Važnost umjetne inteligencije za rudarenje podataka u zdravstvu

Tušek, Karlo

Sveučilište Jurja Dobrile u Puli
Fakultet Informatike u Puli

KARLO TUŠEK

VAŽNOST UMJETNE INTELIGENCIJE ZA RUDARENJE PODATAKA U ZDRAVSTVU

Završni rad

Pula, 2022.

Sveučilište Jurja Dobrile u Puli
Fakultet Informatike u Puli

KARLO TUŠEK

VAŽNOST UMJETNE INTELIGENCIJE ZA RUDARENJE PODATAKA U ZDRAVSTVU

Završni rad

JMBAG: 0303076057, redovni student

Studijski smjer: Sveučilišni preddiplomski studij Informatika

Kolegij: Informacijska tehnologija i društvo

Znanstveno područje: Društvene znanosti

Znanstveno polje: Informacijske i komunikacijske znanosti

Mentorica: doc. dr. sc. Snježana Babić

Pula, rujan, 2022.

IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisani, Karlo Tušek,ovime izjavljujem da je ovaj Završni rad rezultat isključivo mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da niti jedan dio Završnog rada nije napisan na nedozvoljen način, odnosno da je prepisan iz kojega ne citiranog rada, te da i koji dio rada krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

Karlo Tušek

U Puli, 21.09.2022. 2022. godine

IZJAVA

o korištenju autorskog djela

Ja, Karlo Tušek, dajem odobrenje Sveučilištu Jurja Dobrile u Puli, kao nositelju prava iskorištavanja, da moj završni rad pod nazivom „Važnost umjetne inteligencije za rudarenje podataka u zdravstvu“ koristi na način da gore navedeno autorsko djelo, kao cjeloviti tekst trajno objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te kopira u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu s Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnog pristupa znanstvenim informacijama. Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

U Puli, 21.09.2022.

Potpis

Karlo Tušek

SADRŽAJ

1. UVOD

Tema ovog završnog rada je „Važnost umjetne inteligencije za rudarenje podataka u zdravstvu“. Cilj rada je analizirati i definirati važnost umjetne inteligencije za rudarenje podataka u zdravstvu, dok je svrha rada dati teorijski doprinos istraživanju navedene tematike, kao i bolje razumijevanje iste u praksi.

Tehnološke inovacije su danas napredovale te su značajne za brojna područja ljudskog djelovanja. U tom pogledu veoma su važne inovacije umjetne inteligencije i strojnog učenja, i tehnika umjetne inteligencije pod nazivom rudarenje podataka.

Strojno učenje je danas svugdje oko nas, primjerice kod automatiziranog prevođenja, prepoznavanja slika, glasovnog pretraživanja, navigacijskih sustava u automobilima i slično.

Strojno učenje i rudarenje podataka su danas veoma značajni u poslovanju brojnih sektora, kao sastavni dio kontinuirane evolucije umjetne inteligencije. Razvoj strojnog učenja je danas ubrzan, te je raširen na brojna područja poput pametne proizvodnje, maloprodaje, medicine, zdravstva, igara, poljoprivrede, arheologije te dr. U ovom radu naglasak je na analizi umjetne inteligencije, odnosno strojnog učenja i rudarenja podataka u zdravstvu.

Rad se sastoji od četiri poglavlja: od uvoda i tri glavna poglavlja. Nakon uvoda, u drugom dijelu rada govori se o umjetnoj inteligenciji i strojnom učenju. Najprije se objašnjava opći pojam umjetne inteligencije i strojnog učenja, zatim se objašnjava proces strojnog učenja, te vrste i modeli strojnog učenja.

U trećem dijelu rada biti će riječi o rudarenju podataka. Objašnjavaju se podaci, informacije i znanje, zatim se općenito definira rudarenje podataka, tehnike, metode i modeli rudarenja podataka, objašnjava se odnos umjetne inteligencije i rudarenja podataka, te aplikacije rudarenja podataka.

U četvrtom dijelu rada fokus je na prikazivanju primjera iz prakse, u primjeni umjetne inteligencije za rudarenje podataka u zdravstvu. Najprije se prikazuje primjer otkrivanja lijekova i molekularno modeliranje umjetnom inteligencijom, a zatim dijagnostika raka i donošenje odluka o liječenju pomoću umjetne inteligencije. Nakon toga se objašnjavaju prednosti i nedostaci korištenja umjetne inteligencije za

rudarenje podataka u zdravstvu na prethodno prikazanim primjerima.

U zaključku se nalaze zaključna razmatranja na zadanu tematiku, dok se u popisu literature nalazi popis primarnih i sekundarnih izvora podataka, koji su korišteni pri izradi rada. Za izradu rada koristi se metoda analize i sinteze, povijesna metoda, metoda deskripcije, te metoda komparacije.

2. UMJETNA INTELIGENCIJA I STROJNO UČENJE

U ovom poglavlju će biti riječi o umjetnoj inteligenciji i strojnom učenju. Najprije se definiraju pojmovi umjetne inteligencije i strojnog učenja, a zatim se objašnjava proces strojnog učenja, vrste i modeli strojnog učenja.

2.1. Opći pojam umjetne inteligencije i strojnog učenja

Tehnološke inovacije su danas značajne u svim industrijama, jer omogućavaju održivost i pametnu proizvodnju. U tom smislu su posebno značajne inovacije umjetne inteligencije, nastale zahvaljujući istraživačkim naporima. Jedna od tehnika koja se temelji na umjetnoj inteligenciji je strojno učenje, koje je pokretačka snaga evolucije u pametnoj proizvodnji raznih sektora. Pojam umjetne inteligencije prvi je koristio 1950-ih godina Arthur Samuel, koji je napravio prvi samoučeći sustav za igranje dame, te je primjetio je da što je sustav više igrao, to je bio bolji (MonkeyLearn Inc, 2022). Kasnije je došlo do napretka u statistici i računalnoj znanosti, skupovima podataka, neuronskih mreža i općenito strojnog učenja. Strojno učenje je danas svugdje oko nas, primjerice kod automatiziranog prevođenja, prepoznavanja slika, glasovnog pretraživanja, navigacijskih sustava u automobilima te dr. Pravim tvorcem umjetne inteligencije naziva se Johna McCarthya, koji je 1990-ih godina definirao umjetnu inteligenciju kao znanost i inženjering stvaranja inteligentnih strojeva, posebno inteligentnih računalnih programa (Cioffi i sur., 2022). Općenito govoreći, izraz umjetna inteligencija se koristi kada stroj simulira funkcije koje ljudi povezuju s drugim ljudskim umovima, poput učenja i rješavanja problema.

Umjetna inteligencija se dijeli na dvije vrste (JavaTpoint, 2021) :

- Slaba umjetna inteligencija – naziva se još uskom ili umjetnom uskom inteligencijom (engl. *Artificial Narrow Intelligence* – ANI) koja predstavlja uvježbanu umjetnu inteligenciju fokusiranu na obavljanje posebnih zadataka. Takav oblik umjetne inteligencije se danas nalazi svugdje oko nas, jer pokreće

većinu robusnih aplikacija, poput glasovne Appleove Siri aplikacije, Amazonove Alexe, IBMovog Watsona i autonomnih vozila.

- Jaka umjetna inteligencija – obuhvaća opću umjetnu inteligenciju (engl. Artificial General Intelligence – AGI) kod koje bi umjetna inteligencija stroja bila izjednačena sa umjetnom inteligencijom čovjeka; stroj bi mogao samosvjesno rješavati probleme, učiti i planirati budućnost i umjetnu super inteligenciju (engl. Artificial Superintelligence – ASI) koja bi nadmašila inteligenciju i sposobnost ljudskog mozga, takav oblik inteligencije danas nema praktičnih primjera, međutim, teorijski se istražuje njen razvoj.

U 21. stoljeću je umjetna inteligencija veoma značajno područje istraživanja u raznim poljima. Neka od najznačajnijih polja su inženjerstvo, znanost, obrazovanje, medicina, računovodstvo, marketing, ekonomija, poslovanje te brojna druga polja.

Osim navedenog, umjetna inteligencija se klasificira u 16 područja koja se pojavljuju kao zasebna polja znanja, a čine ih (Ayoola, 2008) :

1. razmišljanje,
2. programiranje,
3. umjetni život,
4. revizija uvjerenja,
5. rudarenje podataka,
6. distribuirana umjetna inteligencija,
7. ekspertni sustavi,
8. genetski algoritmi,
9. sustavi,
10. reprezentacija znanja,
11. strojno učenje,
12. razumijevanje prirodnog jezika,
13. neuronske mreže,

14. dokazivanje teorema,
15. zadovoljenje ograničenja i
16. teorija računanja.

Prema navedenome, raspon umjetne inteligencije je širok. Polja koja će se posebno razmatrati u ovom radu, vezana za tematiku rada, su strojno učenje i rudarenje podataka (više u poglavlju 3.). Strojno učenje i rudarenje podataka su danas posebno značajni u poslovanju tehnoloških divova, kao sastavni dio kontinuirane evolucije umjetne inteligencije. Danas je razvoj strojnog učenja ubrzan, i raširen na brojna područja pametne proizvodnje, posebno u zdravstvu, odnosno medicini, farmakologiji, poljoprivredi, arheologiji, igrama, poslovanju (Cioffi i sur., 2022). Strojno učenje je danas puno više od komercijalne primjene metoda za izvlačenje informacija iz podatka, odnosno ono je neophodno za umjetnu inteligenciju, jer da bi se inteligentni sustavi mogli prilagoditi svojoj okolini oni moraju naučiti ponavljati svoje uspjehe, i svoje greške.

Inteligencija nastaje korištenjem jednostavnih algoritama, koji imaju sposobnost učenja na temelju velike količine podataka (Mate d.o.o, 2021, str. 10). Strojno učenje je najčešća metoda za modeliranje poslovnih procesa, pri čemu je najznačajnije simulacijsko modeliranje putem računala, odnosno simulacijskog softvera, te korištenjem kombinacije matematičkih modela za kvantitativnu analizu izvođenja vizualizacije i animacije procesa (Školska knjiga 2008, str. 24). Poslovni procesi se definiraju kao „povezani skupovi aktivnosti i odluka, koji se provode na vanjski poticaj kako bi se postigao neki mjerljivi cilj organizacije, traju određeno vrijeme i troše neke ulazne resurse, pretvarajući ih u specifične proizvode ili usluge od značaja za kupca ili korisnika” (Brumeca 2011, str. 3). Slijedom navedenog, poslovni procesi se modeliraju na temelju strojnog učenja (engl. Machine Learning – ML), koje se definira kao „grana umjetne inteligencije (AI) koja omogućuje računalima da "samo uče" iz podataka o obuci i poboljšavaju se tijekom vremena, bez eksplicitnog programiranja. Algoritmi strojnog učenja sposobni su otkriti uzorke u podacima i učiti iz njih, kako bi napravili vlastita predviđanja. Ukratko, algoritmi i modeli strojnog učenja uče kroz iskustvo” (MonkeyLearn Inc, 2022). Strojno učenje predstavlja proces koji se odvija se po koracima. Više o procesu strojnog učenja slijedi u nastavku.

2.2. Proces strojnog učenja

Strojno učenje je područje istraživanja koje je posvećeno razumijevanju i izgradnji metoda koje “uče”, odnosno metode koje iskorištavaju podatke za poboljšanje izvedbe na određenom nizu zadataka. Algoritmi strojnog učenje izgrađuju modele temeljene na uzorcima podataka.
U Tablici 1. prikazana je glavna razlika između strojnog učenja i umjetne inteligencije.

Tablica 1: Razlika između strojnog učenja i umjetne inteligencije

Umjetna inteligencija	Strojno učenje
Fokus je na povećanju uspjeha, a ne na točnosti.	Glavni fokus je postizanje maksimalne točnosti.
Cilj umjetne inteligencije je imitirati ljudsku inteligenciju koja će se koristiti za rješavanje složenih problema.	Primarni cilj strojnog učenja je biti obučen na temelju podataka o određenom zadatku, kako bi se maksimalno iskoristile performanse stroja pri provođenju zadatka.
Umjetna inteligencija vodi do inteligencije.	Strojno učenje vodi do znanja.
Napreduje kako bi se izgradio način oponašanja ljudi i sličnih ponašanja u određenim okolnostima.	Uključuje razvoj algoritama za samostalno učenje.

Izvor: (Academic Ebrary, 2022)

Strojno učenje omogućava inteligentnim sustavima da uče nove stvari iz podataka, iz kojih se crpe brojne informacije. Takvi podaci i informacije imaju obilježja velike razine točnosti, pa je strojno učenje značajno u pogledu uštede vremena i novca koji se troše za analize rješavanja problema, te na takav način predstavljaju podršku korisnicima (rast zadovoljstva) i podršku rudarenju podataka iz internih izvora koje Internet nudi.

Proces strojnog učenja, odnosno prenošenja inteligencije strojevima, se odvija u 7 glavnih koraka (Simplilearn, 2022) :

1. Prikupljanje podataka: strojevi uče iz podataka koji se unose, pa je veoma važno prikupiti točne i pouzdane podatke, kako bi model strojnog učenja mogao pronaći ispravne uzorke. O kvaliteti podataka kasnije će ovisiti točnost modela.

   U slučaju zastarjelih ili netočnih podataka nastaju pogrešni i ne relevantni rezultati.
2. Priprema podataka: nakon prikupljanja podataka iste je potrebno pripremiti na način da se spoje i da se nasumično rasporede, da se podaci čiste, odnosno da se uklanjaju nepotrebni podaci, te da se podijele u dva skupa: skup za obuku (iz njega model uči) i skup za testiranje (provjera točnosti korištenog modela nakon obuke).
3. Odabir modela: kod odabira modela je važno odabrati onaj model koji odgovara zadacima. Danas su razvijeni različiti modeli za različite zadatke. Npr. prepoznavanje govora i slika, predviđanje, prikladnost za numeričke i kategoričke podatke te dr.
4. Obuka modela: obuka je najvažniji korak strojnog učenja, jer se tijekom nje prosljeđuju pripremljeni podaci odabranim modelu strojnog učenja s ciljem pronalaska obrazaca i pravljenja predviđanja, a rezultat je učenje modela zbog izvršenja postavljenih zadataka. Uz obuku, s vremenom, model postaje sve bolji u predviđanju.
5. Ocjenjivanje modela: predstavlja korak provjere rada modela nakon obuke, što se izvodi testiranjem na prethodno nevidljivim podacima (testni skup podataka u koji smo ranije podijelili naše podatke).
6. Podešavanje parametara: korak koji se vrši nakon ocjene modela, pri kojem se sagledava da li se točnost modela može poboljšati, na bilo koji način, što se postiže podešavanjem parametara (varijabli korištenih u modelu, na odluku programera) prisutnih u korištenom modelu.

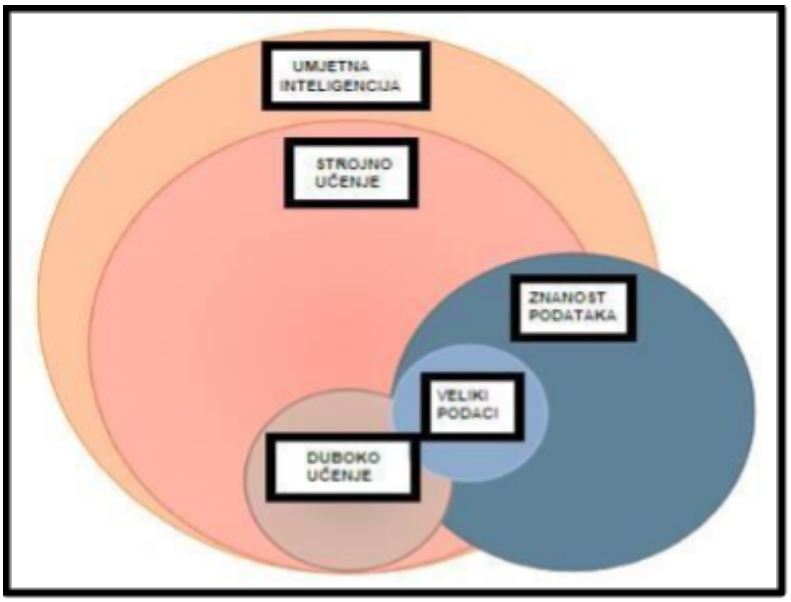
   Točnost je maksimalna na određenoj vrijednosti podešenog parametra. Stoga se podešavanje parametra odnosi na pronalaženje pravih vrijednosti

maksimalne točnosti modela.

7. Izrada predviđanja: zadnji korak procesa strojnog učenja kod kojeg se odabrani model upotrebljava na nevidljivim podacima, s ciljem izrade točnih predviđanja.

Proces strojnog učenja obuhvaćaju navedeni koraci, koji se ne bi mogli odvijati bez prva dva koraka, odnosno prikupljanja podataka i pripreme podataka. Slikovno se položaj strojnog učenja u prostoru umjetne inteligencije i povezanost sa drugim komponentama može prikazati putem Vennovog dijagrama, koji je vidljiv na Slici 1.

Slika 1: Položaj strojnog učenja u prostoru umjetne inteligencije i povezanost sa drugim komponentama – Vennov dijagram



Izvor: (Pandian, 2020)

Prema Slici 1. se vidi da je strojno učenje u prostoru umjetne inteligencije utkano na temelju znanosti podataka, odnosno na temelju unosa velikih količina podataka, koji rezultiraju dubokim učenjem. Prije nego se prijeđe na gore navedene korake procesa strojnog učenja, potrebno je identificirati poslovne probleme i zadatke koji se žele ispuniti, a zatim napraviti implementaciju strojnog učenja. Prikupljanjem

podataka moguće je realizirati ostale korake strojnog učenja, koji će u konačnici dovesti do razvoja i implementacije modela temeljenog na strojnom učenju, finaliziranog za proizvodno okruženje i postizanje rezultata za donošenje poslovnih odluka. U nastavku slijedi više o vrstama i modelima strojnog učenja.

2.3. Vrste i modeli strojnog učenja

Strojno učenje obilježava korištenje algoritama koji uče informacije iz podataka, na temelju matematičkih jednadžbi i modela. „Zadatak algoritma strojnog učenja je pronaći prirodne obrasce i poveznice u podacima te na temelju toga steći uvid te odlučiti i predvidjeti. Već se svakodnevno koriste za donošenje važnih odluka u medicinskoj dijagnostici, trgovanju i mešetarenju dionicama, predviđanju potrošnje energije itd." (Bolf, 2021). U odnosu na vrste zadataka algoritama strojnog učenja nastali su različiti modeli strojnog učenja.

Prema Sl.education (2022), modeli strojnog učenja mogu biti:

„modeli klasifikacije,
regresijski modeli,
grupiranje,
smanjenje dimenzija,
duboko učenje (kao podskup strojnog učenja temeljenog na neuronskim mrežama)".

Strojno učenje se dijele u dvije osnovne vrste, a to su nadzirano i nenadzirano učenje, te učenje s pojačivačem kao dio nenadziranog učenja. Na Slici 2. prikazane su glavne vrste i njima pripadajući modeli strojnog učenja.

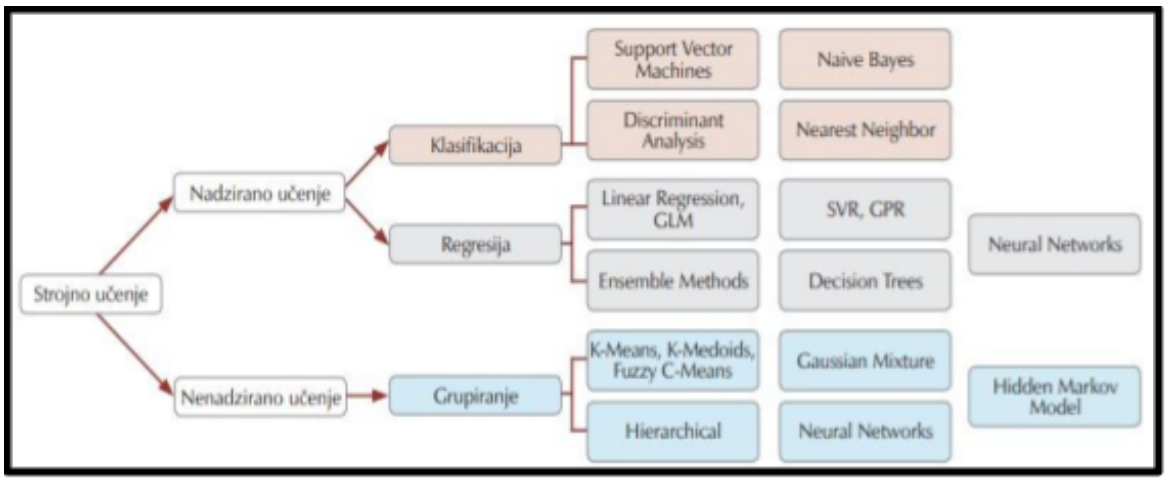
Prema Slici 2. u nadzorno učenje spadaju:

modeli klasifikacije:
model podrška vektorskim strojevima (engl. Support Vector Machine), diskriminantna analiza (engl. Discriminant Analysis), model naivni Bayes (engl.

Naive Bayes), i model Najbliži susjed (engl. Nearest Neighbor), te
modeli regresije:
linearna regresija (engl. Linear Regression), metoda potpornih vektora za regresiju SVR) i Gausov regresijski proces (GPR), metode anasambla (engl. Ensemble Methods), stablo odlučivanja (engl. Decision Trees), te metoda neuronskih mreža (engl. Neural Networks).

Nenadzirano učenje čine modeli grupiranja: model (engl. K – Means), model K- medoids, model (engl. Fuzzy C- Means), Gaussova smjesa (engl. Gaussian Mixture), hijerarhijski model (eng. Hierarchical), model neuronskih mreža, te skriveni Markovljev model (engl. Hidden Markov Model).

Slika 2: Vrste i modeli strojnog učenja



Izvor: (Bolf, 2021)

Navedene vrste strojnog učenja objašnjavaju se na sljedeći način (myservername, 2022) :

1. Nadzirano strojno učenje (nadgledano): Takva vrsta strojnog učenja se vrši pod nadzorom osobe koja nadgleda učenje i provjerava svaku radnju i nastale rezultate. Kod navedenog strojnog učenja algoritam izlaza podataka je poznat te se naziva označenim skupom podataka, a ulaz se napaja sa puno podataka za obuku, a rezultat je brzo učenje s velikom točnošću. Problemi nadgledanja su klasifikacija, odgovori se klasificiraju u razrede, primjerice „da“ i „ne“, pa se takva klasifikacija naziva binarnom, a ako postoji više klasa tada nastupa

klasifikacija sa više klasa i regresija gdje se problemi predviđaju kao vrijednosti koje su kontinuirane, koje se kreću do beskonačnosti.

2. Nenadzirano strojno učenje (nenadgledano):
   učenje koje se odvija na neovisan način, bez nadzora. Izlaz podataka nije mapiran sa ulazom podataka (neobilježeni skup podataka), pa sustav sam uči od unosa podataka, jer su vrijednosti nepoznate. Takav način učenja koristi tehnike za miniranje pravila podataka, grupa i obrazaca podataka sa drugim sličnim vrstama. Treninzi učenja omogućavaju da se ulazni podaci klasteriziraju i udružuju, pa ako ulazni podatak nije prepoznat tada će se isti generirati u novu klasu. Tu spadaju apriorni i K – Means algoritmi poput modela klasteriranja, udruživanja pravila rudarstva, automatski koderi. Takvi modeli sami prilagođavaju svoje parametre, odnosno vrše samoorganizaciju i poboljšanja na način da otkrivaju sličnosti (npr. oblici, cijene, boje, veličine, te dr.) među ulaznim podacima.

Zbog boljeg razumijevanja nadzirnog i nenadzirnog strojnog učenja, u Tablici 2. je prikazana njihova međusobna razlika. Ono što je najvažnije je to da je kod nadzirnog učenja izlaz za zadati ulaz poznat, dok kod nenadzirnog strojnog učenja nije poznat.

Tablica 2.: Razlika između nadziranog i nenadziranog strojnog učenja

Nadzirano strojno učenje	Nenadzirano strojno učenje
U nadzirnim algoritmima učenja izlaz za zadati ulaz je poznat.	U algoritmima nenadglednog učenja izlaz za zadati ulaz je nepoznat.
Algoritmi uče iz označenog skupa podataka. Ovi podaci pomažu u procjeni točnosti podatka o treningu.	Algoritam ima neoznačene podatke gdje pokušava pronaći uzorke i asocijacije između stavki podataka.
To je tehnika prediktivnog modeliranja koja precizno predviđa buduće ishode.	To je tehnika opisnog modeliranja koja objašnjava stvarni odnos između elemenata i povijesti elemenata.
Uključuje algoritme kasifikacije i regresije.	Uključuje algoritme učenja pravila klasterizacije i pridruživanja.
Ova vrsta učenja je relativno složena jer zahtijeva označene podatke.	Manje je složen jer nema potrebe za razumijevanjem i označavanjem podataka.
To je mrežni postupak analize podataka i ne zahtijeva ljudsku interakciju.	Ovo je analiza podataka u stvarnom vremenu.

Izvor: (myservername, 2022)

3. Učenje s pojačanjem – vrsta nenadzirnog strojnog učenja kod kojeg algoritam uči na temelju mehanizma prošlih iskustava i povratnih informacija, a cilj je postići određeni zadatak poduzimanjem novih, sljedećih koraka, da bi se dobio najbolji ishod. Takav način učenja dovodi do brojnih pokušaja i pogrešaka, pa se takvo strojno učenje osnovnog pojačanja naziva Markovljevim procesom odlučivanja ili Markovljevim modelom. Više pokušaja donosi i više povratnih informacija, pa na takav način sustav postaje sve točniji, podaci se učvršćuju. Za primjer se može navesti video igre, kod kojih se učvršćuje učenje kroz završavanje određene razine igrice, nakon koje igrač ima mogućnost ponoviti njenu izvedbu, bolje savladati pogreške i dobiti više bodova. Također, dobar primjer je i treniranje robota, automatsko upravljanje zalihama te drugo.

3. RUDARENJE PODATAKA

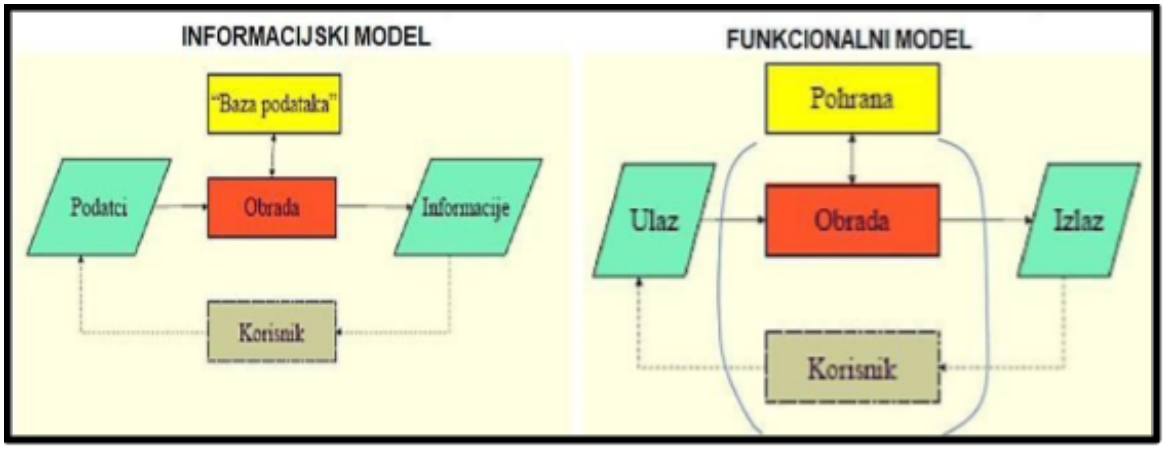
Ovo poglavlje se bavi rudarenjem podataka. Najprije se definiraju pojmovi podataka, informacija i znanja, a zatim se objašnjava rudarenje podataka, metode i modeli rudarenja podataka, odnos umjetne inteligencije i rudarenja podataka, te aplikacije i tehnike rudarenja podataka.

3.1. Podatak, informacija i znanje

Pojmovi podatak, informacija i znanje su međusobno povezani, posebno u području računalstva, odnosno u području umjetne inteligencije. Informacije uključuju pojmove podataka i znanja. Prikupljanjem podataka nastaju informacije, a iz informacija nastaje znanje (Sanders, 2016, str. 1). Nakon Drugog svjetskog rata, u ranim danima razvoja računalstva informacije su bile mjerilo onoga što se može prenijeti preko odašiljača (pošiljatelja) do primatelja. Pohranjivanje takvih informacija činilo je sustav podataka. Bez podataka nema znanja, stoga se „podaci pamte,

zapisuju i bilježe onako kako njima odgovara. Dakle, oblici podataka mogu biti: zvučni, slikovni, brojčani ili tekstualni. Struktura podatka je apstraktna i sastoji se od: značenja (naziv i opis značenja određenog svojstva), vrijednost (mjera i količina) i vremena. Podaci u kontekstu (značenju) i kombinirani unutar strukture čine informaciju" (Phil Papers Org, 2022). Da bi podaci postali informacije, isti se trebaju obrađivati, a da bi informacije postale znanje iste se trebaju interpretirati na način da imaju svoje značenje. Danas računala olakšavaju put do informacija, kojima se bavi nauka pod nazivom informatika. Informacije su informatici obrađuju putem automatskih mašina, koje na takav način postaju nositelji znanja. U tom kontekstu dolazi do interakcije čovjeka i računala. Kod transformacije podataka u informacije koriste se informacijski i funkcionalni model, koji su prikazani na Slici 3.

Slika 3.: Transformacija podataka u informacije: informacijski i funkcionalni model



Izvor: (Mesarić i Šebelj, 2014/2015)

Prema Slici 3. se vidi da se kod informacijskog modela podaci obrađuju, te putem softverskih rješenja čine baze podataka, na takav način nastaju informacije koje upotrebljavaju korisnici. Skupovi informacija se generiraju u ovisnosti o programima koji su instalirani. Kod funkcionalnog modela ulazni skup podataka ide na obradu, čije izvršenje ovisi o pohranjenom modelu. O načinu pohrane ovisi i sposobnostima korisnika ovisi izlazni skup transformiranih podataka. Sposobnost iskorištavanja informacija transformiranih u znanje (primjena podataka i informacija)

čini inteligenciju ili mudrost koja vrednuje razumijevanje znanja. Umjetna inteligencija se koristi u rudarenju podataka, za analizu vanjskih skupova podataka i za otkrivanje korisnih informacija. Više o rudarenju podataka slijedi u nastavku.

3.2. Općenito o rudarenju podataka

Rudarenje podataka (engl. Data Mining) se prema IBM-u definira kao „proces otkrivanja obrazaca i drugih vrijednih informacija iz velikih skupova podataka, odnosno kao proces otkrivanja znanja u podacima" (IBM, 2021). Zbog evolucije tehnologije skladištenja podataka i rasta velikih podataka došlo je od potrebe usvajanja tehnika za rudarenje podatka, koje su pomogle poduzećima transformirati neobrađene podatke u korisno znanje. Rudarenje podataka je pomoglo poduzećima u donošenju odluka na temelju detaljne analize podataka. Proces rudarenja podataka se sastoji od više koraka, koji se kreću od prikupljanja podataka do vizualizacije, a cilj je izdvojiti informacije iz velikih skupova podataka, koji se opisuju kroz opažanje uzorka, asocijacija i korelacija, te se grupiraju i klasificiraju putem metoda klasifikacije i regresije.

Osnovni koraci rudarenja podataka, kao tehnike baza podataka, su (IBM, 2021):

1. Postavljanje poslovnih ciljeva – jedan od najvažnijih koraka svakog poduzeća, jer se putem njega dolazi do definiranja poslovnih problema, te do informiranja po pitanju podataka potrebnih za neki projekt.
2. Priprema podataka – kada su problemi definirani tada se lakše identificira koji će skup podataka pomoći odgovoriti na relevantna pitanja za poslovanje. Nakon prikupljanja podataka, isti se čiste, odnosno uklanjaju se nepotrebni podaci, te se po potrebi smanjuje broj značajki koje bi mogle usporiti izračunavanja. Uvijek se nastoje zadržati oni podaci koji su najvažniji za izračunavanje optimalne točnosti unutar bilo kojeg modela.
3. Izgradnja modela i rudarenje uzoraka podataka – korak koji u ovisnosti od vrste analize podataka istražuje odnose podataka, poput sekvencijalnih obrazaca, pravila povezivanja ili korelacije. Algoritmi dubokog učenja se mogu primijeniti za klasificiranje ili grupiranje skupa podataka, ovisno o dostupnim

podacima.

Ako su ulazni podaci označeni (tj. nadzirano učenje), može se upotrijebiti klasifikacijski model za kategorizaciju podataka ili se alternativno može primijeniti regresija za predviđanje vjerojatnosti. Kod učenja bez nadzora neke se podatkovne točke u skupu za obuku uspoređuju jedna s drugom, s ciljem otkrivanja glavnih sličnosti, na temelju kojih nastaju grupacije podataka.

4. Evaluacija rezultata i primjena znanja – korak kod kojeg se rezultati interpretiraju, pa se utvrđuje valjanost, korisnost i razumljivost podataka. Nakon ispunjavanja ovog kriterija nastaje znanje, koje poduzeća koriste za provedbu novih strategija, te za postizanje svojih namjeravanih ciljeva.

Rudarenje podataka se razlikuje od tradicionalnih tehnika baza podataka ili statističkih metoda po tome što se može koristiti za otkrivanje novih obrazaca ili za potvrdu sumnjivih odnosa korištenjem dva pristupa (Custers 2013, str. 9) :

pristup "odozdo prema gore" ili "pokrenutim podacima" – započinje s podacima, a zatim se grade teorije temeljene na otkrivenim obrascima,

pristup "odozgo prema dolje" ili "teorijom vođenim" pristupom – započinje s hipotezom, a zatim se podaci provjeravaju kako bi se utvrdilo jesu li u skladu s hipotezom.

Obrazac kod rudarenja podataka treba biti istinit, i siguran, točan. Sigurnost obrasca može uključivati čimbenike poput cjelovitosti podataka i veličine uzorka. Rudarenje podataka je nastalo spajanjem strojnog učenja i statistike, pa danas zajednicom rudarenja podataka dominiraju računalni znanstvenici i statističari, koji su ujedinjeni 2001. godine preko Europske konferencije o strojnom učenju (ECML) i Europske konferencije o načelima i praksi otkrivanja znanja u bazama podataka (PKDD). Rudarenje podataka je usmjereno na podatke u svim formatima pa se kao takvo promatra kao aplikacijska domena, a strojno učenje koristi mehanizme na temelju kojih računala mogu učiti, pa se promatra kao tehnologija (Coenen, 2011, str.2). Rudarenje se kao aplikacijska domena pojavljuje kao učinkovit skup tehnika usmjerenih na rudarenje tabularnih podataka.

Neke od aplikacijskih domena rudarenja podataka su (Ibidem, str. 4-5):

Rudarenje teksta (engl. Text Mining) – prije tradicionalnog tabularnog rudarenja podataka koristilo se rudarenje teksta, kod kojeg se primjena vršila izgradnjom klasifikatora za kategoriziranje ili grupiranje velikih zbirki dokumenata (npr. članci s vijestima, web stranice). Osim navedenog primjena se vršila iz slobodnog teksta, u obliku upitnika, te se na takav način dobivalo korisne informacije, te primjena sažimanjem teksta (skraćivale su se informacije, isticale su se samo one koje su bile važnije). Rudarenje je dobar način predstavljanja tekstualnih podataka, primjenom tehnika za rudarenje podataka ili tehnika obrade prirodnog jezika (NLP).

Rudiranje slika (engl. Image Mining) – postoje brojne zbirke digitalnih slika koje su generirane s obzirom na aplikacije. Rudarenje slika se bavi reprezentacijom slike (2D i 3D) primjenom tehnika rudarenja (npr. stvaranje histograma ili stabala/grafova - jedan po slici). Slike se mogu predstaviti i korištenjem tehnika segmentacije, koje imaju ograničen uspjeh, ovisno o prirodi slika, te su predmet istraživanja unutar zajednice za analizu slika. U područjima poput rudarenja medicinskih slika problem se može obuhvatiti na specifičan način, pa je rudarenje medicinskih slika je postiglo određene uspjehe. Za primjer se može navesti klasifikacija podataka slike mrežnice i podataka skeniranja magnetskom rezonancijom (MRI), za prepoznavanje poremećaja.

Rudarenje grafova i stabla (engl. Graph Mining) – predstavlja proširenje učestalog rudarenja uzoraka, pri čemu se analiziraju podgrafovi. Stručnjaci za rudarenje grafova smatraju da se sve može prikazati kao grafikon (npr. entiteti dokumenata, e-pošte i slika). Rudarenje u stablu je jednostavnije jer se mogu iskoristiti inherentna svojstva stabla (bez ciklusa). Rudarenje grafova i stabala zahtijeva neki kanonski oblik s kojim se prikazuju grafovi. Danas je glavni problem rudarenja grafova generiranje podgrafa kandidata i testiranje izomorfizma podgrafa. Najvažniji algoritam za rudarenje podgrafa je gSpan, dok je za suvremeno proširenje rudarenja grafova značajno rudarenje društvenih mreža.

3.3. Tehnike, metode i modeli rudarenja podataka

Učinkovite analize velikih količina podataka u poduzećima su omogućene na temelju korištenja tehnika i modela za rudarenje podataka. Tehnike rudarenja podataka mogu opisati ciljni skup podataka, te mogu predvidjeti ishode korištenjem algoritma strojnog učenja (IBM, 2021). Koriste zbog organiziranja i filtriranja podataka, otkrivanja informacija, prijevara, sigurnosnih proboja, kao i ponašanja korisnika. Kombinacija analitike podataka i alata za vizualizaciju omogućava zadubljivanje u svijet rudarenja podataka, pri čemu su relevantni uvidi postali brži. Takav napredak unutar umjetne inteligencije se danas usvaja velikom brzinom, u svim industrijama.

Tehnike i njima pripadajuće metode i modeli rudarenja podataka su slijedeće (Bharati, 2010) :

Klasifikacija podataka - najčešće primjenjivana tehnika rudarenja podataka, koja koristi model stablo odlučivanja ili klasifikacijske algoritme, koji se temelje na neuronskim mrežama. Proces klasifikacije podataka uključuje učenje (podaci se analiziraju algoritmom klasifikacije) i klasifikaciju. Klasificirani podaci se koriste za procjenu točnosti pravila klasifikacije, pa ako je točnost prihvatljiva, pravila se mogu primijeniti na nove skupove podataka. Neki od klasifikacijskih modela su klasifikacija indukcijom stabla odlučivanja, Bayesova klasifikacija, neuronske mreže, podrška vektorskim strojevima (SVM) i klasifikacija na temelju asocijacija.

Tehnika grupiranja ili klasteriranje – identificira se ukupni obrazac distribucije i korelacije među atributima podataka. Grupiranje se koristi kao pristup pretprocesiranju za odabir i klasifikaciju podskupa atributa, što bi značilo da je potrebno kategorizirati gene sa sličnom funkcionalnošću, ako se, primjerice, želi formirati grupu kupaca na temelju obrazaca kupnje. Neke od metoda klasteriranja su metode particioniranja, hijerarhijske aglomerativne (divizijske) metode, metode temeljene na gustoći, metode temeljene na mreži, te dr.

Predikcija ili predviđanje (prediktivna analiza) - tehnika regresije može se prilagoditi za predikciju. Regresijska analiza se može koristiti za modeliranje odnosa između zavisnih i nezavisnih varijabli (jedne ili više njih). Kod

rudarenja podataka nezavisne varijable su već poznati atributi, a varijable odgovora su ono što želimo predvidjeti. Nedostatak je to što se u stvarnom vremenu puno problema ne može predvidjeti (cijene, stope kvarova, količina prodaje, te dr.), pa se za predviđanje budućih vrijednosti koriste složenije tehnike poput logističke regresije, stablo odlučivanja ili neuronske mreže. Neke od regresijskih metoda rudarenja podataka su linearna regresija, nelinearna regresija, multivarijatna linearna regresija, i multivarijatna nelinearna regresija.

Pravilo pridruživanja, povezivanja, korelacije – odnosi se na pronalaženje čestih skupova stavki među velikim skupovima podataka, čime se pomaže u donošenju odluka (npr. kod unakrsnog marketinga i analize ponašanja kupaca pri kupnji), a neka od pravila su višerazinsko pravilo pridruživanja, pravilo višedimenzionalnog povezivanja, pravilo kvantitativne povezanosti.

Neuronske mreže – predstavljaju skup povezanih ulazno/izlaznih jedinica, te imaju izvanrednu sposobnost izvlačenja značenja iz kompliciranih ili nepreciznih podataka, mogu se koristiti za izdvajanje obrazaca i trendova u podacima, te su vrlo su prikladne za predviđanje.

3.4. Umjetna inteligencija i rudarenje podataka

Umjetna inteligencija se odnosi na stvaranje inteligentnih strojeva koji mogu raditi kao ljudi, na temelju izravno programiranih upravljačkih sustava koji izračunima dolaze do rješenja brojnih problema, a rudarenje podataka je tehnika koju koriste sustavi umjetne inteligencije za stvaranje rješenja. (Softwaretestinghelp, 2022) Prema navedenom rudarenje podatka je tehnika koja je temelj umjetne inteligencije, u obliku programskih kodova koji sadrže informacije i podatke potrebne za sustave umjetne inteligencije.

Umjetna inteligencija ne može sama funkcionirati bez programa, koje omogućava rudarenje podataka, pa u navedenom leži povezanost umjetne inteligencije i rudarenja podataka (Menaga i Saravanan, 2021, str 133-154). Umjetna inteligencija se danas najčešće koristi kod chatbotova (zbog razumijevanja govora i teksta na

prirodnom jeziku, sustavi umjetne inteligencije komuniciraju s ljudima na prirodan, personaliziran način), samovozeći automobili, roboti koji se koriste u proizvodnji, filter neželjene e-pošte, te dr. Između umjetne inteligencije i rudarenja podataka postoje razlike, koje su prikazane u Tablici 3. Prema prikazanoj Tablici 1. se može zaključiti da je umjetna inteligencija put stvaranja inteligentnih strojeva, koji mogu raditi poput ljudi. Funkcioniranje umjetne inteligencije se temelji na korištenju tehnike rudarenja podataka. Važnost povezanosti umjetne inteligencije i rudarenja podataka u suvremenom svijetu ima veliki značaj, u suvremenom svijetu su neizbježni, te će biti tehnologije budućnosti, koje će napredovati.

Tablica 3.: Razlika između umjetne inteligencije i rudarenja podataka

Pojmovi	Umjetna inteligencija	Rudarenje podataka
Koncept	Umjetna inteligencija ima softver koji može razmišljati o unosu i objasniti izlaz podataka. Omogućava interakciju čovjeka sa softverom i nudi podršku pri odlučivanju za određene zadatke, ali nije zamjena za ljude.	Pronalazi uvide i omogućava buduća predviđanja.
Važnost	Mogućnost rada s velikim skupovima podataka, veća brzina, uvođenje inovacije, dizajniranje i razvijanje proizvoda i usluga.	Pomaže u otkrivanju kako su različiti atributi skupova podataka povezani kroz obrasce i tehnike vizualizacije podataka.
Metoda rada	Umjetna inteligencija funkcionira na način da integrira velike količine podataka s brzom, iterativnom obradom i inteligentnim algoritmima.	Kopa duboko u podatke i iz njih izvlači korisne informacije.
Korištenje	Proizvodnja robota Samovozeći automobili Pametni roboti koji pružaju pomoć u obavljanju raznih zadataka Proaktivno upravljanje zdravstvenom zaštitom Mapiranje bolesti Automatizirano financijsko ulaganje Virtualni agent za rezervacije putovanja Praćenje društvenih medija Alati za razgovor među timovima Konverzacijski marketinški bot Alati za obradu prirodnog jezika (NLP)	Web rudarenje Rudarenje teksta Otkrivanje prijevara
Ljudska intervencija	Strojevi temeljeni na umjetnoj inteligenciji su brzi, precizni i logički, ali im nedostaju emocije i nisu kulturno osjetljivi.	Potrebna je manualna tehnika.
Alati	Scikit Learn TensorFlow Theano Caffe MxNet Keras PyTorch CNT	Rapid Miner Oracle Data Mining IBM SPSS Modeler Knime Python Orange Kaggle Rattle
Aplikacije	Umjetna opća inteligencija	Zdravstvo budućnosti

	Planiranje Računalni uvid Opće igranje igre Rasuđivanje znanja Strojno učenje Obrada prirodnog jezika Robotika	Analiza tržišne košarice Proizvodno inženjerstvo CRM Otkrivanje prijevare Otkrivanje upada Segmentacija kupaca Financijsko bankarstvo

Izvor: (JavaTpoint, 2021)

Umjetna inteligencija i rudarenje podataka će posebno biti značajni u području automatizacije, planiranja, povećanja prodaje, te će utjecati na rast profita i poslovanja poduzeća. Osim u prodajnom području posebno će biti značajni u proizvodnji, zdravstvu, financijskim djelatnostima, te dr. Više o aplikacijama rudarenja podataka slijedi u nastavku.

3.5. Aplikacije rudarenja podataka

Rudarenje podataka je relativno nova tehnologija koja se danas koristi u brojnim industrijama. Kao takvo, rudarenje podataka se kombinira često sa statističkim alatima i alatima za prepoznavanje uzoraka (Bharati, 2010. str. 304). Istraživačka zajednica rudarenja podataka proizašla je iz drugih povezanih područja poput strojnog učenja, umjetne inteligencije, vizualizacije, statistike i analitike.

Danas se rudarenje podataka u raznim područjima aplicira kao automatizirano izdvajanje uzoraka koji predstavljaju znanje pohranjeno u velikim bazama podataka, skladištima podataka, web-u te drugim spremištima informacija ili tokova podataka (Custers, 2013, str. 28). S rudarenjem podataka je najviše povezano strojno učenje koje se bavi načinima izvršenja zadataka, dok se rudarenje podataka bavi pronalaženjem znanja iz podataka. Povezanost leži u znanju pohranjenom u sustavu koji je usmjeren na zadatke. Strojno učenje je usmjereno na nadzirane zadatke, dok je rudarenje podataka više fokusirano na nenadzirane zadatke. Rudarenje podataka se aplicira u tehničke, komercijalne i istraživačke svrhe, najčešće u sljedećim područjima (Neha, 2020, str. 3386):

Bioinformatika – predstavlja skup različitih metoda za upravljanje, pohranu i proučavanje bioloških podataka pomoću računala, pa se podaci u ovom

području svakodnevno povećavaju i koriste u istraživačke svrhe. Rudarenje podataka se u bioinformatici koristi za pronalaženje sekvenci gena, analizu sekvenci proteina, konstrukciju komunikacijske mreže gena i proteina, otkrivanje bolesti, sekvenciranje i poravnavanje DNK, te dr. Takvi skupovi podataka se, ovisno o vrsti aplikacije, daju odgovarajućem alatu za rudarenje podataka, zbog dobivanja potrebnih rezultata. Neki od alata za rudarenje podataka u bioinformatici su BLAST (engl. Basic Local Alignment Search Tool - alat za osnovno pretraživanje lokalnog poravnanja), FASTA, CS BLAST za pronalaženje poravnanja sekvenci, GenScan, GeneMark za pronalaženje gena, Pfam, BLOCKS, ProDom za analizu proteina te drugi alati.

Financijsko bankarstvo – danas je bankarstvo digitalizirano pa svakodnevno generira velike količine podataka o transakcijama, a rudarenje podataka se koristi za pronalazak lojalnih kupaca, izdavanje zajmova i kreditnih kartica, na temelju prethodno prikupljenih podataka o klijentima, za identifikaciju rizika na burzi na temelju povijesnih podataka itd. Takve bankarske aplikacije koriste algoritme klasifikacije poput Bayesove klasifikacije,stabla odlučivanja, slučajne šume i dr. Neki od alata koji se koriste za rudarenje podataka u financijama su Rapid Miner, R programming, Weka (Waikato Environment for Knowledge Analysis), Orange, KNIME, NLTK (Natural Language Tool Kit), te drugi.

Obrazovanje – rudarenje podataka u obrazovanju je novi sektor koji je fokusiran na razvoj metoda koje otkrivaju potrebne informacije iz različitih obrazovnih područja; primjerice predviđaju rezultate učenika, ponašanje učenika pri učenju, pronalaze učenike koji imaju slabije rezultate, i sl., a obrasci učenja učenika ili studenata se koriste za razvoj nastavnih metoda. Aplikacija skupa rudarenja podataka u obrazovanju naziva se Education, a neki od alata za rudarenje podataka su SPSS, KEEL, Weka, Spark MLLib i dr.

Kriminalistička analiza – koristi se za prikupljanje kriminalnih podataka za otkrivanje zločina, odnosno povezanosti kriminalaca sa zločinima (npr. cyber zločini, nasilni zločini, otkrivanje prijevara, prekršaji s drogom), a rudarenje podataka je u tom području važno za aplikacije poput terorističkih aktivnosti, usklađivanja zločina, trendova kriminala i sl. Neki od alata za rudarenje podataka u obrazovanju su Weka, H2O, Orange, i drugi alati.

Analiza tržišne košarice – koristi se za predviđanje ponašanja kupaca u maloprodaji. Kod analize tržišne košarice se primjenjuje tehnika rudarenja

pravila pridruživanja, koja pomaže u povećanju prodaje, rasporedu robe na policama u skladu sa ponašanjem potrošača te dr. Neki od alata za rudarenje podataka u tom području su R programming, SAS (Statistical Analysis System), MEXL, XLMINER i drugi. Većina podataka koji se odnose na poslovne transakcije u maloprodaji pohranjena je u skladištima podataka. Ako analitičari takve podatke obrade i pokrenu na pravi način, tada na temelju istih mogu dobiti uvide u odnose sa kupcima, transakcijama, kao i uvide u reguliranje kupnje svojih kupaca (Maksood i Achuthan, 2016, str. 7). Današnji digitalizirani svijet omogućava trgovcima bilježenje svake aktivnosti, svakog podatka, kao i manipulaciju tim podacima, na temelju čega poduzeća ostvaruju uspjeh i svoj rast.

Buduća zdravstvena njega – u zdravstvenim kartonima se nalaze podaci o velikom broju pacijenata, pa se u zdravstvu koriste tehnike rudarenja podataka poput klasifikacije, pravila pridruživanja, i grupiranja, s ciljem otkrivanja odnosa među bolestima, tretmanima, zbog identificiranja novih lijekova, otkrivanja prijevara, smanjenja troškova, te dr. Neki od alata koji se koriste u zdravstvu su Rapid miner, R programming, Weka, Orange, NLTK (Natural Language Tool Kit) te dr.

Proizvodno inženjerstvo – proizvodna poduzeća prikupljaju podatke o proizvodima, pa se tu koriste tehnike rudarenja poput klasifikacija, rudarenje pravila pridruživanja, te regresija, zbog predviđanja vremena i troškova razvoja proizvoda, potreba kupaca, ovisnosti među zadacima itd. Alati rudarenja podataka u proizvodnom inženjerstvu su Rapid miner, Data Melt, Board, i Weka.

Web rudarenje - koristi metode rudarenja podataka s ciljem otkrivanja relevantnih web dokumenata i obrazaca s web stranica. Tehnike rudarenja podataka u tom području su tehnike klasifikacije, grupiranja, i regresije koje se koriste u aplikacijama poput rudarenja web sadržaja zbog izvlačenja korisnih informacija koje nude web dokumenti, rudarenje strukture weba (za otkrivanje informacija o strukturi s web mjesta), te dr. Najznačajniji alati za web rudarenje podataka su SAS (Statistical Analysis System), Scrapy, Page Rank i drugi alati.

Slijedom navedenog se zaključuje da se rudarenje podataka aplicira na brojna

područja ljudskog djelovanja, te da olakšava poslovanje i donosi brojne prednosti, na temelju uvida u pohranjene podatke. Fokus ovog rada je na apliciranju rudarenja podataka, kao jednog od područja umjetne inteligencije, na zdravstvo. U sljedećem poglavlju se prikazuju primjeri umjetne inteligencije za rudarenje podataka u zdravstvu, sa primjerima iz prakse.

4. PRIMJERI PRIMJENE UMJETNE INTELIGENCIJE ZA RUDARENJE PODATAKA U ZDRAVSTVU

U ovom poglavlju će biti prikazani primjeri primjene umjetne inteligencije za rudarenje podataka u zdravstvu: otkrivanje lijekova i molekularno modeliranje umjetnom inteligencijom, te dijagnostika raka i donošenje odluka o liječenju pomoću umjetne inteligencije. Također, objašnjavaju se prednosti i nedostaci korištenja umjetne inteligencije za rudarenje podataka u zdravstvu, na odabranim primjerima. Općenito, rudarenje podataka ima veliku ulogu u medicini i zdravstvu.

Algoritam predviđanja je glavni pristup na koji se usredotočuju stručnjaci medicinske informatike, putem kojeg istraživači prikupljaju podatke o pacijentima te na konzultacijama pronalaze rješenja koja su od najboljeg interesa za pacijentovo zdravlje. Rudarenje podataka u zdravstvu je važno zbog predviđanja učinkovitosti određenih kirurških postupaka, medicinskih testova i lijekova, čime se pomaže u podizanju standarda kliničkog odlučivanja, se se na takav način pridonosi zdravlju i sigurnosti ljudi (Maksood i Achuthan, 2016, str. 7). Podaci u zdravstvu nazivaju se zdravstvenim podacima, koji se pojavljuju kao elektronički zdravstveni zapisi a omogućavaju pružanje informacija o bolesnicima (pacijentima) te praćenje bolničkih aktivnosti upravljanja zdravljem, a dijele na sljedeće vrste (Xiao i Sun, 2021, str. 9-22): strukturirane zdravstvene podatke i nestrukturirane podatke. Strukturirane izvore podataka čine: dijagnostički kodovi, kodovi postupaka, laboratorijski rezultati, bolnički recepti za lijekove (ATC): skup propisanih lijekova u obliku ATC kodova, kodovi onkološke patologije (CODAP) - u obliku CODAP kodova koji predstavljaju sustav kodova za opisivanje abnormalnog rasta tkiva, analiziranog nakon biopsije, medicinske specijalnosti vezane za boravak pacijenata, demografski podaci, šifre postupaka koje opisuju medicinski postupak ili intervenciju (npr. MRI). Nestrukturirani

izvori podataka su primjerice kliničke bilješke, medicinske slike, izvješća o operaciji, otpusna pisma i pisma za usmjeravanje pacijenta specijalistu, protokoli (tekstualni prikazi rezultata određenih postupaka; npr. tekstualna interpretacija snimanja magnetskom rezonancijom (MRI)), potvrde, zahtjevi te dr. (Scheurwegs, 2016).

Rudarenje podataka se pokazalo posebno značajnim kod otkrivanja lijekova i molekularnog modeliranja umjetnom inteligencijom, te kod dijagnosticiranja raka, o čemu će više biti u sljedećim poglavljima.

4.1. Primjer 1 / Dijagnostika raka i donošenje odluka o liječenju pomoću umjetne inteligencije

U zdravstvu se umjetna inteligencija, odnosno sustavi strojnog učenja, pojavljuju u analizi raznih bolesti, a jedna od njih i dijagnosticiranje raka i donošenje odluka o liječenju istog. Koriste se ekspertni sustavi AIBDS (dijagnostički sustavi temeljeni na umjetnoj inteligenciji), koji se sastoje od „velikih skupova podataka koji u suvremenoj medicini neprestano rastu i koji su sve dostupniji. AIBDS ima pristup i mogućnost konzultacije stotina ili tisuća medicinskih knjiga i znanstvenih radova, kao i drugih dijagnostički korisnih izvora, poput zbirki radioloških snimki, snimki magnetske rezonancije, uzoraka krvi i sl. Nijedan ljudski liječnik neće raspolagati tolikim znanjem i podacima, a pogotovo neće biti u stanju raspolagati s njima gotovo trenutno, odnosno jednakom brzinom i pouzdanošću postavljati dijagnoze. Ako se AIBDS temelji na rezultatima različitih područja umjetne inteligencije, poput računalnog vida (omogućujući, na primjer, usporedbu rendgenskih slika) i obrade prirodnog jezika (radi, primjerice, izravne komunikacije s pacijentima), njihov ključan i najintrigantniji aspekt je strojno učenje (machine learning)” (Bracanović, 2021, str. 64). Strojno učenje omogućava pristupe velikim skupovima podataka, koje liječnici koriste za precizno dijagnosticiranje bolesti. Sustavi umjetne inteligencije sadrže velike baze podataka na temelju kojih se analiziraju slike te se detektiraju tumorske stanice. Posebno je značajno to što radiolozima omogućavaju otkrivanje tumora u ranom stadiju (pregledavanje MR i RT nalaza), na temelju čega mogu uspostaviti uspješan

tretman za ozdravljenje pacijenta. Putem elektronskih zdravstvenih kartona dostupni su svi podaci o povijesti bolesti pacijenta, na temelju čega se definiraju načini budućih liječenja.

Rak je danas „drugi vodeći uzrok smrti u svijetu i ima značajan utjecaj na gospodarstvo, kao i na ljudske živote. Istraživači iz cijelog svijeta posvećuju cijelu svoju karijeru liječenju raka; svake se godine ostvaruju ogromni pomaci u dijagnostici i liječenju lijekovima. U nadi da će spasiti nebrojene života, istraživači su se okrenuli futurističkim dijagnostičkim metodama pomoću inteligentnih strojeva. Znanstvenici vjeruju da će im porast tehnologije strojnog učenja pomoći ne samo da pronađu lijek za rak, već i potpuno zaustaviti njegov razvoj" (FOZZ UNIPU, 2019, str. 71).

Od raka obolijevaju žene i muškarci, a najčešće je to rak pluća, rak maternice i dojke kod žena, rak debelog crijeva, rak prostate, te dr. Problematika kod pregledanih mjesta raka je pogrešna dijagnoza, koja može nastati čak i kod korištenja višestrukih testova, a tu najčešće spadaju biopsija, MRI i rendgen. Analiziraju se abnormalne stanice na tkivima, u čemu je doprinijela umjetna inteligencija i njeni računalni programi, uz posredstvo djelovanja čovjeka i njegovog iskustva i znanja. Radiolozi analiziraju slike, obrasce, patološke procese, promjene u tkivu, uz pomoć pametnih strojeva. Stroj ne može funkcionirati bez čovjeka.

Takva tehnologija umjetne inteligencije utječe na bolje izvršavanje složenih zadataka dijagnosticiranja i liječenja raka, štedi vrijeme, segmentira tumore, te predviđa vjerojatnosti za malignost (zloćudnost) tumora. „Segmentiranje i tumora i zdravog tkiva vrlo je važno za terapiju zračenjem, ali ljudski čitatelji jednostavno nemaju sposobnost primijetiti suptilne razlike u teksturi ili obliku. Stanice raka mogu biti slične veličine, oblika i strukture, što ih čini težim prepoznati tehnikama snimanja.

Tehnologija naprednog učenja može pomoći u pronalaženju postojećih obrazaca u staničnoj morfologiji povezanih s određenim mutacijama i biološkim procesima" (Ibidem, str. 89). Danas postoji veliki broj pacijenata sa dijagnosticiranim rakom, u cijelom svijetu, pa prevladava veliki interes za korištenje umjetne inteligencije u dijagnosticiranju i liječenju istog. Cilj je postaviti točnu dijagnozu korištenjem patoloških slajdova, radioloških slika, predviđanja ishoda pacijenata te optimizirati odluke o liječenju. Umjetna inteligencija ima potencijal za rješavanje problema nejednake distribucije medicinskih resursa, te poboljšati liječenje raka. Za analizu

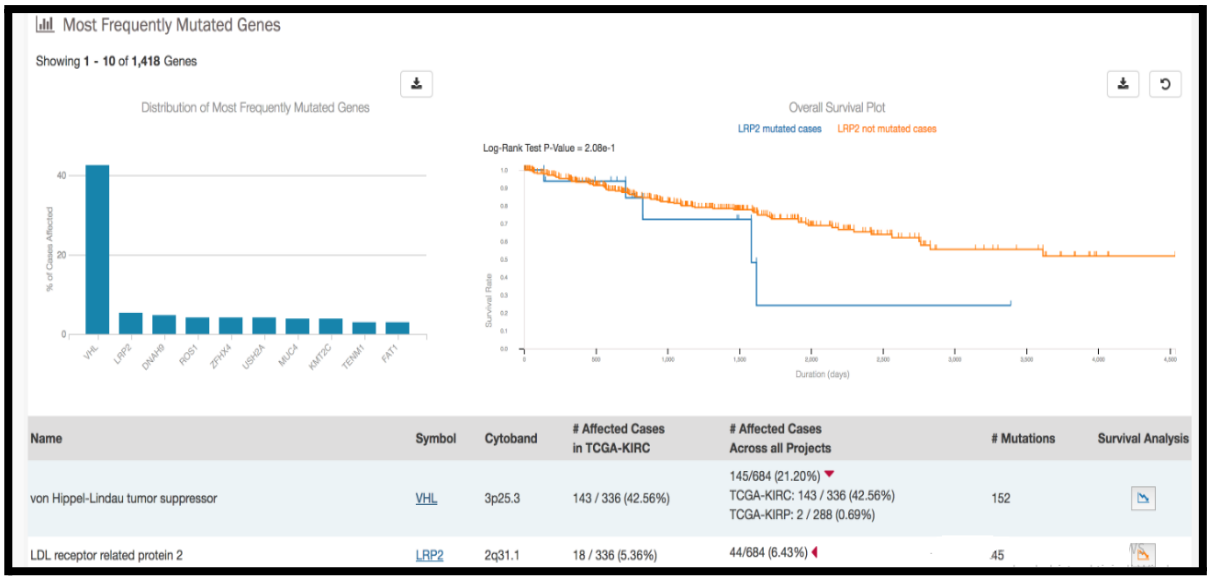
različitih vrsta podataka i za predviđanja, duboko učenje koristi neuronske mreže, putem čega algoritmi opskrbljuju stroj potrebnim podacima koji najbolje odgovaraju zadacima poput prepoznavanja slika, uzoraka, te dr.

Modeli dubokih neuronskih mreža, poput konvolucijske neuronske mreže (engl. Convolutional Neural Networks - CNN) su postali najpopularniji u liječenju lezija raka, prepoznavanja, segmentacije i klasifikacije medicinskih slika (CNN transformira izvorne slike, sloj po sloj, do konačnih rezultata predviđanja; konvolucijski slojevi se sastoje od kombiniranja ulaznih podataka (mapa značajki) s konvolucijskim jezgrama (filterima), a cilj im je formiranje transformirane karte značajki) (Chen, 2021). Filtri se prilagođavaju putem naučenih parametara, te izdvajaju značajke koje su najkorisnije za određeni zadatak. Takav način probiranja utjecao je na smanjenje smrtnosti od nekih vrsta raka, a posebno se ističe identifikacija prekanceroznih lezija cervikalne intraepitelne neoplazije ili CIN-a, koji je uzrok razvoja raka na grliću maternice kod žena. Liječenje putem kolorektalnog probira dovodi do smanjenja incidencije invazivnog raka.

Za probir raka vrata maternice, Wentzensen i sur. su razvili klasifikator dubokog učenja za dvostruko obojene citološke slajdove obučene na zlatnim standardima, temeljene na biopsiji, a u neovisnom testiranju temeljene na umjetnoj inteligenciji, što je imalo puno veći učinak od Papa testa i ručnog tumačenja dvostruko obojenih citoloških slajdova. Dvostruko obojeni citološki slajdovi temeljeni na umjetnoj inteligenciji utjecali su na smanjenje izvođenja kolposkopije za jednu trećinu, te na povećanje razine prepoznavanja CIN-a visokog stupnja, čime je omogućeno pravovremeno liječenje (Ibidem, 2021). Osim raka maternice, za primjer se može navesti i rak debelog crijeva, kod kojeg je umjetna inteligencija značajna u potpomaganju kolonoskopiji, pri čemu je povećala stope otkrivanja adenoma po pacijentu, u odnosu na konvencionalnu kolonoskopiju. Na takav način je smanjena stopa smrtnosti od raka debelog crijeva. Umjetna inteligencija je značajna i kod liječenja raka pluća, kod kojeg je utjecala na točnost klasifikacije (do 95%), čime je dokazan potencijal u probiru raka pluća. Korištenje CNN modela iz sustava umjetne inteligencije ima veliki potencijal u dijagnosticiranju i liječenju svih vrsta raka, pa se u budućnosti očekuje puno od istih, stoga je potrebno ulagati u njihov razvoj, te na takav način smanjiti broj oboljelih, odnosno spasiti brojne ljudske živote. Danas se za dijagnosticiranje raka koriste koriste razni alati za rudarenje podataka, a neki od njih

su GDC (engl. Genomic Data Commons) alati za analizu, vizualizaciju i istraživanje (DAVE), putem kojih se utječe na promicanje stvaranja baza znanja o genomici raka (GDC, 2022.). Primjer korištenja alata GDC DAVE u otkrivanju somatskih mutiranih gena prikazan je na Slici 5. S obzirom na to da je rak bolest genoma, uzrokovana promjenama u DNA, RNA, identificiranje genomskih promjena raka pomaže istraživačima da dekodiraju razvoj raka, poboljšaju dijagnozu i liječenje na temelju različitih molekularnih abnormalnosti. Podaci GDC-a služe u pristupu standardiziranim kliničkim, proteomskim, epigenomskim i genomskim podacima iz studija raka kako bi se omogućila eksplorativna analiza identificiranja promjena u stanicama raka. Dakle, GDC alati su važni u otkrivanju raka, pa se na takav način utječe na suzbijanje njegova razvoja.

Slika 4.: Otkrivanje mutiranih gena putem alata GDC DAVE



Izvor: GDC (2022.)

Alati GDC-a omogućavaju pristupe datotekama *Variant Calling Format* (VCF) i *Mutation Annotation Format* (MAF) koje identificiraju somatske mutacije (primjerice točkaste mutacije, *missense* mutacije, nukleotidi u DNK, i dr.), kvantifikaciju ekspresije gena, identifikaciju događaja genomskog preuređivanja (npr. fuzije, duplikacije), metilacije za identifikaciju epigenomskih modifikacija na DNK, ekspresiju proteina za prepoznavanje promjena u post-translacijskim modifikacijama, te dr.

(GDC, About the Data, 2022.). Podaci se jednostavno pohranjuju, i dijele suradnicima za potrebe daljnjih istraživanja.

4.2. Primjer 2 / Otkrivanje lijekova i molekularno modeliranje umjetnom inteligencijom

Molekularno modeliranje umjetnom inteligencijom je značajno za područje kliničke mikrobiologije, koja je specijalizirano područje zdravstva. Stručnjaci kliničke mikrobiologije „daju liječničko mišljenje o dijagnozi, liječenju i prevenciji bolesti uzrokovanih mikroorganizmima i parazitima, znanstvene osnove za laboratorijsku dijagnozu, liječenje i prevenciju zaraznih bolesti, izrađuju protokole i održavaju standarde u laboratoriju, izvode mikrobiološku dijagnostiku najčešćih uzročnika zaraznih bolesti iz humanih kliničkih uzoraka, preuzimaju brigu o kontroli i prevenciji bolničkih infekcija, predlažu mjere za racionalnu primjenu antibakterijskih, antivirusnih, antifungalnih i antiparazitnih lijekova u bolnici, sudjelovati u istraživanjima i razvoju iz područja kliničke mikrobiologije i zaraznih bolesti.“ Klinička mikrobiologija se bavi strukturom, genetikom i taksonomijom bakterija i virusa, otkrivanjem antibakterijskih i antivirusnih lijekova, sterilizacijom i dezinfekcijom, te brojnim drugim aktivnostima. U području mikrobiologije se koristi prediktivna analiza (o kojoj je bilo više u potpoglavlju 3.3.), koja u zdravstvu podrazumijeva podatke vezane za rudarenje genoma. Provedena su brojna istraživanja na DNA mikronizovima koji imaju tisuće gena, s cilj takvih istraživanja je dijagnosticiranje raznih bolesti. Istraživači kod takvog pristupa nastoje dati odgovore na biološka pitanja iterativnim rudarenjem tisuća genomskih skupova podataka, obuhvaćajući različite molekularne aktivnosti, tehnološke platforme i modelne organizme. Cilj rudarenja podataka povezanih s genomom je revolucioniranje zdravstvene skrbi intenziviranjem znanja o molekularnoj razini bolesti (Maksood i Achuthan, 2016, str. 7). Prikupljeni podaci omogućavaju lakše određivanje bolesti, odnosno njene razine.

Umjetna inteligencija je uključena u razvoj farmaceutskih proizvoda, lijekova, u određivanje prave terapije za pacijenta, uključujući personalizirane lijekove, te

pomaže u upravljanju generiranim kliničkim podacima koje koristiti za budući razvoj lijekova (Debleena, 2021, str. 80-91). Umjetna inteligencija se koristi u otkrivanju i razvoju lijekova jer omogućava prepoznavanje uspješnih i vodećih spojeva, brzu provjeru valjanosti lijeka te optimizaciju dizajna strukture lijeka. Navedeno je bilo nemoguće izvesti korištenjem napredne tehnologije, koja je proces otkrivanja lijekova činila dugotrajnim i skupim zadatkom. Međutim, umjetna inteligencija je omogućila razvoj velikog broja molekula lijekova, u velikom kemijskom prostoru, koji se sastoji od >1060 molekula.

Umjetna inteligencija se odnosi na sposobnost računala da uče iz postojećih podataka, pa računalno modeliranje temeljeno na umjetnoj inteligenciji znači obećavajuću metodu za procjenu bioloških aktivnosti i toksičnosti spojeva. Modeli koji su dostupni komercijalnom softveru za otkrivanje lijekova imaju sposobnost predviđanja jednostavnih fizikalno – kemijskih svojstava, te su precizni u predviđanju farmakokinetičkih svojstava novih spojeva s jednostavnim mehanizmima (Zhu, 2020, str. 574). Noviji pristupi umjetne inteligencije za unapređenje suvremenog otkrivanja lijekova temelje se na prediktivnom modeliranju, koje je pogodno za analizu velikih podataka i novijih vrsta podataka poput slika, koje su specifične za područje zdravstva.

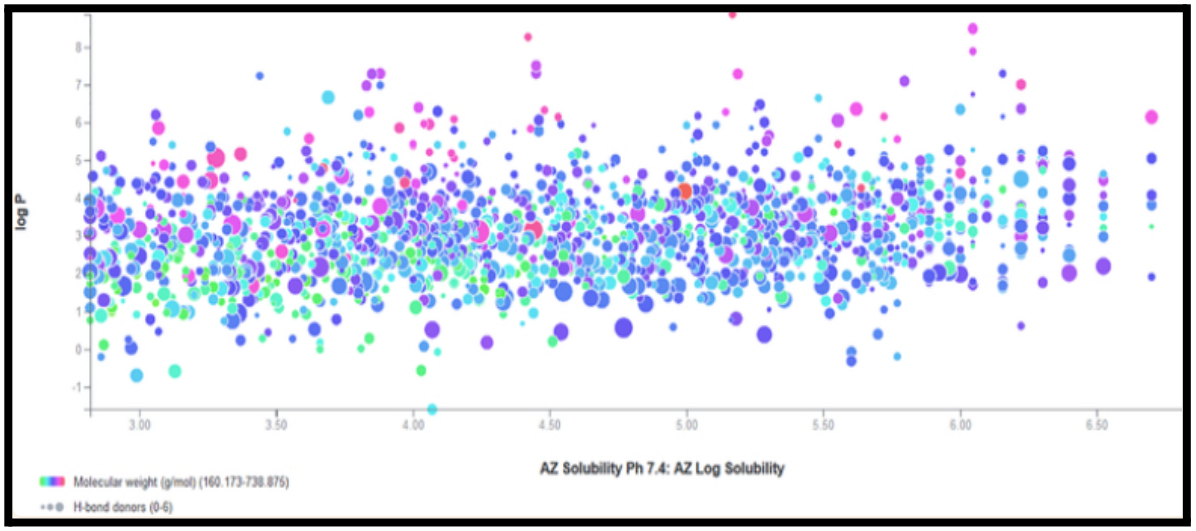
Modeli dubokog učenja pokazuju najbolju prediktivnost za skupove podataka o metabolizmu, izlučivanju i toksičnosti lijekova na kandidate na kojima se lijekovi testiraju, te skupove podataka o apsorpciji i distribuciji lijekova. S razvojem pristupa neuronskih mreža, duboko učenje se naširoko primjenjuje u otkrivanju lijekova u eri velikih podataka, koji su sve više značajni u kliničkim studijama. Potreba za novim tehnikama, poput rudarenja podataka donosi nove izazove i prilike u istraživačkoj zajednici (Ibidem, str. 576). Rudarenje podataka omogućava predviđanje ciljanog lijeka, modeliranje metaboličke mreže, i identifikaciju uzoraka populacijske genetike, oslanjajući se na računalno modeliranje lijekova i molekula. Jedan od takvih programa

za računalno modeliranje je program Genomic Data Commons, koji osigurava skladište podataka za razmjenu u genomskim studijama (tehnikom rudarenja podataka, koje je temelj umjetne inteligencije), koje su potpora preciznoj medicini.

Poduzeće Collaborative Drug Discovery Inc. (CDD) je nedavno razvilo sofisticiranije mogućnosti vizualizacije podataka u biotehnološke svrhe, istraživačke

bolnice, akademske laboratorije te dr., putem CDD Vault-a, kao cjelovite jedinstvene informatičke platforme koja pruža pakete modernih alata temeljenih na webu, koji generiraju grafiku podataka kvalitete objave (CDD, 2022.). Takvi alati omogućavaju vizualizaciju i manipuliranje stotinama tisuća točaka u grafičkom prikaz, a to su WebGL (engl. Web Graphics Library; za 2D i 3D grafičke prikaze) i SVG (engl. Scalable Vector Graphics). Alat SVG je vektorska datoteka koja pohranjuje slike putem matematičkih formula temeljenih na točkama i linijama na mreži, što bi značilo da se mogu značajno mijenjati bez gubitka kvalitete, pa su idealne za složenu grafiku, poput one kod molekuliranja lijekova (Adobe, 2022.). Na Slici 4. nalazi se jednostavan prikaz odabranih molekula i podataka CDD Valuta korištenjem alata za vizualizaciju fizikalno - kemijskih svojstava lijeka Astra Zenca, putem dijagrama uzorka ChEMBL-a na 1763 spoja koji pokazuju odnos s izračunatim molekularnim svojstvima.

Slika 5:. Vizualizacija fizikalno - kemijskih svojstava lijeka Astra Zeneca (jednostavan prikaz) putem alata CDD Vault-a



Izvor: (Ekins, et al., 2018.)

Ukoliko istraživači žele saznati više o određenoj molekuli ili nekom spoju tada vrše podatkovni odabir u tablici CCD modela. Nakon odabira željenog dijagrama, isti se može izvesti u pdf-u, te se može koristiti ua razvoj modela strojnog učenja i dijeliti među suradnicima. Nedostatak otkrivanja lijekova je taj što podaci često mogu biti nepravilni i višedimenzionalni, pa su alati CCD Vault-a pogodni za brzo poništavanje i

ponavljanje radnji. Takvo rudarenje je korisnički usmjereno i učinkovito.

4.3. Prednosti i nedostaci korištenja umjetne inteligencije za rudarenje podataka u zdravstvu na odabranim primjerima

Umjetna inteligencija za rudarenje podataka u zdravstvu, na prethodno prikaznim primjerima, ima brojne prednosti (Ahmad et. al., 2021.) :

prvenstveno, promijenila je način dijagnosticiranja i liječenja raka,
utjecala je na smanjenje medicinskih troškova,
utjecala je na uštedu vremena,
utjecala je na povećanje kvalitete života pacijenata (ranije utvrđena dijagnoza utječe na brže liječenje, zbog brzog prepoznavanja stanja bolesti),
umjetna inteligencija utječe na povećanje vjerojatnosti preživljavanja oboljelih od raka,
umjetna inteligencija se pokazala kao potencijal za rješavanje izazovnih problema koje ljudi jednostavno ne mogu riješiti.

Algoritmi dubokog učenja koji se koriste za automatsko izdvajanje značajki i medicinskih podataka, za izradu modela, su važni jer mogu predvidjeti rizike od nastanka tumorskih stanica, kao i reakcije pacijenata na određene tretmane, poput imunoterapije i kemoterapije (na temelju rezultata predviđanja tretmani su precizniji i prikladniji). Na takav način nastaju točni prediktivni testovi za informiranje o odabiru pacijenata, a navedeno je moguće ostvariti kombinacijom velikih podataka i umjetne inteligencije.

Prediktivni modeli umjetne inteligencije mogu identificirati slikovne tipove raka, i njihovu povezanost sa mutacijom (da li je invazivan ili nije). Točna predviđanja su važna jer omogućavaju pravovremeno i ispravno liječenje, čime se izbjegava toksičnost i odgađanje operacija. CNN modeli nude prilagođen način dijagnosticiranja i prilagođavanja doziranja lijekova na pojedinačne ili kombinirane terapije, koristeći podatke koji su se o bolesti prikupili tijekom određenog vremena (Shao, et. al., 2021.). Algoritmi strojnog učenja i umjetne inteligencije su prilika za rješavanje

problema u zdravstvu, na temelju korištenja tehnike rudarenja velikih podataka. Umjetna inteligencija i strojno učenje su najvažnije tehnologije današnjice koje transformiraju društvo, i gospodarstvo. Cilj gospodarstava je ulagati u sektor umjetne inteligencije jer je usmjeren na stvaranje inovacija koje pridonose budućim istraživanjima koja će utjecati na unapređenje zdravstva i na rast njegove produktivnosti.

Kod otkrivanja lijekova i molekularnog modeliranja umjetnom inteligencijom korištenje prediktivne analize ima brojne prednosti, jer predviđanje omogućava visoku razinu točnosti, ukoliko se koriste složene tehnike, poput logističke regresije, stabla odlučivanja i neuronskih mreža. Duboke neuronske mreže imaju mogućnost klasificiranja velikih integriranih skupova podataka, te pohrane, čime se sprječava gubitak podataka (Debleena et. al., 2021.). Rudarenje podataka omogućava predviđanje ciljanog lijeka, modeliranje metaboličke mreže, i identifikaciju uzoraka populacijske genetike, oslanjajući se na računalno modeliranje lijekova i molekula. Umjetna inteligencija se koristi u otkrivanju i razvoju lijekova jer omogućava prepoznavanje uspješnih i vodećih spojeva, brzu provjeru valjanosti lijeka te optimizaciju dizajna strukture lijeka. Navedeno je bilo nemoguće izvesti korištenjem napredne tehnologije, koja je proces otkrivanja lijekova činila dugotrajnim i skupim zadatkom. Međutim, umjetna inteligencija je omogućila razvoj velikog broja molekula lijekova.

Nedostatak korištenja umjetne inteligencije za rudarenje podataka u zdravstvu donosi brojna pitanja koja se odnose na etičnost. „S obzirom na širu primjenu umjetne inteligencije u zdravstvu i istraživanju, Vijeće za bioetiku Nuffield (Nuffield Council on Bioethics) ističe da, ako se takvi sustavi koriste za postavljanje dijagnoze ili izradu plana liječenja, a zdravstveni djelatnik ne može objasniti kako su nastali, te time mogu ograničavati pravo pacijenata da slobodno i informirano donose odluke o svome zdravlju” (Bracanović 2021, str. 68). Prema navedenom, liječnici primjenu umjetne inteligencije u zdravstvu trebaju dobro razumjeti, i znati interpretirati podatke i ponoviti postupak, a ne slučajno doći do rezultata, koji su im u konačnici nerazumljivi. Na takav način će se moći pravilno skrbiti o pacijentu kojemu je dijagnosticirana bolest, ali i budućim pacijentima sa istom dijagnozom (npr. kod oboljenja poput raka).

Osim navedenog nedostatak umjetne inteligencije u zdravstvu je u tome što se

ista oslanja na velike količine podataka. Ograničeni podaci mogu utjecati na prekomjerno prilagođavanje, te na takav način rezultirati inferiornom izvedbom u vanjskoj ispitnoj kohorti. Također, važno je pitanje zaštite podataka o pacijentima, koji su u vlasništvu pojedinačnih ustanova, a kojima nedostaju sustavi za razmjenu podataka, odnosno za povezivanje institucija (Chen, 2021). Naveden prepreke je moguće prevladati distribucijom dubokog učenja koje omogućava privatnost na temelju sporazuma o dijeljenju lokalnih skupova podataka. Razmjenjuju se podaci radioloških, mikrobioloških i drugih studija, kliničke slike i drugi podaci koji su potrebni zdravstvenim djelatnicima, a ostali podaci o pacijentima ostaju zaštićeni. U budućnosti je potrebno usmjeriti pažnju na bolju sigurnost dijeljenja podataka, kao i na pitanja etičnosti, i to posebno ona koja će definirati tko je odgovoran za iznošenje netočnih odluka u provođenju aktovnosti u zdravstvu korištenjem umjetne inteligencije.

ZAKLJUČAK

U 21. stoljeću je umjetna inteligencija veoma značajno područje istraživanja u raznim poljima. Neka od najznačajnijih polja su inženjerstvo, znanost, obrazovanje, medicina, zdravstvo, računovodstvo, marketing, ekonomija, poslovanje te brojna druga polja. Rudarenje podataka i strojno učenje su samo dva od ukupno 16 zasebno klasificiranih polja umjetne inteligencije. Strojno učenje je danas puno više od komercijalne primjene metoda za izvlačenje informacija iz podatka, odnosno ono je neophodno za umjetnu inteligenciju, jer da bi se inteligentni sustavi mogli prilagoditi svojoj okolini oni moraju naučiti ponavljati svoje uspjehe, i svoje greške. Inteligencija nastaje korištenjem jednostavnih algoritama, koji imaju sposobnost učenja na temelju velike količine podataka. Strojno učenje je najčešća metoda za modeliranje poslovnih procesa, pri čemu je najznačajnije simulacijsko modeliranje putem računala, odnosno simulacijskog softvera, te korištenjem kombinacije matematičkih modela za kvantitativnu analizu izvođenja vizualizacije i animacije procesa. Strojno učenje je proces kod kojeg računalni inženjeri putem tradicionalnog programiranja unose upute računalu o pretvaranju ulaznih podataka u željeni izlaz, ili je pak automatizirani proces putem kojeg strojevi rješavaju probleme bez ljudskog unosa, te poduzimaju radnje na temelju prošlih aktivnosti i opažanja. Strojno učenje omogućava inteligentnim sustavima da uče nove stvari iz podataka, iz kojih se crpe brojne informacije. Takvi podaci i informacije imaju obilježja velike razine točnosti, pa je strojno učenje značajno u pogledu uštede vremena i novca koji se troše za analize rješavanja problema, te na takav način predstavljaju podršku korisnicima (rast zadovoljstva) i podršku rudarenju podataka iz internih izvora koje Internet nudi. Strojno učenje u prostoru umjetne inteligencije utkano na temelju znanosti podataka, odnosno na temelju unosa velikih količina podataka, koji rezultiraju dubokim učenjem, koje je jedan od modela umjetne inteligencije. Rudarenje podataka se razlikuje od tradicionalnih tehnika baza podataka ili statističkih metoda po tome što se može koristiti za otkrivanje novih obrazaca ili za potvrdu sumnjivih odnosa. Obrazac kod rudarenja podataka treba biti istinit, i siguran, točan. Sigurnost obrasca može uključivati čimbenike poput cjelovitosti podataka i veličine uzorka. Rudarenje

podataka je nastalo spajanjem strojnog učenja i statistike, pa danas zajednicom rudarenja podataka dominiraju računalni znanstvenici i statističari. Rudarenje se kao aplikacijska domena pojavljuje kao učinkovit skup tehnika usmjerenih na rudarenje teksta, slika i grafova. Tehnike rudarenja podataka su klasifikacija podataka, grupiranje podataka, predikcija ili predviđanje, pravilo pridruživanja te neuronske mreže. Rudarenje podatka je tehnika koja je temelj umjetne inteligencije, u obliku programskih kodova koji sadrže informacije i podatke potrebne za sustave umjetne inteligencije. Rudarenje podataka se aplicira u tehničke, komercijalne i istraživačke svrhe, najčešće u području bio-informatike, financijskog bankarstva, obrazovanja, kriminalističke analize, u analizi tržišne košarice, te buduće zdravstvene njege. U zdravstvu se koriste tehnike rudarenja podataka poput klasifikacije, pravila pridruživanja, grupiranja, te neuronske mreže, s ciljem otkrivanja odnosa među bolestima, tretmanima, zbog identificiranja novih lijekova, otkrivanja prijevara, smanjenja troškova, te dr. Neki od alata koji se koriste u zdravstvu su Rapid miner, R programming, Weka, Orange, NLTK te dr. Primjeri primjene umjetne inteligencije za rudarenje podataka u zdravstvu su otkrivanje lijekova i molekularno modeliranje umjetnom inteligencijom, te dijagnostika raka i donošenje odluka o liječenju pomoću umjetne inteligencije. Rudarenje podataka u zdravstvu je važno zbog predviđanja učinkovitosti određenih kirurških postupaka, medicinskih testova i lijekova, čime se pomaže u podizanju standarda kliničkog odlučivanja, se se na takav način pridonosi zdravlju i sigurnosti ljudi. Molekularno modeliranje umjetnom inteligencijom je značajno za područje kliničke mikrobiologije, koja je specijalizirano područje zdravstva. Klinička mikrobiologija se bavi strukturom, genetikom i taksonomijom bakterija i virusa, otkrivanjem antibakterijskih i antivirusnih lijekova, sterilizacijom i dezinfekcijom, te brojnim drugim aktivnostima. U području mikrobiologije se koristi prediktivna analiza koja u zdravstvu podrazumijeva podatke vezane za rudarenje genoma. Provedena su brojna istraživanja na DNA mikronizovima koji imaju tisuće gena, s cilj takvih istraživanja je dijagnosticiranje raznih bolesti. Istraživači kod takvog pristupa nastoje dati odgovore na biološka pitanja iterativnim rudarenjem tisuća genomskih skupova podataka, obuhvaćajući različite molekularne aktivnosti, tehnološke platforme i modelne organizme. Cilj rudarenja podataka povezanih s genomom je revolucioniranje zdravstvene skrbi intenziviranjem znanja o molekularnoj razini bolesti. Prikupljeni podaci omogućavaju lakše određivanje bolesti, odnosno njene razine. Umjetna inteligencija je uključena u razvoj farmaceutskih proizvoda,

lijekova, u određivanje prave terapije za pacijenta, uključujući personalizirane lijekove, te pomaže u upravljanju generiranim kliničkim podacima koje koristiti za budući razvoj lijekova.

Noviji pristupi umjetne inteligencije za unapređenje suvremenog otkrivanja lijekova temelje se na prediktivnom modeliranju, koje je pogodno za analizu velikih podataka i novijih vrsta podataka poput slika, koje su specifične za područje zdravstva. Modeli dubokog učenja pokazuju najbolju prediktivnost za skupove podataka o metabolizmu, izlučivanju i toksičnosti lijekova na kandidate na kojima se lijekovi testiraju, te skupove podataka o apsorpciji i distribuciji lijekova. S razvojem pristupa neuronskih mreža, duboko učenje se naširoko primjenjuje u otkrivanju lijekova u eri velikih podataka, koji su sve više značajni u kliničkim studijama. Potreba za novim tehnikama, poput rudarenja podataka donosi nove izazove i prilike u istraživačkoj zajednici. Rudarenje podataka omogućava predviđanje ciljanog lijeka, modeliranje metaboličke mreže, i identifikaciju uzoraka populacijske genetike, oslanjajući se na računalno modeliranje lijekova i molekula. U zdravstvu se koriste ekspertni sustavi AIBDS koji se sastoje od velikih skupova podataka potrebnih medicinskim stručnjacima za određivanje dijagnoza, uzoraka, snimki, određivanja tretmana pri liječenju i sl. Sustavi umjetne inteligencije sadrže velike baze podataka na temelju kojih se analiziraju slike te se detektiraju tumorske stanice. Posebno je značajno to što radiolozima omogućavaju otkrivanje tumora u ranom stadiju, na temelju čega mogu uspostaviti uspješan tretman za ozdravljenje pacijenta. Kod raka se analiziraju abnormalne stanice na tkivima, u čemu je doprinijela umjetna inteligencija i njeni računalni programi, uz posredstvo djelovanja čovjeka i njegovog iskustva i znanja. Radiolozi analiziraju slike, obrasce, patološke procese, promjene u tkivu, uz pomoć pametnih strojeva. Stroj ne može funkcionirati bez čovjeka. Takva tehnologija umjetne inteligencije utječe na bolje izvođenje složenih zadataka dijagnosticiranja i liječenja raka, štedi vrijeme, segmentira tumore, te predviđa vjerojatnosti za malignost tumora. Danas postoji veliki broj pacijenata sa dijagnosticiranim rakom, u cijelom svijetu, pa prevladava veliki interes za korištenje umjetne inteligencije u dijagnosticiranju i liječenju istog. Umjetna inteligencija ima potencijal za rješavanje problema nejednake distribucije medicinskih resursa, te poboljšati liječenje raka. Za analizu različitih vrsta podataka i za predviđanja, duboko učenje koristi neuronske mreže, putem čega algoritmi opskrbljuju stroj potrebnim

podacima koji najbolje odgovaraju zadacima poput prepoznavanja slika, uzoraka, te dr. U Budućnosti je potrebno razriješiti pitanja etičnosti korištenja umjetne inteligencije u zdravstvu, kao i zaštitu podataka koji se dijele među zdravstvenim institucijama.

POPIS LITERATURE

Knjige:

1. Bosilj - Vukšić, V., Hernaus, T., Kovačić, A., (2008.), *Upravljanje poslovnim procesima – organizacijski i informacijski pristup*, Školska knjiga, Zagreb

E – knjige:

1. Alpaydin, E., *Introduction to Machine learning, MIT Press, 2004.*; za hrvatsko izdanje: Debić, B., Njavro, V (ur.), *Strojno učenje – nova umjetna inteligencija*, Mate d.o.o., Zagreb, 2021., dostupno na: https://wdn2.ipublishcentral.com//mate/viewinsidehtml/501708461513431 (29.07.2022.)
2. Binu, D., Rajakuma, B. R. (ur.), (2021.), *Artificial Intelligence in Data Mining: Theories and Applications*, Academic Press, dostupno na: https://www.sciencedirect.com/book/9780128206010/artificial-intelligence-in-data-mining#book-description (12.08.2022.)
3. Brumeca, J., *Modeliranje poslovnih procesa*, 2021., Koris, Zagreb/Varaždin, dostupno na: https://koris.hr/preuzmi/koris-uvod-u-modeliranje-poslovnihprocesa.pdf (30.07.2022.)
4. Custers, B. et. al., (2013.) Studies in Applied Philosophy, Epistemology and Rational Ethics, Springer, dostupno na: https://www.researchgate.net/publication/278661450_What_Is_Data_Mining_and_How_Does_It_Work (6.08.2022.)

Članci:

1. Ahmad, Z., Rahim, S., Zubair, M., Abdul – Ghafar, J. (2021.) *Artificial intelligence (AI) in medicine, current applications and future role with special*

emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review), Diagnostic Pathology volume 16, Article number: 24, dostupno na: https://diagnosticpathology.biomedcentral.com/articles/10.1186/s13000-021-01085-4 (15.09.2022.)

2. Ayoola Oke, S., (2008.) *Artificial Intelligence: A Rewiew of tehe Literature*, International Journal of Information and Management Sciences, Volume 19, Number 4, pp. 535-570, dostupno na: https://www.researchgate.net/publication/228809837_ARTIFICIAL_INTELLIGENCE_A_REVIEW_OF_THE_LITERATURE (14.07.2022.)
3. Academic Ebrary, (2022.) *The Relationship between Data Mining, Machine Learning, and Artificial Intelligence*, dostupno na: https://ebrary.net/190208/health/relationship_data_mining_machine_learning_artificial_intelligence (30.07.2022.)
4. Bolf, N., (2021), *Strojno učenje,* Osvježimo znanje, Kem. Ind. 70 (9-10), 591–593, dostupno na: www.hrcak.srce.hr (30.07.2022.)
5. Bharati, M. R., (December 2010.) *Data mining techniques and applications*, Indian Journal of Computer Science and Engineering 1(4), (301 – 305), Project:
6. Bharati M. Ramageri, (2011.) Data Mining Techniques and Applications, dostupno na: https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_applications (6.08.2022.)
7. Bracanović, T., (2021.) *Umjetna inteligencija, medicina i autonomija*, Nova prisutnost Vol. XIX, No. 1, (63-76), dostupno na: www.hrcak.srce.hr (13.08.2022.)
8. Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., De Felice, C., (January 2020.) *Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions*, Sustainability 12(2):492, dostupno na: https://www.mdpi.com/2071-1050/12/2/492 (14.07.2022.)
9. Coenen, F., (March 2011.) *Data mining: Past, present and future*, The Knowledge Engineering Review 26(01), dostupno na: https://www.researchgate.net/publication/220254364_Data_mining_Past_pres

ent_and_future (7.08.2022.)

10. Chen, Z. H., et. al., (2021.) *Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine,* Cancer Commun (Lond). 41(11):
1100–1115., dostupno na:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8626610/ (14.08.2022.)
11. Debleena, P., et. al., (2021.) *Artificial intelligence in drug discovery and development*, Drug Discov Today. 26(1):80–93., dostupno na:
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7577280/ (13.08.2022.)
12. Ekins, S. et. al. (2018.) *Data Mining and Computational Modeling of High Throughput Screening Datasets*, Methods Mol Biol. 2018; 1755: 197–221., dostupno na: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6181121/ (10.09.2022.)
13. JavaTpoint, (2021.) *Data Mining vs Artificial Intelligence*, dostupno na: https://www.javatpoint.com/data-mining-vs-artificial-intelligence (14.07.2022.)
14. Maksood, F. Z., Achuthan, G., (April 2016.) *Analysis of Data Mining Techniques and its Applications*, International Journal of Computer Applications 140(3):614, dostupno na:
https://www.researchgate.net/publication/301325506_Analysis_of_Data_Mining_Techniques_and_its_Applications (13.08.2022.)
15. MonkeyLearn Inc., (2022.) *An Introduction to Machine Learning*, dostupno na: https://monkeylearn.com/machine-learning/ (14.07.2022.)
16. myservername, (2022.) *Vrste strojnog učenja: Nadzirano protiv nenadgledanog učenja*, dostupno na:
https://hr.myservername.com/typesmachine-learning (30.07.2022.)
17. Mesarić, J., Šebelj, D., (2014./2015.) *Upravljanje informacijskim resursima*, Ekonomski fakultet u Osijeku, dostupno na:
https://slideplayer.com/slide/13987749/ (6.08.2022.)
18. Neha, K., Maramreddy, Y. R., (February 2020.) *A Study On Applications Of Data Mining*, International Journal of Scientific & Technology Research 9(2), (33853388), dostupno na:
https://www.researchgate.net/publication/344459744_A_Study_On_Applications_Of_Data_Mining (6.08.2022.)
19. Pandian, S., (2020.) *Understand Machine Learning and Its End-to-End*

Process, Analytics Vidhya, dostupno na: https://www.analyticsvidhya.com/blog/2020/12/understand-machine-learningand-its-end-to-end-process/ (30.07.2022.)

20. Phil Papers Org., (2022.) *Podatak, informacija, informatika, znanje, mudrost*, dostupno na: https://philpapers.org/archive/MICHRO.pdf (6.08.2022.)
21. Sanders, J. D., (July 2016.) *Defining Terms: Data, Information and Knowledge*, Conference: SAI LondonAt: Excel Centre, Project: Artificial Intelligence, dostupno na: https://www.researchgate.net/publication/305474792_Defining_Terms_Data_Information_and_Knowledge (6.08.2022.)
22. Scheurwegs, E., et. al., (2016.) *Data integration of structured and unstructured sources for assigning clinical codes to patient stays,* J Am Med Inform Assoc. 23(e1): e11–e19., dostupno na: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4954635/ (14.08.2022.)
23. Shao, D., Dai, Y., Li, N., Cao, X., Zhao, W., Cheng, L., Rong, Z., Huang, L., Wang, Y., Zhao, J. (2021.) *Artificial intelligence in clinical research of cancers*, Briefings in Bioinformatics, Volume 23, Issue 1, dostupno na: https://academic.oup.com/bib/article/23/1/bbab523/6470966 (15.09.2022.)
24. Sl.education, (2022.) *Modeli strojnog učenja*, dostupno na: https://hr.educationwiki.com/3480854-machine-learning-models (30.7.2022.)
25. Softwaretestinghelp, (2022.) *Data Mining Vs Machine Learning Vs Artificial Intelligence Vs Deep Learning*, dostupno na: https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/ (12.08.2022.)
26. Zhu, H., (2020.) *Big Data and Artificial Intelligence Modeling for Drug Discovery*, Annual Review of Pharmacology and Toxicology, Vol.60:573-589, dostupno na: https://www.annualreviews.org/doi/abs/10.1146/annurev-pharmtox-010919023324 (13.08.2022.)
27. Xiao, C., Sun, J., (2021.) *Introduction to Deep Learning for Healthcare, Health Data*, (Pages 9-22), Springer, dostupno na: https://link.springer.com/book/10.1007/978-3-030-82184-5 (13.08.2022)

Vodiči:

1. Simplilearn, (2022.) *The Complete Guide to Machine Learning Steps*, dostupno na: https://www.simplilearn.com/tutorials/machine-learning-tutorial/machinelearning-steps (30.07.2022.)

Priručnici:

1. FOZZ UNIPU, *3DandVRforVET, (*2019.) *Savjeti za poboljšanje 3D iskustva; Praktični priručnik za medicinsku i industrijsko - obrtničku školu*, Pula, dostupno na: https://fooz.unipu.hr/_download/repository/%5BHR%5D_Handbook__craft_medical_schools_IO1.pdf (13.08.2022.)

Internetski izvori:

1. Adobe (2022.) *SVG files*, dostupno na: https://www.adobe.com/creativecloud/file-types/image/vector/svg-file.html (10.09.2022.)
2. CDD (2022.) *CDD VAULT, Modern Research Informatics*, https://www.collaborativedrug.com/benefits/ (10.09.2022.)
3. GDC (2022.) *GDC Dave Tools*, dostupno na: https://gdc.cancer.gov/analyze-data/gdc-dave-tools (10.09.2022.)
4. GDC (2022.) *About the Data*, dostupno na: https://gdc.cancer.gov/about-data (10.09.2022.)
5. IBM, (2021.) *Data Mining*, dostupno na: https://www.ibm.com/cloud/learn/datamining (6.08.2022.)
6. *Naziv specijalizacije; Kliničke mikrobiologija*, dostupno na: https://narodnenovine.nn.hr/clanci/sluzbeni/full/dodatni/421792.pdf (13.08.2022.)

POPIS SLIKA

POPIS TABLICA

SAŽETAK

U radu je bilo riječi o važnosti umjetne inteligencije za rudarenje podataka u zdravstvu. Za potrebe definiranja i analiziranja navedenog definirala se umjetna inteligencija, strojno učenje, proces strojnog učenja, te vrste i modeli strojnog učenja. Nadalje, bilo je riječi o rudarenju podataka, pri čemu su se definirali podaci, informacije i znanje, rudarenje podataka, tehnike, metode i modeli rudarenja podataka, odnos umjetne inteligencije i rudarenja podataka, te aplikacije rudarenja podataka. U zadnjem dijelu rada su se prikazivali primjeri primjene umjetne inteligencije za rudarenje podataka u zdravstvu: otkrivanje lijekova i molekularno modeliranje umjetnom inteligencijom, te dijagnostika raka i donošenje odluka o liječenju pomoću umjetne inteligencije. Nakon toga su objašnjene prednosti i nedostaci korištenja umjetne inteligencije za rudarenje podataka u zdravstvu na prethodno prikazanim primjerima.

Ključne riječi: umjetna inteligencija, strojno učenje, rudarenje podataka, aplikacije rudarenja podataka, tehnike, metode i modeli rudarenja podataka, umjetna inteligencija za rudarenje podataka u zdravstvu, otkrivanje lijekova i molekularno modeliranje umjetnom inteligencijom, dijagnosticiranje raka i donošenje odluka o liječenju pomoću umjetne inteligencije

SUMMARY

The paper discussed the importance of artificial intelligence for data mining in healthcare. For the purposes of defining and analyzing the above, artificial intelligence, machine learning, the process of machine learning, and types and models of machine learning were defined. Furthermore, data mining was discussed, defining data, information and knowledge, data mining, data mining techniques, methods and models, the relationship between artificial intelligence and data mining, and data mining applications. In the last part of the paper, examples of the application of artificial intelligence for data mining in healthcare were presented: drug discovery and molecular modeling with artificial intelligence, and cancer diagnosis and treatment decision-making using artificial intelligence. After that, the advantages and disadvantages of using artificial intelligence for data mining in healthcare are explained using the previously presented examples.

Keywords: artificial intelligence, machine learning, data mining, data mining applications, data mining techniques, methods and models, artificial intelligence for healthcare data mining, drug discovery and molecular modeling with artificial intelligence, cancer diagnosis and treatment decision making using artificial intelligence