Steinberg,2019). Prema Newcomeru (2017), Uber Technologies koristi „cjenovni sustav temeljen na ruti“, koji potrošačima naplaćuje različite iznose ovisno o očekivanju koliko bi željeli platiti. Ovaj sustav razlikuje se od tradicionalnog modela cijena koji potrošače naplaćuje samo na temelju količine usluge koju kupuju. Cjenovni sustav koji je Uber ranije koristio izračunavao je tarife zbrajanjem vremena, udaljenosti i množitelja na temelju geografske potražnje. Novi „cjenovni sustav temeljen na ruti“ naplaćuje nekim potrošačima višu cijenu ne zato što se voze do dalje lokacije, već zato što se smatra da bi platili višu cijenu.

2.3.3. Sigurnosni rizici

Pohrana i obrada velikih podataka može povećati rizik od sigurnosnih prijevара odnosno „curenja“ informacija o građanima. Osobni podaci mogu se koristiti kako bi se stekao pristup zaštićenim domenama i imovini podataka subjekta, poput osobnih računa, privatne imovine ili čak cijelog identiteta subjekta. Kibernetički napadi su danas sve učestaliji te predstavljaju ozbiljan rizik za poduzeća ili pojedince. Postoje brojne metode kibernetičkih napada, a neke od poznatijih su (Coursera,2024):

- *Phishing* koji se oslanjaju na komunikacijske metode poput e-pošte kako bi uvjerali pojedinca da otvori poruku i slijedi upute unutar nje. Ako se slijede napadačeve upute, oni dobivaju pristup osobnim podacima, poput kreditnih kartica i mogu instalirati zlonamjerni softver na uređaj pojedinca.
- *Malware* gdje cyber napadači koriste štetne softvere poput spyware-a, virusa, ransomware-a i crva poznatog kao malware kako bi pristupili podacima tuđeg sustava. Kada se klikne na zlonamjerni privitak ili poveznicu, malware se može instalirati i postati aktivan na uređaju pojedinca.
- *Spoofing* preko kojeg cyber napadači pokušavaju imitirati ljude ili tvrtke kako bi prevarili pojedinca i otkrili njegove osobne informacije. Česta strategija prijave uključuje korištenje lažnog broja pozivatelja, gdje osoba koja prima poziv ne vidi da je broj krivotvoren. Druge metode prijave uključuju podvaljivanje sustava za prepoznavanje lica, korištenje lažnog naziva domene ili stvaranje lažne web stranice.

Napadači na internetu često ciljaju industrije poput zdravstva, vlade, neprofitnih organizacija i financijskih tvrtki. Posebno su podložne napadima zdravstvene ustanove, jer imaju pristup osobnim podacima mnogih ljudi. Budući da je zdravstvena infrastruktura ključna, napadači koji koriste *ransomwar* (oblik malware-a gdje se podaci drže zaključanima dok se ne plati otkupnina napadaču) razumiju da će ove organizacije vjerojatno brzo platiti njihove zahtjeve. Povjerljive informacije, poput brojeva socijalnog osiguranja, također dovode do toga da vlade postanu žrtve hakera. Neprofitne organizacije su posebno ciljane jer posjeduju financijske podatke o donatorima i napore za prikupljanje sredstava, što ih čini idealnim metama za cyber napade. U financijskoj industriji, institucije poput banaka i osiguravajućih društava česti su ciljevi ucjene i krađe zbog njihovog pristupa značajnim iznosima novca (Coursera, 2024).

Sve više se spominje i krađa identiteta što predstavlja jednu od najrigoroznijih prijetnji digitalnog doba. Različite vrste osobnih podataka, uključujući datume rođenja, brojeve putovnice, brojeve socijalnog osiguranja, telefonske brojeve, adrese i brojeve financijskih računa, mogu pomoći kriminalcima u krađi identiteta ljudi i obavljanju aktivnosti u njihovo ime, poput otvaranja računa, obavljanja kupovina ili podnošenja prijave u poreznoj prijavi. I dok je krađa identiteta dugo bila zabrinutost, sada se sve učestalije dešava jer je broj prijavljenih slučajeva naglo porastao u posljednjih nekoliko godina. Procijenjene lozinke i odgovori na sigurnosna pitanja (npr. djevojačko prezime majke) mogu se iskoristiti za preuzimanje online računa osobe - ne samo radi financijskog obogaćivanja i prikupljanja osobnih podataka, već i za slanje poruka i potencijalno prijetnji drugim ljudima u ime žrtve. Neke uobičajene posljedice za žrtve krađe identiteta su tužbe, odbijanje kredita i javnih povlastica, uznemiravanje od strane naplatitelja dugova, sramota, stres i anksioznost. Nove oblike krađe identiteta čak iskorištavaju biometrijske uzorke kako bi umjetno imitirali otiske prstiju ili glas žrtve kako bi ih oponašali i prevarili

biometrijske autentifikacijske mehanizme (Bertino,2015).

S napretkom tehnologije, napadači postaju sve složeniji i inovativniji u svojim naporima krađe osobnih podataka, što dodatno povećava sigurnosni rizik velikih podataka. Stoga je ključno redovito provođenje procjene rizika, primjenjivanje sigurnosnih mjera, antivirusnih programa i sustava za otkrivanje prijevare te dodatno educiranje društva sigurnosnim procedurama kako bi se smanjio rizik od kibernetičkih napada.

2.3.4. Utjecaj na ponašanje

Utjecaj na ponašanje pojedinca putem velikih podataka je blisko povezan s gubitkom privatnosti. Kao što je već rečeno brojne tvrtke prikupljaju razne informacije o svojim potrošačima putem „klikova“ kako bi stekli bolje znanje o njihovim preferencijama, navikama i izborima. Pomoću tih podataka svojim kupcima daju personalizirane usluge i proizvode koje u tom trenutku odgovaraju potrebama potrošača što može potaknuti pojedinca na kupnju proizvoda ili usluge koje inače možda i ne bi otkrio. Veliki podaci mogu utjecati na donošenje odluka pružajući pojedincu određene informacije ili preporuke koje ih mogu potaknuti na određene odluke ili izbore npr. u vrijeme političkih izbora građani često koriste društvene medijske platforme kao sredstvo za političku raspravu. Digitalni tragovi koje ostavljaju kroz svoje interakcije pružaju bogat izvor informacija što može stvoriti potrebne uvjete za targetiranje društva nakon odgovarajuće algoritamske obrade podataka. Odnosno tehnike rudarenja podataka omogućuju prikupljanje informacija o općem mišljenju ljudi i preferencijama stranaka. Kroz primjenu algoritama s velikom količinom prikupljenih podataka moguće je grupirati pojedince prema zajedničkim atributima i provesti profiliranje kako bi se identificirale skupine onih na koje se može potencijalno utjecati. Nadalje, svi smo zasigurno imali iskustva gdje bi nam mobilni uređaj preporučavao određene proizvode koji nam baš u tom trenutku trebaju ili one proizvode koje smo možda spomenuli u razgovoru s drugom osobom toga dana, tehnologija je otišla korak naprijed te sada ima mogućnost „razumijevanja i slušanja“ naših želja na temelju kojih može utjecati na naše ponašanje, izbor ili potaknuti potencijalnu kupnju.

2.3.5. Loš rezultat predviđanja

U svijetu tehnologije velikih podataka, obećanje preciznih predviđanja često se susreće s izazovima loših rezultata. Iako su tehnike analize podataka napredovale nevjerojatnom brzinom, veliki podaci nisu imuni na nedostatke. Ponekad, unatoč obilju podataka i složenim analitičkim algoritmima, predviđanja mogu biti nepouzdana, nedosljedna ili čak kontraproduktivna.

Mnogi od nas koriste Google tražilicu kako bi pronašli informacije korisne u našem svakodnevnom životu. Na primjer, kada se razbolimo, koristimo Google kako bismo pronašli informacije o simptomima, lijekovima, nuspojavama itd. Često koristimo Google za dostupne hotele kada planiramo putovanje. Google je pretpostavio da pretrage njihovih korisnika odražavaju trendove u bolestima, putovanjima,

modi, prodaji proizvoda i drugim stvarima.

Godine 2008., analitičari Google-a istražili su mogućnost detektiranja često korištenih pojmova u pretraživanju kako bi predvidjeli širenje gripe u SAD-u. Pretpostavka Google-a bila je da kada osobe obolijevaju od simptoma gripe, koriste Google-ovu tražilicu i upisivati pojmove relevantne za simptome gripe. To bi Google-u pružilo odmah signal o prevalenciji gripe u određenom geografskom području SAD-a. Kada je Google započeo svoje istraživanje, Centar za kontrolu i prevenciju bolesti-CDC (engl. Centers for Disease Control and Prevention), odjel američke vlade, pratio je pojavu gripe u različitim područjima prikupljajući podatke od liječnika, bolnica i ljekarni. Prikupljanje podataka je dugotrajan proces i zbog toga je bilo kašnjenje u otkrivanju širenja gripe. Google je objavio rad u časopisu Nature u kojem je pokazao da pregledom pojmova koje ljudi koriste u Google pretraživanjima i usklađivanjem s informacijama o praćenju gripe koje pruža CDC može pružiti točnu procjenu prevalencije gripe u različitim područjima SAD-a dva tjedna unaprijed u odnosu na CDC. Google-ov program „Flu Trends“ koristio se širom svijeta od strane Google-a pružajući ne samo trendove gripe već i predviđanje širenja groznice (Lazer i sur., 2014).

Metodologija koju je koristio Google bila je spektakularna demonstracija moći analitike velikih podataka u predviđanju, sve dok nije potpuno zakazala 2013. godine. Godine 2013., Google Flu Trends previsoko je predvidio prevalenciju gripe za 140%. Inženjeri Googlea nisu pokušavali shvatiti što uzrokuje što. Oni su samo pronalazili statističke obrasce u podacima. Brinuli su se o korelaciji umjesto uzročnosti. To je uobičajeno u analizi velikih podataka. Shvaćanje što uzrokuje što je teško (neki kažu nemoguće). Shvaćanje što je povezano s čime je puno jeftinije i lakše. Zato, prema knjizi Viktora Mayer-Schönbergera i Kennetha Cukiera, Big Data, „uzročnost neće biti odbačena, ali je skinuta kao primarni izvor značenja“. Jedno objašnjenje neuspjeha Flu Trends-a je da su vijesti bile pune zastrašujućih priča o gripi u prosincu 2012. i da su te priče potaknule internetske pretrage ljudi koji su bili zdravi. Još jedno moguće objašnjenje je da je vlastiti algoritam pretraživanja Googlea pomaknuo ciljeve kada je počeo automatski sugerirati dijagnoze kada su ljudi unijeli medicinske simptome (Lazer i sur. 2014).

Google Flu Trends će se oporaviti, kalibriran novim podacima, ali ako se ne nauče lekcije iz ovog slučaja, ponovno će se ponavljati. Podaci su danas veći, brži i jeftiniji - ali ne smijemo se pretvarati da su sve zamke postale sigurne. Veliki podaci mnogo obećavaju znanstvenicima, poduzetnicima i vladama, međutim „postoji mnogo problema malih podataka koji se pojavljuju u velikim podacima; oni ne nestaju jer imate mnogo tog materijala. Postaju gori,“ prema Davidu Spiegelhalteru, profesoru Javnog razumijevanja rizika na sveučilištu u Cambridgeu (Harford, 2014).

Drugi primjer lošeg rezultata velikih podataka je aplikacija za pametne telefone „Street Bump“ u Bostonu, koja koristi akcelerometar telefona za otkrivanje rupa na cesti bez potrebe da gradski radnici obilaze ulice. Gradonačelnik grada želio je koristiti analitiku velikih podataka kako bi pronašao rupe na cestama koje

treba popraviti što je predstavljalo jedno od jeftinijih i lakših načina prikupljanja informacija o stanju na cestama. Dok građani Bostona preuzimaju aplikaciju i voze se, njihovi telefoni automatski obavještavaju gradsku vijećnicu o potrebi popravka površine ceste. Rješavanje tehničkih izazova koji su uključeni proizvelo je, prilično lijepo, informativno ispuštanje podataka koje rješava problem na način koji bi bio nezamisliv prije nekoliko godina. Grad Boston s ponosom tvrdi da „podaci pružaju gradu informacije u stvarnom vremenu koje koristi za rješavanje problema i planiranje dugoročnih investicija." Međutim, ono što Street Bump zaista proizvodi jest karta rupa na cestama koja sistematično favorizira mlade, bogate dijelove gdje više ljudi posjeduje pametne telefone. Street Bump nudi „N = Svi" u smislu da se svako udaranje sa svakog omogućenog telefona može zabilježiti. To nije isto kao zabilježiti svaku rupu na cesti. Kako ističe istraživačica Microsofta Kate Crawford, pronađeni podaci sadrže sustavne pristranosti i potrebno je pažljivo razmišljati kako bismo ih uočili i ispravili. Veliki skupovi podataka mogu djelovati sveobuhvatno, ali „N = Svi" često je zavodljiva iluzija (Harford, 2014).

U ovom slučaju samo su mladi bogati građani s pametnim telefonima preuzeli aplikaciju i rupe u bogatim područjima Bostona su otkrivene i popunjene. Iz siromašnijih područja (koja su vjerojatno imala više rupa), stiglo je samo nekoliko poruka jer većina stanovništva u tim područjima nije imala pametne telefone, a čak i među onima koji su ih imali, mnogi nisu preuzeli aplikaciju, s rezultatom da je većina rupa ostala nepopunjena.

3. REZULTATI ISTRAŽIVANJA

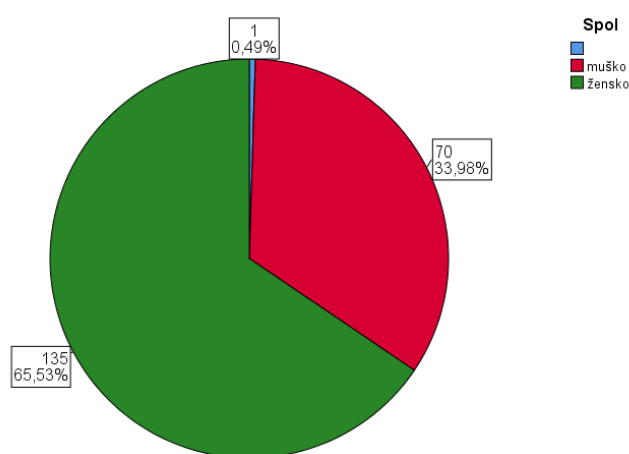
Za potrebe provedbe ove studije proveden je anketni upitnik. Anketa se sastojala od ukupno 24 pitanja gdje su ona bila podijeljena u 5 grupa. Svaka grupa ispitivala je određene karakteristike bitne za provođenje istraživanja pa su se tako 4 pitanja odnosila na demografske podatke, 5 pitanja na razumijevanje velikih podataka, 7 pitanja na spremnost dijeljenja osobnih podataka, 4 pitanja na percepciju društvene dobrobiti te 4 pitanja na percepciju štetnih posljedica koje veliki podaci nose. Korišten je Google Forms alat u svrhu izrade ankete te je provođenje istraživanja trajalo od 26.03.2024. do 09.04.2024. U svrhu prikupljanja odgovora, postavljena je poveznica na društvenim mrežama Facebook, LinkedIn, WhatsApp, te je korištena i metoda snježne grude kako bi se došlo do što većeg broja ispitanika. Ukupan broj ispitanika je 206.

Struktura odgovora na anketna pitanja prezentirana je apsolutnim i relativnim postotnim frekvencijama metoda deskriptivne statistike, i to aritmetičke sredine, te standardne devijacije.

U svrhu pružanja odgovora na prethodno postavljena istraživačka pitanja, korištena je metoda linearnog regresijskog modela, dok se značajnost procijenjenog modela kao cjeline ispituje F omjerom kod ANOVA testa. Analiza je rađena u statističkom softveru SPSS 25.

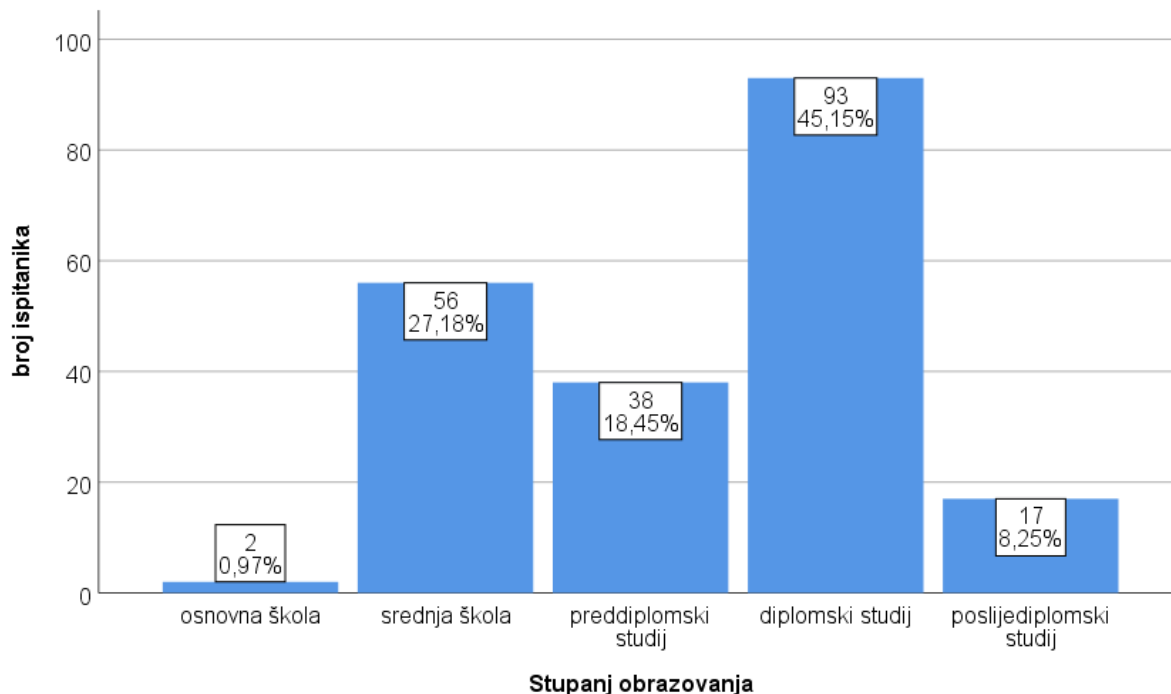
3.1. Demografski rezultati

Prema spolu je veća zastupljenost ispitanica kojih je 135 (65,53%) u odnosu na zastupljenost ispitanika kojih je 70 (33,98%), dok jedan ispitanik nije podijelio podatak o spolu (0,49%). Navedeno je grafički prikazano u nastavku.



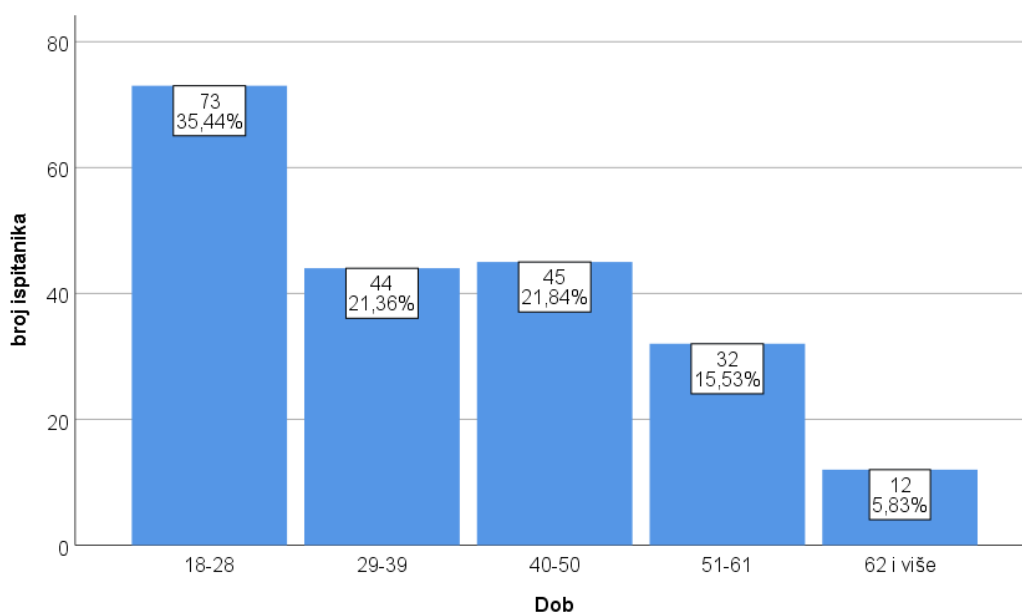
Grafički prikaz 1. Ispitanici prema spolu

Što se tiče strukture obrazovanja među ispitanicima, najveći broj ispitanika ima završen diplomski studij i to njih 93 (45,15%), dok su tek 2 ispitanika sa najviše završenom osnovnom školom (0,97%). Detalji su prikazani grafički u nastavku.



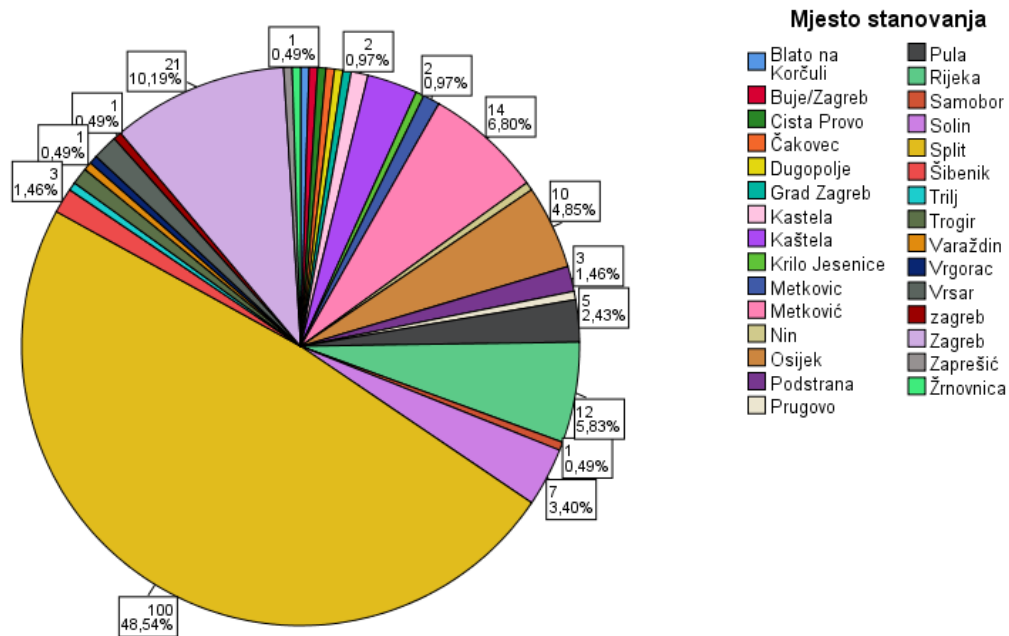
Grafički prikaz 2. Ispitanici prema razini obrazovanja

Prema dobi ispitanika najveći broj je mladih u rasponu dobi između 18 i 28 godina (n=73; 35,44%), dok je najmanja zastupljenost ispitanika dobne skupine 62 i više godina (n=12; 5,83%).



Grafički prikaz 3. Ispitanici prema dobi

Prema mjestu stanovanja se može utvrditi da je najveći broj ispitanika iz Splita (n=100; 48,54%), nakon čega slijede Kaštela sa 21 ispitanikom (10,19%).



Grafički prikaz 4. Ispitanici prema mjestu stanovanja

3.2. Stajališta ispitanika o velikim podacima

Sve percepcije i stavovi ispitanika mjereni su uz pomoć Likertove mjerene ljestvice. Korišten je raspon vrijednosti od 1 do 5, gdje vrijednost 1 upućuje na potpuno neslaganje s tvrdnjom, dok vrijednost 5 upućuje na potpuno slaganje s ponuđenom tvrdnjom.

Što se tiče stavova i percepcija o **razumijevanju korištenja osobnih podataka**, prosječna razina stava je 3,11 sa prosječnim odstupanjem od aritmetičke sredine 0,99. Najmanja razina slaganja je utvrđena kod tvrdnje „Jasno mi je tko ima pristup takvim podacima.“ Kod koje je utvrđena prosječna razina slaganja 2,60, dok je najviša razina slaganja utvrđena kod tvrdnje „Potpuno sam svjestan koje podatke dijelim putem interneta.“ Kod koje je utvrđena prosječna razina slaganja 3,57. Detalji su prikazani u Tablici 2.

	1		2		3		4		5		AS	SD
	n	%	n	%	n	%	n	%	n	%		
Upoznat/a sam s pojmom veliki podaci (engl. BIG DATA).	34	16,50	31	15,05	58	28,16	47	22,82	36	17,48	3,10	1,31

Razumijem za šta se takvi podaci koriste i koja je njihova svrha.	44	21,36	46	22,33	44	21,36	44	21,36	28	13,59	2,83	1,34
Jasno mi je tko ima pristup takvim podacima.	47	22,82	55	26,70	58	28,16	26	12,62	20	9,71	2,60	1,24
Potpuno sam svjestan koje podatke dijelim putem interneta.	11	5,39	27	13,24	49	24,02	68	33,33	49	24,02	3,57	1,15
Osjećam se ugodno učeći nove tehnologije te smatram da sam u tijeku s razvojem istih.	11	5,37	28	13,66	64	31,22	63	30,73	39	19,02	3,44	1,11
razumijevanje korištenja osobnih podataka											3,11	0,99

Tablica 2. Stavovi o razumijevanju korištenja osobnih podataka

Spremnost dijeljenja osobnih podataka je ispitana upotrebom 7 tvrdnji. Prosječna razina spremnosti dijeljenja osobnih podataka je 2,75 sa prosječnim odstupanjem od aritmetičke sredine 0,69. Najmanja razina stava o dijeljenju osobnih informacija je utvrđena kod tvrdnje „Vjerujem velikim tehnološkim tvrtkama i vladi kad su u pitanju moji osobni podaci.“ Kod koje je utvrđena prosječna razina stava 2,18, dok je najveća razina stava utvrđena kod tvrdnje „Uvijek prihvaćam sve uvjete web kolačića (engl. cookies).“ Kod koje je utvrđena prosječna razina stava 3,42.

	1		2		3		4		5		AS	SD
	n	%	n	%	n	%	n	%	n	%		
Vjerujem velikim tehnološkim tvrtkama i vladi kad su u pitanju moji osobni podaci.	76	37,25	53	25,98	48	23,53	17	8,33	10	4,90	2,18	1,16
Smatram sustav preporuke na Internetu korisnim jer mi preporučuje proizvode koji me zanimaju.	23	11,17	47	22,82	82	39,81	45	21,84	9	4,37	2,85	1,02
Općenito, imam pozitivan stav prema tehnologiji velikih podataka.	15	7,28	55	26,70	101	49,03	23	11,17	12	5,83	2,82	0,93

Na mobilnom uređaju redovito isključujem lokacijske usluge kako me ne bi mogli pratiti.*	40	19,51	44	21,46	34	16,59	47	22,93	40	19,51	3,01	1,42
Uvijek prihvaćam sve uvjete web kolačića (engl. cookies).	19	9,27	28	13,66	56	27,32	52	25,37	50	24,39	3,42	1,25
Bez straha unosim svoj broj kartice prilikom online kupovine.	72	34,95	40	19,42	42	20,39	33	16,02	19	9,22	2,45	1,35
Kod potrebne registracije na web shopu lako podijelim svoje osobne podatke.	60	29,13	45	21,84	53	25,73	31	15,05	17	8,25	2,51	1,28
spremnost dijeljenja osobnih podataka											2,75	0,69

*Kod izračuna ukupne vrijednosti konstrukta koriste se inverzne vrijednosti

Tablica 3. Spremnost dijeljenja osobnih podataka

Percepcija društvene dobrobiti je ispitana upotrebom seta od 4 tvrdnje, te je utvrđena prosječna razina stava 2,96 sa prosječnim odstupanjem od aritmetičke sredine 0,73. Najmanja razina stava je utvrđena kod tvrdnje „Upotreba velikih podataka povećava moj osjećaj sigurnosti jer ima mogućnost praćenja i predviđanja kriminala.“ Kod koje je utvrđena prosječna razina stava 2,69, dok je najveća razina stava utvrđena kod tvrdnje „Upotreba velikih podataka doprinosi razvoju prometne infrastrukture (npr. brojači prometa).“ Kod koje je utvrđena prosječna razina stava 3,25; umjerena, ali blago niža.

	1		2		3		4		5		AS	SD
	n	%	n	%	n	%	n	%	n	%		
Upotreba velikih podataka doprinosi razvoju zdravstvenog sektora i pruža mi bolju kontrolu trenutnog, prijašnjeg i budućeg zdravstvenog stanja.	23	11,17	53	25,73	81	39,32	35	16,99	14	6,80	2,83	1,06

Upotreba velikih podataka omogućava mi veću financijsku sigurnost jer takva tehnologija ima mogućnost prepoznavanja anomalija ukoliko netko neovlašteno upravlja mojim računom.	12	5,83	44	21,36	82	39,81	56	27,18	12	5,83	3,06	0,97
Upotreba velikih podataka doprinosi razvoju prometne infrastrukture (npr. brojači prometa).	5	2,44	42	20,49	82	40,00	48	23,41	28	13,66	3,25	1,01
Upotreba velikih podataka povećava moj osjećaj sigurnosti jer ima mogućnost praćenja i predviđanja kriminala.	29	14,08	53	25,73	88	42,72	25	12,14	11	5,34	2,69	1,03
percepcija društvene dobrobiti											2,96	0,73

Tablica 4. Percepcija društvene dobrobiti

Percepcija štetnih posljedica ispitana je upotrebom seta od 4 tvrdnje. Prosječna razina percepcije štetnih posljedica je 4,16 sa prosječnim odstupanjem od aritmetičke sredine 0,79. Najmanja razina stava je utvrđena kod tvrdnje „Kompanija koja koristi podatke o kandidatima poput spola i dobi prije finalne odluke o zaposlenju može dovesti do diskriminacije prema određenim skupinama ljudi.“ Kod koje je utvrđena prosječna razina stava 3,89, dok je najviša razina stava utvrđena kod tvrdnje „Kroz neovlašteni pristup moji podaci mogu biti ukradeni i na taj način može biti narušena moja osobna sigurnost.“ Kod koje je utvrđena prosječna razina stava 4,37.

	1		2		3		4		5		AS	SD
	n	%	n	%	n	%	n	%	n	%		
Korištenje mojih osobnih podataka može narušiti moju privatnost.	7	3,40	7	3,40	29	14,08	51	24,76	112	54,37	4,23	1,04
Kroz neovlašteni pristup moji podaci mogu biti ukradeni i na taj način može biti narušena moja osobna	5	2,44	8	3,90	22	10,73	42	20,49	128	62,44	4,37	0,99

sigurnost.

Kompanija koja koristi podatke o kandidatima poput spola i dobi prije finalne odluke o zaposlenju može dovesti do diskriminacije prema određenim skupinama ljudi.	7	3,40	11	5,34	58	28,16	52	25,24	78	37,86	3,89	1,08
Politička kampanja koja koristi podatke o biračima kako bi ciljano slala poruke koje bi utjecale na njihovo glasanje može dovesti do manipulacije i zlouporabe mojih podataka.	3	1,46	6	2,93	44	21,46	55	26,83	97	47,32	4,16	0,96
percepcija štetnih posljedica											4,16	0,79

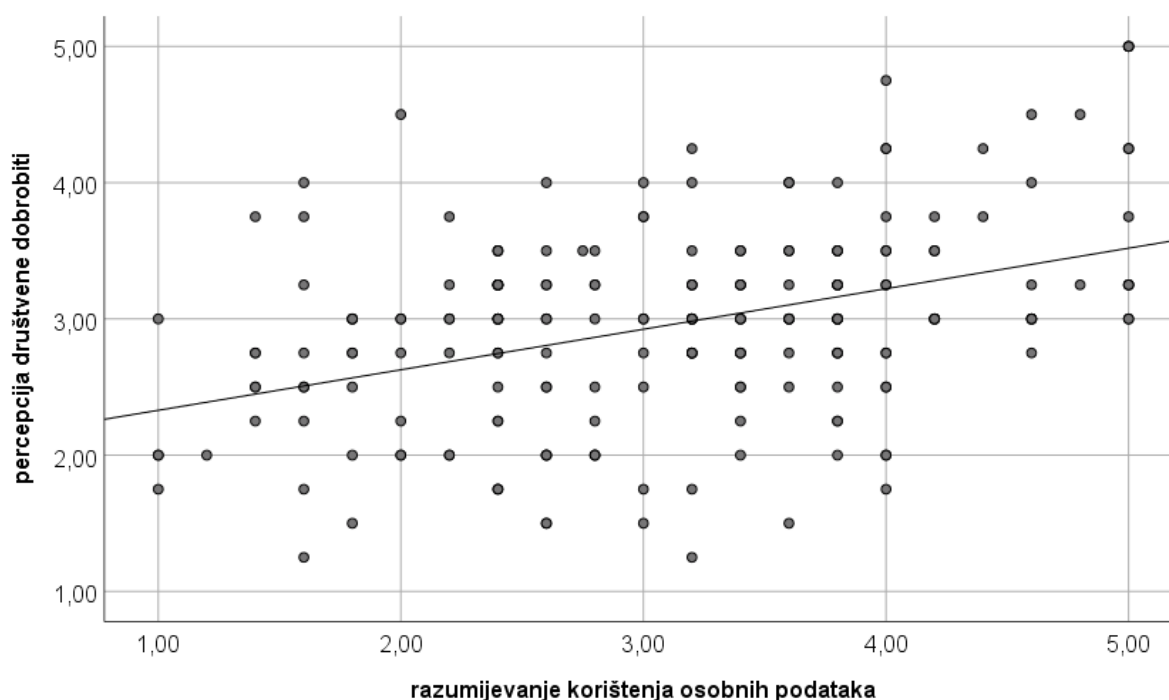
Tablica 5. *Percepcija štetnih posljedica*

3.3. Odgovori na postavljena istraživačka pitanja

IP 1 Utječe li razina razumijevanja korištenja osobnih podataka na percepciju društvene dobrobiti koju stvara upotreba tehnologije velikih podataka?

Odnos između *razumijevanja korištenja osobnih podataka* i *percepcije društvene dobrobiti* koju stvara upotreba tehnologije velikih podataka se prezentira dijagramom rasipanja.

Iz dijagrama se može utvrditi pozitivan odnos između razumijevanja korištenja osobnih podataka i percepcije društvene dobrobiti koju stvara upotreba tehnologije velikih podataka (Grafički prikaz 5.).



Grafički prikaz 5. *Dijagram odnosa između razumijevanja korištenja osobnih podataka i percepcije društvene dobrobiti*

Utjecaj razumijevanja korištenja osobnih podataka na percepciju društvene dobrobiti koju stvara upotreba tehnologije velikih podataka ispitan je linearnom regresijom. Parametar uz nezavisnu varijablu razumijevanja korištenja osobnih podataka ima vrijednost 0,297, odnosno svakim porastom razumijevanja korištenja osobnih podataka za jedan bod se može očekivati porast percepcije društvene dobrobiti u prosjeku za 0,297 bodova i suprotno. Parametar je statistički značajan ($P < 0,001$).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,032	,152		13,327	,000
	razumijevanje korištenja osobnih podataka	,297	,047	,407	6,362	,000

Tablica 6. Prikaz rezultata dobivenih linearnom regresijom

Prema procijenjenim modelu nezavisna varijabla razumijevanja korištenja osobnih podataka tumači 16,6% sume kvadrata odstupanja percepcije dobrobiti od aritmetičke sredine, dok je prosječno odstupanje utvrđene razine percepcije društvene dobrobiti od procjene modela je 0,67 bodova.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,407 ^a	,166	,161	,66690
a. Predictors: (Constant), razumijevanje korištenja osobnih podataka				

Tablica 7. Pokazatelji reprezentativnosti modela

Prema rezultatima ANOVA testa procijenjeni model je statistički značajan ($F=40,479$; $P<0,001$).

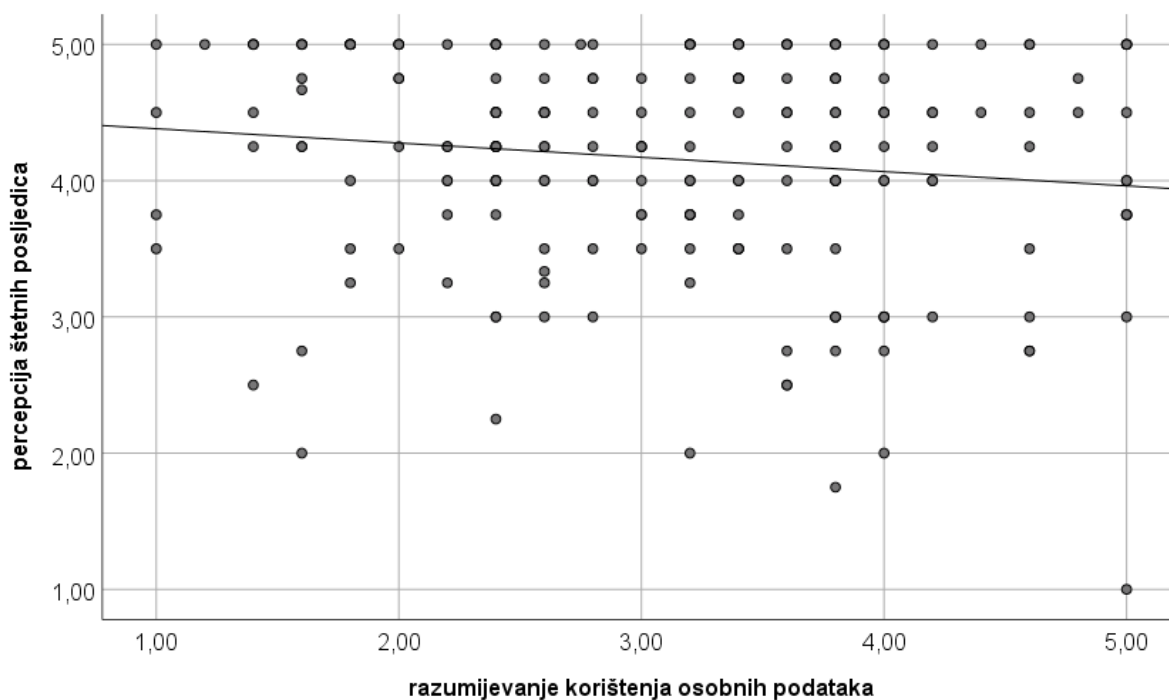
ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	18,003	1	18,003	40,479	,000 ^b
	Residual	90,729	204	,445		
	Total	108,732	205			
a. Dependent Variable: percepcija društvene dobrobiti						
b. Predictors: (Constant), razumijevanje korištenja osobnih podataka						

Tablica 8. Prikaz rezultata ANOVA testa

IP 2 Utječe li razina razumijevanja korištenja osobnih podataka na percepciju štetnih posljedica koju stvara upotreba tehnologije velikih podataka?

Odnos između *razumijevanja korištenja osobnih podataka* i *percepcije štetnih posljedica* koju stvara upotreba tehnologije velikih podataka se prezentira dijagramom rasipanja.

Iz dijagrama se može utvrditi negativan odnos između razumijevanja korištenja osobnih podataka i percepcije štetnih posljedica koje stvara upotreba tehnologije velikih podataka.



Grafički prikaz 6. Dijagram odnosa između razumijevanja korištenja osobnih podataka i percepcije štetnih posljedica

Utjecaj razumijevanja korištenja osobnih podataka na percepciju štetnih posljedica koju stvara upotreba tehnologije velikih podataka ispitan je linearnom regresijom. Parametar uz nezavisnu varijablu razumijevanja korištenja osobnih podataka ima vrijednost $-0,105$, odnosno svakim porastom razumijevanja korištenja osobnih podataka za jedan bod se može očekivati smanjenje percepcije štetnih posljedica u prosjeku za $0,105$ bodova i suprotno. Parametar je statistički značajan pri graničnoj razini signifikantnosti od 10% ($P=0,059$).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,486	,180		24,880	,000
	razumijevanje korištenja osobnih podataka	-,105	,055	-,132	-1,899	,059

Tablica 9. Prikaz rezultata dobivenih linearnom regresijom

Prema procijenjenim modelu nezavisna varijabla razumijevanja korištenja osobnih podataka tumači 1,7% sume kvadrata odstupanja percepcije štetnih posljedica od aritmetičke sredine, dok je prosječno odstupanje utvrđene razine percepcije štetnih posljedica od procjene modela je 0,79 bodova.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,132 ^a	,017	,013	,79
a. Predictors: (Constant), razumijevanje korištenja osobnih podataka				

Tablica 10. Pokazatelji reprezentativnosti modela

Prema rezultatima ANOVA testa procijenjeni model je statistički značajan pri graničnoj razini signifikantnosti od 10% (F=3,605; P=0,059).

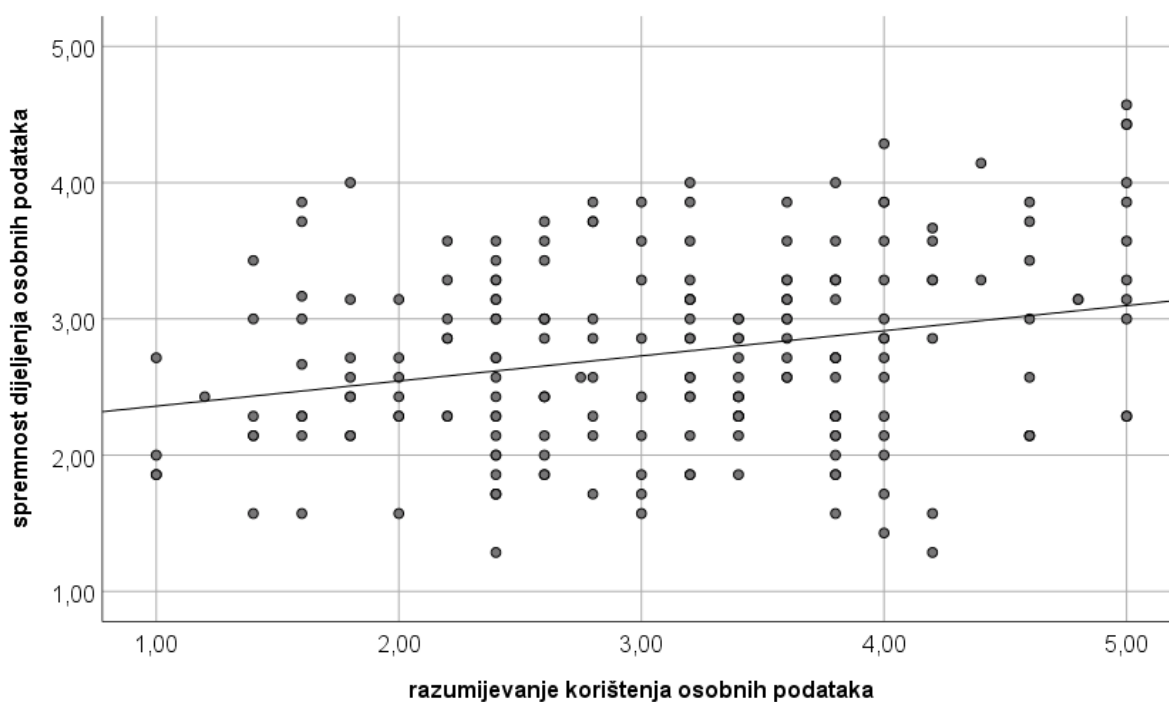
ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2,242	1	2,242	3,605	,059 ^b
	Residual	126,861	204	,622		
	Total	129,102	205			
a. Dependent Variable: percepcija štetnih posljedica						
b. Predictors: (Constant), razumijevanje korištenja osobnih podataka						

Tablica 11. Prikaz rezultata ANOVA testa

IP3 Utječe li razina razumijevanja korištenja osobnih podataka na spremnost dijeljenja takvih podataka?

Odnos između *razumijevanja korištenja osobnih podataka* i *spremnosti dijeljenja podataka* koju stvara upotreba tehnologije velikih podataka se prezentira dijagramom rasipanja.

Iz dijagrama se može utvrditi pozitivan odnos između razumijevanja korištenja osobnih podataka i spremnosti dijeljenja podataka koje stvara upotreba tehnologije velikih podataka.



Grafički prikaz 7. dijagram odnosa između razumijevanja korištenja osobnih podataka i spremnosti dijeljenja podataka

Utjecaj razumijevanja korištenja osobnih podataka na dijeljenje osobnih podataka koje stvara upotreba tehnologije velikih podataka ispitan je linearnom regresijom. Parametar uz nezavisnu varijablu razumijevanja korištenja osobnih podataka ima vrijednost 0,266, odnosno svakim porastom razumijevanja korištenja osobnih podataka za jedan bod se može očekivati porast spremnosti dijeljenja osobnih podataka u prosjeku za 0,266 bodova i suprotno. Parametar je statistički značajan ($P < 0,001$).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2,176	,153		14,246	,000
	razumijevanje korištenja osobnih podataka	,184	,047	,266	3,939	,000

Tablica 12. prikaz rezultata dobivenih linearnom regresijom

Prema procijenjenim modelu nezavisna varijabla razumijevanja korištenja osobnih podataka tumači 7,1% sume kvadrata odstupanja dijeljenja osobnih podataka od aritmetičke sredine, dok je prosječno odstupanje utvrđene razine percepcije štetnih posljedica od procjene modela je 0,67 bodova.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,266 ^a	,071	,066	,67
a. Predictors: (Constant), razumijevanje korištenja osobnih podataka				

Tablica 15. pokazatelji reprezentativnosti modela

Prema rezultatima ANOVA testa procijenjeni model je statistički značajan (F=15,51; P<0,001).

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6,921	1	6,921	15,512	,000 ^b
	Residual	91,012	204	,446		
	Total	97,932	205			
a. Dependent Variable: spremnost dijeljenja osobnih podataka						
b. Predictors: (Constant), razumijevanje korištenja osobnih podataka						

Tablica 14. prikaz rezultata dobivenih ANOVA testom

4. RASPRAVA

U raspravi će se odgovoriti na postavljena istraživačka pitanja te ustanoviti jesu li dobiveni rezultati istraživanja u skladu sa očekivanjima ili ne. Također, željelo se usporediti rezultate ovog istraživanja s postojećima, međutim rezultati tih postojećih istraživanja odnosno temeljna pitanja koja su se proučavala se nisu podudarala s ovim radom, ali se izučavanjem teorijske literature došlo do određenih tvrdnji koje su se mogle rezultatima istraživanja ovog diplomskog rada prihvatiti ili opovrgnuti. Prvo istraživačko pitanje glasi: „**Utječe li razina razumijevanja korištenja osobnih podataka na percepciju društvene dobrobiti koju stvara upotreba tehnologije velikih podataka?**“ te se može zaključiti da razina razumijevanja korištenja osobnih podataka pozitivno utječe na percepciju društvene dobrobiti koju stvara upotreba tehnologije velikih podataka odnosno što je društvo više informirano i educirano o korištenju osobnih podataka to će više rast i percepcija društvene dobrobiti. Za ovakav pozitivan odnos možemo reći da je u skladu s očekivanjima jer pojedinci razumiju načine korištenja osobnih podataka i sve dobrobiti koje tehnologija velikih podataka može donijeti, dok u slučaju nerazumijevanja korištenja osobnih podataka pada i percepcija društvene dobrobiti velikih podataka. Veće razumijevanje i znanje svakako donose pozitivniju i bolju percepciju velikih podataka. Kao što je već rečeno više nego ikad ostavljamo digitalni trag dok pretražujemo internet te je ovo i utvrđeno putem istraživanja gdje je društvo u cjelini svjesno koliko različitih vrsta osobnih podataka ostavlja za sobom online.

Na drugo istraživačko pitanje: „**Utječe li razina razumijevanja korištenja osobnih podataka na percepciju štetnih posljedica koju stvara upotreba tehnologije velikih podataka?**“ statističkom obradom podataka je zaključeno da razina razumijevanja korištenja osobnih podataka negativno utječe na percepciju štetnih posljedica koju stvara upotreba tehnologije velikih podataka odnosno što je društvo više informirano i educirano o korištenju osobnih podataka smanjivat će se percepcija štetnih posljedica, dok u slučaju nerazumijevanja korištenja osobnih podataka povećavat će se percepcija štetnih posljedica. Prema Gruschki i suradnicima (2018) mnogi ljudi nisu upoznati s načinima prikupljanja podataka putem algoritama i tehnologije velikih podataka te ne razumiju koje su posljedice ove prakse, dok se u ovom istraživanju utvrdilo suprotno jer je percepcija štetnih posljedica značajna što nadalje znači da je društvo upoznato s negativnom stranom tehnologije velikih podataka i posljedicama koje ima za njih. Također utvrđeno je da je kod građana RH percepcija društvene dobrobiti manja od percepcije štetnih posljedica što nam pokazuje da se građani više povezuju sa štetnim posljedicama nego s dobrobitima, odnosno postoji nesigurnost i zabrinutost korištenja osobnih podataka u vidu krađe i gubitka privatnosti.

Za treće istraživačko pitanje: „**Utječe li razina razumijevanja korištenja osobnih podataka na spremnost dijeljenja takvih podataka?**“ statističkom analizom je utvrđeno da razina razumijevanja korištenja osobnih podataka pozitivno utječe na spremnost dijeljenja takvih podataka odnosno što je društvo više informirano i educirano o korištenju osobnih podataka to će više rast i spremnost dijeljenja osobnih

podataka. Kako raste razumijevanje korištenja osobnih podataka među pojedincima to će oni biti spremniji dijeliti svoje podatke. Ovakav rezultat je u skladu s očekivanjima ukoliko su osobni podaci građana kvalitetno osigurani i zaštićeni. Međutim valja napomenuti da je istraživanjem utvrđena prosječno niska razina spremnosti dijeljenja što znači da su građani RH oprezni prilikom dijeljenja svojih osobnih podataka. Možemo reći da je jedan od razloga niske spremnosti dijeljenja nepovjerenje prema velikim tehnološkim tvrtkama i vladi kada su u pitanju osobni podaci građana.

Ograničenja ovog istraživanja čine podaci koji su prikupljeni iz uzorka koji je uključivao značajan udio studenata diplomskih studija odnosno mlađe populacije koja dobno pripada u skupinu od 18 do 28 godina dok je starija populacija onih koji pripadaju dobnoj skupini od 51 godine manja. Također većina ispitanika pripada Splitsko-dalmatinskoj županiji odnosno gradu Splitu s obzirom da je bilo teško prikupiti jednak broj anketnih odgovora ispitanika iz ostalih dijelova Hrvatske.

5. ZAKLJUČAK

Ovaj diplomski rad detaljno analizira funkcionalnosti tehnologije velikih podataka u vidu prikupljanja osobnih podataka od strane građana fokusirajući se na koristi i prijetnje koje društvo može imati. Kroz anketni upitnik prikupljeni su stavovi i percepcije hrvatskih građana o velikim podacima te se putem empirijskog istraživanja i prikazanog teorijskog izučavanja uz obradu prikupljenih podataka donose određeni zaključci.

Provedeno istraživanje naglašava važnost razumijevanja korištenja osobnih podataka u kontekstu tehnologije velikih podataka te utjecaj na percepciju društvenih dobrobiti i štetnih posljedica, kao i spremnost pojedinaca za dijeljenje svojih podataka. U skladu s očekivanjima, rezultati pokazuju da veće razumijevanje koristi tehnologije velikih podataka pozitivno utječe na percepciju društvenih dobrobiti, dok istovremeno smanjuje percepciju štetnih posljedica. Osim toga, veće informiranje i edukacija pojedinaca o korištenju osobnih podataka potiče njihovu spremnost za dijeljenje tih podataka, pod uvjetom da su osigurani i zaštićeni. Ovi zaključci naglašavaju važnost transparentnosti, edukacije i zaštite privatnosti kako bi se maksimizirale prednosti tehnologije velikih podataka uz minimaliziranje rizika i negativnih posljedica.

Utjecaj velikih podataka na društvo, posebno u kontekstu osobnih podataka, je složen i nužno uzima u obzir i negativne posljedice i pozitivne aspekte koji su detaljno prikazani u ovom diplomskom radu. Moć ovakve tehnologije je vrlo snažna pa je važno naći balans između zaštite privatnosti pojedinaca i korištenja podataka za ostvarivanje društvenih dobrobiti. To se može postići primjenom strogih regulativa o zaštiti podataka, edukacijom građana o njihovim pravima i rizicima, te promicanjem transparentnosti i odgovornosti u korištenju velikih podataka.

LITERATURA

1. Abadi, D. (2009). Data Management in the Cloud: Limitations and Opportunities. IEEE Bulletin of the Technical Committee on Data Engineering.
2. Aho, B. & Duffield, R. (2020). Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China, Economy and Society. Taylor & Francis Group.
3. Alharthi, A., Krotov, V., Bowman, M. (2017). Addressing barriers to big data. Kelley School of Business, Indiana University. Published by Elsevier Inc.
4. Alalawneh, A.A.F. & Alkhatib, S.F. (2020). The barriers to big data adoption in developing economies. Business Administration Department, Faculty of Economic and Administrative Sciences, Yarmouk University, Irbid. John Wiley & Sons Ltd.
5. Ali, M. M. (2013). Role of data mining in education sector. International Journal of Computer Science and Mobile Computing, 2(4).
6. Bakken S., & Reame N. (2016). The promise and potential perils of Big Data for advancing symptom management research in populations at risk for health disparities.
7. Bazina, O.O. (2020). Human rights and biometric data protection; Social credit system. Moscow State Institute of International Relations – MGIMO-University. Przegląd Europejski, ISSN: 1641-2478 vol. 2020, no. 4.
8. Bay, S., Kumaraswamy, K., Anderle, M. G., Kumar, R., Steier, D. M. (2006). *Large Scale Detection of Irregularities in Accounting Data*. Sixth International Conference on Data Mining (ICDM'06), IEEE, 1-4.
9. Bennett, C. J., & Bayley R. M. (2016). *Privacy protection in the era of 'big data': regulatory challenges and social assessments* In van der Sloot, B., Broeders, D., & Schrijvers, E. *Exploring the Boundaries of Big Data*. Amsterdam University Press, 205-208.
10. Bertino, E. (2015). Big Data – Security and Privacy. IEEE International Congress on Big Data
11. Bottis, M. & Bouchagiar, G.(2018). Personal Data v. Big Data in the EU: Control Lost, Discrimination Found. Open Journal of Philosophy, 8, 192-205.
12. Bhadani, A., & Jothimani, D. (2016). Big data: Challenges, opportunities, and realities. Pennsylvania, PA: IGI Global.
13. Brayne S. (2017). Big Data surveillance: the case of policing. Am Sociol Rev. 82(5):977–1008.

14. Calo, R. (2014). Digital market manipulation. *George Washington Law Review*, 82(4), 995–1051.
15. Citron, D. & Pasquale, F. (2014). The Scored Society: Due Process for Automated Prediction, *Washington Law Review* 89: 1.
16. Chen, J., Chen, Y., Du, X., Li, C., Lu, J. (2013). Big data Challenge: a data management perspective, Key Laboratory of Data Engineering and Knowledge Engineering, School of Information, Renmin University of China.
17. Creemers, R. (2018). China's Social Credit System: An Evolving Practice of Control. *SSRN Electronic Journal*, 1–32.
18. Creswell, J. W., Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th edition). Sage, 206-207.
19. Desai, B.C. (2015). Technological Singularities. Concordia University. Montreal, Canada
20. Dumbill, E., Liddy, E., Stanton, J., Mueller, K., & Farnham, S. (2013). Educating the next generation of data scientists. Mary Ann Liebert, Inc.
21. Eurostat (2018). *Cloud computing services used by more than one out of four enterprises in the EU*. Eurostat Press Office, 2.
22. Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
23. Favaretto, M., De Clercq, E., & Elger, B.S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. *J Big Data* 6, 12.
24. George, B., & Paul, J. (2020). *Digital transformation in business and society*. New York, NY: Springer International Publishing, 16-17.
25. Goold, B. J. (2020). Not just about privacy: Police body-worn cameras and the costs of public area surveillance. Taylor and Francis.
26. Government Office for Science (2020). Evidence and scenarios for global data systems. *Foresight*, 1- 4.
27. Gutierrez, D. (2014). Worldwide big data market revenues for software and services in the next 20 years.
28. Gruschka, N., Mavroeidis, V., Vishni, K., Jensen, M. (2018). *Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR*. IEEE, 5027-5029.

29. Harford, T. (2014). Big data: Are we making a big mistake?
30. Hashem, I. A. T., Yaqoob, I., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2015). Big data: From beginning to future. *International Journal of Information Management*
31. Herschel, R., Miori, V. (2017). *Ethics & Big Data*, *Technology in Society*, 5.
32. Hu, X., & Sastry, N. (2019). Characterising third party cookie usage in the EU after GDPR. In *Proceedings of the 10th ACM Conference on Web Science (WebSci'19)*. ACM, 137–141.
33. Hrvatski sabor (2023). *Strategija digitalne Hrvatske za razdoblje do 2032. godine*. Dostupno na:
34. Hrvatska Poštanka Banka (HPB) (2022). *INFORMACIJE O OBRADI OSOBNIH PODATAKA U SVRHU*
35. Kache, F., Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management*.
36. Kocijan, K. (2014). *Big Data: kako smo došli do Velikih podataka i kamo nas oni vode*. Zavod za informacijske studije
37. Kopal, R., & Karas, Ž. (2017). *BIG DATA IN LAW ENFORCEMENT: FROM REACTIVE TO PROACTIVE*. Ministarstvo unutarnjih poslova Republike Hrvatske, 2.
38. Kostka, G. (2019). China's social credit systems and public opinion: Explaining high levels of approval. *New Media & Society*, 21(7), 1565-1593.
39. Kovacic, Z. (2012). Predicting student success by mining enrolment data. *Research in Higher Education*, 15.
40. Kretschmer, M., Pennekamp, J., & Wehrle, K. (2021). Cookie Banners and Privacy Policies: Measuring the Impact of the GDPR on the Web. *ACM Trans. Web* 15, 4, Article 20
41. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics, and the path from insights to value. *MIT Sloan Management Review*.
42. Lazarsfeld, P.F. (2019). Notes on the History of Quantificain in Sociology- Trends, Sources and Problems. Columbia University
43. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, Vol 343.
44. Leibold, J. (2020). Surveillance in China's Xinjiang region: ethnic sorting, coercion, and inducement. La Trobe. Journal contribution.

45. Majnarić, L.T., Babić, F., O'Sullivan, S., & Holzinger, A. (2021). AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. *10*, 766.
46. Mayer-Schönberger, V., Cukier, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcour.
47. Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big data: Prospects and challenges. *Vikalpa*, 40(1), 74.
48. Morozov, E. (2012). *The Net Delusion: The Dark Side of Internet Freedom*, New York. Public Affairs.
49. Meng, W., Ou, W., Chandwani, S., Chen, X., Black, W., & Cai, Z. (2019). Temporal phenotyping by mining healthcare data to derive lines of therapy for cancer, *Journal of Biomedical Informatics*.
50. Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of 'datification'. *J Strategic Inf Syst*.
51. Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. A. (2016). Convolutional net for medical records. *IEEE J. Biomed. Health Inform*.
52. Official Journal of the European Union (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. L 119/1
53. Pardes, A. (2018). What is GDPR and why should you care? *Wired*. Dostupno na: <https://www.wired.com/story/how-gdpr-affects-you/> (Datum pristupa: 05.12.23.)
54. Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2017). Predicting healthcare trajectories from medical records: A deep learning approach. *J. Biomed. Inform*.
55. Rahmati V. (2016). Big Data: Now and Then. *International Journal of Emerging Computing Methods in Engineering* , Vol. 1, No. 2
56. Russom, P. (2011). Big data analytics. TDWI (The Data Warehousing Institute™), a division of 1105 Media, Inc.
57. Tao, H., Bhuiyan, Z.A., Rahman, A., Wang, G., Wang, T., Ahmed, M., Li, J. (2019). *Economic perspective analysis of protecting big data security and privacy*, *Future Generation Computer Systems*, 661.
58. Tankard, C. (2016). What the GDPR means for businesses. *Digital Pathways*.
59. Van der Sloot, B., Broeders, D., & Schrijvers E. (2016). *Exploring the Boundaries of Big Data*. wrr/Amsterdam University Press, 27-30.

60. Vlada Republike Hrvatske (2018). *PRIJEDLOG ZAKONA O PODACIMA I INFORMACIJAMA U ZDRAVSTVU*. 4.
61. Zelenika, R. (2000). *Metodologija i tehnologija izrade znanstvenog i stručnog djela*. Rijeka: Ekonomski fakultet.
62. Sánchez-Monedero, J., & Dencik, L. (2019). The politics of deceptive borders: 'biomarkers of deceit' and the case of iBorderCtrl. University of Cordoba, Spain.
63. Satra, H. S. (2019). Freedom under the gaze of Big Brother: Preparing the grounds for a liberal defence of privacy in the era of Big Data. Østfold University College, Norway. Elsevier Ltd.
64. Schumpeter, J.A., 1934 (2008). *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, New Brunswick (U.S.A) and London (U.K.): Transaction Publishers.
65. Shi, Y. (2022) "*Digital Economy: Development and Future*," Bulletin of Chinese Academy of Sciences (Chinese Version): Vol. 37 : Iss. 1 , Article 11., 1-3.
66. Steinberg, E. (2019). Big Data and Personalized Pricing. *Business Ethics Quarterly* 30(1):1-21
67. Strielkowski, W., Faminskaya M., & Potekhina E. (2021). Urban infrastructure via Big Data. University of California. EDP Sciences.
68. Vandenhole, W. J. A. (2005). Non-Discrimination and Equality in the View of the UN Human Rights Treaty Bodies. Intersentia.
69. Van Rijmenam, M. (2014). *Think bigger: Developing a successful big data strategy for your business*. New York: AMACOM.
70. Zorić, A.B. (2020). Benefits of educational data mining. *Journal of International Business Research and Marketing* . Volume 6, Issue 1.
71. Zuboff, S. (2015). Big Other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*.
72. Wiedemann, G. (2013). Opening up to Big Data: Computer-Assisted Analysis of Textual Data in Social Sciences. *Forum: Qualitative Social Research*, 14(2), Art. 23
73. Xiao, C., Choi E., & Sun J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, 0(0), 2018, 1–10.

IZVORI S INTERNETA

1. Awati, R., Charles, M., & DelVecchio A. (2024). Clinical decision support system (CDSS). Dostupno na: <https://www.techtarget.com/searchhealthit/definition/clinical-decision-support-system-CDSS> (Datum pristupa: 10.01.24.)
2. Bay Atlantic University (2023). Characteristics of Big Data: Types&Examples. Dostupno na: <https://bau.edu/blog/characteristics-of-big-data/> (Datum pristupa: 25.11.23.)
3. Botelho, B., Bigelow, S.J. (2022). Big data. Dostupno na: <https://www.techtarget.com/searchdatamanagement/definition/big-data> (Datum pristupa: 22.11.23.)
4. Copping, R., & Li, M. (2016). The Promise and Challenge of Big Data for Pharma. Harvard Business Review. Dostupno na: <https://hbr.org/2016/11/the-promise-and-challenge-of-big-data-for-pharma> (Datum pristupa: 18.01.24.)
5. Coursera (2024). 10 Common Types of Cyberattacks and How to Prevent Them. Dostupno na: <https://www.coursera.org/articles/types-of-cyber-attacks> (Datum pristupa: 25.02.24.)
6. Curtis, A.(2023). 7 V's of Big Data Explained with Infographic. BizTech. Dostupno na: <https://yourtechdiet.com/blogs/7vs-of-big-data/> (Datum pristupa: 20.11.23.)
7. Duhigg, C. (2012). How companies learn your secrets. New York Times Magazine. Dostupno na: <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html> (Datum pristupa: 22.02.24.)
8. Europska komisija (2024). Što su osobni podaci? Dostupno na: https://commission.europa.eu/law/law-topic/data-protection/reform/what-personal-data_hr (Datum pristupa: 14.11.23.) i The Digital Services Act. Dostupno na: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en (Datum pristupa 23.05.24.)
9. Službena internetska stranica Europske Unije (2020). U izvješću Komisije o pravilima EU-a za zaštitu osobnih podataka zaključeno je da ta pravila osnažuju građane i da su primjerena za digitalno doba. Dostupno na: https://croatia.representation.ec.europa.eu/news/u-izvjescu-komisije-o-pravilima-eu-za-zastitu-osobnih-podataka-zakljuceno-je-da-ta-pravila-osnazuju-2020-06-25_hr (Datum pristupa: 15.11.23.)
10. European Council(2024). Data protection in the EU. Dostupno na: <https://www.consilium.europa.eu/en/policies/data-protection/> (Datum pristupa: 15.11.23.)

11. E-građani (2024). What is the General data Protection Regulation (GDPR)? Dostupno na: <https://gov.hr/hr/sto-je-opca-uredba-o-zastiti-podataka-eng-general-data-protection-regulation-gdpr/1868?lang=en> (Datum pristupa: 15.11.2023.)
12. Furman, J., & Simcoe T. (2015). The Economics of Big Data and Differential Pricing. Dostupno na: <https://obamawhitehouse.archives.gov/blog/2015/02/06/economics-big-data-and-differential-pricing> (Datum pristupa: 20.02.24.)
13. Gallagher, R. (2015). KARMA POLICE: GCHQ's plan to track every Web user in the world. Dostupno na: <https://boingboing.net/2015/09/25/karma-police-gchqs-plan-to.html> (Datum pristupa: 22.02.24)
14. Narodne novine (2023). Strategija digitalne Hrvatske za razdoblje do 2032. Hrvatski sabor. Dostupno na: https://narodne-novine.nn.hr/clanci/sluzbeni/2023_01_2_17.html (Datum pristupa: 10.04.23.)
15. ODOBRAVANJA RIZIČNIH PROIZVODA. Dostupno na: <https://www.hpb.hr/hr/zastita-osobnih-podataka/2970> (Datum pristupa: 02.04.23.)
16. Hvistendahl, M. (2017). Inside China's Vast New Experiment in Social Ranking. Real Clear Politics. Dostupno na: https://www.realclearpolitics.com/2017/12/19/inside_china039s_vast_new_experiment_in_social_ranking_429523.html (Datum pristupa: 16.12.23.)
17. Newcomer, E. (2017). Uber starts charging what it thinks you're willing to pay. Bloomberg. Dostupno na: <https://economictimes.indiatimes.com/news/international/world-news/uber-starts-charging-what-it-thinks-youre-willing-to-pay/articleshow/59241678.cms?from=mdr> (Datum pristupa: 25.02.24.)
18. Padma, M. (2023). Big Data Analytics In Healthcare Market - Forecast(2023 - 2028). Dostupno na: <https://www.linkedin.com/pulse/big-data-analytics-healthcare-market-forecast2023-2028-padma/> (Datum pristupa: 17.01.24.)
19. Thornhill, J. (2017). The Big Data revolution can revive the planned economy. Dostupno na: <https://www.ft.com/content/6250e4ec-8e68-11e7-9084-d0c17942ba93> (Datum pristupa: 17.12.23.)
20. Sundararajan B. (2023). Harnessing the Power of Now: The Art of Real-Time Big Data Analytics. Dostupno na: <https://medium.com/@bragadeeshs/harnessing-the-power-of-now-the-art-of-real-time-big-data-analytics-38658be7c8a8> (Datum pristupa: 01.12.23.)

POPIS SLIKA, TABLICA I GRAFOVA

Slika 1. Karakteristike velikih podataka

Slika 2. Izvori velikih podataka

Slika 3. Podaci o građanima

Slika 4. Proces obrade osobnih podataka

Slika 5. primjer prihvaćanja „kolačića“ na internetskoj stranici tportal.hr

Tablica 1. „System of carrots and sticks“

Tablica 2. stavovi o razumijevanju korištenja osobnih podataka

Tablica 3. spremnost dijeljenja osobnih podataka

Tablica 4. percepcija društvene dobrobiti

Tablica 5. percepcija štetnih posljedica

Tablica 6. prikaz rezultata dobivenih linearnom regresijom

Tablica 7. pokazatelji reprezentativnosti modela

Tablica 8. prikaz rezultata ANOVA testa

Tablica 9. prikaz rezultata dobivenih linearnom regresijom

Tablica 10. pokazatelji reprezentativnosti modela

Tablica 11. prikaz rezultata ANOVA testa

Tablica 12. prikaz rezultata dobivenih linearnom regresijom

Tablica 15. pokazatelji reprezentativnosti modela

Tablica 14. prikaz rezultata dobivenih ANOVA testom

Grafički prikaz 1. ispitanici prema spolu

Grafički prikaz 2. ispitanici prema razini obrazovanja

Grafički prikaz 3. ispitanici prema dobi

Grafički prikaz 4. ispitanici prema mjestu stanovanja

Grafički prikaz 5. dijagram odnosa između razumijevanja korištenja osobnih podataka i percepcije društvene dobrobiti

Grafički prikaz 6. dijagram odnosa između razumijevanja korištenja osobnih podataka i percepcije štetnih posljedica

Grafički prikaz 7. dijagram odnosa između razumijevanja korištenja osobnih podataka i spremnosti dijeljenja podataka

SAŽETAK

Ovaj diplomski rad obuhvaća proučavanje upotrebe tehnologije velikih podataka u vidu društvenih dobrobiti i štetnih posljedica koje takvi podaci imaju za društvo u cjelini. Fokus će biti stavljen na osobne podatke koji se prikupljaju na različite načine gdje njihova analiza može pozitivno i negativno utjecati na pojedinca. Uvod u rad ima za cilj objasniti koncept velikih podataka, njihovu svrhu i važnost u današnjem digitaliziranom društvu. Obuhvatit će se i teorijski pregled literature gdje će se pobliže obrazložiti tehnologija i upotreba ovakvih podataka. Naglasit će se pozitivne i negativne strane kako bi se dobio realan uvid u korisnost odnosno štetnost velikih podataka prema „običnom“ građaninu. Metodologija istraživanja obuhvaćat će prikupljanje relevantnih podataka o stavovima građana RH o velikim podacima, razumijevanju i spremnosti dijeljenja osobnih podataka kroz anketu gdje će se putem analize deskriptivnom statistikom doći do konačnih rezultata kako bi se uvidjelo stajalište hrvatskih građana prema velikim podacima. Na kraju će se obraditi rasprava, donijet će se odgovori na istraživačka pitanja i zaključiti o daljnjim mogućnostima i preporukama u području tehnologije velikih podataka kroz analizu osobnih podataka pojedinca.

Ključne riječi: tehnologija velikih podataka, osobni podaci, društvene dobrobiti, štetne posljedice, pojedinac, digitalizirano društvo

SUMMARY

This master's thesis encompasses the study of big data technology usage in terms of its societal benefits and adverse effects on society as a whole. The focus will be on personal data collected through various means, where its analysis can positively or negatively impact individuals. The introduction aims to elucidate the concept of big data, its purpose, and significance in today's digitized society. It will also include a theoretical literature review, elaborating on the technology and its use. Both positive and negative aspects will be emphasized to provide a realistic insight into the utility or harm of big data to the "average" citizen. The research methodology will involve gathering relevant data on the attitudes of Croatian citizens towards big data, their understanding, and willingness to share personal data through surveys, with final results derived through descriptive statistical analysis to discern the viewpoint of Croatian citizens towards big data. Finally, a discussion will be conducted, answering research questions and concluding on further possibilities and recommendations in the field of big data technology through individual data analysis.

Key words: big data technology, personal data, societal benefits, adverse effects, individual, digitized society

PRIMJENA TEHNOLOGIJE VELIKIH PODATAKA: STAJALIŠTA HRVATSKIH GRAĐANA O DRUŠTVENIM DOBROBITIMA I ŠTETNIM POSLJEDICAMA



Poštovani,

pozivam Vas za sudjelovanje u anonimnom istraživanju koje se provodi u svrhu izrade diplomskog rada. Ovom anketom istražuje se stajalište Hrvatskih građana o tehnologiji velikih podataka gledajući društvene dobrobiti i štetne posljedice za pojedinca. Rezultati će se koristiti isključivo u svrhu istraživanja. Anketa se sastoji od 24 pitanja te će Vam za njeno ispunjavanje trebati otprilike 5 minuta.

* Sudjelovanje u anketi je dobrovoljno, te imate mogućnost prekinuti anketu u bilo kojem trenutku. Prilikom obrade podataka, pridržavat ćemo se svih propisa koji osiguravaju anonimnost ispitanika i povjerljivost prikupljenih podataka.

Zahvaljujem na izdvojenom vremenu.

Demografska pitanja



Ova sekcija ima za cilj prikupiti osnovne podatke o ispitanicima.

Spol

- muško
- žensko



Dob

- 18-28
- 29-39
- 40-50
- 51-61
- 62 i više

Mjesto stanovanja

Tekst kratkog odgovora



Stupanj obrazovanja

- bez završenog osnovnog obrazovanja
- osnovna škola
- srednja škola
- preddiplomski studij
- diplomski studij
- poslijediplomski studij

Razumijevanje velikih podataka



Ova sekcija istražuje stupanj razumijevanja velikih podataka.

Ocijenite stupanj slaganja s slijedećim tvrdnjama:

Upoznat/a sam s pojmom veliki podaci (engl. BIG DATA).

	1	2	3	4	5	
U potpunosti se ne slažem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	U potpunosti se slažem

Razumijem za šta se takvi podaci koriste i koja je njihova svrha.

	1	2	3	4	5	
U potpunosti se ne slažem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	U potpunosti se slažem

Jasno mi je tko ima pristup takvim podacima.

	1	2	3	4	5	
U potpunosti se ne slažem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	U potpunosti se slažem



Potpuno sam svjestan koje podatke dijelim putem interneta.

	1	2	3	4	5	
U potpunosti se ne slažem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	U potpunosti se slažem

Osjećam se ugodno učeći nove tehnologije te smatram da sam u tijeku s razvojem istih.

	1	2	3	4	5	
U potpunosti se ne slažem	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	U potpunosti se slažem

Spremnost dijeljenja osobnih podataka



Ova sekcija istražuje spremnost dijeljenja osobnih podataka građana.

Vjerujem velikim tehnološkim tvrtkama i vladi kad su u pitanju moji osobni podaci.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Smatram sustav preporuke na Internetu korisnim jer mi preporučuje proizvode koji me zanimaju.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem



Općenito, imam pozitivan stav prema tehnologiji velikih podataka.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Na mobilnom uređaju redovito isključujem lokacijske usluge kako me ne bi mogli pratiti.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Uvijek prihvaćam sve uvjete web kolačića (engl. cookies).

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Bez straha unosim svoj broj kartice prilikom online kupovine.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

...

Kod potrebne registracije na web shopu lako podijelim svoje osobne podatke.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Odjeljak 5 od 6

Društvene dobrobiti



Ova sekcija istražuje percepciju društvene dobrobiti ispitanika.

Upotreba velikih podataka doprinosi razvoju zdravstvenog sektora i pruža mi bolju kontrolu trenutnog, prijašnjeg i budućeg zdravstvenog stanja.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

Upotreba velikih podataka omogućava mi veću financijsku sigurnost jer takva tehnologija ima mogućnost prepoznavanja anomalija ukoliko netko neovlašteno upravlja mojim računom.

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

...

Upotreba velikih podataka doprinosi razvoju prometne infrastrukture (npr. brojači prometa).

1 2 3 4 5

U potpunosti se ne slažem U potpunosti se slažem

⋮

Upotreba velikih podataka povećava moj osjećaj sigurnosti jer ima mogućnost praćenja i predviđanja kriminala.

U potpunosti se ne slažem 1 2 3 4 5 U potpunosti se slažem

Odjeljak 6 od 6

Štetne posljedice

✕ ⋮

Ova sekcija istražuje percepciju štetnih posljedica koje veliki podaci nose.

Korištenje mojih osobnih podataka može narušiti moju privatnost.

U potpunosti se ne slažem 1 2 3 4 5 U potpunosti se slažem

Kroz neovlašteni pristup moji podaci mogu biti ukradeni i na taj način može biti narušena moja osobna sigurnost.

U potpunosti se ne slažem 1 2 3 4 5 U potpunosti se slažem

Kompanija koja koristi podatke o kandidatima poput spola i dobi prije finalne odluke o zaposlenju može dovesti do diskriminacije prema određenim skupinama ljudi.

U potpunosti se ne slažem 1 2 3 4 5 U potpunosti se slažem

Politička kampanja koja koristi podatke o biračima kako bi ciljano slala poruke koje bi utjecale na njihovo glasanje može dovesti do manipulacije i zlouporabe mojih podataka.

U potpunosti se ne slažem 1 2 3 4 5 U potpunosti se slažem

