

Što skrivaju masovni podaci u zdravstvu?

Mario Somek

Zdravstveno veleučilište u Zagrebu, Zagreb, Hrvatska

e-pošta: mariosomek@gmail.com

Sažetak: Složenost zdravstvenog sustava i ubrzana digitalizacija u raznim područjima impliciraju stvaranje velike količine raznolikih podataka. Prema tradicionalnom shvaćanju podataka, takvi podaci se razlikuju i zahtijevaju poseban pristup pri upravljanju i analizi te ih možemo nazvati masovnim podacima. U radu su opisana obilježja masovnih podataka u zdravstvu, izazovi u radu, navedeni su izvori masovnih podataka, primjeri analize i smjernice u radu. Masovni podaci predstavljaju višedisciplinarno područje, zahtijevaju poznavanje podatkovne tehnologije, analitičkih alata i uz primjenu informacijske tehnologije razvoj standarda u području prikupljanja podataka.

Ključne riječi: masovni podaci; obilježja; izvori; standardi; analiza podataka

Uvod

Intenzivna digitalizacija poslovanja omogućava u pojedinim područjima i djelatnostima stvaranje podataka koje, u odnosu na tradicionalne podatke, obilježavaju brzina prikupljanja, količina i raznovrsnost te takve podatke možemo nazvati masovni podaci (engl. big data). Prema definiciji Europskog parlamenta pojam masovni podaci odnosi se na prikupljene skupove podataka koji su toliko veliki i složeni da ih moraju obrađivati nove tehnologije, kao što je umjetna inteligencija (1).

Procjenjuje se da je u 2018. godini više od 5 milijardi korisnika svakodnevno razmjenjivalo podatke i da će do 2025. taj broj korisnika biti 6 milijardi ili 75% svjetske populacije, a zahvaljujući milijardama povezanih internetskih uređaja (engl. Internet of Things, IoT) širom svijeta očekuje se u 2025. stvaranje preko 90 ZB (Zetabajta) podataka (2).

Multipliciranje podataka ubrzava i pojava događaja poput pandemije, što dovodi do intenzivnijeg i ekstenzivnijeg prikupljanja i obrade podataka s ciljem promptnog i točnog izvještavanja upravljačkog osoblja i javnosti. Masovni podaci predstavljaju potencijal za prepoznavanje bolesti u začecima koje se stoga mogu efikasnije tretirati na dobrobit pojedinca i čitave populacije (3).

Prema jednom od vodećih svjetskih pružatelja usluga revizije i osiguranja, procjenjuje se da će u 2021. god. u svijetu biti više od 400 milijuna virtualnih posjeta liječnicima i da će taj broj porasti 5% prema procijenjenih 1% u 2019. godini (4). Podaci su opsežni, složeni, zanimljivi i skrivaju informacije pogodne za rješavanje javnozdravstvenih problema, predviđanje i sprječavanje bolesti (5). Veliki istraživački centri i fondacije uložile su mnogo u područje istraživanja masovnih podataka pa su tako Američki nacionalni zavodi za zdravlje izdvojili 100 milijuna američkih dolara za inicijativu „Masovnim podacima do znanja“ (6).

Područje masovnih podataka zadnjih godina intenzivno se razvija u zdravstvu i predstavlja potencijal koji može pridonijeti poboljšanju zdravstvene skrbi.

U radu se daje pregled obilježja masovnih podataka u zdravstvu, izazova u manipuliranju, prikazani su izvori podataka, opisuju se standardi u prikupljanju podataka, mogućnosti analize te iznose smjernice u radu s masovnim podacima.

Obilježja masovnih podataka

U počecima opažanja posebnosti masovnih podataka i predviđanja razvoja, Laney D. 2001. godine definira tri obilježja: volumen, brzinu i raznolikost (7). Panimalar A. i dr., 2017. godine u svom radu opisuju sedamnaest (8), a Hussein A.A. 2020. opisuje pedeset i šest obilježja masovnih podataka (9). Trend pojavljivanja obilježja upućuje na složenost podataka koja zahtijeva poseban pristup pri manipulaciji i upotrebi u istraživačke svrhe i zasigurno dinamičan razvoj informacijske tehnologije i podatkovne znanosti pridonijet će otkrivanju novih obilježja. Na Slici 1 prikazana su obilježja masovnih podataka u području zdravstva s kratkim opisom.

VOLUMEN	BRZINA	RAZNOLIKOST	ISTINITOST	VRIJEDNOST	POVJERLJIVOST
					
Količina za manipulaciju	Dinamika prikupljanja i obrade	Oblik, strukturiranost	Pouzdanost izvora	Važnost obzirom na istraživanje	Tajnost, legislativa
Velik	Velika	Velika	Visoka	Visoka	Visoka

Slika 1. Obilježja masovnih podataka u području zdravstva, prilagođeno prema (10)

- Volumen predstavlja količinu podataka kojom se manipulira. Prema jednom od vodećih svjetskih mrežnih sjedišta za analizu podataka, u istraživanju iz 2020. godini provedenom kod korisnika koji obavljaju analizu podataka, najčešća veličina podataka kreće se u rasponu 1,1 – 10 GB (11).
- Brzina opisuje dinamiku stvaranja i obrade. U području zdravstvene skrbi ona je velika zbog različitih izvora koji su svakodnevno u funkciji od 0 do 24 sata (laboratoriji, klinike, administracija, razne slikovne pretrage itd.).
- Raznolikost, osim različitosti u izvornom obliku (npr. brojčane vrijednosti, datumski intervali, slikovni oblici u radiološko dijagnostičkom području i sl.), podrazumijeva tri skupine podataka: strukturirani, nestrukturirani i polustrukturirani. Pripadnost skupini određuje jesu li prikupljeni prema definiranom modelu i pohranjeni u odgovarajućem obliku (12).
- Istinitost predstavlja kvalitetu podataka s obzirom na provjeren izvor (npr. laboratorijski informacijski sustav predstavlja provjerljiv, a društvene mreže neprovjerljiv izvor).
- Vrijednost je povezana s procesom analize i pretpostavlja prepoznavanje reprezentativnih podataka važnih za istraživanje (npr. vrijednosti podataka važnih za postavljanje dijagnoze).
- Povjerljivost u području zdravstvene skrbi važna je jer su podaci osobne prirode i podložni zakonskoj regulativi.

Izazovi u radu

Količina masovnih podataka u zdravstvenom sustavu svakodnevno raste i u kratkom vremenu dostiže vrijednosti koje se iskazuju gigabajtima i terabajtima što predstavlja primarni izazov u smislu osiguranja diskovnog prostora za spremanje, izrade sigurnosne kopije i pristupa podacima. Standardni alati za rad s podacima postaju nedostatni te se primjenjuju tehnologije sažimanja, posebno u području slikovnih zapisa (13).

Pohranu medicinskih podataka potrebno je planirati zbog upravljanja povijesnim podacima o pacijentima i operacijama u budućnosti te je važno razlikovati sigurnosnu kopiju, arhivu i aktivnu arhivu. Stvaranjem sigurnosne kopije sprječava se gubitak važnih podataka. Arhiviranjem se oslobađaju skupi diskovni kapaciteti premještanjem manje upotrebljivih podataka na isplativija mjesta za pohranu, a kod aktivnog arhiviranja koriste se skalabilni sustavi s inteligentnim softverom koji koristi metapodatke, indeksiranje i posebno označavanje kako bi se osigurao dohvat podataka uz visoke performanse (14).

Pri upotrebi standardnih tehnologija za obradu podataka, veličina i heterogenost masivnih podataka čine ih manje upotrebljivim i korisnim. Najčešće softverske platforme za analizu bazirane su na računalnim klasterima velike snage i koriste tzv. distribuirano procesiranje podataka te im se pristupa putem mrežne infrastrukture. Neki od najpoznatijih sustava otvorenog koda za analizu masovnih podataka jesu Hadoop i Apache Spark (12).

Podaci u oblaku

U vrijeme kada milijuni digitalnih uređaja komuniciraju međusobno diljem svijeta koristeći razne internetske mreže pristup podacima omogućen je u svakom trenutku. Upotrebom tehnologije spremanja podataka u oblak i uz primjenu informacijske infrastrukture zdravstvene organizacije smanjuju troškove spremanja i povećavaju pouzdanost (12). Pružatelji usluga u oblaku imaju kapacitete i razvili su kvalitetne resurse da omoguće različitim sektorima efikasniju i ekonomičniju pohranu masovnih podataka. Pohrana u oblaku omogućava medicinskim stručnjacima dohvat i transfer informacija, analizu podataka u stvarnom vremenu, daljinsku dijagnostiku i razvoj drugih usluga poput telemedicine (15). Centraliziran pristup podacima omogućavat će liječnicima i drugom medicinskom osoblju da brzo i jednostavno dohvate sve potrebne podatke i tako uštede vrijeme koje u procesu liječenja mogu iskoristiti za rad s pacijentima. Prema smjernicama Međunarodne podatkovne organizacije (engl. International Data Corporation), migracija podataka u oblak je strategija pri planiranju razvoja tvrtke te se pretpostavlja da će do 2022. godine biti više podataka pohranjenih u javnom oblaku nego u tradicionalnim podatkovnim centrima (2).

Izvori masovnih podataka

U sustavu zdravstva izvore podataka predstavljaju bolnički zapisi, medicinski zapisi pacijenata, rezultati medicinskih istraživanja i razni digitalni uređaji povezani putem interneta (engl. Internet of Things) (12). U Tablici 1 prikazani su izvori podataka u zdravstvu s opisom i primjerima.

Tablica 1. Izvori zdravstvenih podataka, prilagođeno prema (16).

Izvor	Opis	Primjer
Klinički podaci	Podaci o ispitivanjima, razna opažanja	Dijagnoze, indikacije
Pacijentovi podaci	Biometrijske osobine, podaci s društvenih mreža	Otisak prsta, komentari pacijenata
Senzorni podaci	Podaci iz mjernih uređaja	Uzorci krvi, dijagnostičke pretrage
Genomski podaci	Tipizacija gena	Ekspresija gena, DNA sekvencioniranje
Podaci kliničkih istraživanja	Podaci o zdravstvenim proizvodima	Lijekovi, cjepiva
Vanjski podaci	Osiguranje, administrativni podaci	Osnovno i dopunsko osiguranje

Sustavima za e-naručivanje pacijenti generiraju nove zahtjeve, pametni telefoni primaju i šalju podatke za udaljene terapije, zdravstveno osoblje generira podatke za elektroničke zdravstvene zapise, slikovni podaci s pomoću raznih senzornih uređaja spremaju se u oblaku. Objedinjavanje podataka iz različitih izvora dovodi do novih osobina. Primjerice, podaci iz jednog područja su manjeg opsega u odnosu na podatke iz drugog područja. Pojavljuje se i raznolika skalabilnost (vrijednosti se iskazuju u različitim rasponima), neskladnost i nepotpunost podataka (17).

Standardi u prikupljanju podataka

Razvoj i primjena perifernih medicinskih uređaja i pripadajućih aplikacija olakšava i ubrzava prikupljanje podataka i istovremeno zahtijeva razvoj pripadajućih standarda koji će osigurati jedinstven model za prikupljanje i razmjenu masovnih podataka.

Hijerarhijski format podataka, verzija 5 (engl. Hierarchical Data Formats 5, HDF5), format je datoteke otvorenog koda koji podržava velike, složene, heterogene podatke. Format osigurava mogućnost organiziranja podataka u datoteci na razne strukturirane načine s opcijom samo-opisa zahvaljujući metapodacima (18).

U dijagnostičkom području prevladavaju razni slikovni podaci, međunarodno prihvaćeni format za pregled, pohranu, preuzimanje i dijeljenje medicinskih slika je DICOM (engl. Digital Imaging and Communications in Medicine). Ovaj format sadrži protokole za provjeru točnosti informacija i standardno sadrži dva dijela: zaglavlje u kojem se nalazi opis slike s podacima o pacijentu i drugi dio koji sadrži sliku (19).

Velika raznolikost i izvori podataka u zdravstvu zahtijevaju razvoj specifičnih standarda (normi) za svako pojedino područje. Pregled važnijih standarda u zdravstvu s navođenjem područja primjene i organizacije koja razvija standard objavili su autori Schulz S. i dr. u radu Standards in Healthcare Data (20).

Važan dio zdravstvenog informacijskog sustava koji omogućava sistematizirano i standardizirano prikupljanje podataka u području zdravstva je Elektronički zdravstveni zapis (EZZ). EZZ objedinjuje sve podatke prikupljene u procesu prevencije i liječenja pacijenta u

zdravstvenom sustavu od rođenja do smrti uz odgovarajuću razinu dostupnosti i sigurnosti podataka. Osigurava brže prikupljanje podataka i olakšava izvještavanje o ključnim pokazateljima kvalitete zdravstvene zaštite, a također pridonosi boljem nadzoru javnog zdravstva omogućavajući izradu izvješća neposredno po izbivanju bolesti (12). O važnosti organiziranog prikupljanja podataka zasnovanog na standardu govore autori Girardi F. i dr. u svom radu gdje ističu da primjena tehnologije zasnovane na EZZ može pomoći u sprječavanju pogrešaka kod nekih dijagnostičkih i terapijskih postupaka (21). Mnoge zdravstvene ustanove implementirale su EZZ, no još uvijek i kod najboljih svjetskih zdravstvenih sustava postoji razlika u nazivima i terminologiji (22) jer se često EZZ prilagođava potrebama ustanove u kojoj se koristi. Prioritet bi trebale biti nacionalne i globalne potrebe s težnjom stvaranja interoperabilnog EZZ koji će omogućavati objedinjeno korištenje podataka više ustanova istog područja djelovanja (npr. klinički podaci). Interoperabilnost dvaju EZZ podrazumijeva nesmetanu razmjenu i korištenje podataka za što je potrebno da poruka koja se prenosi sadrži standardizirane kodirane podatke kako bi ih sustav prijema mogao interpretirati (22).

U uvođenju standardizacije podataka važnu ulogu imaju strukovne udruge koje u suradnji s upravnim tijelima trebaju težiti donošenju regulative koja će omogućavati primjenu normi i standarda kako bi podaci iz jedne zemlje bez tehničkih i tehnoloških prepreka mogli postati dio globalnih podataka.

Analiza masovnih podataka

Prikupljanje masovnih podataka nema neku posebnu vrijednost ukoliko ne postoji mogućnost analize i dobivanja rezultata na temelju kojih se mogu poduzeti korisne radnje (23).

Analiza predstavlja proces izdvajanja informacija s ciljem potpore poslovnom odlučivanju pri čemu se koriste razni postupci analize koji se prema obliku učenja dijele u tri skupine: nadzirano, nenadzirano i polunadzirano učenje (17). Pripadnost skupini ovisi o tome jesu li unaprijed očekivane vrijednosti rezultata analize nepoznate ili su u cijelosti ili djelomično poznate. Za kvalitetnu i uspješnu provedbu analize potrebno je prethodno pripremiti podatke što podrazumijeva poznavanje tehnika predprocesiranja poput: integracije, čišćenja, normalizacije, skaliranja (24). Na predprocesiranje rijetko kada se utroši manje od 70% vremena cjelokupne analize (25).

Ciljevi analize velikih podataka u području zdravstva su predikcija, modeliranje i zaključivanje te se prema tome često koriste postupci klasifikacije, klasteriranja i regresije (17). Klasifikacijski postupci (26) predstavljaju nadzirano učenje pomoću kojih se stvaraju predviđajući modeli sa svrhom brže i preciznije dijagnostike. U svom radu Zriqat I. A. i dr. opisuju klasifikacijski postupak analize sa svrhom predviđanja srčanih oboljenja (27). Klasteriranje (26) spada u nenadzirano učenje s ciljem stvaranja klastera sličnih obilježja. U svom radu Baeka J. W. i dr. uz pomoć algoritma srednjih vrijednosti s ciljem uvođenja dijete kod bolesnika s kroničnim bolestima, formiraju klaster na temelju bolesti pacijenata te prema pravilima nutritivne struke za svaki klaster preporučuju odgovarajuću prehranu (28). Regresija (26) spada u nadzirano učenje i koristi se za utvrđivanje odnosa između promatranih podataka te omogućava prikaz trenda. Pilot-studija autora Ayyoubzadeh S. M. i dr. usmjerena na dubinsku analizu, temeljem prikupljenih podataka predviđa incidenciju Covid-19 u Iranu (29). Postupci analize zasnovani su na pripadajućim algoritmima, a istraživači dobiveni rezultat nastoje evaluirati raznim tehnikama validacije s ciljem dobivanja optimalnog rezultata.

Provođenje analize podrazumijeva poznavanje specifičnih tehnika i alata za što su potrebni odgovarajući ljudski resursi. Pretpostavka je da u današnje vrijeme s podacima u zdravstvu još uvijek najviše manipuliraju profesionalci iz područja zdravstvenih struka, a kako je područje

masovnih podataka višedisciplinarno, postavlja se pitanje educiranosti kadrova za obavljanje tih poslova. U prepoznavanju potencijala masovnih podataka i potreba tržišta rada obrazovne ustanove organiziraju edukacije za stjecanje certifikata (30) za obavljanje poslova medicinskog podatkovnog analitičara, a visokoškolske ustanove uvode i specijalističke studije iz područja podatkovne znanosti (31).

Smjernice u radu s masovnim podacima

Prema navedenom u radu, autor preporučuje važnije smjernice u radu s masovnim podacima:

- Upravljanje masovnim podacima zahtjeva primjenu tehnologija pohrane podataka poput aktivnog arhiviranja i spremanja podataka u oblak (engl. Cloud).
- Prikupljanje podataka standardizirati primjenom obrazaca zasnovanih na sadržaju elektroničkog zdravstvenog zapisa.
- Na izvoru podataka težiti generiranju strukturiranih podataka jer su kao takvi spremni za analizu i nije potrebno poznavanje posebnih tehnika pripreme podataka.
- U planiranju razvoja zdravstvenog informacijskog sustava primijeniti međunarodno priznate standarde.
- U radu s masovnim podacima nastojati iskoristiti potencijal „novog znanja“. Koristiti provjerene tehnike analize s ciljem smanjenja pogrešaka i povećanja kvalitete u zdravstvenoj skrbi.
- S obzirom na potrebu višedisciplinarnog znanja u radu s masovnim podacima, obrazovne ustanove trebale bi pokrenuti odgovarajuću edukaciju i/ili studijske programe u području podatkovne znanosti.
- Upravljačko osoblje zdravstvenih ustanova u suradnji s podatkovnim stručnjacima za provedbu istraživačkih projekata zasnovanih na masovnim podacima treba predvidjeti odgovarajuća financijska sredstva.

Zaključak

Masovni podaci u zdravstvu zahtijevaju poseban pristup pri manipuliranju i upravljanju, poznavanje odgovarajuće tehnologije i tehnika za rad uz istovremeno osiguranje visoke povjerljivosti. Postojeći protokoli i standardi neizostavni su za kvalitetno prikupljanje i sekundarnu upotrebu podataka te ih je potrebno razvijati u tom smjeru. Poznavanje postupaka analize i odgovarajuća primjena dobivenih rezultata ubrzava i pospješuje dijagnostičke postupke, smanjuje troškove poslovanja i pridonosi kvalitetnoj zdravstvenoj skrbi. Razvoj znanosti u području podataka zahtjeva odgovarajuću edukaciju i kadrove koji će samostalno ili u suradnji sa zdravstvenim profesionalcima biti sposobni upravljati masovnim podacima i iskoristiti njihov potencijal.

Literatura

1. Veliki podaci: definicija, koristi i izazovi (infografika). Dostupno na: <https://www.europarl.europa.eu/news/hr/headlines/society/20210211STO97614/veliki-podaci-definicija-koristi-izazovi-infografika> (pristupljeno 11.5.2021.)
2. The Digitization of the World From Edge to Core. Dostupno na: <https://resources.moredirect.com/white-papers/idc-report-the-digitization-of-the-world-from-edge-to-core> (pristupljeno 20.5.2021.)
3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;7;2:3. doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3).
4. Deloitte Global tech, media and telecom predictions: growth in digital reality, virtual doctor visits, cloud, and sports tech intensify. Dostupno na: <https://www2.deloitte.com/global/en/pages/about-deloitte/press-releases/deloitte-global-tech-media-and-telecom-predictions.html?nc=1> (pristupljeno 18.6.2021.)
5. Holgado J L, Lopez C, Fernandez A, Sauri I, Uso R, Trillo J L, Vela S, Nuñez J, Redon J, Ruiz A. Acute kidney injury in heart failure: a population study. ESC Heart Failure. ESC Heart Fail. 2020;7(2):415-422. doi.org/10.1002/ehf2.12595.
6. Wang W, Krishnan E. Big data and clinicians: a review on the state of the science. JMIR Med Inform. 2014;2(1):e1. doi: [10.2196/medinform.2913](https://doi.org/10.2196/medinform.2913).
7. 3d Data Management: Controlling Data Volume, Velocity and Variety. Dostupno na: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an> (pristupljeno 18.6.2021.)
8. Panimalar A, Shree V, Kathrine V. The 17 V's Of Big Data. International Research Journal of Engineering and Technology. 2017;4(9):329-333. Dostupno na: <https://www.irjet.net/archives/V4/i9/IRJET-V4I957.pdf>
9. Hussein A. A. How Many Old and New Big Data V's Characteristics, Processing Technology And Applications (BD1). International Journal of Application or Innovation in Engineering & Management. 2020;9(9):15-27. Dostupno na: <https://www.ijaiem.org/Volume9Issue9/IJAIEM-2020-09-17-9.pdf>
10. Accelerating the Path from Data to Foresight for Healthcare. Dostupno na: <https://cdn.ymaws.com/hisa.site-ym.com/resource/resmgr/bigdata2013/DrLauraHaas2.pdf> (pristupljeno 15.6.2021.)
11. Largest Dataset Analyzed – Poll Results and Trends. Dostupno na: <https://www.kdnuggets.com/2020/07/poll-largest-dataset-analyzed-results.html> (pristupljeno 10.6.2021.)
12. Dash S, Shakyawar S K, Sharma M et al. Big data in healthcare: management, analysis and future prospects. J of Big Data 2019;6:54:1-25. doi: 10.1186/s40537-019-0217-0.
13. Sin K, Muthu L. Application Of Big Data In Education Data Mining And Learning Analytics – A Literature Review. Journal On Soft Computing 2015;5(4):1035-1049. doi: 10.21917/ijsc.2015.0145.
14. Health Data Volumes Skyrocket, Legacy Data Archives On the Rise. Dostupno na: <https://www.harmonyhit.com/health-data-volumes-skyrocket-legacy-data-archives-rise-hie/> (pristupljeno 18.6.2021.)
15. What Is Telemedicine? Dostupno na: <https://vsee.com/aboutus> (pristupljeno 20.6.2021.)
16. Raja R, Mukherjee I, Sarkar B K. A Systematic Review of Healthcare Big Data. Scientific Programming 2020;(6):1-15. doi: 10.1155/2020/5471849
17. Choong H L, Hyung-Jin Y. Medical big data: promise and challenges. Kidney Res Clin Pract, 2017;36(1):3-11. doi: 10.23876/j.krcp.2017.36.1.3.

18. Hierarchical Data Formats - What is HDF5? Dostupno na: <https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5> (pristupljeno 2.7.2021.)
19. What is DICOM in Medical Imaging? Dostupno na: <https://www.postdicom.com/en/blog/handling-dicom-medical-imaging-data> (pristupljeno 2.7.2021.)
20. Schulz S, Stegwee R, Chronaki C. Fundamentals of Clinical Data Science: Standards in Healthcare Data. Springer Open 2019 eBook. Dostupno na: <https://link.springer.com/book/10.1007/978-3-319-99713-1#about> (pristupljeno 3.7.2021.)
21. Girardi F, Gennaro GD, Colizzi L, Convertini N. Improving the Healthcare Effectiveness: The Possible Role of EHR, IoMT and Blockchain. Electronics. 2020;9(6):884. doi: 10.3390/electronics9060884.
22. Reisman M. EHRs: The Challenge of Making Electronic Data Usable and Interoperable. P & T journal for formulary management. 2017;42(9):572-575.
23. Ketchersid T. Big Data in Nephrology: Friend or Foe? Blood Purif. 2013;36(3-4):160-4. doi:10.1159/000356751.
24. Misra P, Yadav A S. Impact of Preprocessing Methods on Healthcare Predictions. Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE). 2019:144-150. doi:10.2139/ssrn.3349586.
25. Gibert K, Sánchez-Marrè M, Izquierdo J. A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. AI Communications. 2016;29(6):627-663. doi:10.3233/AIC-160710.
26. Hong L, Luo M, Wang R, Lu P, Lu W, Lu L. Big Data in Health Care: Applications and Challenges. Data and Information Management. 2018;2(3):175-197. doi:10.2478/dim-2018-0014.
27. Zriqat I A, Altamimi A M, Azzeh M. A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods. International Journal of Computer Science and Information Security (IJCSIS). 2016;14(12):868-879. Dostupno na: <https://arxiv.org/ftp/arxiv/papers/1704/1704.02799.pdf>
28. Baeka J W, Kima J C, Chunb J, Chung K. Hybrid clustering based health decision-making for improving dietary habits. Technology and Health Care. 2019;27(5):459-472. doi:10.3233/THC-191730.
29. Ayyoubzadeh S M, Ayyoubzadeh S M, Zahedi H, Ahmadi M, Niakan Kalhori S R. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill. 2020;14;6(2):e18828. doi:10.2196/18828.
30. Certifications. Dostupno na: <https://www.ahima.org/certification/chda> (pristupljeno 21.6.2021.)
31. Podatkovna znanost. Dostupno na: <https://www.algebra.hr/visoko-uciliste/studij/diplomski-studij/podatkovni-smjer/> (pristupljeno 22.6.2021.)